

Université de Montréal

**Étude de la variante d'histone H2A.Z et
du cycle de phosphorylation de l'ARN polymérase II
chez *Saccharomyces cerevisiae***

Par Alain R. Bataille

Thèse présentée à la Faculté de Médecine
en vue de l'obtention du grade de Docteur
en Biologie moléculaire

Février 2012

© Alain R. Bataille, 2012

Université de Montréal
Faculté de Médecine

Cette thèse intitulée :

Étude de la variante d'histone H2A.Z et du cycle de phosphorylation
de l'ARN polymérase II chez *Saccharomyces cerevisiae*

Présentée par :
Alain R. Bataille

a été évaluée par un jury composé des personnes suivantes :

Dr Muriel Aubry, président-rapporteur
Dr François Robert, directeur de recherche
Dr Jacques Côté, membre du jury
Dr Jacques Greenblatt, examinateur externe
Dr Sébastien Carréno, représentant de la doyenne

Résumé

La chromatine est plus qu'un système d'empaquetage de l'ADN ; elle est le support de toutes les réactions liées à l'ADN dans le noyau des cellules eucaryotes et participe au contrôle de l'accès de l'ARN polymérase II (ARNPolII) à l'ADN. Responsable de la transcription de tous les ARNm des cellules eucaryotes, l'ARNPolII doit, suivant son recrutement aux promoteurs des gènes, transcrire l'ADN en traversant la matrice chromatinienne. Grâce au domaine C-terminal (CTD) de sa sous-unité Rpb1, elle coordonne la maturation de l'ARNm en cours de synthèse ainsi que les modifications de la chromatine, concomitantes à la transcription. Cette thèse s'intéresse à deux aspects de la transcription : la matrice, avec la localisation de la variante d'histone H2A.Z, et la machinerie de transcription avec le cycle de phosphorylation du CTD de l'ARNPolII.

Suivant l'introduction, le chapitre 2 de cette thèse constitue un protocole détaillé et annoté de la technique de ChIP-chip, chez la levure *Saccharomyces cerevisiae*. Cette technique phare dans l'étude *in vivo* des phénomènes liés à l'ADN a grandement facilité l'étude du rôle de la chromatine dans les phénomènes nucléaires, en permettant de localiser sur le génome les marques et les variantes d'histones. Ce chapitre souligne l'importance de contrôles adéquats, spécifiques à l'étude de la chromatine.

Au chapitre 3, grâce à la méthode de ChIP-chip, la variante d'histone H2A.Z est cartographiée au génome de la levure *Saccharomyces cerevisiae* avec une résolution d'environ 300 paires de bases. Nos résultats montrent que H2A.Z orne un à deux nucléosomes au promoteur de la majorité des gènes. L'enrichissement de H2A.Z est anticorrélé à la transcription et nos résultats suggèrent qu'elle prépare la chromatine pour l'activation des gènes. De plus H2A.Z semble réguler la localisation des nucléosomes.

Le chapitre suivant s'intéresse à la transcription sous l'angle de la machinerie de transcription en se focalisant sur le cycle de phosphorylation de l'ARN polymérase II. Le domaine C-terminal de sa plus large sous-unité est formé de répétitions d'un heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$ dont les résidus peuvent être modifiés au cours de la

transcription. Cette étude localise les marques de phosphorylation des trois résidus sérine de manière systématique dans des souches mutantes des kinases et phosphatases. Nos travaux confirment le profil universel des marques de phosphorylations aux gènes transcrits. Appuyés par des essais *in vitro*, ils révèlent l'interaction complexe des enzymes impliqués dans la phosphorylation, et identifient Ssu72 comme la phosphatase de la sérine 7. Cet article appuie également la notion de « variantes » des marques de phosphorylation bien que leur étude spécifique s'avère encore difficile.

La discussion fait le point sur les travaux qui ont suivi ces articles, et sur les expériences excitantes en cours dans notre laboratoire.

Mots clefs : Immunoprécipitation de la chromatine; Chromatine; H2A.Z; Transcription; ARN polymérase II; CTD; Kin28; Bur1; Ssu72; Fcp1.

Abstract

Chromatin is more than just the eucaryotic DNA packaging system; it is the substrate of all reactions involving DNA in eukaryotic cells and actively regulates RNA Polymerase II (RNAPolIII) access to DNA. Responsible for all mRNA transcription in eucaryotes, the RNAPolIII must, following its recruitment to the pre-initiation complex, overcome the chromatin barrier in order to transcribe genes. The RNAPolIII CTD allows for the co-transcriptional coordination of mRNA maturation and chromatin modifications. The work covered in this thesis addresses two aspects of transcription: the chromatin substrate, with the localization of H2A variant, H2A.Z, and the transcription complex with the phosphorylation cycle of the RNAPolIII CTD.

Following the introduction, chapter 2 constitutes a detailed and annotated *Saccharomyces cerevisiae* ChIP-chip protocol, from the culture to the hybridization of the array, with an emphasis on the proper controls required for chromatin study. This technique, extremely powerful for the *in vivo* study of all DNA transactions, leads to a better understanding of chromatin function in nuclear phenomena, thanks to the localization of histone variants and modifications.

The third chapter maps the H2A.Z variant across the yeast genome at ~300 base pairs resolution using ChIP-chip. Our data shows that H2A.Z is incorporated into one or two promoter-bound nucleosomes at the majority of genes. H2A.Z enrichment is anticorrelated with transcription, and the results suggest that it configures chromatin structure to poise genes for transcriptional activation. Furthermore, we have shown that H2A.Z can regulate nucleosome positioning.

The next chapter focuses on the transcription machinery and, more precisely, on the phosphorylation cycle of RNAPolIII. The CTD contains repetitions of a heptapeptide (Y₁S₂P₃T₄S₅P₆S₇) on which all serines are differentially phosphorylated along genes in a prescribed pattern during the transcription cycle. Here, we systematically profiled the location of the RNAPII phospho-isoforms in wild-type cells and mutants for most CTD modifying enzymes. The results provide evidence for

a uniform CTD cycle across genes. Together with results from *in vitro* assays, these data reveal a complex interplay between the modifying enzymes, identify Ssu72 as the Ser7 phosphatase and show that proline isomerization is a key regulator of CTD dephosphorylation at the end of genes. Moreover, it reinforces the notion of variants of the phosphorylation marks, even though the exact nature of the variant is still difficult to identify.

The discussion introduces the studies that followed this work, including new projects conceived in our lab.

Key words: Chromatin Immunoprecipitation; Chromatin; H2A.Z; Transcription; RNA polymérase II; CTD; Kin28; Bur1; Ssu72; Fcp1;

Table des matières

Résumé	iii
Abstract	v
Table des matières.....	vii
Liste des Figures.....	xiii
Liste des Figures Supplémentaires.....	xiv
Liste des Tableaux.....	xv
Liste des abréviations	xvi
Remerciements	xix
1 Introduction	1
1.1 Chromatine	2
1.1.1 Bref historique de la chromatine	2
1.1.2 Organisation	2
1.1.2.1 Les états de condensation de la chromatine	2
1.1.2.2 Nucléosomes	8
1.1.2.3 Histones.....	11
1.1.3 Mécanismes de régulations	17
1.1.3.1 Positionnement des nucléosomes.....	17
1.1.3.2 Modifications covalentes des histones	20
1.1.3.3 Variantes d’histones.....	24
1.2 Transcription.....	38
1.2.1 Bref historique de la transcription	38

1.2.2	ARN polymérase II	39
1.2.2.1	Structure	39
1.2.2.2	Domaine C-terminal.....	43
1.2.3	Cycle de la transcription.....	45
1.2.3.1	Pré initiation.....	46
1.2.3.2	Initiation	51
1.2.3.3	Pause lors de l'élongation précoce.....	52
1.2.3.4	Élongation.....	52
1.2.3.5	Terminaison	55
1.2.3.6	Recyclage de la polymérase	56
1.2.4	Le cycle de phosphorylation du CTD.....	57
1.2.4.1	Le cycle.....	58
1.2.4.2	La fonction du CTD	62
1.3	Rationnel de la thèse.....	66
2	Protocole de localisation génomique des modifications et des variantes d'histones par la méthode de CHIP-chip chez <i>S. cerevisiae</i>	67
2.1	Mise en contexte de l'article.....	67
2.2	Contribution des auteurs.....	67
2.3	Manuscrit.....	68
2.3.1	Summary	68
2.3.2	Introduction	69
2.3.3	Materials.....	70
2.3.3.1	Growing and Crosslinking Cells.....	70
2.3.3.2	Breaking Cells.....	70
2.3.3.3	Immunoprecipitation.....	71

2.3.3.4	DNA Purification	71
2.3.3.5	Amplification and labeling.....	71
2.3.3.6	Hybridization	72
2.3.4	Methods	73
2.3.4.1	Growing and Crosslinking Cells.....	73
2.3.4.2	Breaking Cells.....	74
2.3.4.3	Preparation of Soluble Chromatin Fragments.....	75
2.3.4.4	Immunoprecipitation.....	75
2.3.4.5	DNA Purification	76
2.3.4.6	Amplification and Labeling	77
2.3.4.7	Hybridization	79
2.3.5	Notes.....	80
2.4	Références.....	85
3	La variante d’histone H2A.Z est globalement localisée au promoteur des gènes inactifs de la levure et régule le positionnement des nucléosomes	86
3.1	Mise en contexte de l’article.....	86
3.2	Contribution des auteurs.....	86
3.3	Manuscrit.....	87
3.3.1	Summary	87
3.3.2	Introduction	88
3.3.3	Results	91
3.3.3.1	Genome-wide location analysis of H2A.Z.....	91
3.3.3.2	H2A.Z is associated with small genomic regions present all across the yeast genome	92
3.3.3.3	H2A.Z-containing nucleosomes are essentially located within promoter regions.....	94

3.3.3.4	H2A.Z is not enriched in the promoter regions of highly active genes.....	97
3.3.3.5	H2A.Z regulates nucleosome positioning.....	99
3.3.3.6	H2A.Z is unusually localized at HZAD genes.....	102
3.3.4	Discussion	105
3.3.5	Materials and Methods.....	108
3.3.5.1	Yeast strains and genetic methods	108
3.3.5.2	Antibodies	108
3.3.5.3	Chromatin immunoprecipitations	109
3.3.5.4	DNA microarrays.....	109
3.3.5.5	DNA labeling and hybridization.....	109
3.3.5.6	Data Analysis	110
3.3.5.7	Chromatin analysis.....	110
3.3.6	Supplementary Documents.....	111
3.3.7	Acknowledgments	121
3.4	References.....	122
4	Le cycle de la l'ARN polymérase II est universel et orchestré le long des gènes par des interactions complexes entre enzymes, kinase, phosphatase et isomérase ...	128
4.1	Mise en contexte de l'article.....	128
4.2	Contribution des auteurs.....	128
4.3	Manuscrit.....	129
4.3.1	Abstract	129
4.3.2	Introduction	130
4.3.3	Results	133
4.3.3.1	Phospho-CTD profiling on active genes using multiple antibodies.....	133

4.3.3.2	A similar CTD cycle operates on genes of different size classes	136
4.3.3.3	The kinase activity of Kin28 down regulates the activity of Bur1 against P-Ser5 and 7 along the coding region	138
4.3.3.4	Ctk1 is the major serine 2 kinase during transcription <i>in vivo</i>	141
4.3.3.5	Kin28 and Bur1 have antagonistic roles in early steps of transcription	142
4.3.3.6	Ssu72 dephosphorylates both serine 5 and serine 7 prior to termination	142
4.3.3.7	The prolyl isomerase Ess1 plays a critical role in CTD dephosphorylation by Ssu72	145
4.3.3.8	Two variants of P-Ser2 are dephosphorylated with different kinetics	146
4.3.3.9	Fcp1 partially dephosphorylates Ser2 prior to termination	147
4.3.3.10	Ssu72 enables complete dephosphorylation of Ser2 by Fcp1 after termination	150
4.3.4	Discussion	150
4.3.4.1	A universal CTD cycle	150
4.3.4.2	Is CTD phosphorylation important for transcription?	151
4.3.4.3	Phosphorylation of Ser7	152
4.3.4.4	Phosphorylation of Ser2 by Bur1/2	153
4.3.4.5	Different variants of the phospho-marks	154
4.3.5	Experimental procedures	154
4.3.5.1	Yeast strains, antibodies, cell extracts, purification of protein complexes and recombinant proteins	154
4.3.5.2	ChIP-chip assay	155
4.3.5.3	Accession numbers	155

4.3.5.4	<i>In vitro</i> CTD phosphatase assay.....	155
4.3.6	Supplemental data	156
4.3.7	Supplemental experimental procedures.....	165
4.3.8	Acknowledgments	173
4.4	References.....	174
5	Discussion	181
5.1	H2A.Z	181
5.1.1	Localisation de H2A.Z	181
5.1.2	Implication de H2A.Z dans la mémoire transcriptionnelle	184
5.1.3	Dissection des domaines de H2A.Z	186
5.2	Transcription.....	191
5.2.1	Conclusion.....	191
5.2.2	Perspectives	195
6	Conclusion.....	196
7	Bibliographie	198

Liste des Figures

Figure 1.1 : Modèle de compaction de la fibre de 30 nm.....	4
Figure 1.2 : Modèle d'organisation en polymère liquide de la chromatine.	6
Figure 1.3 : Modèle d'assemblage de l'octamère d'histones	15
Figure 1.4 : Interactions Protéine-ADN et Protéine-Protéine au sein du nucléosome.	16
Figure 1.5 : Comparaison des séquences primaires d'H2A et H2A.Z	29
Figure 1.6 : Superposition des structures du nucléosome canonique et du nucléosome incorporant H2A.Z.	30
Figure 1.7 : Représentation schématique de l'ARNPolIII et de son site catalytique. .	42
Figure 1.8 : Le domaine C-terminal de l'ARN polymérase II de levure.	44
Figure 1.9 : Représentation graphique du cycle de phosphorylation du CTD.	61
Figure 3.1 : Genome-wide location analysis of H2A.Z and H2A; a zoom-in on chromosome III.	93
Figure 3.2 : H2A.Z is predominantly localized within promoter regions.	96
Figure 3.3 : H2A.Z/H2B ratios are inversely correlated with transcription rate.....	99
Figure 3.4 : H2A.Z regulates nucleosome positioning.....	101
Figure 3.5 : H2A.Z is unusually localized in HZAD genes.	104
Figure 4.1 : A uniform CTD cycle at most transcribed genes.....	135
Figure 4.2 : Complex interplays between CTD kinases.....	140
Figure 4.3 : Ssu72 dephosphorylates Ser5 and Ser7 at the ends of genes.....	144
Figure 4.4 : Ess1 is required for the dephosphorylation of Ser5 and Ser7 at the ends of genes.	146
Figure 4.5 : Two variants of P-Ser2 are removed in two steps by the concerted action of Fcp1 and Ssu72.	149
Figure 5.1 : Collection de souches chimériques H2A/H2A.Z.....	189
Figure 5.2 : Essai sur plaques de la collection de souches chimériques H2A/H2A.Z	190

Liste des Figures Supplémentaires

Figure Sup. 3.1: High-resolution mapping of H2A.Z at the SRB8 locus by Q-PCR	117
Figure Sup. 3.2: Q-PCR analysis of H2A.Z ChIP experiments.....	118
Figure Sup. 3.3: H2A.Z occupancy at HZAD10.	119
Figure Sup. 3.4: Z loci are wider around telomeres.....	120
Figure Sup. 4.1: Gene groups previously identified as having non-canonical profiles by Tietjen <i>et al.</i> , (2010) are populated by genes with biased gene length.	156
Figure Sup. 4.2: <i>In vitro</i> and <i>in vivo</i> evidence for Bur1/2 being an opportunistic kinase.....	158
Figure Sup. 4.3: The data from Figure 4.3 is shown for the total length of the metagene.....	160
Figure Sup. 4.4: The data from Figure 4 is shown for the total length of the metagene.....	161
Figure Sup. 4.5: RNAPII CTD is masked in the <i>fcp1-1</i> strain.	162

Liste des Tableaux

Table 1.1: Liste, non exhaustive, des modifications d'histones.	21
Table Sup. 3.1: Raw data for the H2A.Z/H2B, H2A/H2B and H2A.Z/H2A.Z ChIP-chip experiments.....	111
Table Sup. 3.2: Smoothed data for the H2A.Z/H2B ChIP-chip experiment.	111
Table Sup. 3.3: Smoothed data for the H2A/H2B ChIP-chip experiment.....	111
Table Sup. 3.4: Smoothed data for the H2A.Z/H2A.Z ChIP-chip experiment.....	111
Table Sup. 3.5: List of Z loci	111
Table Sup. 3.6: Genome Browser-ready file.....	111
Table Sup. 3.7: Sequences of the oligonucleotides used to quantify ChIPs by PCR in this study.....	112
Table Sup. 3.8: Sequences of the oligonucleotides used in the LM-PCR assay to map nucleosomes within the <i>GALI</i> promoter	113
Table Sup. 4.1: Subunit composition of the TAP-purified complexes shown in Figure Sup. 4.2A based on mass spectrometry data	164
Table Sup. 4.2: List of strains used in this study.	172

Liste des abréviations

3C	: <i>Chromatin conformation capture</i>
ADN	: Acide déoxyribonucléique
ARN	: Acide ribonucléique
ARNm	: ARN messenger
ARNPII	: ARN Polymérase II
ARNpolII	: ARN Polymérase II
ARNt	: ARN de transfert
ATP	: Adénosine triphosphate
ATPase	: ATP hydrolase
BRE	: <i>TFIIB recognition element</i>
CAK	: <i>Cyclin-activating kinase complex</i>
CBP	: <i>Calmodulin binding peptide</i>
CDK	: <i>Cyclin dependant kinase</i>
ChIP	: <i>Chromatin immunoprecipitation</i>
ChIP-chip	: <i>Chromatin immunoprecipitation on chip</i>
ChIP-seq	: <i>Chromatin immunoprecipitation sequencing</i>
CPF	: <i>cleavage/polyadenylation factor</i>
cryo-EM	: cryo microscopie électronique
CTD	: <i>Carboxy-terminal domain</i>
C-terminal	: Carboxy-terminal
FGT	: Facteur généraux de la transcription
FRET	: <i>Fluorescence resonance energy transfer</i>
GST	: Glutathion-S-transférase
HA	: Hemagglutinin (HA tag)
HAT	: Histone acétyltransférase
HDAC	: Histone déacétylase
HMR	: <i>Hidden MAT Right</i>
HMT	: Histone méthyltransférase
HZAD	: <i>H2AZ affected domain</i>
H α	: Hélice alpha
H α C	: Hélice alpha carboxy terminale
H α N	: Hélice alpha amino terminale
INO80	: <i>Inositol requiring 80</i>
KAT	: Lysine acétyltransférase
KDM	: Lysine déméthylase
KMT	: Lysine méthyltransférase
LM-PCR	: <i>Ligation mediated PCR</i>
MNase	: <i>Micrococcal Nuclease</i>
NAD ⁺	: Nicotinamide adénine dinucléotide oxydée
NER	: <i>Nucleotide excision repair</i>

NFR	: <i>Nucleosome free region</i>
N-terminal	: <i>Amino-terminal</i>
pb	: <i>paire de base</i>
PCR	: <i>Polymerase chain reaction</i>
PIC	: <i>Preinitiation complex</i>
P-Ser	: <i>Phospho serine</i>
Q-PCR	: <i>Quantitative polymerase chain reaction</i>
Ser	: <i>Sérine</i>
snoRNA	: <i>Small nucleolar RNA</i>
snRNA	: <i>Small non coding RNA</i>
SWI-SNF	: <i>switching defective/sucrose nonfermenting</i>
TAF	: <i>TBP-associated protein</i>
TAP	: <i>Tandem Affinity Purification</i>
TEV	: <i>Tobacco etch virus</i>
TBP	: <i>TATA binding protein</i>
TC-NER	: <i>transcription coupled nuclear excision repair</i>
TES	: <i>Transcription end site</i>
TF	: <i>Transcription factor</i>
TSS	: <i>Transcription start site</i>
UAS	: <i>Upstream activating sequence</i>

À ce qui nous échappe.

Remerciements

J'aimerais tout d'abord remercier François, mon superviseur, de m'avoir donné l'opportunité de travailler à ses côtés. Je lui suis redevable pour son soutien et sa confiance, maintes fois renouvelés. Je retiendrai nos discussions scientifiques et son optimisme.

J'aimerais remercier les membres de mon jury, le D^r Muriel Aubry, le D^r Jacques Côté, le D^r Jacques Greenblatt et le D^r Sébastien Carréno pour l'évaluation de ma thèse, et Vivianne Jodoin aux programmes de biologie moléculaire pour son efficacité et sa bonne humeur.

Je tiens à remercier mes coauteurs, Benoît et Célia, car travailler avec eux fut un réel plaisir.

Merci à mes collègues et compagnons : Lorelei, Josette, Célia, P-E, Sara, Audrey, Marie-Ève, Griselle et plus particulièrement Louise et Simon, pour leur collaboration, leur appui et leurs conseils. Sans eux, mon expérience et mes expériences au laboratoire n'auraient pas été les mêmes.

Merci à mes amis, de Montréal et d'ailleurs : Nadège, Barbara, Simon, Geneviève, Lisette, Daniel, Kathy, Lorelei... pour leur soutien, et peut-être surtout, pour m'avoir permis d'oublier le laboratoire, de temps en temps.

Merci à ma famille, ma Mère et ma Sœur, pour avoir été à mes côtés, malgré la distance. Leur soutien inconditionnel a souvent été essentiel à la réussite de mes projets.

Enfin, merci à Montréal qui fut le décor idéal de toutes ces années.

1 Introduction

La transcription est la première étape menant à l'expression des gènes, et donc un des mécanismes centraux de la cellule. L'ADN sert de matrice à la machinerie de transcription pour la synthèse des ARN qui à leur tour peuvent servir de matrice à la machinerie de traduction pour la synthèse des protéines. Chez les eucaryotes, ce n'est toutefois pas un filament d'ADN double brin nu qui sert de matrice aux ARN polymérases, mais la chromatine, c'est-à-dire un complexe nucléoprotéique dans lequel les histones et d'autres protéines entrent en contact étroit avec l'ADN. Longtemps, la chromatine fut considérée uniquement comme le support inerte de l'empaquetage de l'ADN ; son étude n'a donc soulevé que peu d'intérêt. Il faudra attendre que son implication dans la régulation des phénomènes associés à l'ADN soit démontrée pour que son étude suscite l'enthousiasme des chercheurs. Les principes et les mécanismes de la transcription sont très largement semblables à travers l'évolution. On observe cependant une complexification synchronique à l'évolution, aboutissant à la diversification des complexes et des mécanismes de régulation de la machinerie de transcription. Par exemple, les organismes procaryotes n'ont qu'une ARN polymérase responsable de l'ensemble de la transcription, alors que les eucaryotes possèdent 3 polymérases spécialisées pour des sous-types d'ARN particuliers. La machinerie de transcription des eucaryotes a aussi dû acquérir de nouvelles propriétés afin de surmonter l'obstacle que représente la chromatine, et ainsi tirer avantage des nombreuses possibilités de régulation que celle-ci leur offre.

Les travaux rassemblés dans cette étude abordent deux acteurs de la transcription. La variante d'histone H2A.Z incorporée à la chromatine, et la machinerie de transcription de l'ARN polymérase II : plus particulièrement le rôle du domaine C-terminal (CTD) de sa sous-unité principale (Rpb1) dans la transcription d'ARN messenger (ARNm). Ces études ont été réalisées chez la levure *S. cerevisiae* (dans ce texte le terme « levure » fait référence à l'espèce *Saccharomyces cerevisiae* sauf précision contraire).

1.1 Chromatine

1.1.1 Bref historique de la chromatine

La chromatine est spécifique aux organismes eucaryotes. Bien avant la reconnaissance de l'ADN comme support de l'hérédité, les premières observations de la chromatine sont de nature histochimique. En effet, au 19^e siècle, Walther Flemming développe une technique de coloration pour mettre en évidence la substance fibreuse présente dans le noyau des cellules eucaryotes qu'il nomme chromatine en 1879 (dérivé du grec : chroma, signifiant: couleur) [1, 2]. Au début du 20^e siècle, les observations d'Emil Heitz sur la coloration des chromosomes polythènes de drosophile font ressortir des patrons de bandes plus ou moins colorées, observables au microscope. Ces bandes différencient deux types de chromatine : une chromatine moins intensément colorée (moins dense) nommée euchromatine et postulée comme étant génétiquement active et une chromatine plus intensément colorée (plus dense) postulée comme génétiquement inactive nommée hétérochromatine [2, 3]. Dans la deuxième partie du 20^e siècle, la résolution de la structure de l'ADN, sa reconnaissance comme support de l'information génétique [4], ainsi que l'identification du nucléosome comme unité répétitive de la chromatine [5] ont précisé le rôle, la composition et la structure de la chromatine. Enfin, l'analyse biochimique et la résolution de la structure cristallographique des nucléosomes ont permis de faire d'immenses progrès dans la compréhension de la chromatine et de ses fonctions.

1.1.2 Organisation

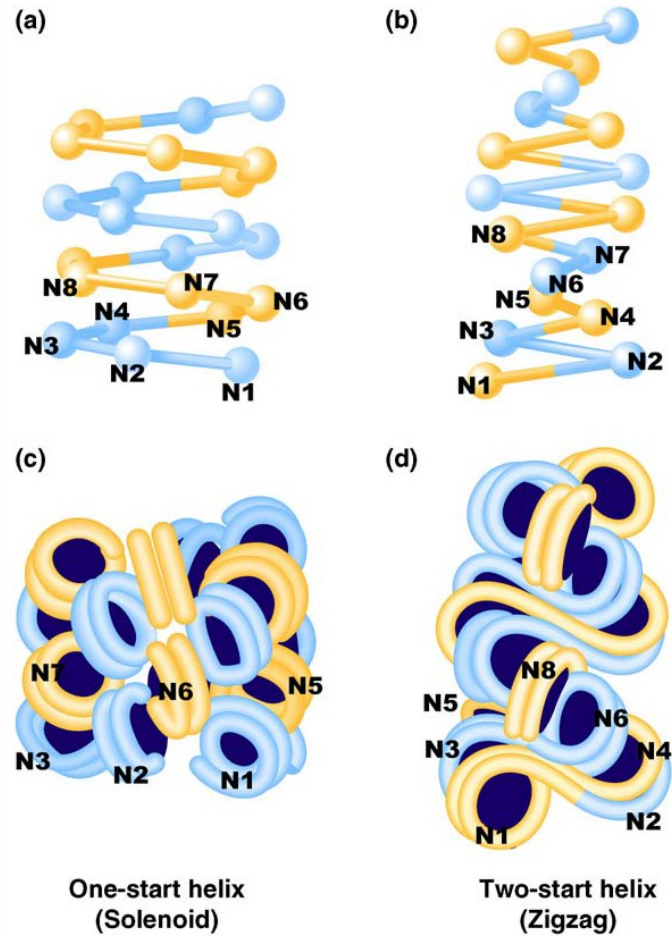
1.1.2.1 Les états de condensation de la chromatine

La chromatine est un complexe nucléoprotéique dont le rôle principal est l'empaquetage de l'ADN dans le noyau des organismes eucaryotes. Les cellules eucaryotes modulent le degré de condensation de la chromatine afin, par exemple, de la décondenser pour permettre le passage des machineries de transcription et de réplication ou, au contraire, de la condenser d'avantage, sous forme de chromosomes

mitotiques lors de la division cellulaire. Ces états extrêmes de compaction de la chromatine mettent en évidence une de ses caractéristiques principales : la plasticité. De nombreuses études ont démontré que le degré de compaction participait de manière active à la régulation de l'expression génique et que, par conséquent, la vision selon laquelle l'euchromatine correspondait à des domaines de transcription active et l'hétérochromatine à des domaines transcriptionnellement inactifs s'avérait exacte. Les eucaryotes, au cours de leur évolution, ont donc développé les moyens de modifier et remodeler la compaction et la composition de la chromatine de manière à remplir les fonctions d'empaquetage tout en maintenant l'accès à l'ADN [6].

À son niveau le plus fondamental, la chromatine est organisée en une succession de nucléosomes visibles par microscopie électronique sous forme d'un collier de perles, appelé fibre de 10 nm, correspondant au premier niveau de condensation de l'ADN qui compacte l'ADN d'un facteur six [7]. Dans le modèle classique de la chromatine, l'agencement des nucléosomes les uns par rapport aux autres, forme le second niveau de compaction, la fibre de 30 nm. Au fil des années, plusieurs modèles ont été proposés pour expliquer l'agencement des nucléosomes et leurs interactions le long de la fibre de 30 nm [8, 9]. Deux modèles principaux ont émergé i) le solénoïde formant une hélice à un point de départ, dans laquelle chaque nucléosome est lié à son voisin ou interdigitalisé comme représenté dans la (**Figure 1.1**) [8, 10, 11]. ii) le zigzag à deux points de départ, dans lequel les nucléosomes sont liés à leurs seconds voisins [8, 12] (**Figure 1.1**). La fibre de 30 nm est elle-même repliée et agencée en des niveaux de compactations supérieurs incorporant des protéines de structure additionnelles [6].

À l'inverse des eucaryotes supérieurs, la majorité de la chromatine de la levure *Saccharomyces cerevisiae* se trouve sous forme ouverte et les seules régions d'hétérochromatine réelle sont les régions sous-télomériques et le locus de conjugaison. En plus, chez les eucaryotes supérieurs, outre l'hétérochromatine constitutive, on peut observer une hétérochromatine dite « facultative » dont le degré de compaction varie, par exemple en fonction du type cellulaire.



Current Opinion in Cell Biology

Figure 1.1 : Modèle de compaction de la fibre de 30 nm.

Maeshima *et al.*, (2010) [13].

(a)(b) Représentation schématiques des deux principaux modèles, le solénoïde et le zigzag respectivement. **(a)** Chaque tour du solénoïde est coloré alternativement en jaune ou en bleu. **(b)** Les paires de nucléosomes sont alternativement colorées en bleu et jaune. **(c)(d)** Représentation des deux modèles en y incorporant les nucléosomes afin de mieux visualiser leur interaction au sein de la fibre.

À cause des limitations imposées par les techniques d'investigation, peu de choses sont réellement connues sur l'organisation des niveaux de condensation supérieurs à la fibre de 30 nm. Les expériences démontrant la formation d'une fibre de 30 nm sont basées uniquement sur des méthodes *in vitro*, dans lesquelles la dilution de chromatine favorise majoritairement les interactions entre les nucléosomes appartenant à la même fibre. En fait, l'organisation en fibre de 30 nm n'a jamais pu être prouvée dans les chromosomes en interphase ou au sein des chromosomes en mitose [14]. Sur la base d'expériences de capture de conformation de la chromatine (3C) et cryo microscopie électronique (cryo-EM), l'existence de la fibre de 30 nm a même été remise en question au profit d'une organisation dite en polymère liquide (*polymer melt*) (**Figure 1.2 a-d**) [7, 13]. Les auteurs de l'étude proposent un modèle prédisant que les interactions nucléosomiques entre des segments distants d'une même fibre ou interfibres deviennent majoritaires dans les conditions extrêmement compactes du noyau [13, 14]. Les gènes réprimés se trouveraient alors à l'intérieur de domaines organisés en gouttes visqueuses, tandis que les gènes transcrits se trouveraient à la surface de ces gouttes (**Figure 1.2 e,f**) [13]. Ce modèle nous force à reconsidérer nos conceptions sur l'organisation de la chromatine en interphase et même dans les chromosomes mitotiques.

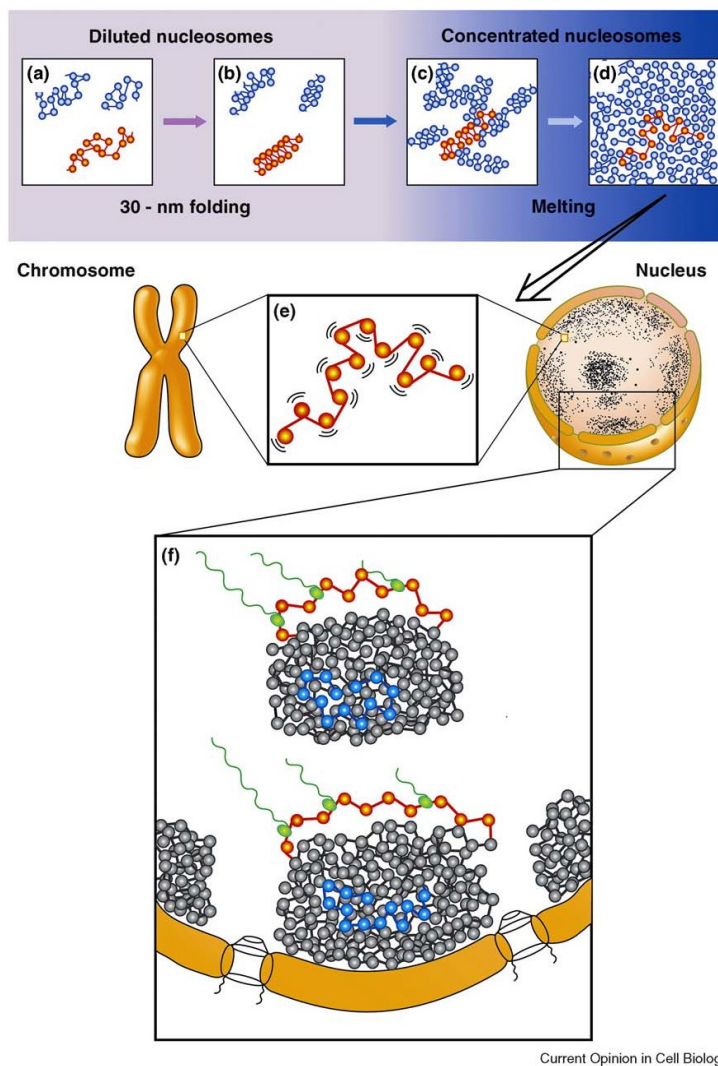


Figure 1.2 : Modèle d'organisation en polymère liquide de la chromatine.

Maechima *et al.*, (2010) [13].

(a)(b) Dans des conditions diluées, la fibre nucléosomale se compacte par l'intermédiaire de contacts avec les nucléosomes voisins appartenant à la même fibre, ce qui aboutit à la formation de la fibre de 30 nm. **(c)(d)** Lorsque la concentration en nucléosomes augmente, les contacts interfibrilles augmentent et interfèrent avec les contacts intrafibrilles jusqu'au stade ultime de polymère liquide. **(e)** Le concept de polymère liquide implique un réarrangement constant des fibres de nucléosomes. **(f)** Les « gouttes » visqueuses de chromatine permettent l'organisation des régions transcrites, avec la chromatine active (représentée en orange) en surface des gouttes visqueuses, et la chromatine réprimée au cœur des « gouttes » (représentée en bleu).

En interphase, les chromosomes, bien que décondensés, ne sont pas désorganisés dans le noyau, mais occupent des territoires distincts [15, 16]. On remarque d'ailleurs que les régions chromosomiques situées à proximité dans le noyau corrélient avec les translocations observées dans les tumeurs [17]. L'observation de la structure des territoires chromosomiques suggère que les gènes actifs se trouvent à la surface de régions globulaires formées de chromatine plus condensée [16]. L'organisation spatiale du génome peut servir à rassembler dans une même région du noyau des gènes corégulés, afin d'augmenter la concentration locale des facteurs de régulation, comme c'est le cas, par exemple, pour les ARNt au niveau du nucléole [15, 16]. La membrane nucléaire joue un rôle important dans la localisation des séquences nucléaires [17]. Chez les mammifères comme chez la levure, l'activation ou la répression de certains gènes peut entraîner leurs relocalisations relativement à la membrane nucléaire. Toutefois, il est important de noter que les observations à ce sujet diffèrent entre ces deux types d'organismes [17]. De façon générale, la relocalisation à la membrane semble liée à l'inactivation chez les mammifères et à l'activation chez la levure [18, 19]. Chez la levure, les protéines du pore nucléaire sont associées préférentiellement avec les gènes transcrits, par exemple, le recrutement des gènes GAL au pore nucléaire corrèle avec leur activation [20]. Chez les mammifères, un certain nombre de gènes migrent de la périphérie vers le centre du noyau, consécutivement à leur activation. C'est par exemple le cas de la β -globine lors de la différenciation des cellules érythroïdes de souris [21] ou de *HoxB1* et *HoxB9* dans les embryons de souris (bien que ceux-ci ne soient pas localisés à la membrane nucléaire lors de leur répression) [22]. Chez la levure, les séquences télomériques sont généralement en association avec la membrane nucléaire (chez les eucaryotes supérieurs elle n'est que transitoire, lors de la mitose) [17]. Il y a donc, chez la levure, deux domaines de localisation périnucléaire aux propriétés distinctes. Le premier, en association avec le complexe des pores nucléaires, est en lien avec l'activation des gènes alors que le second, situé entre les pores nucléaires, est associé aux régions réprimées [23]. Les raisons de l'association aux pores nucléaires ne sont pas encore très bien comprises, mais cette localisation pourrait

favoriser l'export de l'ARNm, et a été mise en cause dans la mémoire de la transcription [24].

L'organisation de la chromatine en trois dimensions dans le noyau cellulaire représente une nouvelle frontière dans la compréhension du génome. Chez la levure, on observe une interaction entre le promoteur et le terminateur de certains gènes [25]. Chez les eucaryotes supérieurs, on observe de nombreuses interactions entre activateurs et promoteurs pouvant se faire sur de grandes distances. Une carte des interactions de la chromatine permettrait par exemple d'associer activateurs et promoteurs. Une percée décisive dans cette direction fut réalisée grâce à la mise au point de la technique de capture de conformation de la chromatine (3C) par Job Dekker [26]. Les perfectionnements successifs de cette méthode s'appuient sur l'hybridation de puces à ADN et le séquençage à haut débit pour obtenir une carte non biaisée des interactions de la chromatine [27, 28]. Cette approche a servi pour modéliser la structure en trois dimensions du génome de la levure, révélant la complexité dans l'organisation d'un génome pourtant relativement simple [29].

1.1.2.2 Nucléosomes

Le nucléosome forme l'unité répétitive fondamentale de la chromatine. Chaque nucléosome est constitué de 147 pb enroulées en approximativement 1,7 tour autour d'un octamère d'histones. L'espace entre chaque nucléosome peut varier de 10 à 90 pb en fonction de l'organisme et du locus considéré [30]. Les nucléosomes sont majoritairement incorporés dans la chromatine durant la réplication, lorsque la synthèse des brins complémentaires d'ADN réclame un apport massif. L'intégration des nucléosomes est en partie assumée par le complexe CAF-1 et les chaperonnes d'histones NAP1 [31].

Le positionnement des nucléosomes dans le génome est important pour l'expression des gènes car ils déterminent l'accès des facteurs d'expression au génome. Les premières expériences portant sur le positionnement des nucléosomes sont des expériences de reconstitution *in vitro* de la chromatine et d'étude de la position des nucléosomes à certains promoteurs, notamment lors de l'activation des gènes *PHO5* et *GAL10* de levure [32-35]. Le traitement à la nucléase micrococciale

(MNase), permettant de digérer l'ADN de liaison en laissant intacte la séquence protégée par les nucléosomes, est utilisé pour leurs localisations dans des régions particulières du génome [36-38]. L'avènement et le raffinement des techniques génomiques ont permis de coupler cette méthode aux protocoles d'hybridation de puces ADN à haute densité afin d'obtenir la première carte, bien que partielle, de la position des nucléosomes chez la levure [39]. L'augmentation de la densité des puces et le développement du séquençage à haut débit ont, par la suite, abouti à l'obtention des cartes complètes des positions nucléosomales chez la levure et chez les métazoaires [40-44]. Chez la levure, les nucléosomes localisés aux promoteurs des gènes sont plus précisément positionnés que ceux situés dans les régions codantes. En effet, la majorité des promoteurs de levure sont caractérisés par deux nucléosomes, dénommés -1 et +1, précisément positionnés de part et d'autre d'une région dépourvue de nucléosomes, d'environ 200 pb, communément dénommée en anglais *nucleosome free region* (NFR) [45, 46]. Lorsque les promoteurs des gènes sont alignés par rapport au NFR, le nucléosome +1 apparaît très précisément positionné ; les nucléosomes suivants (+2, +3...) sont eux aussi précisément placés, mais la cohérence s'estompe à mesure de la progression dans le corps du gène. Les régions dépourvues de nucléosomes aux promoteurs des gènes sont enrichies en séquences polyA/polyT, ce qui est cohérent avec leur plus grande rigidité, moins propice à l'enroulement autour des nucléosomes [47]. Les NFR sont aussi le site de fixation de la majorité des facteurs de transcription [39, 48]. Les techniques de génomique ont également été utilisées pour la localisation des modifications et des variantes d'histones qui seront évoquées dans les sections suivantes [42, 45, 49, 50].

Enfin, il a été possible de déterminer la dynamique d'échange des histones aux promoteurs indépendamment de la réplication, grâce à des expériences de remplacement d'histones. Ces expériences, basées sur l'immunoprécipitation d'histones étiquetées, permettent de mettre en évidence les régions du génome les plus dynamiques, c'est-à-dire les régions où les nucléosomes sont les plus instables. Leurs résultats ont identifié les nucléosomes aux promoteurs comme des régions plus instables que les régions codantes [51, 52].

Déterminer la position des nucléosomes dans le génome de la levure a aussi permis de s'intéresser aux principes et aux facteurs qui guident leur positionnement. D'autre part, l'influence des séquences d'ADN a été abordée par la détermination de séquences plus favorables à l'assemblage ou à l'exclusion des nucléosomes, afin de pouvoir modéliser une carte des nucléosomes du génome en dehors de toute intervention extérieure [53-56]. De la comparaison entre les simulations bio-informatiques et l'observation de la position des nucléosomes dans le génome, il résulte que si les séquences sous-jacentes de l'ADN sont importantes, elles ne suffisent pas à expliquer le positionnement des nucléosomes [53-56].

De l'ensemble des études, il ressort des facteurs de localisation des nucléosomes qui se complètent, mais dont l'importance respective reste encore aujourd'hui controversée. Le premier facteur est celui de la séquence : les nucléosomes ne forment pas d'interactions spécifiques avec des bases de l'ADN, mais la capacité de courbure des séquences entraîne des variations dans leur affinité pour les nucléosomes [57]. En conséquence, les nucléosomes tendent à occuper les régions du génome les plus favorables comme la position +1 et à être largement absents des régions moins favorables comme le NFR [53, 55, 58]. Le deuxième facteur est l'arrangement des nucléosomes selon un modèle statistique (*statistical*). Dans ce modèle, la séquence d'ADN détermine le positionnement exact du nucléosome +1 en bordure du NFR ; ce nucléosome établit une barrière contre laquelle les nucléosomes en aval vont se positionner, forçant ainsi leur alignement de manière indépendante de la séquence d'ADN sous-jacente. Toutefois, son influence s'estompe avec la distance [59, 60]. Enfin, le dernier facteur fait intervenir des complexes de remodelage de la chromatine dans le maintien de la position des nucléosomes en condition d'équilibre (les remodeleurs de la chromatine seront abordés plus loin dans ce chapitre)[61]. Pour l'équipe de Pugh, les nucléosomes en aval du +1 sont maintenus alignés en permanence par une activité dépendante de l'ATP et indépendante de la transcription, mais non par un effet statistique [61]. Évidemment, lors de l'activation ou de la répression d'un gène, les remodeleurs de la chromatine sont essentiels à la relocalisation des nucléosomes [62].

1.1.2.3 Histones

Composant unique des nucléosomes, les histones sont de petites protéines très basiques, en contact étroit avec l'ADN et dont les charges positives permettent la compaction de l'ADN en neutralisant les charges négatives du squelette phosphate [57]. Les quatre types d'histones, associées en octamère dans le nucléosome (H2A, H2B, H3 et H4) sont parmi les protéines eucaryotes les plus conservées, toutes formées d'un domaine globulaire central encadré de deux queues. Le domaine globulaire (*histone fold*) structuré est constitué de 3 hélices alpha (H α 1 à H α 3), reliées par deux boucles (L1, L2). Les histones H2A et H3 contiennent une hélice- α supplémentaire à leur extrémité N-terminale (H α N), et les histones H2A et H2B une hélice supplémentaire à leur extrémité C-terminale (H α C) qui constituent toutes des extensions du domaine globulaire [57, 63]. Les queues d'histones sont des domaines non structurés dont la séquence est généralement exclue de la structure cristallographique [63]. De nombreux résidus des histones, particulièrement au niveau de la queue N-terminale, peuvent être modifiés de façon covalente par acétylation des lysines, phosphorylation des sérines, méthylation des lysines et des arginines ou ubiquitination, par exemple. David Allis fut le premier à évoquer l'idée d'un code d'histones permettant d'annoter le génome [64, 65]. Si les queues d'histones incluent la majorité des sites de modification des histones, d'autres sites existent aussi dans le domaine globulaire ; l'un des plus étudiés est la méthylation de la lysine 79 de H3 (H3K79me) [66]. Les techniques ChIP-chip et, plus récemment, de ChIP-seq ont permis de localiser les modifications d'histones dans les génomes eucaryotes [49, 67]. La modification d'histone est l'un des trois mécanismes de régulation utilisés par la cellule pour contrôler l'état de la chromatine, avec le remodelage et l'intégration de variantes d'histones. Chacune de ces interventions sur la chromatine nécessite l'intervention de complexes spécialisés qui seront abordés plus loin.

Extérieure aux nucléosomes, l'histone H1, ou histone de liaison, s'associe chez les métazoaires à l'ADN de liaison, à proximité du site d'entrée et de sortie de l'ADN, mais de manière assymétrique et dans une conformation qui reste encore à

élucider [68, 69]. De plus, son association aux nucléosomes semble plus dynamique que celle de l'octamère d'histones [70]. Elle accompagne une large majorité des nucléosomes chez les mammifères [71]. Elle est depuis longtemps associée à la compaction de la chromatine pour les niveaux supérieurs à la fibre de 10 nm où son rôle a longtemps été vu comme un répresseur général de la transcription [68, 72]. Toutefois, chez la majorité des organismes, sa délétion ou sa déplétion ne semble avoir que peu d'effet sur l'expression des gènes, ce qui interroge son implication dans la compaction de la fibre de 10 nm [68]. Chez les mammifères, l'histone H1, moins conservée que les histones nucléosomales, est formée d'un domaine globulaire encadré par une courte queue en N-terminale et une longue queue en C-terminale [68]. La structure de l'histone H1 de *Sacharomyces cerevisiae*, encodée par le gène *HHO1*, est unique et atypique dans l'évolution. Elle contient un deuxième domaine globulaire séparé du premier par un domaine de liaison homologue à la queue C-terminale des métazoaires mais plus court [73, 74]. Contrairement aux eucaryotes supérieurs, chez la levure, elle n'est pas uniformément présente dans le génome, mais est enrichie aux ARN ribosomiaux (ARNr) où elle a pour fonction de limiter la recombinaison homologue [75, 76].

La structure cristallographique par rayon X du nucléosome à basse résolution (7 Å) par Richmond [77], puis à 2.8 Å par Luger [63, 78] et enfin à 1.9 Å par Davey a permis de cartographier les interactions interprotéiques, et histones-ADN. La structure modulaire du nucléosome révèle l'assemblage séquentiel des histones pour former les nucléosomes (**Figure 1.3**). H3 et H4 s'hétérodimérisent de façon antiparallèle en un motif dit en « poignée de main » (*head to tail*) ayant une conformation en croissant. Deux hétérodimères s'associent pour former un complexe (H3/H4)₂ stable, grâce à l'homodimérisation de H3. Dans le nucléosome, ce tétramère coordonne les 60 pb d'ADN au centre de la séquence (30 bp de chaque côté de la dyade) [63, 78]. Enfin deux dimères H2A/H2B, également agencés en poignée de main, vont s'associer indépendamment de chaque côté du tétramère (H3/H4)₂. L'ajout de ces dimères permet à l'octamère de coordonner 30 pb supplémentaires de chaque côté des 60 bp du tétramère [63, 78]. Enfin, environ 13 pb aux points d'entrée

et de sortie de l'ADN sont pris en charge par une portion de H3, nécessitant la fixation des dimères H2A/H2B afin d'interagir avec l'ADN [63, 78]. L'ajout des dimères H2A/H2B est stabilisé par une hétérodimérisation, reposant sur des interactions hydrophobiques engageant quatre hélices des histones H2B et H4 dans leur domaine C-terminal, et l'établissement de ponts hydrogènes entre le domaine d'amarrage (*docking*) de H2A et les hélices des histones H3 et H4 (**Figure 1.4**). De plus, chez les métazoaires, les deux dimères H2A/H2B interagissent par l'intermédiaire de la boucle 1 de H2A [63, 78]; toutefois, cette interaction est absente chez la levure [79]. L'association des dimères H2A/H2B au nucléosome étant plus dynamique de par leur position, ces dimères peuvent se dissocier du nucléosome laissant le tétramère (H3/H4)₂ en complexe stable avec l'ADN, par exemple durant la transcription ; par contre, il semble difficile de dissocier le tétramère (H3/H4)₂ sans dissocier l'intégralité du nucléosome [80]. Les dimères et les tétramères sont stabilisés par des interactions hydrophobiques [57]. La courbure de l'ADN autour du noyau d'histones est maintenue par 142 ponts hydrogènes, répartis sur 14 points d'interactions distincts, distribués environ tous les 10,5 pb autour de l'octamère d'histones (**Figure 1.4**)[78]. Les contacts entre les histones et l'ADN se font principalement entre la chaîne principale des protéines et le squelette phosphate de l'ADN, au niveau du sillon mineur [78]. Chaque dimère d'histones interagit 3 fois avec l'ADN, deux fois aux boucles L1-L2 situées aux extrémités du dimère et une fois au site $\alpha 1-\alpha 1$ formé par les extrémités N-terminales des hélices $\alpha 1$ (**Figure 1.4**). Ces 12 points de contact structurent 120 pb. Lorsque les dimères H2A/H2B s'assemblent à l'octamère, l'hélice α N-terminale et une partie de la queue N terminale de H3 structurent 13 pb aux points d'entrée et de sortie de l'ADN, constituant, ainsi, les deux derniers points de contact [63, 78]. Cette structure confirme que l'enroulement de l'ADN est largement indépendant de la séquence des bases. Cependant comme mentionné plus haut, la séquence, en agissant de manière subtile sur l'ensemble des contacts protéines/ADN par sa facilité à se courber, peut affecter le positionnement des nucléosomes [57]. On doit noter que l'ADN situé aux points d'entrée et de sortie du nucléosome est plus dynamique et se dissocie naturellement du nucléosome, il « respire » [81]. Ce phénomène peut permettre

l'association de facteurs de transcription ou faciliter le désassemblage du nucléosome [82]

Synthétisées durant la phase S du cycle cellulaire afin de couvrir les besoins dus à l'intégration de nouveaux nucléosomes, les histones canoniques représentent la majorité des histones incorporées durant la réplication de l'ADN, et peuvent être considérées comme les histones par défaut du génome [83]. L'organisation génomique des gènes d'histones varie considérablement d'un organisme à l'autre. Chez la levure, les gènes d'histones sont organisés en paires, transcrits à partir d'un promoteur commun divergent : deux paires H3/H4, et deux paires H2A/H2B. Les deux paires H3/H4 et une des paires H2A/H2B sont exprimées sous le contrôle des protéines Hir (Hir1-3) et Hpc2, grâce à la présence d'un élément de répression en CIS (NEG) présent à leurs promoteurs [83, 84]. En effet, la délétion d'une des 3 protéines Hir empêche la répression hors de la phase S. La seconde paire H2A/H2B est soumise à la même régulation temporelle, mais de manière indépendante des protéines HIR, et selon un mécanisme qui reste à élucider [83].

Chez l'humain, les histones sont exprimées à partir de « *clusters* » contenant plusieurs copies des gènes encodant les 5 histones canoniques (H2A, H2B, H3, H4 et H1), sans ordre établi [85]. Chez l'humain, la similarité entre HIRA et les protéines de *S. cerevisiae* Hir1 et 2 suggère que HIRA assume les fonctions de Hir1 et Hir2. Chez les eucaryotes supérieurs, HIRA est uniquement lié au placement de l'histone H3.3 indépendamment de la phase S [86].

Le niveau des histones canoniques dans la cellule est aussi régulé au niveau de l'ARNm, dont le transcrit n'est pas polyadénylé, afin d'être maximum durant la phase S et rapidement dégradé à la fin de celle-ci. Le niveau est en fait ajusté à la progression du cycle cellulaire et à la réplication afin de prévenir l'accumulation d'histones en cas, par exemple, d'interruption de la synthèse d'ADN [87, 88].

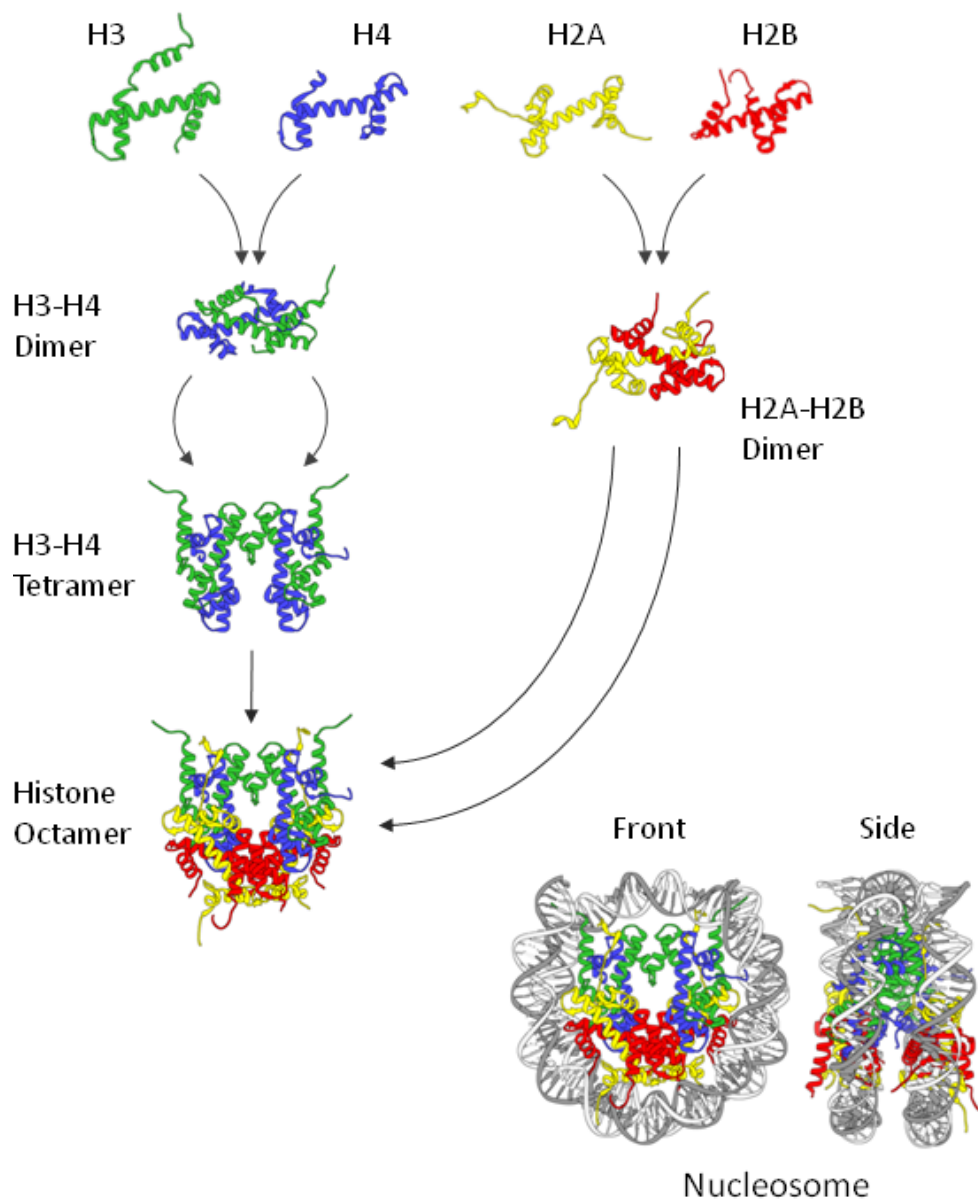


Figure 1.3 : Modèle d'assemblage de l'octamère d'histones

Les histones H3, H4 et H2A, H2B s'hétérodimérisent en un motif dit en « poignée de main » (*head to tail*). Deux dimères H3/H4 se dimérisent en un tétramère auquel 2 dimères H2A/H2B viennent se fixer (un de chaque côté du tétramère) afin de compléter l'octamère d'histones. Le schéma, en bas à droite, illustre l'octamère d'histones (vu de face et latéralement) en complexe avec l'ADN pour former le nucléosome. Structure cristallographique du nucléosome de levure White *et al.*, (2001) [79], visualisation grâce au logiciel Chimera [89].

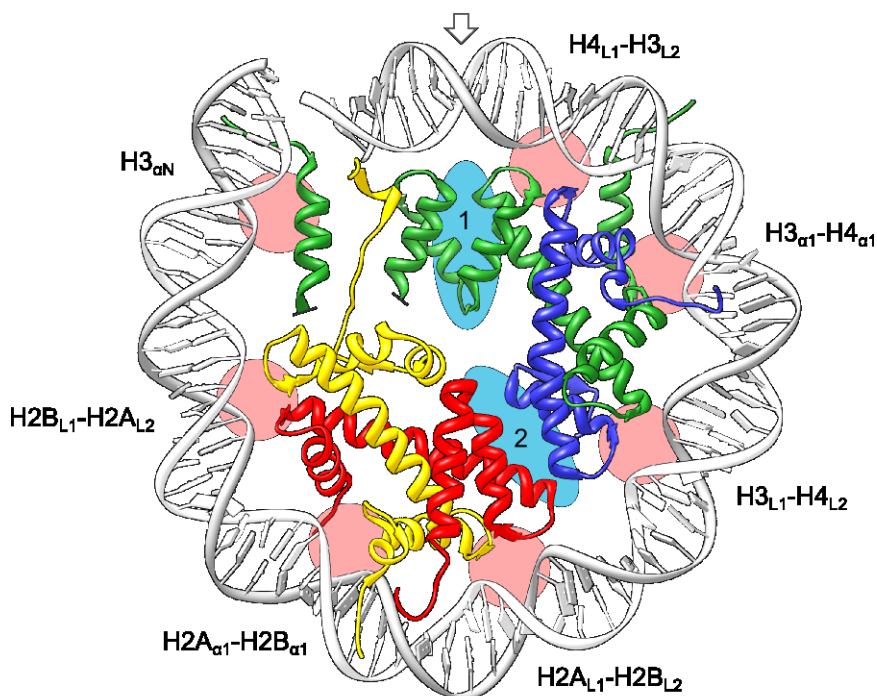


Figure 1.4 : Interactions Protéine-ADN et Protéine-Protéine au sein du nucléosome.

Représentation schématique d'un héminucléosome, la flèche au sommet du nucléosome indiquant la position de la dyade. L'histone H3 est colorée en vert, H4 en bleu, H2A en jaune et H2B en rouge. Les cercles roses localisent 7 des 14 points d'interactions de l'octamère d'histones avec l'ADN, répartis de la dyade au point de sortie de l'ADN. Pour chaque point d'interactions, les structures protéiques associées sont indiquées à l'extérieur du nucléosome. Les ellipses bleues localisent les deux zones d'interactions, formées de quatre hélices, qui sont impliquées dans la cohésion du nucléosome : 1 la zone de dimérisation de H3, stabilisant le tétramère $(H3/H4)_2$, 2 la zone d'interactions entre H4 et H2B, stabilisant l'octamère. Pour des raisons de clarté, les interactions entre le domaine d'amarrage de H2A avec les hélices de H3 et H4 ne sont pas montrées. Contrairement aux métazoaires, chez la levure, les deux histones H2A n'interagissent pas au niveau de la boucle L1. Structure cristallographique du nucléosome de levure White *et al.*, (2001) [79], visualisation grâce au logiciel Chimera [89].

1.1.3 Mécanismes de régulations

Le remodelage de la chromatine n'est pas une propriété intrinsèque aux facteurs de transcription et à l'ARNPolIII, mais nécessite l'action d'enzymes spécialisées afin de réguler l'accès à l'ADN et de vaincre l'obstacle que représente la chromatine durant la transcription. Les organismes eucaryotes ont développé des complexes capables de modifier la structure de la chromatine selon plusieurs modalités : influencer sur le positionnement des nucléosomes, les éjecter, modifier les queues d'histones de manière covalente ou bien échanger les histones canoniques par leurs variantes. Ces complexes interviennent donc dans la majorité des mécanismes du noyau et sont à l'origine de la plasticité de la chromatine. La levure, bien que pourvue d'un génome de petite taille et d'une structure chromatinienne moins complexe que les eucaryotes supérieurs, est un outil important dans notre compréhension des mécanismes de régulation de la chromatine. En effet, chez les organismes multicellulaires, les modifications et les variantes d'histones plus nombreuses rendent la chromatine plus complexe ; toutefois la structure et les principes de régulation restent similaires.

1.1.3.1 Positionnement des nucléosomes

Les eucaryotes ont développé des complexes remodeleurs de la chromatine leur permettant de déplacer les nucléosomes, de les éjecter ou de modifier leur composition en utilisant l'énergie d'hydrolyse de l'ATP pour surmonter la barrière thermodynamique [90, 91]. Les chaperonnes d'histones vont quant à elles s'associer aux histones ou aux dimères d'histones afin notamment de les séquestrer, prévenir les interactions histones-ADN sans formation de tétramères ou d'octamères stables ; elles vont ainsi prendre en charge les histones avant de les déposer à l'ADN ou de les « remettre » aux remodeleurs de la chromatine [92]. La présence de remodeleurs de la chromatine chez la levure témoigne de la nécessité de pouvoir modifier la chromatine dès son apparition dans l'évolution. Ils jouent un rôle dans tous les aspects de la chromatine. Ils interviennent par exemple dans la réplication, le maintien de la

structure chromatinienne aux promoteurs durant l'activation ou la répression des gènes, au cours de la transcription et durant la réparation de l'ADN [61, 90].

Les remodeleurs de la chromatine sont de larges complexes (>1 MDa), formés de 9 à 17 sous-unités, capables de déplacer les nucléosomes en faisant glisser l'ADN autour de l'octamère d'histones [91]. Ils sont recrutés, retenus ou régulés à la chromatine par des modifications d'histones et travaillent donc de concert avec les complexes de modification de la chromatine. À cet effet, ils contiennent des sous-unités avec des bromodomaines reconnaissant les marques d'acétylation et des chromodomaines reconnaissant les lysines méthylées [90].

Le mécanisme d'action des remodeleurs de la chromatine diverge d'une famille à l'autre, on peut cependant en tracer les grandes lignes. Les remodeleurs de la chromatine s'ancrent à l'octamère d'histones, le domaine ATPase/translocase fixe l'ADN afin de « tirer » sur l'ADN de la séquence de liaison pour générer une boucle d'ADN. Cette boucle sera propagée unidirectionnellement en direction de la dyade et autour du nucléosome, brisant les contacts en amont afin de les reformer en aval [93]. Si la plupart des remodeleurs de la chromatine semblent fonctionner sur ce principe, on peut s'interroger sur les mécanismes qui permettent d'aboutir aux différentes activités (déplacement, éjection, échange). Pour faire glisser un nucléosome, la taille de la boucle n'a pas besoin d'être importante, mais la structure du nucléosome suggère différemment pour le remplacement et l'éjection. La taille de la boucle semble donc déterminante pour permettre aux chaperonnes ou à d'autres protéines de venir prendre en charge les dimères désappariés [93]. L'espacement des nucléosomes suggère une autre possibilité : le remodelage d'un nucléosome pourrait favoriser l'éjection d'un dimère ou de l'octamère du nucléosome voisin en « débobinant » (*spool*) l'ADN de celui-ci [93].

Les domaines ATPase des remodeleurs de la chromatine, appartenant tous à la superfamille II (SFII), sont apparentés aux hélicases et composés de deux sous-domaines, DExx et HELICc, séparés par un domaine de liaison. La classification des remodeleurs de la chromatine est basée sur la présence de domaines spécifiques adjacents ou internes au domaine ATPase [94]. On dénombre 4 familles principales de remodeleurs de la chromatine, toutes très conservées chez les eucaryotes. SWI-

SNF (*switching defective/sucrose nonfermenting*), ISWI (*imitation switch*), CHD (*chromodomain, helicase, DNA binding*), and INO80 (*inositol requiring 80*) [90].

Le domaine ATPase de la famille SWI/SNF est caractérisé par la présence d'un domaine HSA (*helicase-SANT*) et d'un bromodomaine C-terminal [90]. Les membres de la famille SWI/SNF font glisser ou éjectent les nucléosomes, mais ne sont pas impliqués dans l'assemblage de la chromatine lors de la réplication [90, 93]. Sur un chapelet de nucléosomes également espacés, SWI/SNF réarrange les nucléosomes aléatoirement [95].

La sous-unité ATPase de la famille ISWI est caractérisée par la présence d'un domaine SANT et d'un domaine SLIDE [90]. Chez la levure, cette famille contient 3 membres ISW1a, ISW1b et ISW2. Les complexes de la famille ISWI altèrent la position des nucléosomes mais ne peuvent pas les éjecter [93]. Le complexe ACF, présent chez l'humain et la drosophile, peut optimiser l'espacement des nucléosomes, facilitant ainsi l'assemblage de la chromatine après le passage de la machinerie de réplication [96, 97]. Isw2 est recruté par certains facteurs de transcription aux promoteurs de gènes pour lesquels il joue un rôle important dans la répression en limitant l'accès à l'ADN [98].

La famille CHD est caractérisée par la présence de deux chromodomains en tandem à l'extrémité N-terminale de la sous-unité ATPase et ne contient, chez la levure, qu'un seul membre, le complexe CHD1 [90]. Certains membres de la famille CHD font glisser ou éjectent les nucléosomes pour favoriser la transcription, alors que d'autres sont capables de réprimer la transcription grâce à la présence de sous-unités déacétylases [90].

La famille INO80 est caractérisée par une large insertion au centre du domaine ATPase [90]. Cette famille comprend, chez la levure, les complexes INO80 et SWR1 contenant 15 et 14 sous-unités respectivement, dont certaines sont communes aux deux complexes [99]. Le complexe INO80 fut identifié par sa capacité à réguler les gènes de réponse à l'inositol [100]. INO80 et SWR1 sont tous les deux recrutés aux sites de cassure double brin de l'ADN sans pour autant y assumer la même fonction, puisque SWR1 semble faciliter la réparation de l'ADN et INO80 faciliter l'éviction des histones [101, 102]. Bien que SWR1 soit connu pour

être le complexe responsable de l'échange de la variante d'histone H2A.Z, INO80 et SWR1 semblent tous les deux pouvoir catalyser l'échange de la variante d'histone H2A.Z ; leurs spécificités seront abordées plus en détail dans la suite de l'introduction [103].

1.1.3.2 Modifications covalentes des histones

Les queues d'histones, particulièrement les queues N-terminales, non structurées et capables de se projeter en dehors du nucléosome, sont les cibles majeures des complexes de modification des histones et peuvent être par exemple acétylées, méthylées, phosphorylées, ubiquitinyllées et sumoylées (**Table 1.1**) [104]. Parmi ces modifications, l'acétylation et la méthylation sont les plus nombreuses ; elles seront donc abordées plus en détail dans cette section. Les modifications covalentes des queues d'histones peuvent agir de deux manières non exclusives : en modifiant les propriétés physiques des nucléosomes ou en annotant la chromatine de manière à recruter ou affecter l'activité de certains complexes. La variété et la complexité des patrons de modification des histones ont donné lieu à la théorie de « code d'histones » énoncée pour la première fois par le groupe de David Allis [64, 65]. Les marques s'influencent les une les autres (*crosstalk*), la déposition d'une ou plusieurs marques pouvant en réguler d'autres [105, 106]. La localisation de nombreuses marques a permis leur association avec certains états de la chromatine. Ces marques ont été utilisées par les chercheurs afin d'annoter le génome dans un type cellulaire pour des conditions précises leur permettant d'obtenir une « photo » de l'état de transcription de chacune des zones du génome. Par exemple, la marque de méthylation H3K36 est généralement associée à l'activation des gènes (présente ou récente), H3K4me3 à l'initiation de la transcription, H3K9me à l'hétérochromatine (**Table 1.1**) [45, 67, 107-109]. En 2007, une nouvelle nomenclature d'une partie des enzymes de modification des histones fut proposée. Cette démarche ayant pour but de clarifier les inconsistances de nomenclature, dues à l'essor rapide des connaissances dans le domaine, prend aussi en compte l'activité de ces enzymes sur des substrats autres que les histones [110].

Table 1.1 : Liste, non exhaustive, des modifications d'histones.

Modification	Histone	Résidu	Complexe(s) de déposition		Fonction(s)	
			levure	mammifère		
méthylation	H3	K4	Set1	MLL, ALL-1, Set9/7, ALR-1/2, ALR, Set1	Transcription	
		K9		Suv39h, G9a, Eu- HMTase I, ESET, SETBD1	Répression	
		K27		E(Z)	Répression	
		K36	Set2	HYPB, Smyd2, NSD1	Transcription	
		K79	Dot1	Dot1L	Transcription, Maintien de l'euchromatine	
		R2		CARM1	Transcription	
		R17		CARM1	Transcription	
		R26		CARM1	Transcription	
		H4	K20		PR-Set7, SET8 SUV4-20H1	Répression Réparation de l'ADN
		R3			PRMT1	Transcription
Acétylation	H3	K56	Rtt109	P300	Réplication Réparation de l'ADN	
		H4	K16	Sas2, NuA4	hMOF	Maintien de l'euchromatine, Réparation de l'ADN
		H2A.Z	K14	NuA4		Transcription
Phosphorylation	H3	S10	Snf1	Aurora B	Transcription, Mitose	
		H2A.X	S129/139	Mec1 Tel1	ATR ATM	Réparation de l'ADN
Ubiquitination	H2B	K120/123	Rad6, Bre1	UbcH6	Transcription	
		H2A	K119		Ring1b	Répression

Bien que la table n'indique qu'une ou deux fonctions, plusieurs de ces marques ont pu être reliées à de nombreux mécanismes nucléaires.

1.1.3.2.1 Acétylation

Les histones acétyltransférases (HAT) ou lysine acétyltransférases (KAT) catalysent l'ajout du groupe acétyle de l'acétyle CoA aux groupes ϵ -amino des résidus lysine des histones. Cette modification neutralise la charge positive des lysines, affaiblissant leur affinité pour l'ADN. L'acétylation des histones génère aussi une marque qui pourra être reconnue de manière spécifique par des complexes effecteurs [111]. On distingue deux groupes d'HAT selon leur localisation cellulaire et la nature de leur substrat. Les membres du groupe A acétylent les histones à la chromatine, alors que ceux du groupe B acétylent les histones libres dans le cytoplasme, avant leur incorporation. Le groupe B ne contient qu'une HAT, Hat1 (KAT1), alors que le groupe A est subdivisé en familles, dont GNAT (*Gcn5-related N-acetyltransferase*) et MYST (*MOZ*, *Ybf2/Sas3*, *Sas2*, *Tip60*) sont les mieux caractérisées [112]. Le groupe B est responsable de l'acétylation des histones avant leur incorporation à la chromatine par le complexe CAF1 [31]. Ces marques sont effacées peu de temps après la déposition des histones à la chromatine, afin de permettre la mise en place de nouveaux profils [113]. Si les HAT du groupe B ont une activité cytoplasmique caractéristique, elles ne sont cependant pas restreinte au cytoplasme des gènes [114].

Les groupements acétyles sont retirés par les histones déacétylases (HDAC) (non concernées par la nouvelle nomenclature). Les HDAC sont réparties en quatre classes (I à IV) suivant leur homologie [112]. Les membres des classes I, II et IV, catalysent la déacétylation selon un mécanisme dépendant d'un ion Zinc, alors que les membres de la classe III (famille des *sirtuin*), utilisent un mécanisme dépendant du NAD⁺ [112]. On peut noter que, chez la levure, Hda1 et Hos3, les seuls membres des familles II et III peuvent être regroupées au sein de la famille II, faisant ainsi descendre le nombre de classes à trois [115].

Les activités antagonistes des HAT et HDAC font de l'acétylation une marque dynamique, dont le niveau repose sur la balance entre ces deux activités. On observe dans le génome deux types d'activité HAT et HDAC : une activité non ciblée,

responsable du niveau global d'acétylation du génome, et une activité ciblée pour laquelle les HAT sont recrutées en des points précis du génome [116]. Les activateurs recrutent les HAT aux promoteurs des gènes, lors de leur induction ; l'acétylation des nucléosomes favorise l'ouverture du promoteur, l'éjection des chromosomes, et marque les promoteurs [117]. Chez la levure, NuA4 et SAGA, deux complexes contenant une activité acétyltransférase (par leur sous-unité Esa1 et Gcn5 respectivement) semblent être les principales HAT participant à l'activation des gènes [118-122]. Un sous-complexe de NuA4, picNuA4 (*piccolo NuA4*), formé des sous-unités Esa1, Epl1 et Yng2 est, quant à lui impliqué dans l'acétylation globale de la chromatine [123].

À l'inverse, les HDAC sont généralement recrutés aux promoteurs lors de la répression des gènes [117]. Lors de l'élongation, les histones sont acétylées de façon transitoire pour permettre le passage de la polymérase, puis déacétylées par la HDAC Rpd3, afin de reconstituer la structure de la chromatine dans le sillage de l'ARNPolIII [124]. Les protéines Sir (*silent information regulator*), incluant la protéine HDAC Sir2, sont essentielles à la répression épigénétique des régions silencieuses par déacétylation de la marque d'histone H4K16 déposée par les HAT Sas2 (*something about silencing 2*) et Esa1 (*essential Sas2-related acetyltransferase 1*) [117].

1.1.3.2.2 Méthylation

La méthylation est une marque, somme toute, plus stable déposée sur les résidus lysines (lysine méthyl transférase KMT) ou arginines. Contrairement à l'acétylation, elle ne modifie pas la charge des histones et se retrouve sous trois formes (lysine mono-, di- et triméthylée) dont les conséquences peuvent être distinctes. La plus grande demi-vie des marques de méthylation est utile pour annoter des régions du génome plus stables et pour garder une mémoire des événements qui ont abouti à leur déposition [125].

L'ajout de groupes méthyles est catalysé par des méthyltransférases et leur retrait par des déméthylases. Chez la levure, trois histones lysine méthyltransférases sont présentes, Set1(KMT2) et Set2 (KMT3), caractérisées par un domaine SET, et Dot1 (*disruptor of telomeric silencing*, KMT4). Ces trois enzymes catalysent la

mono-, la di- et la triméthylation de leur substrat respectif. Étant impliquées dans la transcription, leur fonction fera l'objet d'une section ultérieure de l'introduction. On peut cependant noter ici la contribution de Dot1 dans la restriction de la propagation des régions silencieuses de la chromatine. La répression épigénétique des régions silencieuses se propage le long de la chromatine à partir d'une zone de nucléation ; la méthylation du résidu H3K79, en empêchant la fixation de la protéine Sir3, prévient l'action du complexe Sir (notamment la déacétylation par Sir2) et fonctionne comme une barrière à l'étalement des régions silencieuses [66].

Bien que considérée comme plus stable que l'acétylation, la méthylation est réversible. Toutefois les protéines aux activités histone déméthylases n'ont été identifiées que récemment, d'abord chez l'humain [126], puis chez la levure [127]. Les dernières études sur cette classe d'enzymes les relient au développement et aux maladies [128].

1.1.3.3 Variantes d'histones

Au cours de l'évolution, les eucaryotes ont diversifié le répertoire des histones incorporées dans les nucléosomes en acquérant de nouvelles variantes aux propriétés spécifiques. Les variantes d'histones sont des formes non alléliques des histones canoniques. Elles remplacent les histones canoniques dans une partie des nucléosomes, modifiant ainsi leurs propriétés sans toutefois altérer de façon drastique la structure du nucléosome. Contrairement aux histones canoniques, leurs ARNm sont généralement polyadénylés et exprimés tout au long du cycle cellulaire. La conservation des variantes d'histones, supérieure à celle de leur contrepartie canonique, confirme que des propriétés spécifiques leur sont attribuées [129]. Toutes les histones n'ont pas le même nombre de variantes, H2A et H3 sont les histones canoniques avec le plus grand nombre de variantes alors que H4 et H2B n'ont pas ou peu de variantes. Parmi les variantes d'histones, celles apparues précocement dans l'évolution des eucaryotes (H2A.Z), ou bien ayant un lien avec la transcription (H2A.Z, H3.3) ou la réparation des dommages à l'ADN (H2A.X) sont les plus étudiées. H2A.Z sera traité plus en détail dans l'introduction et constitue le sujet du chapitre 3 de cette thèse.

H3.3 ne diffère de la forme canonique que par quatre acides aminés, soit chez l'humain, les résidus A31S de la queue N-terminale et les résidus A87S, G90M et S96C de la seconde hélice du domaine globulaire. L'apparition de la variante d'histone H3.3 s'est produite plusieurs fois en des événements distincts au cours de l'évolution [130]. Une grande partie des études sur cette variante ont été réalisées chez la drosophile. H3.3 est incorporée de manière dépendante de la transcription et impliquée dans la condensation des chromosomes sexuels avant la méiose [131, 132].

H2A.X, quant à elle, joue un rôle dans la réparation des cassures double brin de l'ADN. Chez la levure, où peu de variantes ont émergé, la fonction de l'histone H2A.X est occupée par l'histone canonique H2A [133]. En effet, celle-ci possède la séquence SQEL, spécifique de la variante d'histone H2A.X, localisée dans la queue C-terminale de la protéine (**Figure 1.5**). Lorsque le résidu serine est phosphorylé, l'histone est appelée P-H2A ou gamma H2A [133].

1.1.3.3.1 H2A.Z

La majorité des variantes d'histones sont des variations de l'histone canonique H2A qui compte cinq sous-types : H2A, H2A.X, H2A.Z, macroH2A et H2ABbd [130]. Cette abondance de variantes reflète, sans doute, la position unique occupée par H2A dans le nucléosome. En effet, H2A intervient dans la cohésion de l'octamère et participe à une large portion de la surface du nucléosome au niveau de la zone acide (*acid patch*) [63]. H2A.Z est certainement l'une des variantes d'histones dont la structure et les fonctions sont les plus étudiées. Apparue très tôt dans l'évolution des eucaryotes, elle est déjà présente chez la levure *S. cerevisiae*. Bien que non essentielle chez cette dernière, la perte de H2A.Z est associée à une instabilité chromosomique et à un défaut d'expression [134, 135]. H2A.Z est essentielle chez une large majorité des organismes eucaryotes englobant tetrahymena, le xénope, la drosophile et la souris [136-139].

H2A.Z partage environ 60 % d'homologie avec sa contrepartie canonique, mais est conservée à 90 % entre la levure et l'humain. Elle est donc plus conservée entre espèces que l'histone canonique H2A [130]. Très tôt, lors de sa découverte, H2A.Z fut associée à la transcription puisque les premières études chez le protozoaire

cilié *Tetrahymena thermophile* font avis de sa présence dans le macronucleus transcriptionnellement actif, mais de son absence dans le micronucleus associé aux fonctions reproductrices, transcriptionnellement inactif [140]. Cependant, ce n'est que 20 ans plus tard que son implication dans la transcription fut démontrée [134, 135]

1.1.3.3.1.1 H2A.Z et structure du nucléosome

Si les gènes encodant les histones H2A et H2A.Z ont divergé très tôt au cours de l'évolution, la divergence n'est pas uniforme le long des gènes. En effet, l'alignement des séquences d'acides aminés met en évidence cinq régions de divergence considérable incluant les deux queues, moins soumises aux contraintes structurelles du nucléosome, et trois régions situées dans la partie globulaire de l'histone (**Figure 1.6**). La présence de régions plus densément divergentes suggère l'idée d'une variante H2A.Z apparue afin d'assumer des fonctions spécifiques, dépendantes de ces régions. H2A et H2A.Z ont la même organisation globale : un corps globulaire formé de 5 hélices, αN , $\alpha 1-3$ et αC , encadré par deux queues largement non structurées une N-terminale et une C-terminale [141]. Les régions divergentes du domaine globulaire sont localisées au niveau de la boucle 1, de la boucle 2 et du domaine d'amarrage (**Figure 1.6**). À cause de la grande similarité de structure entre H2A et H2A.Z et malgré la présence de régions divergentes, H2A.Z peut se substituer à H2A dans le nucléosome, sans perturber l'arrangement de l'octamère et sans modifier l'enroulement de l'ADN autour du noyau d'histones (**Figure 1.6**) [142].

L'échange aboutit à de subtiles modifications d'interactions, révélées par la reconstitution et l'analyse de la structure cristallographique d'un nucléosome (métazoaire) dans le laboratoire de Luger [142]. Chez les métazoaires, les deux dimères H2A/H2B interagissent uniquement au niveau de la boucle 1. Les différences de topologie de la boucle 1 entre H2A et H2A.Z modifient fondamentalement les interactions entre les deux dimères H2A/H2B du nucléosome [142]. Chez les métazoaires, la structure de la boucle 1 semble à première vue défavorable à l'existence de nucléosomes hétérogènes H2A/H2A.Z à cause de l'encombrement

stérique créé dans cette région. Toutefois, advenant l'accommodation de cette structure dans le nucléosome par un réarrangement des chaînes secondaires, aucune interaction ne stabiliserait les deux dimères H2A/H2B [142]. Pourtant l'existence de nucléosomes hétérogènes a été démontrée *in vitro* et *in vivo*, aussi bien chez les métazoaires que chez la levure [143-146]. Malgré les changements au niveau de la boucle 1, les interactions entre protéines (H2AL1-H2BL2) et l'ADN ne sont pas significativement affectées (**Figure 1.6**) [142]

Les changements au niveau de la zone acide (*acid patch*) sont particulièrement intéressants ; en effet, cette zone forme une partie de la surface du nucléosome, idéalement placée pour moduler ses interactions protéiques. La zone acide de H2A est formée, chez les métazoaires, par 6 acides aminés : Glu 56, Glu 61, Glu 64, Asp 90, Glu 91, Glu 92. (Correspondant aux résidus de levure : Glu 57, Glu 62, Glu 65, Asp 91, Asp 92, Glu 93). Ces résidus sont conservés dans H2A.Z. Chez la levure, les 3 derniers résidus formant le motif DDE sont situés immédiatement à la sortie de l'hélice $\alpha 3$ (**Figure 1.6**) [142, 147]. Le remplacement de l'asparagine 94 (95 chez la levure) par un acide aspartique et le remplacement du résidu basique de la lysine 95 (96 chez la levure) par une sérine aboutissent à l'extension de la zone acide dans un nucléosome incorporant H2A.Z. L'absence du résidu adjacent au résidu 100 de H2A.Z (105 chez la levure) altère encore davantage la surface du nucléosome contenant la région de la zone acide, laquelle joue un rôle important dans la déposition et les fonctions de H2A.Z [142, 147].

Au niveau du domaine d'amarrage (*docking*), la substitution du résidu Glu104 d'H2A par le résidu Gly106 de H2A.Z aboutit à la déstabilisation de l'interaction entre H2A.Z et H3, par la perte de trois ponts hydrogènes. L'histidine 112, absente de H2A, en collaboration avec l'histidine 114, peut coordonner un ion métallique ; cependant, il n'y a pas de preuve de sa présence *in vivo* [142].

La structure du nucléosome de levure incorporant la variante H2A.Z n'a pas encore été publiée. Il est possible qu'elle réserve quelques surprises comme cela a été le cas pour la structure du nucléosome canonique de levure [79].

Les conséquences de l'incorporation de H2A.Z sur la stabilité des nucléosomes ont généré des résultats contradictoires. Bien qu'il soit pratiquement impossible de prédire la stabilité du nucléosome par sa structure cristallographique, le remplacement de H2A par la variante H2A.Z aboutit à la subtile déstabilisation de l'interaction entre le tétramère (H3/H4)₂ et les dimères (H2A/H2B) [142]. Toutefois, des expériences testant la stabilité des nucléosomes en condition croissante de sels semblent indiquer l'inverse [148]. Enfin lorsque la stabilité des nucléosomes est testée par la méthode de transfert d'énergie de résonance (*fluorescence resonance energy transfer : FRET*) les conclusions semblent confirmer celles basées sur la structure cristallographique [149]. Les observations contradictoires sont probablement le résultat de différences méthodologiques et il est important de noter que les modifications post-traductionnelles peuvent moduler la stabilité des nucléosomes comme le suggère Thambirajah *et al.*, (2006) [150].

Deux études réalisées chez des vertébrés se sont intéressées à la localisation et à la stabilité des nucléosomes incorporant simultanément les deux variantes, H3.3 et H2A.Z [151, 152]. Ces nucléosomes existent et sont localisés au promoteur des gènes très transcrits, et aux régions régulatrices [151]. Les études de stabilité montrent qu'ils sont moins stables que ceux avec une seule variante. La deuxième étude a déterminé la répartition génomique des nucléosomes H3.3/H2A.Z ; elle les positionne sur le NFR des gènes, postulant que les protocoles précédents ont abouti à la perte de ces nucléosomes extrêmement instables [152].

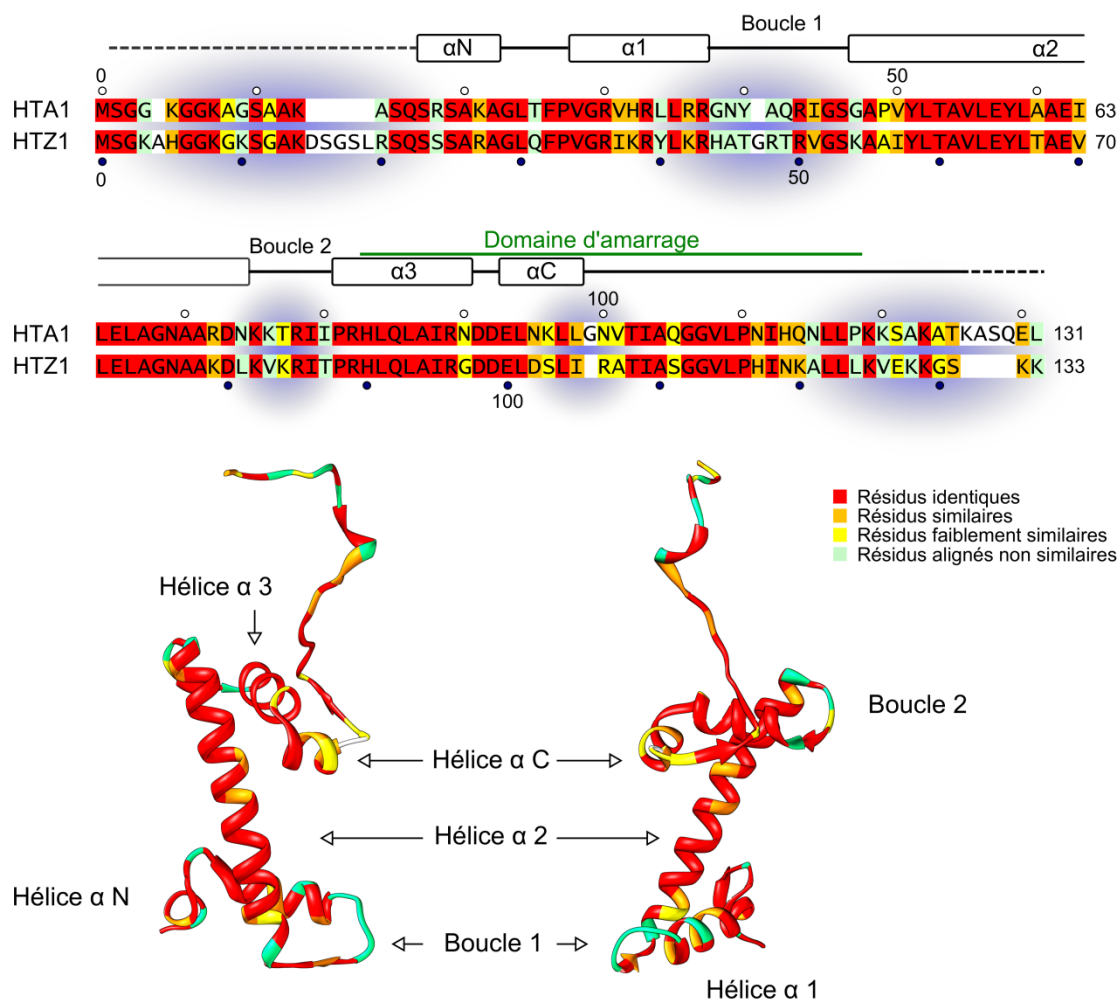


Figure 1.5 : Comparaison des séquences primaires d'H2A et H2A.Z

Panneau supérieur : alignement des séquences de H2A (HTA1) et H2A.Z (HTZ1) de levure. Alignement réalisé grâce au serveur T-coffee [153]. Les acides aminés sont colorés sur la base de la similarité de leurs propriétés entre les deux histones. Les hélices α de la structure secondaire sont indiquées par les rectangles au-dessus de l'alignement et les lignes pointillées indiquent les régions non structurées et donc absentes de la structure cristallographique. Les halos bleus mettent en évidence les segments de divergence principale (incluant les queues, aux extrémités N- et C-terminales). Panneau inférieur : le même code de couleur est appliqué à la structure de H2A, isolée au sein du nucléosome de levure, afin de mieux visualiser la localisation des régions divergentes. Structure cristallographique White *et al.*, (2001) [79] visualisation par le logiciel Chimera [89].

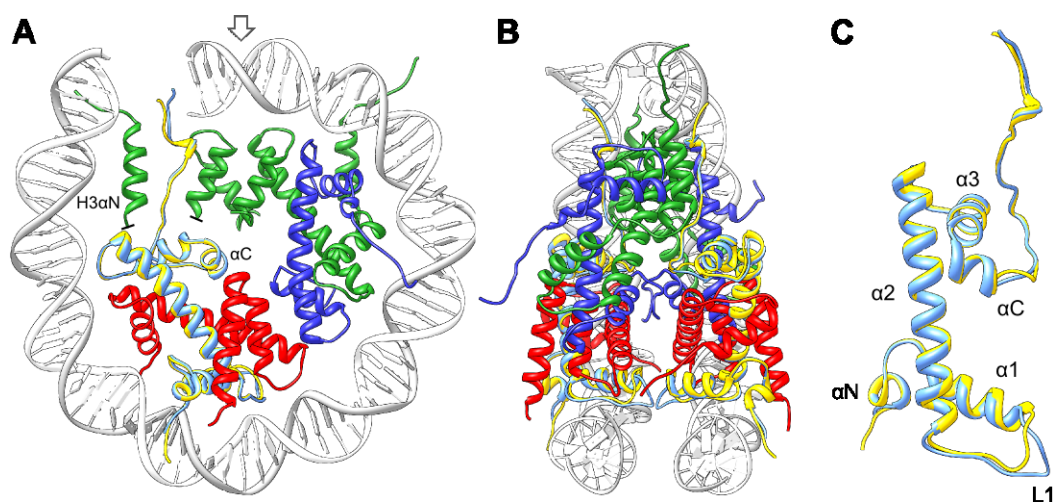


Figure 1.6 : Superposition des structures du nucléosome canonique et du nucléosome incorporant H2A.Z.

Figure adaptée de Suto *et al.*, (2000) [142].

A) Un heminucléosome est visualisé de façon similaire à la **Figure 1.4** avec superposition des structures cristallographiques de H2A et H2A.Z. H3 est colorée en bleu, H4 en vert, H2B en rouge, H2A et H2A.Z sont superposées en jaune et cyan respectivement. **B)** Vue latérale du nucléosome (rotation de 90° sur l'axe de la dyade par rapport à A). L'ADN bloquant la vue des histones a été enlevé pour plus de clarté **C)** Superposition de la structure de H2A et de H2A.Z dans le nucléosome.

1.1.3.3.1.2 Localisation et fonctions de H2A.Z

H2A.Z remplace H2A dans 5 % des nucléosomes chez la levure [154]. Malgré sa présence modeste dans le génome, H2A.Z assume des fonctions souvent essentielles.

Les premiers travaux impliquant H2A.Z dans l'activation des gènes ont montré qu'aux promoteurs des gènes *GAL1* et *PHO5*, H2A.Z est présente en condition répressive, mais disparaît lors de l'activation de ces gènes [134, 135]. Adam *et al.* (2001) [134] ont démontré, par l'utilisation de protéines chimériques contenant le domaine globulaire de H2A.Z et le domaine C-terminal de H2A ou vice-versa, que la région nécessaire à la fonction de H2A.Z au promoteur de GAL résidait dans la partie C-terminale de la protéine englobant le domaine d'amarrage. La région correspondante, chez la drosophile, a été identifiée comme essentielle à la fonction de H2A.Z [137]. Une délétion de *HTZ1* engendre une diminution du recrutement de la polymérase au promoteur du gène *GAL10* [134]. Cette conclusion fut plus tard généralisée à l'échelle génomique chez l'humain [108].

En 2003, une étude des niveaux d'expression par puce ADN, réalisée par le groupe de Madhani, a identifié approximativement 200 gènes dont l'expression est diminuée dans une souche *htz1Δ* [155]. Remarquablement, ces gènes sont enrichis pour les régions sous-téломériques, en bordure de régions d'hétérochromatine dites silencieuses, c'est-à-dire réprimées épigénétiquement par l'action des protéines Sir [155]. Cette étude démontre que H2A.Z antagonise la propagation des protéines Sir (Sir2-4), responsables de la déacétylation et de la formation de l'hétérochromatine au niveau des régions silencieuses. Comme anticipé par ces travaux, le défaut d'expression des gènes peut être restauré, dans la souche *htz1Δ*, par la délétion du gène *SIR2* (HDAC) [155]. Un effet similaire de H2A.Z en bordure du HMR, lui aussi réprimé par les protéines Sir, avait précédemment été observé par Dhillon *et al.* (2000) [156]. Se basant sur la localisation des gènes affectés, le groupe de Madhani définit les domaines chromosomiques dans lesquels l'expression adéquate des gènes est dépendante de H2A.Z, et qui sont nommés HZAD (*Htz1-activated domains*) [155].

En 2005, la localisation génomique de l'histone H2A.Z, chez la levure *S. cerevisiae*, effectuée indépendamment par 5 groupes, a relancé l'intérêt des chercheurs pour cette variante, suggérant un rôle général dans la transcription plus en adéquation avec les résultats obtenus aux gènes *GAL* et à *PHO5* [50, 157-160]. L'article par Guillemette *et al.* (2005) [157] fait l'objet du chapitre 3 de cette thèse. Contrairement à ce que l'on pouvait attendre après l'étude de Madhani, la localisation de H2A.Z est pangénomique, comme remarqué chez la drosophile [161]. Mais surtout la variante est restreinte, pour une grande majorité des gènes, à une zone d'un à deux nucléosomes bordant le NFR (nucléosome -1 et +1) [50, 157]. Sa localisation coïncide donc avec celle des nucléosomes les plus précisément positionnés et les plus dynamiques du génome [51]. Remarquablement, l'enrichissement de H2A.Z aux promoteurs est anticorrélé à l'activité des gènes [157-160]. Cependant, comme observé pour les gènes *GAL*, H2A.Z ne sert pas à l'inactivation des gènes, mais maintient les promoteurs dans un état propice à leur activation [134, 157, 158]. La présence de H2A.Z au promoteur semble aussi réguler l'organisation des nucléosomes en bordure du NFR [157]. Notre groupe s'est aussi intéressé à la localisation de H2A.Z dans les domaines HZAD. Dans ces domaines, le profil de déposition de H2A.Z diffère considérablement des pics étroits observés aux promoteurs, couvrant une zone plus importante, qui pénètre le corps des gènes [157]. Ces profils distincts semblent confirmer les rôles différents de H2A.Z dans l'activation des gènes et la prévention de la propagation de la structure chromatinienne des régions silencieuses.

La localisation aux promoteurs des gènes est conservée chez les métazoaires, [40, 41, 108, 162] mais se limite au nucléosome +1 chez la drosophile et *S. pombe* [41, 163]. Chez les métazoaires, H2A.Z est également présente aux éléments de régulations (tel qu'indiqué par la marque d'histone H3K4) [42, 108, 162]. Contrairement à la levure, chez les eucaryotes supérieurs, l'enrichissement de H2A.Z est corrélé à la présence de l'ARNPolIII [41, 42, 108]. H2A.Z est recrutée de manière dynamique au promoteur des gènes, où elle sert à favoriser le recrutement de la polymérase [108]. Son rôle chez les métazoaires et chez la levure serait donc similaire, bien que son recrutement chez les métazoaires soit dépendant de

l'activation (mais pas de la transcription) [108]. Chez l'humain, la forme hypoacétylée de H2A.Z s'accumule dans les régions d'hétérochromatine (indiqué par la marque d'histone H3K9me2). Toutefois sa présence et sa fonction dans ces régions restent à expliquer [164]. Cette observation est à rapprocher de celle faite dans le même papier sur la relative déplétion de H2A.Z dans les régions transcrites [108]. La localisation complexe de H2A.Z peut donc résulter de 3 mécanismes simultanément en action sur le génome : un mécanisme de déposition aléatoire, un mécanisme de déposition ciblée (par exemple au promoteur des gènes) et un mécanisme d'éviction lié à la transcription [164]. L'ubiquitination de H2A.Z à l'hétérochromatine est aussi étudiée, mais peu de choses sont encore connues [165, 166]. H2A.Z est ubiquitinée, chez l'humain, au résidu lysine 120 faisant partie de la queue C-terminale, cette marque est enrichie à l'hétérochromatine facultative des régions silencieuses comme le chromosome X-inactif [165].

Enfin, il est intéressant de penser que H2A.Z est liée à la mémoire de l'activation des gènes. Récemment, il a été démontré que H2A.Z était nécessaire au maintien des gènes activés à la membrane nucléaire, mais non à leur translocation, chez *S. cerevisiae* [167, 168]. La rétention à la membrane nucléaire fait certainement partie des moyens dont la cellule dispose pour garder une mémoire de la transcription permettant ainsi la réinduction rapide d'un gène récemment transcrit. Ce mécanisme est important dans l'adaptation aux conditions changeantes du milieu. Il est aussi intéressant de noter que le bouclage des gènes (*DNA looping*) est lui aussi probablement associé à la mémoire transcriptionnelle. Les travaux sur la mémoire transcriptionnelle et le bouclage des gènes en relation avec H2A.Z seront revus plus en détail dans la discussion de cette thèse.

1.1.3.3.1.3 Déposition à la chromatine et ciblage génomique

La déposition des variantes d'histones de manière indépendante de la réplication requiert des mécanismes d'échange. Le domaine C-terminal de H2A.Z semble aussi responsable de sa localisation, ainsi la protéine chimérique contenant le

domaine globulaire de H2A.Z et le domaine C-terminal de H2A est moins présente au promoteur des gènes et plus présente dans le corps du gène [169].

L'histone canonique H2A est échangée pour sa variante H2A.Z de manière ATP dépendante par le complexe SWR1, apparenté au complexe SWI/SNF [146, 170, 171]. Son activité a été démontrée *in vitro* et *in vivo* ; en effet, la délétion de SWR1 entraîne la diminution de l'incorporation de H2A.Z aux régions qui bordent l'hétérochromatine silencieuse à proximité des télomères [172, 173]. Chez *S.pombe*, l'absence de Swr1 entraîne une perte globale de H2A.Z dans l'ensemble du génome avec une relative accumulation de H2A.Z aux centromères et dans les régions sous télomériques [163].

SWR1 est un complexe composé de 14 sous-unités chez la levure dont Swr1 est la sous-unité catalytique [146, 171, 173, 174]. SWR1 est assisté dans sa fonction par les chaperonnes Chz1 et Nap1, chargées de transférer le dimère H2A.Z/H2B à la sous-unité Swc2 de SWR1 [172, 173]. Alors que Chz1 a une nette préférence pour le dimère H2A.Z/H2B, la chaperonne Nap1, considérée comme une chaperonne du dimère H2A/H2B, est aussi capable de s'associer au dimère H2A.Z/H2B et dans tous les organismes testés, une fraction des dimères H2A.Z/H2B non incorporés ont été trouvés associés en complexe avec la chaperonne Nap1 [172-175]. Même si Nap1 et Chz1 s'associent aux dimères H2A.Z/H2B, leurs fonctions ne sont peut-être pas complètement superposables. Nap1 faciliterait l'import du dimère H2A/H2B dans le noyau, en stabilisant son interaction avec la protéine d'import nucléaire Kap114, et participerait au maintien d'un réservoir d'histone H2A.Z [176]. Le fait que Nap1, mais non Chz1, s'associe avec H2A.Z dans le cytoplasme des cellules tend à confirmer cette hypothèse [176].

Une fois la localisation génomique de H2A.Z déterminée, la question du mécanisme de ciblage au promoteur des gènes s'est posée. Dans la foulée de son étude de localisation génomique, Raisner propose un mécanisme de déposition qui repose sur l'identification d'une séquence bipartite suffisante pour la formation d'un NFR bordé par des nucléosomes contenant H2A.Z [50]. Ce signal comprend un site pour le facteur Reb1 immédiatement suivi d'une région (dT : dA)₇. Toutefois, si les

sites Reb1 sont extrêmement bien conservés au promoteur des gènes, tous les gènes contenant H2A.Z n'en sont pas pourvus et leur présence ne peut donc pas expliquer l'incorporation de cette variante à tous les promoteurs [50]. La délétion de H2A.Z n'empêche pas la formation d'un NFR au promoteur des gènes mais on ne peut dire clairement si la présence d'un NFR induit l'incorporation d'H2A.Z, ou si les deux évènements sont distincts [50].

Le groupe de Cairns remarque la corrélation avec le facteur Reb1 et 3 autres activateurs, mais surtout, observe une colocalisation avec Bdf1, une protéine à bromodomaine [158]. Bdf1 est une sous-unité de SWR1 qui pourrait donc en théorie guider la déposition de H2A.Z aux nucléosomes acétylés des promoteurs. Aux promoteurs de nombreux gènes, NuA4 acétyle principalement l'histone H4, mais aussi l'histone H2A, par l'intermédiaire de sa sous-unité HAT (Esa1), ce qui aboutit à la formation du complexe de pré-initiation (PIC)[120]. Alternativement, la formation du PIC est initiée par SAGA, qui acétyle la queue des histones H3, *via* sa sous-unité Gcn5. On peut, cependant, noter que le recrutement de Bdf1 aux régions promotrices ne semble pas dépendre de SAGA, ou TFIID [177]. Les expériences d'échange d'histones *in vitro* utilisant une chromatine native et des complexes SWR1 et NuA4 purifiés, démontrent que l'acétylation de H4 et de H2A aux promoteurs stimule l'incorporation de H2A.Z par Swr1 [178]. Bdf1 reconnaîtrait donc les résidus acétylés des queues d'histones H4 et ainsi dirigerait la déposition de H2A.Z au promoteur des gènes [178]. Chez la levure, NuA4 et SWR1 partagent quatre sous-unités ce qui suggère un lien fonctionnel [179]. De plus, chez l'humain, le complexe Tip60 correspond à la fusion des deux complexes SWR1 et NuA4 de la levure [180, 181]. Chez la levure, Bdf1 ne semble pas être en proportion stœchiométrique dans le complexe SWR1 suggérant, peut-être, l'existence de deux complexes, et donc de deux types de déposition : Bdf1 dépendant et indépendant [180]. Parallèlement chez l'humain, le complexe SCRAP, contrairement au complexe TIP60, ne contient pas Brd8, homologue de Bdf1, et pourrait être l'équivalent humain de la portion de SWR1 n'incluant pas Bdf1 [180, 182]. Ces interactions forment aujourd'hui l'hypothèse principale de la déposition de H2A.Z aux promoteurs (revu dans [183]).

À l'image de l'histone canonique, H2A.Z peut aussi être modifiée. Chez la levure, les lysines K3, 8, 10, 14 de la queue N-terminale sont acétylables, avec une préférence pour le résidu 14 *in vivo* [160, 184]. L'acétylation de H2A.Z K14 est entièrement dépendante de NuA4, comme cela fut démontré par l'utilisation d'allèle conditionnelle *esa1*, ou par délétion de sous-unités non essentielles de NuA4 (mais requises pour l'acétylation de H4) [184]. L'acétylation de H2A.Z est essentielle à son action aux régions silencieuses, puisque l'histone H2A.Z non acétylable est correctement recrutée aux domaines sous-télomériques mais s'avère incapable de contrer la propagation des protéines Sir [185]. Dans le mécanisme de maintien de l'expression dans les domaines HZAD chez la levure, il est important de noter que l'incorporation de H2A.Z aux régions sous-télomériques est fonction de l'acétylation de la marque H4K16 catalysée par le complexe SAS (*something about silencing*) [186-188].

Le complexe INO80 possède la spécificité inverse à SWR1 et catalyse l'échange des dimères H2A.Z/H2B d'un nucléosome contre des dimères H2A/H2B libres. Il a donc une fonction antagoniste à celle de SWR1 [102, 103]. L'équilibre entre ces deux complexes pourrait servir à la restriction de H2A.Z aux nucléosomes bordant le NFR aux promoteurs des gènes, par éviction des histones H2A.Z incorrectement incorporées dans le génome [103, 146]. Il est donc intéressant de noter que INO80 échange spécifiquement les histones H2A.Z non acétylées de la chromatine et que cette fonction semble jouer un rôle dans l'intégrité du génome [103].

1.1.3.3.1.4 Implication biologique de H2A.Z

Biologiquement, H2A.Z joue un rôle important lors du développement. Ainsi, l'absence de H2A.Z est létale lors du développement embryonnaire chez la souris [136]. De plus, les études de H2A.Z dans les cellules souches embryonnaires de souris démontrent que H2A.Z est nécessaire à l'établissement d'états de la chromatine qui sont essentiels à l'exécution du programme de différenciation cellulaire [136, 189]. Une augmentation du niveau d'expression de H2A.Z est

observée dans un certain nombre de cancers [190-193]. Les travaux du laboratoire de Luc Gaudreau placent H2A.Z dans la cascade de réponses aux estrogènes permettant de tracer un lien direct entre H2A.Z et le cancer du sein [194]. Les travaux qui en ont découlé suggèrent qu'une hausse du niveau de H2A.Z augmente la prolifération des cellules cancéreuses, notamment en absence d'estrogène α [195]. H2A.Z fait donc partie des cibles possibles dans le traitement des cancers du sein, notamment par l'inhibition de son incorporation [196].

1.2 Transcription

La transcription est l'étape la plus régulée dans l'expression génétique. Chez les eucaryotes, la transcription des ARNm n'est pas une réaction isolée, mais concomitante aux étapes de maturation des ARNm comme le coiffage, l'épissage et la polyadénylation et aux modifications de la chromatine. Ces étapes sont nécessaires à la production d'un ARNm stable pouvant être correctement exporté et traduit dans le cytoplasme au passage de la machinerie de transcription, et à la reconstitution de la chromatine dans le sillon de l'ARNPoIII. Une multitude de protéines et de complexes interviennent donc de manière coordonnée au long de la transcription afin d'en assurer le bon déroulement [197-200]. Centrés autour de l'ARNPoIII les facteurs impliqués dans la transcription sont extrêmement bien conservés dans l'évolution, particulièrement parmi les eucaryotes et quelques fois jusqu'aux archéobactéries. Notre compréhension des mécanismes de transcription a été éclairée au cours des décennies par de nombreuses techniques notamment : la cristallographie, pour la résolution des complexes, la biochimie, pour l'étude des relations entre substrat et enzyme, et plus récemment les techniques de localisation génomique.

1.2.1 Bref historique de la transcription

En 1934, bien avant la reconnaissance de l'ADN comme support de l'information génétique, Thomas Morgan développe dans sa lecture pour le prix Nobel l'hypothèse d'une action du protoplasme sur les gènes afin de réguler l'activation séquentielle des gènes durant le développement embryonnaire [201]. Bien que très spéculative pour l'époque, cette hypothèse est en fait une esquisse de la régulation transcriptionnelle. Ce n'est qu'au début des années 1960, avec le travail de Jacques Monod et François Jacob sur l'opéron lactose, que l'hypothèse d'un ARNm et de la transcription proprement dite est clairement formulée [202, 203]. Bien que difficile à imposer durant les premières années, elle est le fondement de l'étude de la transcription. L'importance des travaux de Jacques Monod et François Jacob dans la

compréhension du contrôle des gènes, sera reconnue par un prix Nobel (physiologie ou médecine) en 1965 [204]. Ils partagent ce prix avec André Lwoff, mentor de Jacques Monod, lui aussi membre de l'institut Pasteur à Paris [202].

Contrairement aux procaryotes qui n'ont qu'une seule ARN polymérase pour la synthèse de l'ensemble de leurs ARN, les eucaryotes possèdent trois ARN polymérases spécialisées. Quelques années après l'ARN polymérase procaryote chez *E. coli* [205], les trois polymérases eucaryotes furent identifiées en 1969 par Roeder et Rutter, à partir d'un fractionnement de chromatine d'embryons d'oursin de mer [206], et, indépendamment en 1970, par le groupe de Pierre Chambon, à partir de thymus de veau [207]. Les polymérases furent nommées suivant leur ordre d'élution I, II et III [206] ou encore A, B, C [207], bien qu'aujourd'hui, seule la première appellation soit restée. Les études qui en découlèrent permirent de déterminer la spécificité de chacune des polymérases. L'ARN polymérase I transcrit les ARNr, l'ARN polymérase II est responsable de la synthèse des gènes encodant les protéines alors que l'ARN polymérase III transcrit les petits ARN stables comme les ARNt ou l'ARNr 5S [208]. Les polymérases sont des complexes à plusieurs sous-unités dont une partie est partagée entre toutes les polymérases, une partie trouve des homologues dans chacune et seulement une minorité est réellement spécifique à chaque polymérase. Malgré la complexification observée chez les eucaryotes tant au niveau du nombre de polymérases que de leur composition, le site catalytique et le mécanisme de synthèse sont conservés dans l'ensemble du vivant [209].

1.2.2 ARN polymérase II

1.2.2.1 Structure

En raison de sa fonction dans la transcription de tous les ARNm de la cellule, l'ARN polymérase II est, sans aucun doute, la plus étudiée des trois ARN polymérases. Elle forme un large complexe d'environ 600 kDa constitué de 12 sous-unités dont la taille varie de 6 à 200 kDa [209]. Au cours de l'évolution, les structures et les interactions des membres du complexe ont été extrêmement bien conservées ;

en conséquence, un complexe similaire peut être purifié chez la plupart des organismes eucaryotes [209].

La contribution du Dr Kornberg à la résolution de la structure cristallographique du complexe de l'ARNPolII fut consacrée par un prix Nobel de chimie en 2006 [210]. Les premières structures cristallographiques de l'ARN polymérase ne comprennent que 10 des 12 sous-unités ; en effet, la présence de quantité sous-stœchiométrique des sous unités Rpb4 et Rpb7 interfère avec la cristallisation du complexe [211]. Quelques années plus tard, l'isolation du complexe en utilisant Rpb4 comme appât ou l'addition des sous-unités Rpb4 et 7 purifiées chez la bactérie ont permis d'obtenir une population homogène contenant les 12 sous-unités de la polymérase [212]. Les sous-unités Rpb4 et Rpb7 forment un hétérodimère qui s'associe de manière réversible à la polymérase et est impliqué dans l'initiation [213]. L'addition de cet hétérodimère à la polymérase purifiée avec 10 sous-unités rétablit sa capacité d'initiation [213]. Contrairement aux sites catalytiques des polymérases procaryotes, celui de l'ARN polymérase est enfoui au cœur du complexe (**Figure 1.7**) [214]. La résolution de la structure cristallographique de l'ARN polymérase II nous permet de mieux comprendre le fonctionnement de l'enzyme. En se basant sur ces résultats, on peut définir les éléments principaux de la structure de la polymérase en élongation, c'est-à-dire contenant l'hybride ARN/ADN [211, 215]. L'ADN double brin s'insère dans le sillon (*cleft*) tapissé de charges positives et bordé par des mâchoires mobiles auxquelles participent la sous-unité Rpb9 (mâchoire supérieure, *upper jaw*) et la sous-unité Rpb5 (mâchoire inférieure, *lower jaw*). Les brins d'ADN sont séparés au niveau de la boucle d'embranchement 2 (*fork 2*), le brin matrice continuant par-dessus l'hélice de pontage (*bridge*) jusqu'au site catalytique alors que le brin non codant est dirigé vers la surface de l'enzyme. L'extrémité aval du sillon est bloquée par le mur (*wall*), un domaine appartenant à la sous-unité Rpb2, qui force le brin matrice (et l'hybride ADN/ARN) en amont du site catalytique à former un angle droit avec l'ADN du sillon. Au pied du mur, le site catalytique, partagé entre les deux sous-unités Rpb1 et Rpb2, est marqué par la présence d'un ion Mg²⁺. Sur la face postérieure de l'enzyme, se trouve une ouverture en entonnoir (*funnel*) dont l'extrémité la plus étroite forme un pore (pore1) d'environ

12Å qui pénètre jusqu'au site actif. Ce pore permet d'évacuer l'ARN par son extrémité 3' lorsque la polymérase fait marche arrière (*backtrack*) et pourrait aussi servir de canal d'entrée pour les nucléotides. Toutefois, le chemin emprunté par les nucléotides jusqu'au site catalytique de la polymérase reste encore aujourd'hui sujet à débat [216]. La structure de la polymérase en solution et celle en cours d'élongation diffèrent dans la position de la pince (*clamp*). Suivant l'ouverture de la bulle de transcription, la pince, formée par les régions N-terminale de Rpb1, C-terminale de Rpb2 et une portion N-terminale de Rpb6, se referme sur le sillon et le site catalytique. À la base de la pince, cinq boucles (*switches*), dont trois sont peu organisées ou désorganisées, associent la rotation de la pince avec la présence de l'hybride ARN/ADN et le maintien de la bulle de transcription. La boucle de la gâchette (*trigger*), située à proximité de l'hélice de pontage (*bridge*), interagit avec celle-ci et intervient dans la sélection des nucléotides en favorisant la catalyse lorsque le nucléotide est correctement apparié et positionné [214, 217, 218]. Lorsque le 11^e nucléotide est ajouté à l'hybride ARN/ADN, la polymérase passe en élongation productive et l'avortement de la transcription n'est alors plus possible. L'encombrement stérique avec le gouvernail (*rudder*) force la séparation de l'ADN matrice et de l'ARN empêchant l'hybride de s'étendre au-delà de 10 nucléotides. Le gouvernail et le capuchon (*lid*) orientent l'ARN en élongation vers le canal de sortie (*exit channel*), situé sous le capuchon, par lequel il émerge hors de l'enzyme. Les brins d'ADN se réappariaient en amont de la fermeture à glissière (*zipper*). La bulle de transcription est maintenue, en aval par la boucle d'embranchement 2 (*fork 2*), et en amont par le capuchon et la fermeture à glissière bien que d'autres structures participent à sa stabilité.

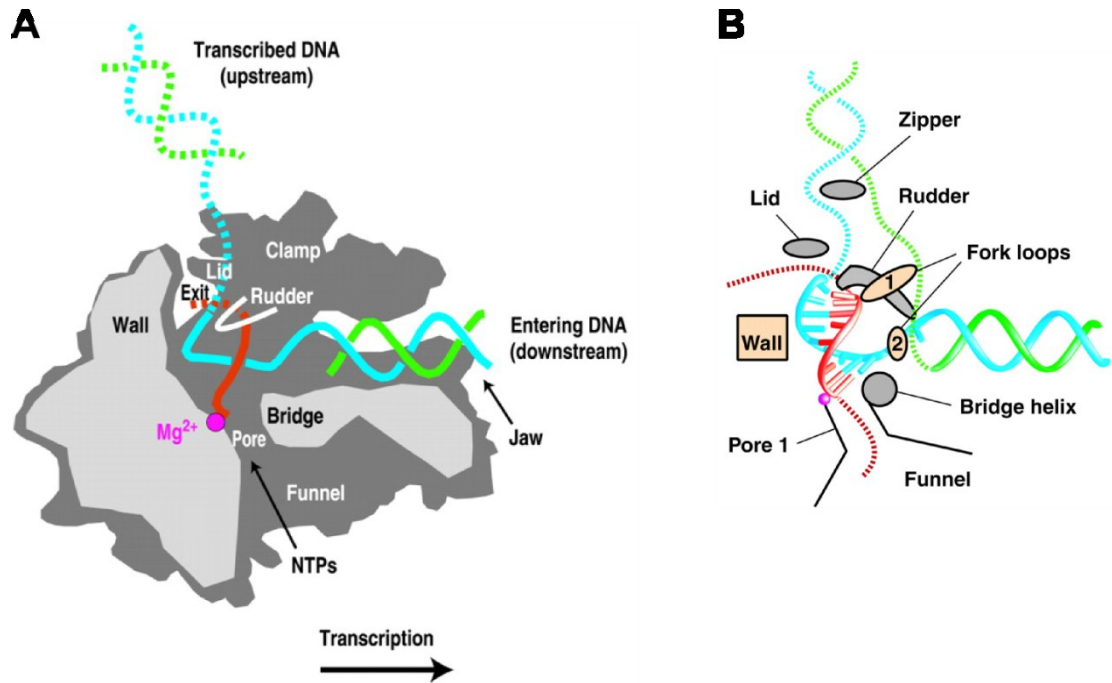


Figure 1.7 : Représentation schématique de l'ARNPolII et de son site catalytique.

Figure adaptée de Bushnell *et al.*, (2002) [219] et de Gnatt *et al.*, (2001) [215].

A) Vue en coupe schématique du complexe de l'ARN polymérase II. Le brin matrice est coloré en bleu, le brin non codant en vert et l'ARN appartenant à l'hybride ARN/ADN en rouge. Les lignes continues représentent le squelette des acides nucléiques dans la structure cristallographique, alors que les lignes pointillées représentent les acides nucléiques non présents dans la structure cristallographique

B) Représentation des acides nucléiques et des éléments protéiques de l'ARN polymérase entrant en contact avec la bulle de transcription. Les rubans représentent les acides nucléiques positionnés selon la structure cristallographique, les lignes pointillées représentent le passage possible des acides nucléiques non déterminés par la structure cristallographique. Les éléments appartenant à Rpb1 et Rpb2 sont indiqués en gris et en jaune respectivement.

1.2.2.2 Domaine C-terminal

Le domaine C-terminal (CTD) de la sous-unité Rpb1 est un domaine spécifique à l'ARN polymérase II. Non structuré, le CTD forme une « queue » constituée de répétitions d'un heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$, dont le nombre et le degré de dégénération varient d'une espèce à l'autre. On dénombre ainsi, 26 répétitions chez la levure et 52 chez l'humain ; la dégénération des répétitions accompagne l'augmentation de la complexité génétique [220, 221]. De plus, elle n'est pas uniformément répartie le long du CTD. En règle générale, les répétitions les plus distales (C-terminales) sont plus divergentes que les répétitions proximales (N-terminales). Ceci est particulièrement évident pour la position Ser7. Le biais est modeste chez la levure, puisque, sur 7 répétitions ne contenant pas la Ser7, 5 sont localisées dans la partie distale. Par contre, chez l'humain, cette tendance est très prononcée, puisque sur les 26 répétitions ne contenant pas de sérine en position 7, 20 sont situées dans la moitié distale du CTD [222]. Chez la levure, au moins huit répétitions sont nécessaires pour assurer la viabilité alors que 10 sont suffisantes pour procurer une croissance proche de l'état sauvage dans les conditions de laboratoire [223]. Bien que les formes tronquées s'avèrent peu délétères pour la croissance, d'un point de vue évolutif, il existe une grande pression sur le nombre de répétitions [222]. Le CTD est stratégiquement localisé à proximité du canal de sortie de l'ARNm (**Figure 1.8 A**) [224]. Les répétitions sont précédées par un domaine de liaison (*linker*) de 91 a.a (chez la levure), également non structuré, les reliant au domaine globulaire de Rpb1. La composition du domaine de liaison est intéressante du fait de sa richesse en sous-motifs YSPx et SPxY, bien que ce ne soit pas le cas pour le CTD de la levure [221]. Les répétitions sont suivies d'une séquence variable jouant un rôle dans la stabilité du CTD, mais dont la séquence importe peu [225]. À leur extension linéaire maximale, le domaine de liaison et le CTD atteignent une longueur de 900 à 930 Å, dont 650 Å pour le CTD seul, ce qui lui permet d'interagir avec de nombreuses protéines et avec des éléments éloignés de la polymérase dont le diamètre est d'environ 150 Å (**Figure 1.8 B**) [226].

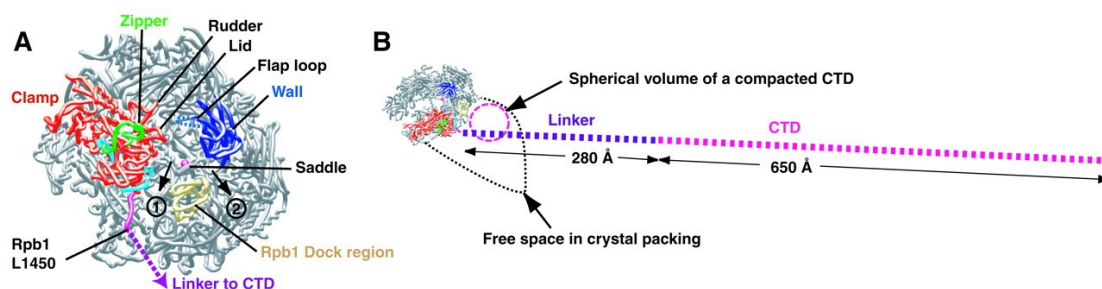


Figure 1.8 : Le domaine C-terminal de l'ARN polymérase II de levure.

Cramer *et al.*, (2001) [226]

A) Représentation sous forme de ruban de la sous-unité Rpb1. Les canaux de sortie de l'ARN précédemment proposés sont annotés 1 et 2. Le canal 1 est maintenant identifié comme le canal de sortie de l'ARN. Le dernier résidu structuré de Rpb1 (L1450) est indiqué. Les 25 résidus N-terminaux de Rpb1 sont représentés en bleu et correspondent au fragment qui forme un pontage avec l'ARN sortant de la polymérase chez la bactérie *E.coli*. La fermeture à glissière (*zipper*) est indiquée en vert, la pince (*clamp*) en rouge, le mur (*Wall*) en bleu **B)** Localisation et taille du CTD. L'espace libre dans le réseau cristallin pour les CTD de quatre polymérases adjacentes est indiqué en pointillé noir. Les tirets représentent la longueur du domaine de liaison et du CTD en extension. Le cercle rose représente la taille d'un repliement aléatoire d'une masse équivalente au CTD.

Les acides aminés formant les heptades peuvent tous être dynamiquement modifiés et les ponts peptidyl-prolyle peuvent être isomérisés entre les conformations Cis et Trans [222, 227, 228]. Le CTD a même été récemment démontré comme méthylé [229]. Si la structure du CTD peut donner l'impression que ses répétitions forment des unités fonctionnellement distinctes, le CTD doit plutôt être envisagé comme une succession continue de motifs chevauchants. Ce dernier point est d'ailleurs illustré par le fait que la levure *S. cerevisiae* peut survivre avec un CTD composé de répétitions de diheptades espacés par des résidus alanine (bien que la dépendance aux nombres de répétitions soit étendue), mais pas avec des monoheptades espacés [230]. De plus, la fixation de protéines aux diheptades favorise la fixation coopérative des protéines interagissant avec le CTD grâce à des interactions protéine-protéine [199]. Le CTD est reconnu par des protéines contenant un domaine d'interaction au CTD [231]. Il a été démontré que, chez l'humain, toutes les répétitions ne sont pas équivalentes quant à leur fonction. Ainsi les répétitions 27 à 52 sont capables de coupler toutes les étapes de la maturation des ARNm à la transcription, alors que les répétitions localisées en N-terminales ne sont fonctionnelles que pour le coiffage [232].

1.2.3 Cycle de la transcription

Le cycle de transcription correspond à l'ensemble des événements entre le recrutement de l'ARN polymérase aux promoteurs des gènes et son recyclage pour un nouveau cycle de transcription. D'un point de vue mécanistique, le cycle de la transcription est une succession d'étapes continues qui s'enchaînent et se chevauchent, il est donc parfois difficile d'en définir les limites précises. Malgré cela, le cycle de la transcription est généralement divisé en étapes. Afin d'en faciliter, ici, la description, le cycle sera divisé en six étapes : pré-initiation, initiation, élongation précoce, élongation, terminaison et recyclage. L'étude du cycle de la transcription a révélé l'intégration des étapes de maturation de l'ARNm en cours de synthèse à la progression de la polymérase le long de l'unité de transcription.

1.2.3.1 Pré initiation

La pré-initiation correspond au recrutement et à l'assemblage du complexe de pré-initiation (PIC) au promoteur basal des gènes, ce qui en fait la cible de nombreux mécanismes de régulation de l'expression des gènes. Le PIC contient les facteurs généraux de la transcription (FGT) TFIID, TFIIA, TFIIB, TFIIF, TFIIE et TFIIH, associé à la polymérase [233]. Durant de nombreuses années, l'assemblage du PIC fut disséqué *in vitro* à partir de facteurs purifiés et d'un promoteur basal. Ces travaux ont abouti au modèle d'assemblage séquentiel « classique » des FGT [234]. Lors du recrutement séquentiel, TFIID est recruté en premier, suivi de TFIIA, TFIIB, de l'ARNPolIII en complexe avec TFIIF. Enfin TFIIE et TFIIH complètent l'assemblage du PIC au promoteur [235]. Toutefois ce modèle n'intègre pas l'activation des gènes.

Le développement des techniques de génomique a enrichi notre connaissance des mécanismes de recrutement de la polymérase au promoteur des gènes. Chez la levure, deux principaux modes de recrutement se partagent l'activation des gènes. Le mécanisme de recrutement dépendant de TFIID, domine pour l'expression des gènes d'entretien (*housekeeping genes*) de la cellule (90 % des gènes) alors que le mécanisme de recrutement dépendant de SAGA (*Spt-Ada-Gcn5 acetyltransferase*) domine pour l'expression des gènes de réponse au stress (10 % des gènes) [177]. TFIID est plus généralement associé aux gènes dont les promoteurs ne contiennent pas de boîte TATA, alors que SAGA s'associe plus généralement aux gènes qui en sont pourvus [236]. Les deux complexes ont pour effet d'amener la protéine de liaison de la boîte TATA, TBP (*TATA binding protein*), au promoteur afin de nucléer l'assemblage du PIC. Le lien entre l'activation spécifique des gènes et la machinerie de transcription, quel que soit le mécanisme de recrutement de TBP au promoteur, est assumé par le complexe coactivateur du médiateur [237]. Le médiateur est indispensable au recrutement de TBP aux promoteurs dépendant de SAGA, mais non pas à ceux dépendant de TFIID. Toutefois, il est essentiel à l'initiation des gènes SAGA et TFIID dépendants [237]. De façon simplifiée, le médiateur est recruté par les domaines d'activation des facteurs de transcription et va interagir avec la machinerie de transcription pour la recruter au promoteur ou l'activer [238].

1.2.3.1.1 TFIID

Dans le modèle classique d'assemblage séquentiel du PIC, TFIID est le premier facteur à s'associer au promoteur, nucléant ainsi l'assemblage du PIC. TFIID est un large complexe d'environ 1.3 MDa composé de TBP et d'environ 14 TAF (*TBP associated factor*) parfois présentes à plus d'une copie et dont 13 sont essentielles chez la levure [239-241]. Reflétant la conservation observée chez les eucaryotes, les TAF sont numérotées de 1 à 15 suivant une nomenclature unifiée permettant une meilleure comparaison entre espèces [242]. Grâce à TBP qui s'associe de manière dynamique à l'holocomplexe formé par les TAF, TFIID est capable de reconnaître et de se fixer à la boîte TATA des promoteurs [240]. Cependant, celle-ci n'est pas présente au promoteur de tous les gènes et les TAF étendent les capacités de reconnaissance des promoteurs de TFIID en fixant des éléments distincts de la boîte TATA. TAF1 et TAF2 vont s'associer à l'élément INR (*initiator*) du promoteur basal dont la présence est indépendante de celle de la boîte TATA [243]. En présence d'une boîte TATA, ces interactions vont aussi stabiliser la liaison de TFIID au promoteur [243]. Le rôle des TAF est également celui de co-activateurs transmettant le signal d'un activateur à la machinerie de transcription de l'ARNPolIII [244]. En plus de sa fonction de reconnaissance des promoteurs et de nucléation du PIC, les activités enzymatiques liées aux TAF impliquent TFIID dans la modification de la chromatine et des facteurs impliqués dans la régulation de la transcription [245]. Fait intéressant, TBP, TAF5, 12, 6, 10 et 9 font également partie du complexe SAGA (détaillé plus loin) [246].

1.2.3.1.2 TFIIA

Chez la levure, TFIIA est constitué de 2 protéines de 32 et 13.5 kDa encodées par deux gènes essentiels *TOA1* et *TOA2*. Chez l'humain, TFIIA est constitué de trois polypeptides α , β , γ de 35, 19 et 12 kDa respectivement. Toutefois, les sous-unités α et β sont encodées sous forme d'un précurseur (55 kDa) par un gène unique et les extrémités N- et C-terminales de *Toa1* de la levure, sont structurellement comparables à ces sous-unités [233, 235]. De plus, la partie centrale de *Toa1*, moins conservée, n'est pas essentielle à sa fonction [241]. *Toa2* stabilise la liaison de TBP au

promoteur en établissant des contacts à l'ADN en amont de la boîte TATA, ainsi qu'avec TBP et TAFII. Comme les autres facteurs généraux de la transcription, TFIIA a d'abord été identifié comme nécessaire à la transcription, mais les études subséquentes ont démontré qu'il n'était pas indispensable lors de l'initiation par TBP mais avait une fonction antirépression [235, 247-250].

1.2.3.1.3 TFIIIB

TFIIB est formé d'un seul polypeptide de 38 kDa chez la levure, encodé par le gène essentiel *SUA7*, et de 35 kDa chez l'humain. TFIIB joue un rôle central dans l'assemblage du PIC. TFIIB s'associe avec TBP stabilisant l'association de TFIID/TBP avec le promoteur [235]. Chez les eucaryotes supérieurs, TFIIB est capable de se fixer à l'élément BRE (*TFIIB responsive element*) du promoteur basal (localisé adjacent à la boîte TATA dans certains promoteurs) [251, 252]. Plus récemment, la capacité de TFIIB à se fixer indépendamment de la polymérase au promoteur et au terminateur des gènes a été démontrée [253]. La fixation de TFIIB au terminateur des gènes est responsable de la formation de boucles d'ADN (*DNA looping*) entre le promoteur et le terminateur des gènes chez la levure [253].

1.2.3.1.4 TFIIIF

TFIIIF est recruté au promoteur en complexe avec la polymérase. Il est formé de deux sous-unités RAP30 et RAP74 (*RNAPolIII-associating protein*) organisées en un hétérotétramère [254]. TFIIIF augmente et stabilise l'affinité de la polymérase pour le complexe TBP/TFIIB au promoteur [255]. En outre, sa sous-unité RAP74 stimule directement l'activité de la phosphatase Fcp1 (une phosphatase du CTD de la polymérase) [256]. TFIIIF intervient également dans la phase d'élongation de la transcription, traitée plus tard dans ce chapitre [257].

1.2.3.1.5 TFIIIE

TFIIIE est constitué d'un hétérodimère de deux sous-unités, α et β . TFIIIE fut initialement purifié à partir de foie de rat par Joan et Ronald Conaway en 1989 [258]. Il est impliqué dans le recrutement de TFIIH aux promoteurs des gènes lors des

événements tardifs de la pré-initiation, et il en stimule les activités enzymatiques (ATPase, kinase et DNA hélicase) [259-261]. TFIIE interagit avec TFIIIF, TFIIIB, l'ARNPolIII et le promoteur. Il est essentiel pour le désappariement au promoteur et la transition de l'initiation à l'élongation [235, 262].

1.2.3.1.6 TFIIH

TFIIH, recruté par TFIIE au PIC, est le dernier FGT à joindre le complexe de pré-initiation. Son incorporation dans le complexe de pré-initiation engendre l'ouverture de la bulle de transcription au promoteur et la phosphorylation des sérines 5 et 7 du CTD [263, 264]. De par son implication dans la réparation de l'ADN par la voie d'excision des nucléotides liée à la transcription (TC-NER pour *transcription coupled nucleotide excision repair*), il est l'un des facteurs de transcription les mieux connus et plus étudiés [265, 266]. TFIIH comprend 11 sous-unités réparties en 2 sous-complexes, nommés le complexe Core et le CAK (*cyclin-activating kinase*) complexe. Au total, TFIIH contient 4 activités enzymatiques : deux activités hélicases, une activité kinase et une activité endonucléase [263]. Le Core est composé, chez la levure, de 8 protéines Rad5, Ssl2, Rad3, Tfb1, Tfb2, Ssl1, Tfb4, Tfb5 (correspondant chez l'humain à : XPG, XPB, XPD, p62, p52, p44, p34, TTDA respectivement). Le CAK est, quant à lui, composé de 3 sous-unités Ccl1, Tfb3, Kin28 (cycline H, MAT1 et cdk7 chez l'humain) [263]. Chez l'humain, la sous-unité XPD est associée aussi bien avec le complexe Core qu'avec le complexe CAK et sert à relier les deux sous-complexes [267]. TFIIH a une structure en anneau encerclant l'ADN [268]. Les deux sous-unités XPB et XPD sont des hélicases aux polarités inverses (de 3' en 5' et de 5' en 3' respectivement). Seul XPB est requise pour la transcription, alors que les deux activités hélicases sont requises pour la réparation de l'ADN [269-271]. L'activité hélicase est importante pour l'évasion du promoteur chez l'humain, le couple Kinase/cycline (cdk7/cycline H) est aussi impliqué dans la régulation du cycle cellulaire [272]. En outre, plusieurs complexes incorporent des sous-unités de TFIIH, par exemple le sous-complexe CAK peut être trouvé seul (50 % du CAK), ce complexe a alors comme substrat préférentiel cdk1, cdk2, cdk4 et cdk6 [263, 273, 274].

1.2.3.1.7 SAGA

SAGA est un large complexe de 1.8 MDa composé de 21 protéines [275]. Chez la levure, 15 des sous-unités sont non essentielles et 6 sont essentielles [237]. SAGA partage les TAF 5, 12, 6, 10, 9 avec TFIID. SAGA possède deux activités catalytiques : une activité HAT, et une activité histone déubiquitylase assumée par les protéines Gcn5 et Ubp8 respectivement [276]. Ces activités jouent un rôle dans la régulation de la transcription par l'acétylation de l'histone H3 aux promoteurs des gènes et la déubiquitination de H2B [118, 119, 245]. Les sous-unités Spt7, Spt20 et Ada1 font partie du noyau de SAGA et maintiennent l'intégrité du complexe [277]. La comparaison des structures de SAGA et TFIID par cryo-EM indique que malgré leurs points communs ils n'interagissent pas de la même manière avec l'ADN au promoteur [278]. Chez l'humain, SAGA contient plusieurs homologues : TFTC, PCAF, STAGA [278].

1.2.3.1.8 Médiateur

Le médiateur est un large complexe (~1 MDa) de plus de 20 sous-unités, conservé de la levure à l'humain [279]. En raison de sa topographie, visible par cryo-EM, et de son organisation modulaire, le complexe est subdivisé en trois parties : la tête, le milieu et la queue. Cependant, un quatrième module kinase lui est quelquefois rattaché : le module cdk8 [238, 280]. Le médiateur n'est pas requis pour la transcription *in vitro*, mais stimule l'activité du complexe de transcription même en absence d'activateurs. Il est requis pour la transcription de la majorité des gènes transcrits par l'ARN polymérase II [281]. Des études génomiques démontrent que le médiateur affecte le recrutement *in vivo* des facteurs de transcription TFIIH et TFIIIE et pourrait stabiliser l'assemblage du PIC. À la lumière des expériences réalisées durant les dernières années, il est devenu clair que le médiateur joue un rôle essentiel dans la transcription [281, 282]. Plus récemment, le médiateur a été également impliqué dans l'élongation [283, 284].

1.2.3.2 Initiation

De manière stricte, l'initiation est définie comme la formation du premier lien phosphodiester entre les acides ribonucléiques de l'ARN. On peut toutefois considérer, dans cette section, l'initiation comme allant de l'ouverture de la bulle de transcription jusqu'à la formation d'une chaîne d'environ 10 nucléotides, point crucial dans la progression de la transcription au-delà duquel la polymérase est engagée vers l'élongation et les liens avec le promoteur basal sont définitivement rompus [285]. L'initiation est une étape de transition qui suit l'assemblage complet du PIC et précède l'élongation le long du gène ; durant cette phase, les facteurs d'initiation sont échangés contre des facteurs d'élongation [286].

Suivant le recrutement de TFIIF au PIC, le complexe subit un changement de conformation entraînant l'ouverture de la bulle de transcription par l'hélicase de TFIIF [287]. Le désappariement d'environ 10 à 15 nucléotides permet l'entrée du brin matrice d'ADN dans le site actif de la polymérase. Dans un premier temps, l'ouverture de l'ADN se fait de la position -9 à +2 (par rapport au premier nucléotide transcrit). Cette ouverture est instable et nécessite donc une hydrolyse constante d'ATP pour son maintien [285]. L'étape suivante est la formation du premier lien phosphodiester. Lorsque la synthèse de l'ARN atteint 4 à 5 nucléotides, le complexe tertiaire se stabilise, l'interaction entre la polymérase et TFIIB se modifie pour permettre l'extension de l'ARN [288]. Lorsque 10 nucléotides ont été polymérisés la région -9 à +2, restée ouverte, collapse pour reformer le double brin, ramenant la taille de la bulle à environ 8 nucléotides, et l'ARNPolIII est incapable de réinitialiser la transcription [285]. L'effondrement de la bulle de transcription facilite l'évasion du promoteur par la restitution de l'énergie accumulée [288]. Après la synthèse de 10-15 nucléotides, lors de l'évasion du promoteur, la polymérase se dissocie des facteurs généraux de la transcription et du complexe resté ancré au promoteur [285, 286]. À l'issue de l'initiation, la polymérase compétente à l'élongation forme un complexe extrêmement stable avec l'ADN et l'ARN [285].

1.2.3.3 Pause lors de l'élongation précoce

Chez les métazoaires, le début de l'élongation est souvent marqué par une pause de la polymérase après la synthèse de seulement 25 à 50 nucléotides. Mis en évidence aux promoteurs des gènes inductibles par choc thermique chez la drosophile, le phénomène de pause de la polymérase a été généralisé à la majorité des gènes chez les métazoaires [289]. Cette pause offre plusieurs avantages dans la régulation de la transcription. La modulation du temps de pause permet de réguler le taux de transcription des gènes ; par exemple, au gène de choc thermique *hsp70*, le relâchement de la pause passe de 10 min à 4 secs lors de l'application d'un choc thermique [289]. La pause permet également de conserver la structure de la chromatine dans un état favorable à la transcription et ainsi de diminuer le temps de réaction lors de l'induction de certains gènes [290]. Deux complexes protéiques sont nécessaires à l'établissement de la pause, DSIF (*D*R*B* *S*e*n*s*i*t*i*v*i*t*y*-*I*n*d*u**c**i*n*g**i*n*g**u**m**F*a*c*to*r*) et NELF (*N*e*g**a*t*iv*e* *E*l*on*g**a*t*io*n* *F*a*c*to*r*) [291]. Le complexe DSIF est un hétérodimère composé des deux sous-unités : Spt4 et Spt5. Le complexe est extrêmement bien conservé chez les eucaryotes et peut même être retrouvé chez les archéobactéries (NusG) [292]. Par opposition, NELF, composé de 4 sous-unités (NELF-A, B, C/D, E) est distinctif des eucaryotes supérieurs et donc absent de la levure *S. cerevisiae* ou du nématode *C.elegans* [290]. DSIF est nécessaire au recrutement de NELF au complexe d'élongation [290]. La kinase P-TEFb entraîne la dissociation de NELF du complexe d'élongation, et stimule ainsi le départ en élongation, par phosphorylation du domaine C-terminal de la sous-unité Spt5 du complexe DSIF. DSIF reste associé à la polymérase durant la phase productive de l'élongation pendant laquelle il joue un rôle positif [289].****

1.2.3.4 Élongation

L'élongation est la phase productive de la transcription. Durant cette phase, la polymérase doit surmonter la barrière que représente la chromatine et être capable de repartir lorsqu'un évènement entraîne l'arrêt de la machinerie de transcription. De nombreux complexes vont intervenir pour faciliter ou stimuler le passage de la polymérase le long de l'unité de transcription. Ces complexes agissent sur deux

aspects de l'élongation : la processivité, c'est-à-dire le nombre de polymérase qui vont achever la transcription avec succès, ou la vitesse d'élongation. On peut aussi diviser ces complexes entre ceux qui agissent directement sur la polymérase lors de la pause ou de l'arrêt de l'élongation et ceux qui agissent sur les histones afin de faciliter le passage à travers la matrice de chromatine.

1.2.3.4.1 DSIF

DSIF, déjà évoqué plus haut pour son rôle lors de la pause en élongation précoce, fut identifié en 1998 lors du criblage de facteurs partiellement purifiés à partir de cellules HeLa, capables de susciter une sensibilité à l'inhibiteur de la transcription DRB (5,6-dichloro-1- β -D-ribofuranosylbenzimidazole), dans un système de transcription reconstitué *in vitro* [293]. Ses sous-unités sont les homologues humains des gènes *SPT4* et *SPT5* de la levure [293]. Les gènes SPT sont ressortis à l'issue d'un crible ayant pour but de trouver des mutations pouvant compenser l'insertion d'un élément transposable (Ty) dans la région 5' non codante du gène *HIS4* [294-296]. Les mutants, en relaxant la reconnaissance du site d'initiation, permettent l'initiation à partir de la séquence de l'élément Ty [297]. De nombreux gènes issus de ce criblage sont associés à la chromatine, et plusieurs d'entre eux seront rencontrés dans la suite de ce chapitre. DSIF affecte la processivité de la polymérase en stabilisant le complexe d'élongation [293]. L'étude cristallographique de la polymérase archéale en complexe avec Spt4/5 suggère que l'hétérodimère se fixe au niveau des mâchoires de l'ARN polymérase, venant ainsi refermer le canal laissé partiellement ouvert par les mâchoires [292]. Cette conformation pourrait prévenir le désengagement de l'ADN et ainsi favoriser l'élongation [298]. Chez la levure, le domaine C-terminal de Spt5 est formé d'une séquence répétée, dont la phosphorylation par la kinase Bur1 serait impliquée dans le recrutement du complexe PAF [299, 300].

1.2.3.4.2 TFIIIS

TFIIIS, facteur général de la transcription, influence la processivité de la polymérase. En plus de son rôle dans l'assemblage du complexe de pré-initiation, il

se réassocie avec la polymérase de manière transitoire lorsque celle-ci est bloquée [301]. Dans cette configuration, il stimule l'activité naturelle de clivage de la polymérase, permettant la génération d'une extrémité 3' fraîche réalignée avec le site catalytique de la polymérase [302]. Les études cristallographiques ont permis de mieux comprendre le fonctionnement de TFIIS. Il insère un motif en ruban de zinc (*zinc ribbon*) par le pore situé sur la face antérieure de la polymérase jusqu'au site catalytique de l'ARN polymérase à proximité de l'ion Mg^{2+} . Ce pore sert aussi, sans doute, à évacuer le transcrit libéré après clivage (lorsque la polymérase a fait marche arrière). Cette insertion entraîne une restructuration du site actif de la polymérase permettant un réalignement de l'ARN avec le site actif et ainsi favorise le clivage par la polymérase [287]. TFIIS est essentiel à la transcription d'une matrice de chromatine, ce qui suggère que la présence des nucléosomes induit des arrêts et nécessite le retour en arrière de la polymérase [303].

1.2.3.4.3 TFIIF

TFIIF, un autre des facteurs généraux de la transcription, stimule la vitesse d'élongation de l'ARNPolII en diminuant le temps de pause [301]. Il ne voyage pas avec la polymérase, mais localise principalement au promoteur [286]. Cependant, il est lui aussi capable de se réassocier avec l'ARNPolII lorsqu'elle pause, et se dissocie de nouveau lorsque celle-ci a repris l'élongation [304].

1.2.3.4.4 Chaperonnes et d'histones

La machinerie de transcription doit traverser la région codante des gènes pourvue de nucléosomes. Les nucléosomes doivent être au moins partiellement désassemblés afin de permettre le passage de la polymérase, mais doivent aussi être reconstitués dans son sillage ou lors de la répression des gènes afin d'assurer l'intégrité de la chromatine. Cette prise en charge de la chromatine par des chaperonnes d'histones est essentielle pour prévenir l'apparition de sites cryptiques de la transcription [305].

FACT (*facilitates chromatin transcription*) fut mis en évidence lors d'une expérience de transcription *in vitro* d'une matrice de chromatine reconstituée, pour laquelle il s'avère essentiel [306]. FACT est une chaperonne d'histones, formée de deux sous-unités Spt16 et Pob3 (SSRP1 : *structure spécifique recognition protein-1*, chez l'humain) essentielles, chez l'humain comme chez la levure. La principale fonction de FACT est de prendre en charge les dimères H2A/H2B durant la transcription, favorisant ainsi le passage de la polymérase [301, 307]. FACT déstabilise l'interaction entre le dimère H2A/H2B et le tétramère [308]. Les mécanismes d'action de FACT entre la levure et l'humain divergent vraisemblablement, car un domaine HMG (*high mobility group*) de liaison à l'ADN est présent dans la sous-unité humaine (SSRP1), mais absent de son homologue chez la levure (Pob3). Ce domaine est apporté au complexe FACT par la protéine Nhp6 qui s'associe faiblement avec Spt16-Pob3. Chez la levure plusieurs protéines Nhp6 vont stimuler le complexe, alors qu'un seul motif HMG est présent chez l'humain [309]. On peut noter que Nhp6 interagit également avec Swi/Snf, RSC et Spt6 [310]. FACT semble agir aussi au niveau du promoteur pour remodeler la chromatine [297]. De plus, le rôle de FACT ne se limite pas à la transcription, mais s'étend à la réplication et à la réparation de l'ADN [311, 312].

Spt6 est également une chaperonne d'histones. Spt6 localise sur les régions codantes en corrélation avec les niveaux d'occupation de l'ARNPolIII; il est requis pour le réassemblage des nucléosomes aux gènes très transcrits [313]. Il interagit avec l'histone H3 et est capable d'assembler les nucléosomes *in vitro* [314]. Les souches mutantes de *spt6* ont un niveau réduit de nucléosomes sur les régions codantes de certains gènes [313, 315].

1.2.3.5 Terminaison

La terminaison de l'ARNPolIII correspond à la libération de l'ARNm en fin de synthèse et au décrochage de polymérase de la chromatine. Suivant l'émergence de la séquence de polyadénylation du canal de sortie, l'ARNm est libéré par le complexe de clivage et de polyadénylation (CPF : *cleavage/polyadenylation factor*) [316]. Le désappariement entre la polymérase et la matrice d'ADN n'intervient pas en un site

spécifique, mais de manière stochastique, dans une fenêtre située en aval de la séquence de polyadénylation. Deux mécanismes principaux de la terminaison ont été proposés. Le premier, appelé allostérique, suppose que l'émergence de la séquence de polyadénylation engendre un changement de composition du complexe de transcription, aboutissant par exemple à l'éviction de facteurs d'élongation ou au recrutement d'un facteur de terminaison [317]. Le second, appelé torpédo, propose que le clivage de l'ARNm génère une nouvelle extrémité 5', non coiffée, permettant le recrutement d'exonucléases (Rat1, Xrn1) capables de déstabiliser le complexe de transcription lorsqu'elles rejoignent la polymérase [318]. Plus récemment, Bentley propose un modèle qui intègre des éléments appartenant à ces deux modèles [317]. Ce modèle hybride suggère que ni la modification allostérique causée par Pcf11, ni la dégradation de l'ARN émergent par les exonucléases (Rat1 et Xrn1) ne sont suffisantes pour engendrer la terminaison dépendante du signal de polyadénylation. Ces événements agiraient ensemble pour entraîner le décrochage de la polymérase en fin de transcription. Il est à noter que les gènes courts, comme les petits ARN nucléolaires (snoRNA), n'utilisent pas la voie de terminaison dépendante de la polyadénylation généralement utilisée par les gènes encodant les protéines [319].

1.2.3.6 Recyclage de la polymérase

Le recyclage de la polymérase permet une réinitiation plus rapide en évitant les étapes de formation du PIC [320]. La réinitiation est facilitée par la présence d'un complexe resté au promoteur après le relâchement de la polymérase, comprenant TFIID, TFIIA, TFIIF, TFIIE et le médiateur, [257, 321]. Le médiateur peut stabiliser le complexe formé par les facteurs généraux de la transcription après l'initiation [322]. Récemment, l'efficacité des mécanismes de terminaison a aussi été proposée comme un facteur favorisant le recyclage efficace de la polymérase au promoteur des gènes [323]. Chez la levure, les extrémités des gènes forment des boucles par contact entre les éléments du terminateur et ceux du promoteur [25]. Bien que l'effet des boucles ne soit pas formellement démontré, ce rapprochement peut faciliter le transfert des polymérases du terminateur au promoteur. La formation des boucles nécessite la transcription et est dépendante de TFIIB, capable de se lier au

terminateur, par un mécanisme différent de sa liaison au promoteur [253]. La liaison de TFIIB au terminateur est dépendante de Ssu72, membre du complexe CPF, suggérant une relation entre le bouclage des gènes et la transcription, puisque le CPF est recruté à la polymérase lors de la terminaison [253].

Le recyclage de la polymérase n'est généralement pas considéré comme une étape en soi dans la transcription, mais son bon déroulement pourrait s'avérer essentiel au maintien du niveau d'expression des gènes.

1.2.4 Le cycle de phosphorylation du CTD

Rapidement mise en évidence, la phosphorylation du CTD de Rpb1 a d'abord été perçue comme un phénomène binaire, dont les formes hypophosphorylée et hyperphosphorylée ont été désignées I_a et I_o respectivement et dont la forme I_o fut associée à l'élongation [324-329]. L'apparition d'anticorps spécifiques à la phosphorylation des différentes sérines a permis la mise en évidence de la nature dynamique de la phosphorylation des résidus Ser5 et Ser2 au cours de la transcription et mené au concept d'un code du CTD, faisant écho au code d'histones [330, 331]. Ce code s'est d'ailleurs complexifié avec la détection, plus récente, de la phosphorylation de la Ser7 [332, 333]. Durant de nombreuses années, la reconnaissance des marques de phosphorylation P-Ser2 et P-Ser5 s'est appuyée sur l'utilisation des anticorps H5 et H14 respectivement [334, 335]. Cependant, les études sur leur spécificité, aboutissent à des résultats divergents mais s'accordent sur la sensibilité de ces anticorps aux groupements phosphates localisés sur les résidus sérines avoisinants [332, 336, 337]. La technique de CHIP s'est avérée précieuse pour l'accumulation des connaissances sur la dynamique du cycle de phosphorylation du CTD de la polymérase, mais est affectée par la spécificité des anticorps. Bien que de nouveaux anticorps aient été développés depuis, ceux-ci ont leurs propres biais ; en conséquence, les résultats des expériences doivent être interprétés avec prudence [332, 336, 337].

1.2.4.1 Le cycle

Durant le déroulement de la transcription, le CTD de la polymérase est modifié par une panoplie d'enzymes, responsables de la phosphorylation et de la déphosphorylation coordonnée des résidus sérines (**Figure 1.9**). Cette partie retrace les événements de phosphorylation de la polymérase au cours de la transcription ; toutefois, pour des raisons de clarté, le recrutement des protéines au CTD et ses conséquences seront abordés dans les sections suivantes.

La polymérase est recrutée hypophosphorylée aux promoteurs lors de la formation du PIC [338]. Le médiateur facilite le recrutement des facteurs généraux de la transcription TFIID et TFIIE [339], et stimule l'activité de la sous-unité kinase de TFIID (Kin28 chez la levure), responsable de la phosphorylation des Ser5 et 7 [264, 340]. La phosphorylation de la Ser5 du CTD favorisant l'évasion du promoteur, le médiateur contribue donc à sa propre dissociation de l'ARNPolIII [260, 340]. La phosphorylation de la Ser7 reste une marque énigmatique qui, pour l'instant, n'a été impliquée que dans la transcription des petits ARN non codants (snRNA), où elle semble affecter le recrutement du complexe intégrateur [341].

Peu après le début de la phase d'élongation, on observe une nette diminution du niveau de phosphorylation de la Ser5 et une augmentation du niveau de la Ser2 [330, 342-344]. Il y a donc une transition de la P-Ser5 à la P-Ser2 ; toutefois, un niveau appréciable de Ser5 reste présent jusqu'à la fin de l'élongation [345]. La déphosphorylation de la Ser5 à proximité du promoteur des gènes est attribuée à l'action de Rtr1. En effet, la délétion de *RTR1* aboutit à une augmentation du niveau de phosphorylation de la Ser5 sur l'ensemble du cadre de lecture [346]. Cependant, elle n'affecte pas le profil de phosphorylation de la Ser7 qui reste donc constante [333].

Chez la levure, la Ser2 est phosphorylée par 2 kinases : Bur1 et Ctk1 [330, 347]. Leur localisation indique qu'elles sont toutes les deux présentes sur la région codante des gènes [344]. Les protéines Bur (1 et 2) ont été isolées en 1993, lors d'un criblage pour des mutations capables de pallier la nécessité d'une séquence activatrice (BUR : *bypass UAS requirement*) [348], mais le couple kinase/cycline Bur1/Bur2 ne

fut identifié que plus tard [349]. Durant longtemps, les chercheurs ont considéré que les fonctions de Bur1 et de Ctk1 étaient assumées par un unique complexe chez l'humain, P-TEFb, comportant la sous-unité kinase Ctk9 [350]. Cette vision a été remise en question lorsque le groupe de Greenleaf identifia Ctk12 comme orthologue de Ctk1 chez les métazoaires, rétablissant ainsi la symétrie entre les organismes (Bur1/Ctk9, Ctk1/Ctk12)[351]. Bur1 est recrutée par la phosphorylation de la Ser5 et serait responsable d'une partie de la phosphorylation du CTD à proximité du promoteur [352]. Ctk1 est, par contre, responsable de la majorité de la phosphorylation du CTD le long de l'unité de transcription [336, 345, 352]. La phosphorylation de la Ser2 n'influence pas la déphosphorylation de la Ser5, car le blocage de la Ser2 n'augmente pas le niveau de Ser5 [300, 352]. S'appuyant sur l'analyse des niveaux de Ser2 dans des souches mutantes *fcp1* et *ctk1*, l'augmentation progressive du niveau de Ser2 phosphorylée, résulterait, d'après Buratowski, des effets antagonistes de la kinase Ctk1 et de la phosphatase Fcp1 [336]. Fcp1 (*TFIIF associating CTD phosphatase*) est identifiée en 1997 chez la levure comme une protéine associée à l'activité CTD phosphatase de la levure interagissant avec Rap74, la sous-unité du FGT TFIIF [353]. Son homologue humain est identifié par double hybride et est associé à l'activité phosphatase de la forme hyperphosphorylée Ilo de la polymérase [354]. Reinberg confirme cette activité et suggère une fonction de Fcp1 dans le recyclage de la polymérase [355]. La spécificité de Fcp1 de *S. pombe* sur des peptides synthétiques semble indiquer une préférence pour la P-Ser2 [356].

Un article récent émet l'hypothèse d'une déphosphorylation de la Ser7 au cours de la transcription par une phosphatase non identifiée (un peu à la manière de la Ser5), suivie d'une rephosphorylation par la kinase Bur1 [343].

Lors de la terminaison, toutes les marques de phosphorylation sont retirées, afin de permettre le recyclage de la polymérase au PIC. Ssu72 est la seconde phosphatase à agir sur la Ser5 et fut même pressentie, avant la découverte de Rtr1, comme la responsable de la déphosphorylation initiale de la Ser5 [357, 358]. *SSU72* (*Suppressor of SUA7, gene 2*) fut isolé dans un criblage ayant pour but d'identifier des mutations compensant la sensibilité au froid engendrée par la mutation *sua7-1* du

gène codant le FGT TFIIB [359]. Ssu72 serait capable d'interagir directement avec Rpb2 (une sous-unité de l'ARNPolIII) [360, 361]. De plus, Ssu72 interagit génétiquement avec kin28 et Ctk1 et Fcp1 [362]. Ssu72 est une sous-unité du complexe de clivage et de polyadénylation (CPF, *cleavage and polyadenylation factor*) recruté à la polymérase en 3' des gènes [363-365] et nécessaire pour la terminaison des petits ARN nucléolaires [361, 362]. En accord avec ces études, la localisation par CHIP place Ssu72 majoritairement en 3' des gènes [364]. En 2004, Hampsey, grâce à l'utilisation de l'anticorps H14 (Ser5-P), démontre *in vitro* et *in vivo* que Ssu72 est une phosphatase spécifique de la sérine 5 [357]. Enfin, comme précédemment évoqué, la fonction de TFIIB dans le bouclage des gènes est dépendante de Ssu72 [253]. La déphosphorylation de la Ser5 par Ssu72 est stimulée par l'activité peptidyl-prolyl isomérase (PPI) d'Ess1 (Pin1, chez les mammifères) [366, 367]. Découverte par Hanes en 1989 [368], Ess1, essentiel chez *S. cerevisiae*, se fixe au CTD de l'ARNPolIII phosphorylé en Ser5, par l'intermédiaire d'un domaine WW, et isomérisse le lien peptidyl-proline Ser5-Pro6 [228, 369, 370]. Fcp1 intervient pour compléter la déphosphorylation de la Ser2 et permettre le recyclage de la polymérase [355, 371]. Il n'est cependant pas clairement établi si cette déphosphorylation a lieu à la chromatine ou bien après le décrochage de la polymérase lors du recyclage au PIC [336, 355].

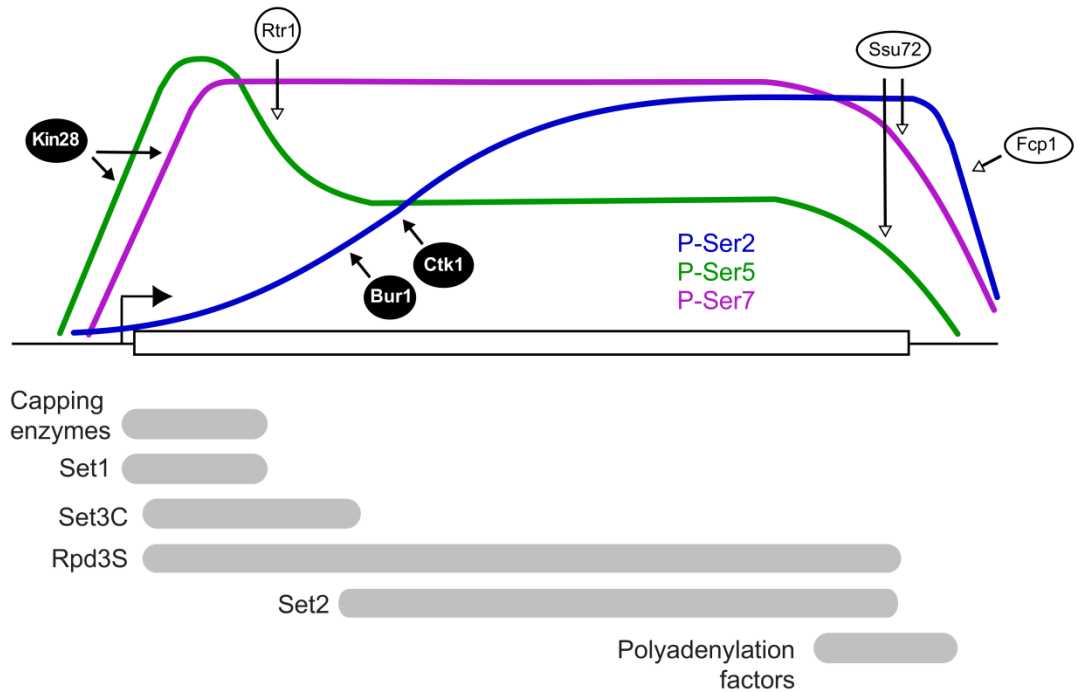


Figure 1.9 : Représentation graphique du cycle de phosphorylation du CTD.

Profil des marques phospho-sérine 2 (bleu), phospho-sérine 5 (vert), phospho-sérine 7 (mauve) le long d'un gène typique. Les kinases et phosphatases responsables de l'établissement des profils sont représentées en noir et en blanc respectivement. Les barres grises dans la partie inférieure représentent les régions de recrutement des facteurs par le CTD.

1.2.4.2 La fonction du CTD

Le rôle principal du CTD est de servir de plateforme de recrutement aux facteurs de maturation des ARNm durant la transcription, et ainsi de coupler les étapes de maturation de l'ARN à la progression de la polymérase le long de l'unité de transcription (**Figure 1.9**) [104, 224, 372]. Ses nombreux états ont mené à l'hypothèse d'un code du CTD permettant d'orchestrer les événements liés à la transcription [227, 331, 373].

1.2.4.2.1 Maturation de l'ARNm

Chez les eucaryotes, la maturation des ARNm comprend l'ajout d'une coiffe, l'épissage des ARN contenant des introns, et l'addition d'une queue polyA.

La phosphorylation de la Ser5 du CTD de la polymérase est responsable du recrutement du complexe de coiffage des ARNm (**Figure 1.9**) [374-376]. Trois activités catalytiques participent à l'ajout d'un groupe 7-méthyl-guanosine à l'extrémité 5' du transcrit par une liaison 5'-5' triphosphates [377]. Cette coiffe prévient la dégradation du transcrit par une exonucléase et est donc essentielle à la stabilité de l'ARNm, mais joue aussi un rôle dans son transport au cytoplasme et sa traduction [372, 378]

Chez les eucaryotes supérieurs, l'épissage alternatif est un mécanisme qui augmente considérablement la complexité du protéome. Il est réalisé par le large complexe du spliceosome. Contrairement au coiffage, l'épissage des ARNm peut être catalysé indépendamment de la transcription et n'a pas besoin de lui être couplé ; il a d'ailleurs, durant longtemps, été étudié en dehors de celle-ci. Toutefois, l'épissage est grandement stimulé par son couplage à la transcription [379]. La cotranscriptionnalité de l'épissage fut d'abord suggérée par l'observation au microscope électronique de boucles d'introns au niveau d'ARNm en cours de synthèse, lors de la transcription de gènes embryonnaires de drosophile [379, 380]. L'interaction entre le CTD de l'ARNPolIII et le spliceosome est moins bien caractérisée que pour les autres étapes de maturation de l'ARNm. Toutefois, le CTD phosphorylé semble stimuler l'épissage des gènes, et s'avère capable d'interagir avec certains facteurs d'épissage [381-383].

La vitesse d'élongation de l'ARNPolIII influence la sélection des sites d'épissage; ainsi, une polymérase plus lente laissera plus de temps aux sites « faibles » pour s'associer au spliceosome avant que les sites forts ne soient transcrits [379].

Enfin l'ARNm est pris en charge par le complexe de clivage et de polyadénylation (**Figure 1.9**). La protéine Pcf11 impliquée dans la polyadénylation se lie préférentiellement au CTD phosphorylé en Ser2 [345]. Rtt103, capable de fixer la P-Ser2 du CTD, peut aider au recrutement de l'exonuxléase Rat1, impliquée dans les modèles de terminaison torpédo ou hybride [345]. On peut remarquer que la voie de terminaison dépendante de Nrd1, utilisée pour les petits ARN, est logiquement inhibée par la présence de la marque de phosphorylation de la Ser2, alors que la phosphorylation de la Ser5 lui est requise [384].

1.2.4.2.2 Modification de la chromatine

Le recrutement de l'ARNPolIII au promoteur des gènes dépend des modifications de la chromatine, mais la machinerie de transcription va, à son tour, participer à la modification de la chromatine le long de l'unité de transcription lors de l'élongation. Tout comme la maturation des ARNm, ces modifications sont majoritairement coordonnées par le CTD de l'ARNPolIII.

Les régions transcrites sont associées avec trois marques de méthylations, H3K4, H3K36 et H3K79, qui sont catalysées par les méthytransférases Set1, Set2 et Dot1 respectivement (**Figure 1.9**) [66, 125, 385]. Parmi ces méthytransférases, deux, Set1 et Set2, sont recrutées aux régions transcrites de manière dépendante de la phosphorylation du CTD [386]. Le recrutement de Dot1 lors de la transcription est plus mystérieux. On sait, cependant, que la méthylation de H3K79, localisée dans le domaine globulaire de H3, est dépendante de l'ubiquitination du résidu H2BK123 par Rad6 et Bre1, elle-même dépendante du recrutement du complexe PAF [387] [388]. La transcription est aussi associée avec une acétylation transitoire des histones pour permettre le passage de la polymérase, et une déacétylation dans son sillage. Toutefois, le mécanisme de recrutement de l'HDAC responsable de la déacétylation a été revu récemment [198, 389].

La méthyltransférase Set1 est spécifique de la marque d'histone H3K4, puisqu'une délétion de Set1 abolit les trois formes de cette marque (mono-, di- et triméthylation). Elle fait partie d'un complexe de huit sous-unités, appelé COMPASS (*Complex Proteins Associated with Set1*), dont le recrutement est stimulé par la phosphorylation de la Sérine 5 du CTD (**Figure 1.9**) [390]. Le recrutement de COMPASS aboutit à la déposition d'un pic de méthylation H3K4me3 aux promoteurs des gènes en aval duquel se trouvent les marques H3K4me2 puis H3K4me1 [45, 125]. Fait intéressant, les marques di- et triméthyl de H3K4 sont dépendantes de l'ubiquitination de l'histone H2B à la lysine 123 par le complexe Rad6/Bre1 (comme la marque H3K79), mais la marque de monométhylation n'est que modérément affectée [391-393]. La kinase Bur1, impliquée dans la phosphorylation du CTD de l'ARNPolIII, catalyse également la phosphorylation de Rad6, et est nécessaire au recrutement de PAF1. En conséquence, elle régule indirectement l'ubiquitination de l'histone H2B et la déposition des marques H3K4 et H3K79 [299, 394]. La marque H3K4me est impliquée dans la mémoire de la transcription [125]. Elle pourrait faciliter le recrutement de facteurs d'épissage par l'intermédiaire du remodelleur de la chromatine CHD1 [395] et, selon Buratowski, recruterait le complexe Set3 en 5' des gènes [396]. Set3, bien que possédant un domaine SET, n'a pas encore été associée à une activité méthyltransférase ; par contre, son complexe possède une activité déacétylase grâce à la présence des sous unités HDAC, Hos2 et Hst1 [397]. Peu de choses sont connues quant aux fonctions du complexe Set3, mais il pourrait agir de manière partiellement redondante avec le complexe HDAC Rpd3S [198] ou être impliqué dans la déposition de la variante d'histone H2A.Z, comme semblent l'indiquer ses interactions génétiques. La phosphorylation du CTD a aussi été impliquée dans le recrutement à la chromatine du complexe SET3 [398].

Set2 est spécifique de la marque d'histone H3K36 [399]. Set2 est recrutée à la chromatine des gènes transcrits de manière dépendant de la marque P-Ser2 du CTD [385, 400, 401]. La méthylation par Set2 est associée à la répression de la transcription [399]. On lui a attribué le recrutement du complexe Rpd3S [402], mais

cette hypothèse fut revisitée. Deux papiers récents ont démontré que, contrairement à ce qui était communément admis, le recrutement du complexe HDAC Rpd3S était dû à l'action conjuguée de DSIF et du CTD phosphorylé et non à la présence de la marque H3K36 [198, 389]. Indépendamment de son mode de recrutement, la présence de l'activité déacétylase du complexe Rpd3S à la chromatine est importante pour « refermer » la chromatine dans le sillage de la polymérase. En son absence, la structure de la chromatine devient trop permissive et de la transcription cryptique apparaît aux sites d'initiation non spécifiques [402].

1.3 Rationnel de la thèse

Comme développé dans l'introduction, la machinerie de transcription et la chromatine sont en interaction continue afin de garantir le bon déroulement de l'expression des gènes.

La variante d'histone H2A.Z, que de précédentes études avaient impliquée dans la régulation de la transcription, est essentielle pour de nombreux organismes, mais n'avait cependant pas été localisée. Durant mon doctorat, j'ai appliqué la technique de CHIP-chip, dont le protocole détaillé constitue le chapitre 2 de cette thèse, afin de cartographier la présence de l'histone H2A.Z dans le génome. Les résultats de cette étude, présentés au chapitre 3 de cet ouvrage, ont mis en évidence la présence de H2A.Z aux promoteurs de la majorité des gènes de la levure. Mes travaux, ainsi que les résultats similaires d'autres groupes publiés à la même époque, ont orienté la recherche vers un rôle général de H2A.Z à la transcription.

Le CTD de l'ARN polymérase II est un élément central dans la coordination des mécanismes périphériques à la transcription, incluant la maturation des ARN messagers et la modification de la chromatine. L'étude des modifications du CTD débutée depuis plus d'une décennie, a néanmoins laissé des zones d'ombre dans les mécanismes enzymatiques mis en jeu. Le deuxième objectif de mes recherches était de compléter, en tirant partie du développement de nouveaux anticorps, les connaissances accumulées sur la déposition des marques de phosphorylations et leur dynamique, ainsi que de mettre en évidence les interactions entre les enzymes responsables. Mes expériences de localisation génomique des Phospho sérines 2, 5 et 7 appuyées par des essais biochimiques ont permis de faire de nombreuses observations et d'enrichir l'image de ce mécanisme complexe.

Ces travaux sur la transcription, initiés durant mon doctorat, ont de plus débouché sur de nouveaux projets au sein du laboratoire.

2 Protocole de localisation génomique des modifications et des variantes d'histones par la méthode de ChIP-chip chez *S. cerevisiae*

2.1 Mise en contexte de l'article

La méthode d'immunoprécipitation de chromatine (ChIP) s'est imposée comme une technique extrêmement puissante dans l'étude des phénomènes nucléaires. Ceci est particulièrement vrai de son association aux puces à ADN (ChIP-chip), et plus récemment au séquençage à haut débit (ChIP-seq), lui permettant une couverture génomique. Comme toute méthode, le ChIP-chip pose la question des contrôles à utiliser et l'étude de la chromatine impose des stratégies distinctes de celles généralement adoptées pour l'étude des facteurs de transcription. En effet, la densité des histones sur le génome n'est pas constante; il est donc impératif, afin de ne pas créer d'artéfact, de compenser ces écarts. Dans le cas d'une marque d'histone, la comparaison se fait à cette même histone indépendamment de toutes marques. Dans le cas d'une variante d'histone, la comparaison se fait, idéalement, à celle qui lui est associée en dimer. Ce chapitre est un protocole détaillé de la localisation d'une variante d'histone par la méthode de ChIP-chip, de la culture des cellules à l'hybridation de la puce à ADN. Il contient de nombreuses notes destinées à guider l'expérimentateur dans les étapes critiques. Une stratégie similaire a été appliquée dans l'étude de H2A.Z chez la levure au chapitre suivant.

2.2 Contribution des auteurs

A. R. Bataille et **F. Robert** ont corédigé l'article

2.3 Manuscrit

Profiling Genome-Wide Histone Modifications and Variants by ChIP–chip on Tiling Microarrays in *S. cerevisiae*

Alain R. Bataille and François Robert

Article publié dans *Methods in Molecular Biology*, (2009); 543:267-79.

Key words: Chromatin immunoprecipitation, Location analysis, ChIP on chip, DNA microarrays, Chromatin, Histone modifications, Histone variants.

2.3.1 Summary

Chromatin immunoprecipitation coupled to DNA microarray has become a widely used method to study transcription factors and chromatin structure. Here, we provide a detailed protocol for the localization of the variant histone Htz1 in the *S. cerevisiae* genome. This protocol can easily be adapted to fit other purposes such as profiling histone modifications.

2.3.2 Introduction

Chromatin is an important player in the regulation of nuclear processes such as gene transcription, DNA replication, recombination, and DNA repair. This is possible because chromatin is a very plastic structure *in vivo*. Chromatin can be remodeled, that is to say that its level of condensation and the position of the nucleosomes can be modulated via the action of cellular factors. In addition, the protein moiety of the nucleosomes, the histones, are subject to a plethora of post-translational modifications that can occur either prior to or after chromatin assembly. Finally, many nonallelic variants exist for most histone genes. The incorporation of these variants into chromatin is another way for the cell to modify chromatin structure and function.

Tools to study chromatin *in vitro* have been around for a long time, but the study of chromatin *in vivo* has long been a challenge. A first generation of techniques to study chromatin *in vivo* was making use of nucleases [1-3]. These assays were based on the principle that nucleosomes protect DNA from the action of nuclease. Information about the composition and the epigenetic state of the nucleosomes cannot be obtained from those assays. The development of antibodies against specific histone modifications, coupled to a technology called Chromatin immunoprecipitation (ChIP), has made it possible to study the epigenetic states of specific DNA loci *in vivo* [4]. More recently, the ChIP assay has been coupled to DNA microarrays, enabling the study of chromatin on a large scale [5, 6].

In a ChIP experiment, an immunoprecipitation is made from crosslinked chromatin fragments, and the coimmunoprecipitated DNA is monitored by quantitative or semiquantitative PCR. The enrichment of a given locus in the immunoprecipitated material, compared to a control sample, is read as an indication that the immunoprecipitated protein was bound to that locus *in vivo*. The control sample is often a nonimmunoprecipitated extract or a mock immunoprecipitation. When the immunoprecipitation is performed using an epitope-tagged protein, the best control is perhaps an immunoprecipitation done on a control cell line that carries no tag [7]. When monitoring histone modifications or the presence of a histone variant, however,

these kinds of controls are not appropriate. This is because nucleosome density is not uniform along the genome. As a consequence, a high enrichment at a specific locus using an antibody directed against a given histone modification could be the sole consequence of a higher nucleosome density at that locus. Inversely, a low enrichment may be caused by a depletion of nucleosomes at the locus. To circumvent the problem, an immunoprecipitation performed against total nucleosomes is usually used as a control. This normalizes for nucleosome density and a more accurate measure of the modification state of chromatin is therefore achieved.

Here we detail a protocol used to map the variant histone H2A.Z (called Htz1 in yeast) across the *Saccharomyces cerevisiae* genome. In this experiment, a yeast strain was engineered to express a myc-tagged version of Htz1 and an HA-tagged version of H2B. Immunoprecipitations using anti-myc and anti-HA antibodies are performed from the same crosslinked extract, and both derived DNAs are labeled and hybridized on the same DNA microarray. The H2B (HA) IP is therefore used to normalize nucleosome density. A similar strategy was originally used and published in [8].

2.3.3 Materials

2.3.3.1 Growing and Crosslinking Cells

1. 50 mL of fresh logarithmic growth yeast culture in YPD. The strain we use has the following genotype (*MATa ura3-1, leu2-3,112, ade2-1, his3-11,15, trp1-1, can1-100, HTZ1::3xMyc, HTB1::3xHA*).
2. 37 % Formaldehyde.
3. Ice-cold TBS.

2.3.3.2 Breaking Cells

Lysis buffer (50 mM HEPES-KOH of pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate, 1 mM PMSF, 1 mM Benzamidine, 10 µg/mL Aprotinin, 1 µg/mL Leupeptin, 1 µg/mL Pepstatin).

1. Acid-washed glass beads (400–600 µm) (Sigma).

2. Bead beater.

2.3.3.3 Immunoprecipitation

1. Dynabeads Pan Mouse IgG (Dyna).
2. Anti-Myc antibody (9E10 from Santa Cruz Biotech.).
3. Anti-HA antibody (F-7 from Santa Cruz Biotech.).
4. Hematology Chemistry Mixer 346 (Fisher) or equivalent.
5. Magnet system MPC-S (Dyna).
6. Lysis buffer (Same as for breaking the cells).
7. PBS + BSA (PBS supplemented with 0.5 % Bovine Serum Albumin).
8. Wash buffer (10 mM Tris-HCl of pH 8.0, 250 mM LiCl, 0.5 % NP40, 0.5 % Na-deoxycholate, 1 mM EDTA).
9. TE (10 mM Tris of pH 8.0, 1 mM EDTA).

2.3.3.4 DNA Purification

1. TE/SDS (10 mM Tris of pH 8.0, 1 mM EDTA, 1 % SDS).
2. TE (10 mM Tris of pH 8.0, 1 mM EDTA).
3. DNase-free RNaseA solution (10 mg/mL in water).
4. 10 % SDS.
5. Proteinase K solution (20 mg/mL from Gibco).
6. Glycogen (20 mg/mL from Roche).
7. Phenol/chloroform/isoamyl alcohol (25:24:1 from Sigma).
8. 5 M NaCl.
9. 100 % freezer-cold ethanol.
10. 70 % freezer-cold ethanol.

2.3.3.5 Amplification and labeling

1. T₄ DNA polymerase (NEB).
2. Buffer NEB2.
3. BSA (10 mg/mL from New England Biolabs).

4. dNTP mix (20-mM each of dATP, dCTP, dGTP, and dTTP).
5. 3 M NaOAc of pH 5.2 (Sigma).
6. 0.1 M NaOAc of pH 5.2.
7. Glycogen (20 mg/mL from Roche).
8. Phenol/chloroform/isoamyl alcohol (25:24:1 from Sigma).
9. Freezer-cold 100 % ethanol.
10. Freezer-cold 70 % ethanol.
11. 5x ligase buffer (Invitrogen).
12. Annealed linker (see protocol later).
13. T₄ DNA ligase high concentration (Invitrogen).
14. 10x ThermoPol buffer (New England Biolabs).
15. 5-(3-aminoallyl)-2'-deoxyuridine-5'-triphosphate (aa-dUTP) (Ambion).
16. aa-dUTP dNTP mix (5 mM each of dATP, dCTP, dGTP, 3 mM dTTP, 2 mM aminoallyl-dUTP).
17. Monoreactive Cy-5 NHS Ester (GE Healthcare).
18. Monoreactive Cy-3 NHS Ester (GE Healthcare).
19. Oligo 1 (GCGGTGACCCGGGAGATCTGAATTC) (40 μ M stock).
20. Oligo 2 (GAATTCAGATC) (40 μ M stock).
21. *Taq* DNA polymerase (Invitrogen).
22. *Pfu* DNA polymerase (Fermentas).
23. Qiaquick PCR purification kit (Qiagen).
24. Phosphate wash buffer (80 % ethanol / 5 mM KPO₄ of pH 8.5 buffer).
25. Phosphate elution buffer (4 mM KPO₄ of pH 8.5 buffer).
26. 0.1 M Na₂CO₃ of pH 9.0 (from 1 M stock solution made fresh every few weeks to a month).

2.3.3.6 Hybridization

1. Ozone-free enclosure.
2. Microarray hybridization chamber (Agilent).
3. Hybridization oven with rotator slide rack (Agilent).
4. *S. cerevisiae* tiling array (Agilent).

5. Microarray gasket slide (Agilent).
6. Hybridization solution (0.05 mM Mes NaOH of pH 6.9, 30 % Formamide, 650 mM NaCl, 0.5 % sodium N-Lauroylsarcosinate, 6 mM EDTA, 250 ng/slide salmon sperm DNA, 80 μ g/slide yeast tRNA).
7. Wash solution A (6x SSPE, 0.005 % N-Lauroylsarcosine).
8. Wash solution B (0.06x SSPE).
9. Heat blocks at 95°C and 40°C.
10. Three staining dishes.
11. Slide rack.
12. Lab rotator orbital shaker (Barnstead) or equivalent.

2.3.4 Methods

2.3.4.1 Growing and Crosslinking Cells

In a ChIP experiment, cells are first treated with formaldehyde. Formaldehyde is applied directly to living cells and acts very quickly, which allows for the design of very sharp time course experiments if needed. Virtually any growth condition can be used depending on experimental needs. We typically use 50 mL of exponentially growing cells (i.e., OD_{600} : 0.6–1.0), which represents around $3 \times 10^8 - 5 \times 10^8$ cells. The crosslinking time is one of the first factors that should be optimized if initial attempts to ChIP a protein fail. It has been reported that excessive exposure to formaldehyde can make the subsequent sonication steps more difficult.

1. Grow 50 mL of yeast in YPD medium until OD_{600} reaches 0.6–1.0.
2. Add 1.4 mL 37 % formaldehyde directly to the flask or transfer the contents of the flask into a 50-mL conical tube containing 1.4 mL formaldehyde.
3. Incubate 30 min at room temperature with agitation.
4. Centrifuge in 50-mL conical tubes for 5 min at 3,000 rpm (2047 g) in a refrigerated (4°C) table top centrifuge.

5. Pour off the supernatant (discard as chemical waste) and wash the cells by resuspending into 40 mL of ice-cold TBS (vigorous shaking is required) and centrifuge again as in **step 4**.
6. Repeat **step 5** for a total of two washes.
7. Using the remaining liquid, resuspend the cell pellet and transfer into a 1.5-mL screw cap tube.
8. Centrifuge 1 min at maximum speed, remove the supernatant with a pipette, and snap freeze in liquid nitrogen.
9. Samples can either be used directly or stored at -80°C .

2.3.4.2 Breaking Cells

Yeast cells are broken by vigorous agitation in the presence of glass beads (often referred to as bead beating). We do not recommend using regular vortexers since they tend to break rapidly due to overuse. We use a bead beater from Biospec Products sold in an 8-tube or 48-tube format.

1. Thaw cell pellet on ice.
2. Resuspend in 700 μL lysis buffer.
3. Add the equivalent of a 0.5-mL PCR tube of glass beads.
4. Screw cap tightly and secure tubes to the bead beater.
5. Beat for 5 min, and then put into ice-water bath for 5 min.
6. Repeat **step 5** three more times (see **Note 1**).
7. Punch a hole at the bottom using an 18G1 needle (do not go all the way through. Stop as soon as liquid or air is going through) and set up over a 2-mL tube.
8. Loosen the cap and spin 3-4 s. This should allow the material to transfer into the 2-mL tube while the beads stay trapped in the top tube. This operation may not work with all types of centrifuge. We use an Eppendorf 5417C with a 30-slot rotor.
9. Resuspend the chromatin pellet with its own supernatant and transfer into a new 1.5-mL tube (see **Note 2**).

2.3.4.3 Preparation of Soluble Chromatin Fragments

Just as for cell breakage, there is a lot of variation in the efficiency of the sonication due not only to the type of sonicator used, but also the type and age of the probe used. This step therefore needs to be optimized, which can be done by looking at the size of the sheared DNA on an agarose gel. We suggest omitting the formaldehyde treatment when optimizing the sonication in order to remove the need for reversing the crosslinking reaction. The protocol described here was optimized for a Fisher dismembrator 100 equipped with a microprobe.

1. Sonicate the sample four times for 20 s at power 1.5 (for an output of 7 Watts). Put the tube on ice for at least a minute between every round of sonication (see **Note 3**).
2. Centrifuge 5 min at maximum speed and transfer the supernatant (which contains solubilized chromatin fragments) to a new 1.5-mL tube. From now on, this will be referred to as the whole-cell extract (WCE).

2.3.4.4 Immunoprecipitation

Immunoprecipitation can be done with a great variety of systems. Both monoclonal and polyclonal antibodies have been shown to work for ChIP. All sorts of matrices such as agarose, Sepharose, or magnetic beads were used as well. The protocol described here uses magnetic beads coupled to an anti-myc and anti-HA antibody.

1. Set up a tube with 500 μ L of WCE and 30 μ L of anti-Myc coupled-magnetic beads (see **Note 4**).
2. Incubate overnight at 4°C with agitation (we use a Hematology Chemistry Mixer 346 from Fisher).
3. Wash the beads twice with 1 mL lysis buffer, twice with 1 mL lysis buffer supplemented with 360 mM NaCl, twice with 1 mL Wash buffer, and once with 1 mL TE (see **Note 5**).

4. After the last wash (TE), centrifuge briefly and remove the last bit of liquid using a P200 pipette. This should leave a dry bead pellet.

2.3.4.5 DNA Purification

The crosslink between DNA and the proteins needs to be reversed. This is done here by the action of Tris which, with heat, attacks the chemical bonds generated by the formaldehyde and liberates free DNA. The DNA is purified by a succession of RNase A and proteinase K extractions, followed by phenol extractions and ethanol precipitation.

1. To the bead pellets (from **subheading 3.4, step 4**), add 50 μ L of TE/SDS, vortex, and incubate overnight at 65°C (use an oven instead of a water bath in order to minimize evaporation and condensation in the cap of the tubes).
2. Centrifuge 3 min at 3,000 rpm (300 g) and transfer the supernatant to a new tube.
3. Add 350 μ L of a mixture (345 μ L TE, 3 μ L RNase A, 2 μ L glycogen), mix by vortexing, and incubate 2 h at 37°C.
4. Add 15 μ L of 10 % SDS and 7.5 μ L of proteinase K solution, mix by vortexing, and incubate two more hours at 37°C.
5. Extract twice with 400 μ L phenol/chloroform/isoamyl alcohol.
6. Add 12–16 μ L 5 M NaCl (200 mM final concentration), vortex; add 1 mL ethanol, vortex, and centrifuge 20 min at maximum speed at 4°C.
7. Pour off liquid and wash with 70 % ethanol.
8. Centrifuge 5 min and pour off liquid.
9. Centrifuge briefly, remove the last bit of liquid with a pipette, and let dry 1 min (Do not overdry the pellet).
10. Resuspend the pellet (which should be very small but still visible) in 50 μ L TE.
11. Store at -20°C.

2.3.4.6 Amplification and Labeling

The immunoprecipitated DNA fragments are amplified and labeled by ligation-mediated PCR (LM-PCR), where aminoallylmodified dUTP is incorporated into the product. The labeling is done post-PCR using monoreactive Cy-dye NHS esters that will react specifically with the aminoallyl-modified dUTP (see **Note 6**). The source of reagents is indicated here as a suggestion (see **Note 7**).

2.3.4.6.1 Blunting

1. Transfer 40 μL of immunoprecipitated DNA to a 0.5-mL PCR tube. Keep the sample on ice.
2. Add 70 μL of blunting mix (11 μL 10x NEB2 buffer, 0.5 μL BSA, 0.5 μL dNTPs, 0.2 μL T₄ DNA pol, and 57.8 μL ice-cold water).
3. Mix by pipetting and incubate 20 min at 12°C (we use a thermocycler block).

2.3.4.6.2 Ligation

1. Place on ice and add 12 μL of (11.5 μL 3 M NaOAc and 0.5 μL glycogen).
2. Vortex and add 120 μL of phenol/chloroform/isoamyl alcohol.
3. Vortex hard and centrifuge 5 min at maximum speed.
4. Transfer 110 μL to a new 1.5-mL tube and add 230 μL ethanol.
5. Vortex and centrifuge 20 min at 4°C.
6. Pour off supernatant and add 500 μL 70 % ethanol.
7. Vortex and centrifuge 5 min at 4°C.
8. Pour off supernatant, centrifuge briefly, and remove the remaining liquid with a pipette.
9. Let air dry 1 min, add 25 μL ice-cold water, and put on ice for about 30 min.
10. Vortex, centrifuge briefly, and put on ice.
11. Add 25 μL of ligase mix (8 μL water, 10 μL 5x ligase buffer, 6.7 μL annealed linkers, 0.5 μL T₄ DNA ligase) (see **Note 8** for linker preparation).
12. Mix by pipetting and incubate overnight at 16°C (a waterbath is fine for this step).

2.3.4.6.3 PCR

1. Add 6 μL of 3 M NaOAc, vortex, and add 130 μL EtOH.
2. Vortex to mix and centrifuge 20 min at 4°C, maximum speed.
3. Pour off supernatant and add 500 μL 70 % ethanol.
4. Vortex and centrifuge 5 min at 4°C, maximum speed.
5. Pour off supernatant centrifuge briefly and remove remaining liquid with a pipette.
6. Let air dry for a minute (do not overdry the pellet), and resuspend in 25 μL ice-cold water by pipetting up and down and rinsing the side of the tube (the pellet at this stage is usually diffused on the side of the tube).
7. Leave on ice for about 30 min, vortex, centrifuge briefly, and put on ice.
8. Add 15 μL of labeling mix (4 μL 10x ThermoPol buffer, 7.75 μL water, 2 μL aa-dUTP dNTP mix, 1.25 μL oligo 1).
9. Mix by pipetting and transfer to a new 0.5-mL PCR tube.
10. Put in thermocycler and start the program (see later).
11. When temperature reaches 55°C, add 10 μL of enzyme mix (8 μL water, 1 μL 10x ThermoPol buffer, 1 μL *Taq* polymerase, 0.01 μL *Pfu* polymerase) and mix by pipetting (see **Note 9** for PCR program).
12. Purify PCR reactions on Qiaquick PCR purification kit. Use the manufacturer's protocol, but replace buffer PE with the Phosphate wash buffer and buffer EB with the Phosphate elution buffer. Do the wash step twice. Elute twice with 30 μL phosphate elution buffer (60 μL total volume).
13. Dry the eluates in a speed-vac.
14. Resuspend the pellets in 4.5 μL fresh 0.1 M Na_2CO_3 of pH 9.0 buffer.
15. Add 4.5 μL of the appropriate Cy dye (we use Cy5 for the IP and Cy3 for the control) (see **Note 10**).
16. Incubate 1 h in the dark at room temperature.
17. Add 35 μL 0.1 M NaOAc of pH 5.2.
18. Add 250 μL buffer PB (Qiagen) and purify on Qiaquick PCR purification columns, using the manufacturer's protocol and using the manufacturer's

buffers. Do the wash step twice. Elute twice with 30 μ L buffer EB (60 μ L total volume) (see **Note 11**).

19. Dry the eluates in a speed-vac (see **Note 12**).
20. Store the dry colored pellets at -20°C in the dark.

2.3.4.7 Hybridization

Hybridization is done in Agilent hybridization chambers. This hybridization technology ensures a homogenous hybridization of the slide surface. All the process and especially the washes and scanning are done in an ozone-free chamber to minimize the degradation of the Cy dyes (see **Note 13**). More details concerning Agilent chamber manipulation and Agilent hybridization technology can be found in the Agilent microarray hybridization chamber user guide (provided with each chamber or on their web site).

2.3.4.7.1 Hybridization

1. Resuspend Cy3-labeled sample in 4 μ L of water and transfer to the corresponding Cy5-labeled sample (see **Note 14**).
2. Add 450 μ L of fresh hybridization solution to each combined sample (volumes from 400 μ L to 500 μ L will work) and mix by vortexing (see **Note 15**).
3. Spin briefly and denature samples for 3 min in a 95°C heat block.
4. Spin briefly and transfer to a 40°C heat block for another 10 min.
5. Assemble the gasket slide and the base of the hybridization chamber, and check for misalignment.
6. Dispense the sample on the gasket slide starting close to one end and slowly moving to the opposite end without touching the surface of the gasket slide (avoid bubbles).
7. Carefully cover with the array and check for misalignment (see **Note 16**).
8. Assemble the hybridization chamber upper part.
9. Slip the clamp assembly onto the chamber and tighten the thumbscrew so that it is fully hand tight.

10. Vertically rotate the slide to wet the coverslip and ensure that the bubbles are moving freely (see **Note 17**).
11. Incubate in rotating oven at 40°C, 20 rpm, overnight (don't forget to balance the rotor both horizontally and vertically).

2.3.4.7.2 Wash

1. Pour buffer A into a standard Staining dish with a slide rack (around 250 mL) and into another dish that will be used as a “disassembly dish” (see **Note 18**).
2. Remove the hybridization chamber from the oven, disassemble all metal parts, submerge the slide-gasket slide sandwich into the “disassembly dish” and separate the slide from and gasket slide using plastic tweezers (see **Note 19**).
3. Briefly rinse the slide in the “disassembly container” and transfer to the slide rack in the washing dish with buffer A.
4. Proceed to next slide (see **Note 20**).
5. Wash with buffer A on an orbital shaker for 5 min (start the timer when the last slide has been transferred).
6. Transfer the slide rack to a dish filled with buffer B and wash on the orbital shaker for another 5 min. Slowly remove the rack from the liquid (see **Note 21**).
7. Store the slides in the dark and proceed to scanning rapidly.

2.3.5 Notes

1. The amount of beating that is required may vary depending on the equipment, the yeast strain, the age of the culture, and the sample/beads ratio used. Cell breakage can be easily monitored by looking at cells under a light microscope.
2. Alternatively, the supernatant can be discarded and the chromatin pellet resuspended into 700 μ L of fresh lysis buffer. For proteins with a prominent soluble form (not associated with chromatin), this may prevent unbound proteins from competing for antibody during the subsequent immunoprecipitation step.

3. When doing multiple samples, sonicate each sample once and go back to the first one. Rinse the probe with distilled water between every sample in order to avoid crosscontamination. Avoid contact between the probe and the side of the tube. This would reduce the efficiency of the sonication.
4. The beads should be prepared the day prior to use in the following way. Take enough beads to get 50 μ L per immunoprecipitation. Put into a conical plastic tube and centrifuge 3 min at low speed to pellet the beads. Wash twice with PBS + BSA and incubate them overnight at 4°C with agitation in the presence of anti-Myc antibody (we use 2 μ g of 9E10 antibody per immunoprecipitation reaction) or anti-HA antibody (we use 2 μ g of F-7 antibody per immunoprecipitation reaction.) and enough PBS + BSA to ensure a good agitation. The next day (prior to use) wash twice in lysis buffer to remove nonbound antibodies and resuspend beads in 30 μ L of lysis buffer per immunoprecipitation.
5. The washes can be done quickly using a magnetic system (MCP-S from Dynal). Put the tubes to the magnetic system and wait a few seconds for the beads to get captured by the magnet. By keeping the tubes in the system, remove the liquid by aspiration or simply by inverting the tubes. Add the next washing solution, close the tubes, remove the magnet, and agitate to resuspend the beads. Put the magnet back on and remove the liquid as earlier. Process as so for the necessary steps.
6. This method is less expensive and yields better results than the direct incorporation of Cy-dye-modified dNTPs during PCR. We found that dye swap experiments are not necessary with this protocol.
7. These reagents represent a combination that gives good results in our laboratory. While it probably does not matter for most of the other reagents, we

found that using ligase buffers which contains polyethylene glycol and the high concentration ligase makes an improvement.

8. The annealed linker is prepared by annealing oligos 1 and 2 in the following way. Mix 250 μL 1 M Tris of pH 7.9, 375 μL oligo 1, and 375 μL oligo 2. Make 50- μL aliquots in 1.5-mL tubes. Place in a 95°C heat block for 5 min. Transfer tubes to a 70°C heat block with water in holes. Place the block at room temperature and let it cool down to 25°C. Transfer the block to 4°C and allow it to cool down overnight. Store annealed linkers at -20°C. The T_m of the linker is very low. Always keep the linkers on ice and use ice-cold buffers when manipulating them to avoid deannealing.
9. Use the following PCR program for amplification: step 1: 4 min at 55°C; step 2: 5 min at 72°C; step 3: 2 min at 95°C; step 4: 30 s at 95°C; step 5: 30 s at 55°C; step 6: 1 min at 72°C; step 7: Go back to step 4, 31 more times; step 8: 4 min at 72°C; step 9: 4°C forever.
10. Resuspend new tubes of the Cy3 and Cy5 dyes in 73 μL of DMSO (kept in the dark, under desiccant and vacuum). Store at -80°C.
11. The color should be visible on the filter and the eluate should have a faint but visible taint of blue/pink.
12. Optionally, the DNA amount and dye incorporation can be estimated by spectrometry prior to speed-vac using 50 μL of the undiluted elution (which is recovered from the cuvet after measurement). For this we measured the absorbance at 260, 320, 550 and 650 nm for Cy3 and 260, 320, 650 and 750 nm for Cy5 and apply the following formulas:

$$\text{DNA(mg)} = (A_{260} - A_{320}) 50\mu\text{g/mL} \times \text{Volume (mL)} \times \text{dilution factor}$$

$$\text{Base/Cy3} = (((A_{260} - A_{320}) - ((A_{550} - A_{650}) \times 0.08)) \times 150\,000) / ((A_{550} - A_{650}) \times 6\,600)$$

$$\text{Base/Cy5} = (((A_{260} - A_{320}) - ((A_{650} - A_{650}) \times 0.05)) \times 250\,000) / ((A_{650} - A_{750}) \times 6\,600)$$

$$\text{Dye molecules / 100bp} = 100 / (\text{base} / \text{dye})$$

Expect about 5 μg of DNA (typically between 4 and 6 μg) and an incorporation of 1.5 dyes per 100 bp (typically between 1.2 and 1.8 dyes per 100 bp).

13. Cyanide dyes (Cy5 particularly) are sensitive to even low levels of atmospheric ozone. Working under an ozone-free enclosure is especially important when the dyes are dry and thus directly exposed to the atmospheric ozone (after the last wash). Alternatively, Agilent sells a solution saturated with ozone scavengers that will help protect the slide from ozone during the scan.
14. Samples are easily resuspended in a small volume, but a larger volume can be used (adjust water volume in hybridization solution in consequence).
15. The hybridization solution is made fresh each time. While it probably does not matter for most of the other reagents, we found that using Sodium N-Lauroylsarcosinate from Fluka makes an improvement compared to other brands.
16. The active face of the array is the face reading “Agilent”; thus, this face should face the gasket slide while you are facing the side reading the slide number.
17. It is ideal to obtain just one “mixing” bubble; however, small bubbles are fine as long as they can move freely between the slides. Static bubbles at the edge of the gasket can be dislodged by gently tapping the corner of the assembled chamber on a firm surface.
18. We use a glass staining dish for the disassembly dish, but black plastic ones for the subsequent wash in order to protect the slides from the light as much as possible.

19. Insert one end of the tweezers between the slide and the gasket slide in the barcode area and then gently turn the tweezers to separate the slides. Let the gasket slide drop to the bottom of the dish while you are still holding the array with the tweezers. To transfer the slide into the rack it is easier and safer to manipulate the slide by holding it at the barcode area with your hand.
20. Up to ten slides can be processed in parallel in the same slide rack.
21. It is important to slowly remove the slides from the liquid. The surface of the slides will come out dry and usually don't need to be centrifuged or air blown.

2.4 Références

1. Wu, C., *Analysis of hypersensitive sites in chromatin*. Methods Enzymol., 1989. **170**: p. 269-89.
2. Bellard, M., et al., *Nuclease digestion of transcriptionally active chromatin*. Methods Enzymol., 1989. **170**: p. 317-46.
3. Lutter, L.C., *Digestion of nucleosomes with deoxyribonucleases I and II*. Methods Enzymol., 1989. **170**: p. 264-9.
4. Kuo, M.H. and C.D. Allis, *In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment*. Methods Enzymol., 1999. **19**: p. 425-33.
5. Kurdistani, S.K., S. Tavazoie, and M.R. Grunstein, *Mapping global histone acetylation patterns to gene expression*. Cell, 2004. **117**: p. 721-33.
6. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast*. Cell, 2005. **122**: p. 517-27.
7. Drouin, S. and F. Robert, *Genome wide location analysis of chromatin-associated protein by ChIP on CHIP: Controls Matter*. In Press, 2006.
8. Guillemette, B., et al., *Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning*. PLoS Biol., 2005. **3**: p. e384.

3 La variante d'histone H2A.Z est globalement localisée au promoteur des gènes inactifs de la levure et régule le positionnement des nucléosomes

3.1 Mise en contexte de l'article

Dès sa découverte qui l'associe au macronucleus, transcriptionnellement actif, de tetrahymena, H2A.Z est liée à la transcription. Les études subséquentes ont permis de renforcer ce lien (notamment aux gènes GAL) jusqu'aux expériences d'expression par puce à ADN qui ont établi le rôle de H2A.Z pour contrer la propagation de l'hétérochromatine silencieuse chez la levure. Toutefois, ce n'est que fin 2005 que plusieurs groupes s'attacheront à cartographier H2A.Z dans le génome de la levure. Ces travaux, dont fait partie l'article présenté à ce chapitre, suggèrent qu'elle prépare la chromatine pour l'activation et ont relancé l'intérêt des chercheurs pour cette variante. La localisation de H2A.Z en bordure du NFR, de la majorité des gènes, a permis de mieux comprendre la structure de la chromatine au promoteur et le rôle de H2A.Z dans la transcription. Les études chez les eucaryotes supérieurs qui ont suivi révèlent la conservation de la fonction.

3.2 Contribution des auteurs

F. Robert et L. Gaudreau ont conçu les expériences. **B. Guillemette, A.R. Bataille** ont réalisé les expériences de ChIP-chip. **N. Gévry et M. Adam** ont réalisé les expériences de positionnement des nucléosomes. **A.R. Bataille et M. Blanchette** ont analysé les données. **F. Robert et L. Gaudreau** ont rédigé le papier.

3.3 Manuscrit

Variant Histone H2A.Z is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning

Benoît Guillemette^{1#}, Alain R. Bataille^{2#}, Nicolas Gévry¹, Maryse Adam¹, Mathieu Blanchette³, François Robert^{2*}, and Luc Gaudreau^{1*}

[#]These authors made equal contributions to this work, ^{*}Corresponding authors

Article publié dans PLoS Biology (2005) (12):e384.

3.3.1 Summary

Background: H2A.Z is an evolutionarily conserved histone variant involved in transcriptional regulation, anti-silencing, silencing and genome stability. The mechanism(s) by which H2A.Z regulates these various biological functions remains poorly defined, in part due to the lack of knowledge regarding its physical locations along chromosomes as well as the bearing it has in regulating chromatin structure. **Methodology/Principal Findings:** Here we mapped H2A.Z across the yeast genome at a ~300 bp resolution using chromatin immunoprecipitation combined with tiling microarrays. We have identified 4862 small regions -typically one or two nucleosomes wide - decorated with H2A.Z. Those “Z loci” are predominantly found within specific nucleosomes in the promoter of inactive genes all across the genome. Furthermore, we have shown that H2A.Z can regulate nucleosome positioning at the *GALI* promoter. Within HZAD domains, the regions where H2A.Z shows an anti-silencing function, H2A.Z is localized in a wider pattern, suggesting that the variant histone regulates anti-silencing and transcriptional activation via different mechanisms. **Conclusion/Significance:** Our data suggests that the incorporation of H2A.Z into specific promoter-bound nucleosomes configures chromatin structure to poise genes for transcriptional activation. The relevance of these findings to higher eukaryotes is discussed.

3.3.2 Introduction

Eukaryotic chromosomes are packaged into a complex nucleoprotein structure termed chromatin, from which the fundamental unit is the nucleosome. Nucleosomes are composed of octamers of histone proteins designated as H2A, H2B, H3, and H4 that wrap ~146 bp of DNA [1, 2]. Chromatin is generally highly repressive to cellular processes that involve DNA transactions such as gene transcription [3]. Cells have thus devised at least three schemes to overcome the nucleosomal barrier: the first involves ATP hydrolysis to mechanically displace nucleosomes [4, 5], and the second consists of chemically modifying the tails of histones such as by acetylation, phosphorylation, methylation, and ubiquitination [2]. This last scheme can take the form of a “code” where modifications act as signals for subsequent chromatin modifications [6, 7]. A third scheme that has been devised by cells to overcome the nucleosomal barrier is to alter the composition of nucleosomes through the incorporation of histone variants that can – directly or indirectly – alter the permissiveness of chromatin to gene expression [8, 9].

Non-allelic variants have been described for many classes of histones, and one of the best-studied examples is the Z variant of H2A. H2A.Z, formerly called H2A.F/Z, is evolutionarily conserved from yeast to mammals [10, 11]. H2A.Z is essential for the viability of *Tetrahymena*, *Drosophila* and mice [12-14] and constitutes approximately 10 % of all cellular H2A molecules in S-phase [15, 16]. Experiments carried out in many organisms have shown that H2A.Z has a function distinct from the canonical H2A [13, 17, 18]. H2A variant histones differ mainly in their N- and C-terminal portions as compared to S-phase H2A, whereas the core region is highly conserved [16]. In fact, experiments carried out in *Drosophila* and in yeast have revealed that the unique feature of H2A.Z important for its specific function resides in the carboxyl-terminal region of the variant histone and not in its histone fold [18, 19].

In *Tetrahymena thermophila*, H2A.Z has been found to be associated with the transcriptionally active macronucleus thus implying a positive role for H2A.Z in gene transcription [20]. Subsequently, experiments carried out in yeast have shown that

H2A.Z could regulate transcription [19, 21], and that its function is partially redundant with nucleosome remodeling complexes such as SAGA and Swi/Snf [21]. H2A.Z has been shown to be quickly remodeled soon after gene induction [19, 21-23], a result that implies that the role of H2A.Z in positive gene transcription occurs at an early step. In line with this idea, we have shown that the transcriptional machinery is not efficiently recruited to the *GALI* (SGD:S00000224) promoter in *htz1* Δ cells [19]. Another mechanism by which H2A.Z could contribute to gene regulation has been suggested by Farris et al. [23] where they have shown at the *c-myc* gene that transcription-induced chromatin transactions involves the exchange of H2A.Z for canonical H2A.

In addition to its well established role in transcriptional activation, a recent important piece of work by the Madhani group showed that H2A.Z positively regulates clusters of genes that are proximal to telomeres [24]. The gene expression defect observed at those gene clusters (called HZAD domains) in H2A.Z mutants can be suppressed by the deletion of silencing proteins such as Sir2 (SGD:S000002200). It has therefore been suggested that H2A.Z functions as an anti-silencing factor by antagonizing the spread of heterochromatin to euchromatic regions [24]. The expression of other H2A.Z-affected genes, however, is not sensitive to Sir proteins suggesting H2A.Z positively regulates gene expression via both anti-silencing and other mechanisms.

Evidence also suggests that H2A.Z can negatively regulate gene expression, notably through gene silencing. In *Drosophila* and in yeast, H2A.Z was reported to be involved in heterochromatic silencing [25, 26]. In mammalian cells, H2A.Z co-localizes with, and directly interacts with, heterochromatin protein 1 (HP1- α) (LocusLink:23468) as well as with INCENP (LocusLink:3619) – a protein involved in chromosome segregation – in pericentric heterochromatin [27, 28]. In addition, HP1- α and H2A.Z may function together to establish a specialized conformation at constitutive heterochromatin [27]. Thus, from these findings it is clear that H2A.Z is important both for the positive and negative regulation of gene expression.

To add to all of these recent findings, there has been the co-discovery of an ATP-dependent chromatin remodeling complex that specifically loads H2A.Z into

chromatin and exchanges it with H2A [29-31]. This complex, from which the catalytic subunit is Swr1 (SGD: S000002742), also shares essential subunits with the NuA4 histone acetyltransferase complex [29]. In addition to their importance in gene regulation, the Swr1 complex, H2A.Z, as well as NuA4 are all involved in the regulation of chromosome stability in yeast [32]. This is interesting and important since it was also demonstrated that in mammalian cells, knockdown of H2A.Z causes the genome to become unstable, a process that is attributed to defects in chromosome segregation [33]. Of equal importance is the fact that H2A.Z is required for early mouse and *Xenopus* development [13, 14].

Despite all the work reviewed above, the mechanisms by which H2A.Z regulates transcription, silencing, genome stability and development remain largely unknown. A key piece of data that is missing is a global view of the distribution of H2A.Z across the genome. Polytene chromosome staining has shown that *Drosophila* H2A.Z is widely distributed in the genome, but with non-uniform distribution that is characterized by a complex banded pattern [34]. From this data, it is thus conceivable that certain regions of the genome are preferentially occupied by variant histones and interspersed with regular histones. The resolution provided by this assay, however, is not high enough to allow one to correlate H2A.Z localization with its effect on gene expression. Experiments aimed at localizing proteins on defined chromatin segments *in vivo* – Chromatin Immunoprecipitation (ChIP) – have been performed at specific loci, but these experiments do not allow determining whether H2A.Z assembles preferentially within large chromosomal domains, whether it is randomly spread out across the genome, or whether it is associated with specific loci along chromosomes. Answering this question is a crucial step toward understanding the various functions of this important protein.

Here we make use of a genome-wide location assay (ChIP-chip) to map H2A.Z across the yeast genome with a 300 bp resolution. Our data show that H2A.Z preferentially and specifically associates with small regions within the promoters of inactive yeast genes. We also show that replacement of H2A.Z with regular H2A within the *GAL1* promoter region perturbs nucleosome positioning, a situation that might account for the transcription defect at the latter gene in the absence of H2A.Z.

Interestingly, the very strict promoter-specific association of H2A.Z that is otherwise observed elsewhere is not observed for genes contained within the HZAD domains. At those genes, the presence of H2A.Z can be detected within wider domains encompassing both the promoter region and part of the open reading frame (ORF).

3.3.3 Results

3.3.3.1 Genome-wide location analysis of H2A.Z

In order to investigate the genomic distribution of H2A.Z in yeast, we performed genome-wide location experiments (also known as ChIP-chip) on whole-genome tiling microarrays. The assay consists of first performing ChIP assays to immunoprecipitate (IP) our target DNA molecules of interest. Thus, yeast cells are first cross-linked with formaldehyde prior to being lysed. Chromatin fragments (~400 bp) are next obtained by sonication and specific antibodies are then used to IP our protein(s) of interest along with its target DNA. Purified DNA is next amplified by LM-PCR, labeled, and hybridized to an array that contains a total of 44,290 60-MER probes (including controls) for an average of one probe every 266 bp (see Materials and Methods for more details). Thus, ChIPs were performed using yeast strains bearing Myc epitopes on H2A.Z and on the canonical H2A. In every case, a ChIP for histone H2B (performed using an HA tagged-H2B strain) was hybridized together with the H2A.Z (or H2A) ChIP to control for nucleosome density. In addition, other control experiments where two Myc-H2A.Z ChIPs are hybridized against each other were also performed. All experiments were performed in duplicate and combined using a weighted average method. A Gaussian filter was then applied to each dataset in order to create a smooth interpolation occupancy curve. This method allowed for the identification of 4862 loci enriched in H2A.Z, referred to below as “Z loci”. The enrichment ratio for every probe on the microarray is available in **Table Sup. 3.1** whereas the smoothed data can be found in **Table Sup. 3.2-4**. The coordinates of all Z loci are listed in **Table Sup. 3.5**. All datasets were deposited into ArrayExpress under accession number E-MEXP-396.

3.3.3.2 H2A.Z is associated with small genomic regions present all across the yeast genome

The 4862 loci identified here are widely distributed across all chromosomes and no regions show any bias (see **Figure 3.1** for a map of chromosome III and **Table Sup. 3.6** to see the whole genome). Typical Z loci are less than 400 bp-long, suggesting that the variant histone occupies regions that are generally no larger than 2 nucleosomes. This estimation is based on four observations. First, the probes enriched in the H2A.Z experiments were not depleted in the H2A experiments (data not shown). Because the average size of the DNA fragments used in our ChIP assays is 400 bp, this suggests that most 400 bp fragments within the genome that contain H2A.Z will also contain H2A. Since, in principle, a nucleosome cannot contain both H2A and H2A.Z simultaneously [35], the data suggest that H2A.Z occupies regions that are smaller than 400 bp. Second, the sizes of the Z loci as they appear from the microarray data (see **Figure 3.1** for example) are consistent with H2A.Z occupying small regions, especially considering the fact that ChIP experiments necessarily overestimate the size of the region bound by proteins due to the size of the chromatin fragments generated. Third, high resolution Q-PCR analysis of ChIP experiments using small overlapping PCR probes allowed us to map the Z loci within the promoter of the *SRB8* genes (SGD:S000000677) to a ~260 bp region using TFIIB (SGD:S000006290) as a reference [36] (**Figure 3.1 D**; see **Text Sup. 3.1** and **Figure Sup. 3.1** for more details). Lastly, alignment of our data with global nucleosome positioning information from Yuan et al [37] revealed that H2A.Z maps to one or two nucleosomes in promoters (see below). Taken together, these analyses show that H2A.Z occupies very small regions scattered across the genome, rather than being concentrated within large chromosomal domains.

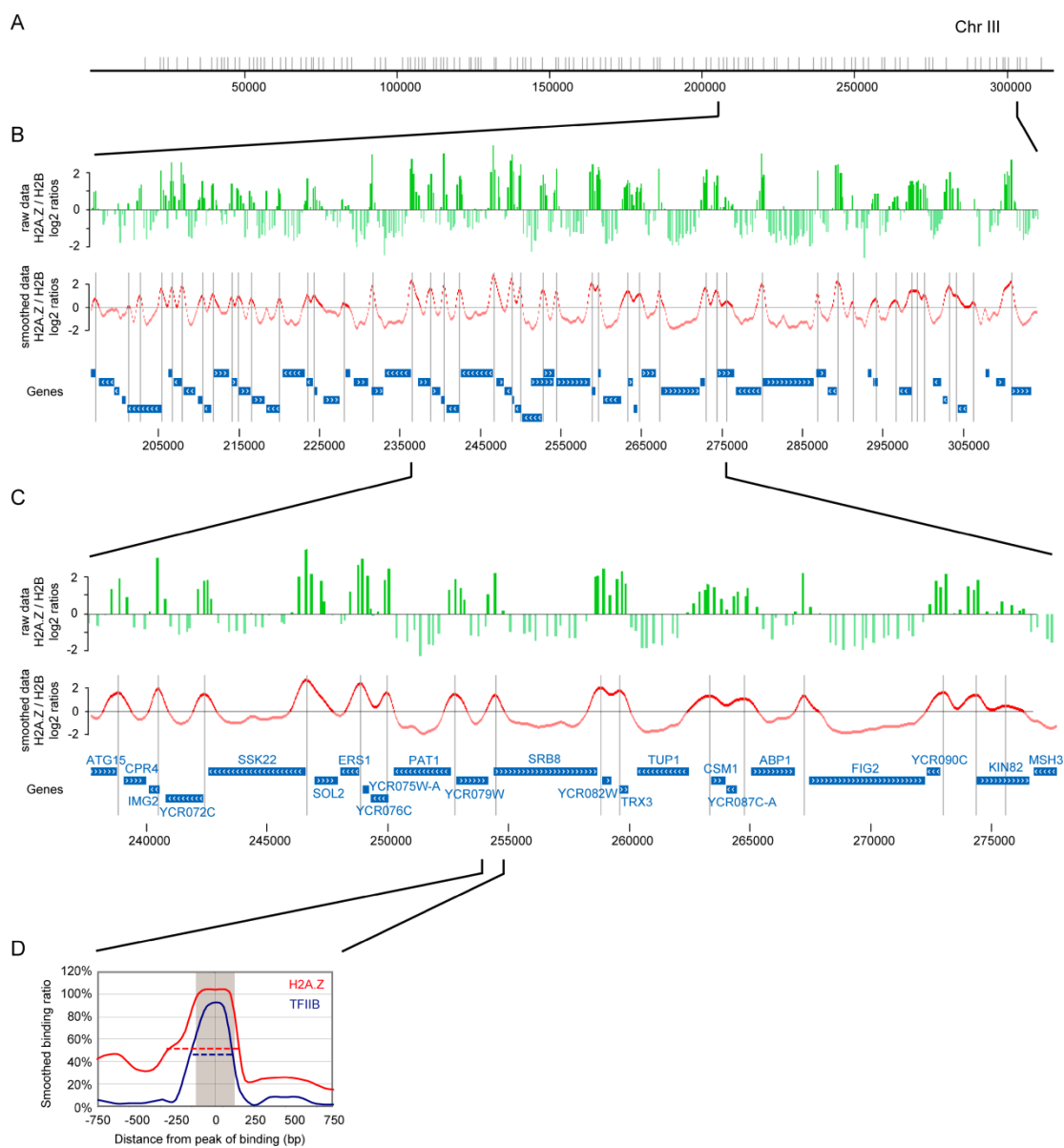


Figure 3.1: Genome-wide location analysis of H2A.Z and H2A; a zoom-in on chromosome III.

(A) The location of the Z loci along chromosome III is depicted by grey bars. **(B)** A zoom-in on a 120 KB region (positions 196611-316611) shows the raw H2A.Z/H2B log₂ ratios for each probe in that region (green bars) and the smoothed data resulting from the Gaussian plot analysis of the raw data for H2A.Z/H2B log₂ ratios (red line). The positions of the Z loci are shown by grey lines. The genes present in

this region are shown in blue. **(C)** Same as in (B) but for a zoom-in on region 237700-277700 (40 KB). **(D)** The size of the Z loci within the promoter of the *SRB8* gene as determined by Q-PCR analysis of ChIP experiments. The binding ratios for H2A.Z (red) and TFIIB (blue) are shown relative to the center of the probe that generates the maximum enrichment. The data were smoothed by a sliding median over 3 probes (see **Figure Sup. 3.1**). The size of the observed H2A.Z domain is about 250 bp larger than that of TFIIB. Since TFIIB covers about 10 bp of DNA, we can infer that H2A.Z covers about 260 bp of DNA at that locus (shaded area). More details can be found in the **Text Sup. 3.1** and **Figure Sup. 3.1**.

3.3.3.3 H2A.Z-containing nucleosomes are essentially located within promoter regions

We next examined the actual location of the Z loci relative to genes. Manual inspection of the data suggests that Z loci are predominantly located within promoter regions. This feature is illustrated in **Figure 3.1** that shows i) only one open reading frame (ORF) within the 40 KB window shown in **Figure 3.1 C** contains enrichment for H2A.Z; ii) none of the seven intergenic regions that do not contain any promoter in that region (e.g. between *SOL2* (SGD:S000000718) and *ERS1* (SGD:S000000671)) contain a Z locus; iii) there is a Z locus upstream of every ORF shown in **Figure 3.1 C**; and iv) Interestingly, the intergenic region between the divergently transcribed *YCR090C* (SGD:S000000686) and *KIN82* (SGD:S000000687) ORFs contains two separate Z loci, supporting the idea that each promoter is decorated with its own H2A.Z nucleosome(s). In order to evaluate the H2A.Z preference for promoter regions systematically, we analyzed the relation between the H2A.Z/H2B ratio and the distance to the closest 5' gene boundary (e.g. the ATG for protein-coding genes) for all probes on the microarray. As shown in **Figure 3.2 A**, H2A.Z binding (green line) is higher around the 5' boundary with a small bias toward the regions located just upstream compared to downstream (see inset for a zoom into the 5' region). This relationship is not observed on random data (**Figure 3.2 A**, black line) or when the proximity to the 3' ends of genes is examined

(not shown). Because 5'UTRs are very small in yeast, this data is consistent with H2A.Z being localized over promoters, although we cannot rule out the possibility that H2A.Z is, rather, present within the 5'UTR, since only a few transcriptional start sites have been annotated in yeast. For the sake of simplicity the word “promoter” will be used here to refer to the regions localized between 500 bp upstream and 100 bp downstream of the 5' boundaries of transcribed elements (this includes ORFs, but also other transcribed elements such as Ty elements, tRNAs etc.). This interval was chosen based on a recent genome-wide analysis of yeast promoters [38]. As shown in **Figure 3.2 B**, 74 % (3639 out of 4862) of the Z loci identified in this study are located within the promoters of annotated genes. Considering that promoters account for about 31 % of the yeast genome, this translates into a 2.4-fold enrichment over a random uniform distribution (p-value < 10^{-308}). In addition, we find that 63 % (4742 out of 7488) of yeast promoters are decorated with H2A.Z (**Figure 3.2 C**). Conversely, only 2 % of the H2A.Z loci are located within intergenic regions that contain no promoters (where two genes are transcribed convergently).

In a recent elegant study, the Rando group mapped nucleosomes over 482 KB of yeast DNA, including most of the entire chromosome III. They showed that the scored yeast promoters contain a nucleosome-free region flanked with well-positioned nucleosomes [37]. A median sliding window of their data is reproduced in **Figure 3.2 D** (blue) along with our genome-wide location data for H2A.Z/H2B (green) and H2A/H2B (gold) for a composite of 1850 genes [37]. Although the two assays have different resolution, the Figure clearly shows that H2A.Z is preferentially localized downstream of the nucleosome-free region in yeast promoters. Importantly, data from the Figure suggests that H2A.Z may occupy the first, and perhaps also the second, well-positioned nucleosome downstream of the nucleosome-free region (see cartoon in **Figure 3.2 D**). Taken together, these data suggest that H2A.Z targets specific nucleosomes within promoters across the genome and therefore contributes to the creation of a promoter-specific chromatin structure.

In addition to class II promoter regions, we found H2A.Z loci at other transcribed units such as tRNAs, snoRNAs, rDNAs, and Ty elements, suggesting that the role of H2A.Z in transcription is not limited to protein-coding genes (data not

shown). Interestingly, we also found H2A.Z loci centered around centromeres and replication origins (data not shown). The association of H2A.Z with centromeres has also been reported previously [32], and the role of H2A.Z in DNA replication and chromosomal segregation will be studied elsewhere (Bataille et al, manuscript in preparation).

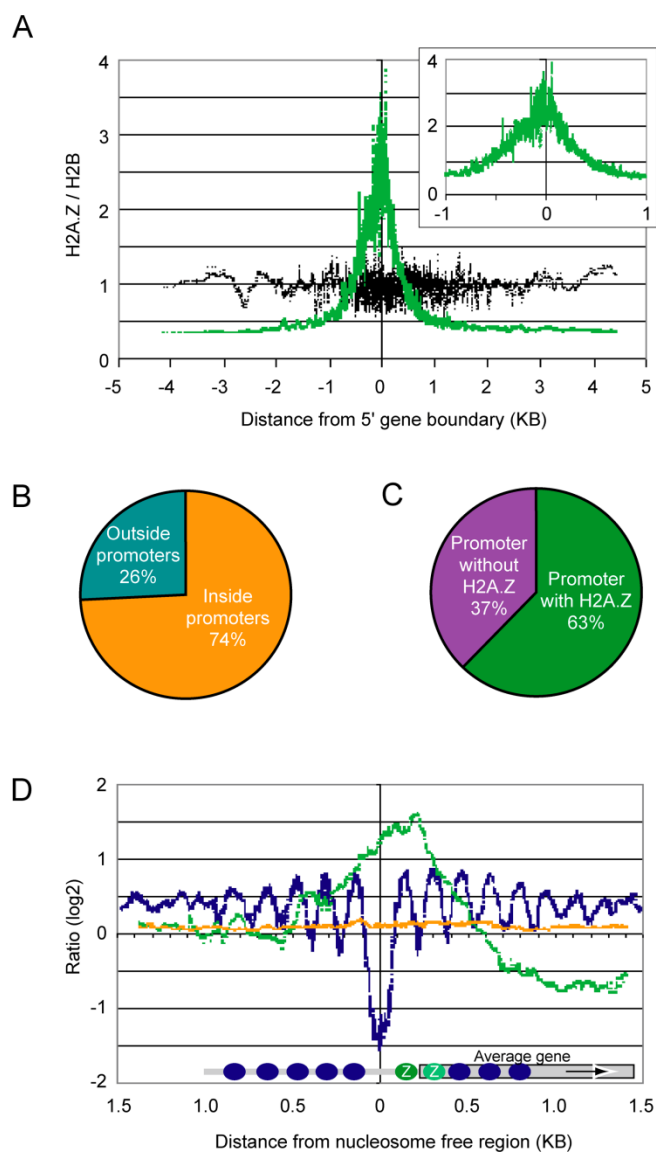


Figure 3.2: H2A.Z is predominantly localized within promoter regions.

(A) A sliding median of the H2A.Z/H2B ratio for all probes on the microarray is plotted against the distance from the 5' boundary of their closest transcribed element (including ORF, tRNA, transposon etc.) (green). Randomized data (where the H2A.Z/H2B ratios were scrambled prior to calculating the sliding median) are plotted in black. (B) Most Z loci are found within promoters. The fraction of the 4862 Z loci present within promoters (gold) and non-promoter regions (aqua) is shown. Promoter regions are defined as the -500/+100 interval relative to the 5' boundary of transcribed elements as annotated in the SGD database. (C) The fraction of the 7571 promoter regions (defined as in (B)) containing (green) or not containing (purple) a Z loci is shown. (D) H2A.Z is preferentially localized downstream of the nucleosome-free region in yeast promoters. A sliding median of the ratio for H2A.Z/H2B (green) and H2A/H2B (gold) was plotted against the distance from the nucleosome-free region (NFR) (defined as a linker of more than 100 bp located less than 500 bp upstream of a 5' gene boundary) for regions where the NFR can be unambiguously assigned to a single gene (i.e. NFR that map to two diverging genes were filtered out). A similar sliding window was applied to the nucleosome positioning data from Yuan et al. [37] (blue). A cartoon representation of the nucleosomal organization of a typical gene is shown at the bottom.

3.3.3.4 H2A.Z is not enriched in the promoter regions of highly active genes

Data from **Figure 3.1** and **Figure 3.2** have clearly shown that H2A.Z has a preference for promoter regions. However, **Figure 3.2 B** also shows that up to 37 % of promoters are not associated with a Z locus. In order to investigate what determines the association of H2A.Z with some promoters but not others, we investigated the correlation between H2A.Z promoter occupancy and the transcription rate for all genes. As shown in **Figure 3.3**, there is an inverse correlation between the transcription rate and H2A.Z occupancy. Accordingly, when the genome-wide location of H2A.Z was determined for cells grown in galactose, we noticed that the occupancy of H2A.Z was reduced, compared to its occupancy in glucose, at the promoters of 147 out of 173 genes whose expression levels have been demonstrated

to be induced the most under those conditions according to the study of Ren et al. [39]. This data is consistent with H2A.Z being dynamically associated with inactive genes. These genomic observations are also in line with our ChIP/Q-PCR data (**Text Sup. 3.2** and **Figure Sup. 3.2**) and with previously published evidence at specific promoters [19, 21, 22]. We would like to point out that data represented in the experiment of **Figure 3.3** were collected from a population of cells that are heterogeneous in the sense that each cell is in a different physiological state and phase of the cell cycle. For that reason, and also because promoters have inherently different strengths, a given gene is never completely off or on within the population. The observed transcription rate across all genes is, therefore, a continuum ranging from near zero to about 300 mRNA per hour for genes that are the most transcriptionally active. The transcription rate being a continuum, it is expected that the H2A.Z occupancy will also be a continuum. In that regard, RNA pol II or general transcription factors like TFIIB also show continuum-like binding patterns in those assays [36, 40, 41]. We also propose that the low enrichment of H2A.Z at the promoters of highly active genes most likely reflects the low abundance of nucleosomes in those regions, since recent evidence shows that the promoters of highly active genes are depleted of nucleosomes [42-46]. Collectively, these data show that H2A.Z specifically associates with one or a few nucleosomes within the promoter, of many inactive genes but is generally absent from the promoters of highly active genes.

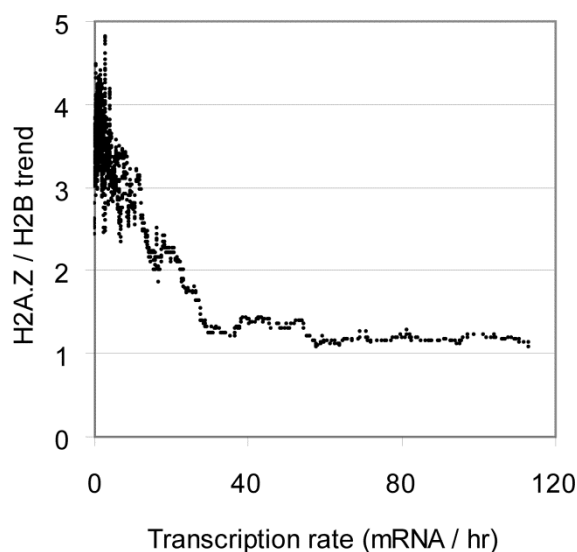


Figure 3.3: H2A.Z/H2B ratios are inversely correlated with transcription rate.

The binding trend was calculated by computing the moving median of the H2A.Z/H2B flat ratios over a sliding window of 50 genes across all genes ordered by transcription rate as described previously [36, 40, 41, 59-61]. Each probe was assigned to its closest ATG and the probe with the highest H2A.Z/H2B value for each gene was used. The transcription rate for yeast genes was determined previously [62].

3.3.3.5 H2A.Z regulates nucleosome positioning

Having established that H2A.Z is present at most yeast promoters, we sought to investigate the functional significance of the promoter-specific localization of H2A.Z. First, we investigated nucleosome positioning on genes that have Z loci (**Figure 3.4 A**) versus genes that are devoid of those loci (**Figure 3.4 B**), as previously defined in **Figure 3.2** using data presented in Yuan et al [37]. The Figure clearly shows that nucleosomes within the genes that bear Z loci are well positioned, whereas nucleosome positioning on genes that do not have Z loci are significantly less well organized. This observation is consistent with the fact that H2A.Z-free

promoters are less covered by nucleosomes than the more frequent H2A.Z-containing promoters (not shown). The Figure further suggests that H2A.Z may play an important role in organizing chromatin structure at promoters. Alternatively, however, H2A.Z may associate with promoters as a consequence of their well-organized chromatin structure. In order to test directly whether H2A.Z can regulate chromatin structure at promoters, we studied nucleosome positioning at a model gene in wild type (WT) and H2A.Z null (*htz1Δ*) cells. The *GAL1* gene was chosen for these experiments since H2A.Z is required for its full expression, and since the variant histone preferentially binds the repressed promoter of that gene [19, 21]. Furthermore, chromatin structure at the *GAL1* promoter has been thoroughly investigated [47-50], so any changes in chromatin structure should be experimentally tractable. First, we performed indirect end-labeling experiments on chromatin samples that were partially digested by MNase. **Figure 3.4 C**, left panel, shows that the gross nucleosome positioning pattern is not significantly altered in *htz1Δ* cells when the latter are grown in media containing glucose. However, we do note a slight shift of nucleosome C (nucC) in a downstream position as well as a few other nucleosomes. The right panel of that figure shows the cleavage pattern of MNase-digested naked *GAL1* promoter DNA. The UAS_G region can be easily identified as it is quite resistant to MNase cleavage [49]. Since indirect end-labeling experiments only have a resolution of ~20-30 bp, we decided to carry out high-resolution LM-PCR assays of MNase-digested mononucleosomes at two positioned nucleosomes within the *GAL1* promoter region (nucB and nucC; see **Figure 3.4 D**). Thus, LM-PCR linkers were ligated to mononucleosome-purified DNA and PCR reactions were carried out with external oligonucleotides corresponding to the linkers in combination with oligonucleotides corresponding approximately to the center of the nucleosome-protected DNA as previously determined by Fisher-Adams and Grunstein [49]. These last oligonucleotides are designed to either amplify the right end portion of nucleosome-embedded DNA (R) of nucB and nucC, or the left end portion (L) (**Figure 3.4 D**, above). The results show that nucB does not shift position in the *htz1Δ* mutant as compared to WT cells when PCR reactions were performed at either the right or left boundaries of the nucB. However, in the case of nucC, there is a ~20 bp

shift to a downstream position in *htz1Δ* cells as compared to the WT case (**Figure 3.4 D**, right panel). Importantly, this shift was observed with both the right and left boundaries of *nucC*. These results confirm our observations in indirect end-labeling experiments, and suggest that the absence of H2A.Z induces a specific shift of *nucC* to a downstream position but has no effect on the positioning of *nucB*. Our results suggest that H2A.Z may be enriched at the promoters of inactive genes in order to modulate the position of key regulatory promoter-associated nucleosomes. Because we found that most Z loci are located within well-positioned nucleosomes and because H2A.Z-free promoters have a less organized chromatin structure than H2A.Z-containing genes, it is tempting to speculate that the variant histone also plays a role in nucleosome positioning at other genes.

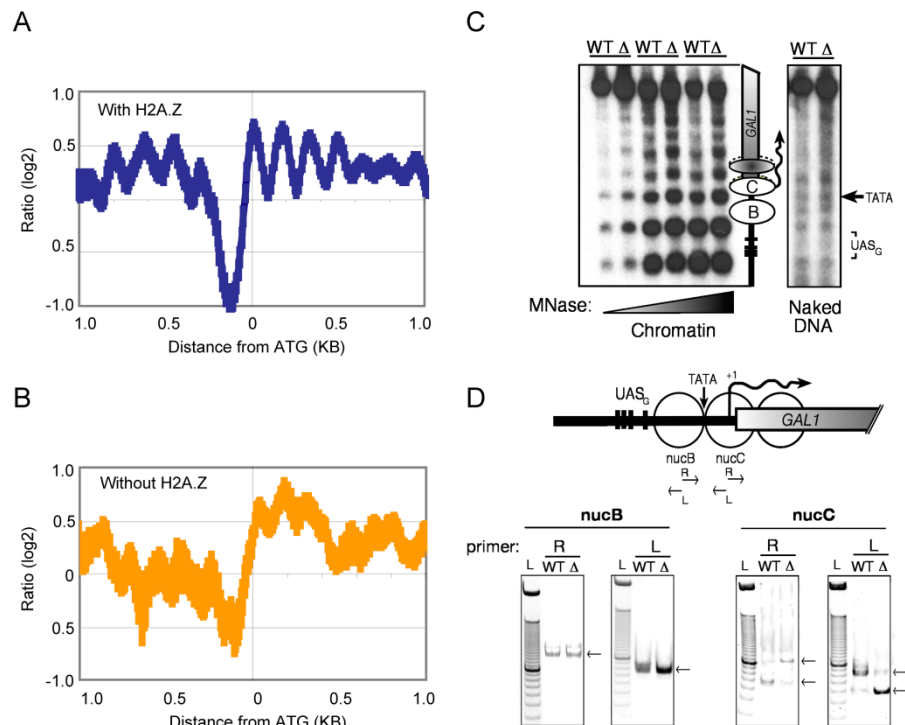


Figure 3.4: H2A.Z regulates nucleosome positioning.

(A) Nucleosome positioning map of all genes associated with a Z locus. The data from Yuan et al. [37] were used to compute the nucleosome occupancy curve for all genes containing a Z locus aligned on their ATG. Peaks represent nucleosomes and

valleys represent linker regions. A nucleosome-free region is detected ~200 bp upstream of ATGs as described by Yuan et al. [37]. The vertical thickness of the curves contains one-standard-deviation error bars for the mean \log_2 ratio. **(B)** Same as **(A)** but for genes containing no Z locus. **(C)** Indirect end-labeling of MNase-digested chromatin from WT and *htz1* Δ cells grown in the presence of glucose. **(D)** High resolution LM-PCR analysis of MNase digested nucB-C mononucleosomes. Upper part: structure of the *GAL1* promoter and PCR probes used; Left: nucB analysis probing the right (R) and left (L) boundaries in WT and *htz1* Δ (Δ) cells; Right: same as left part of the Figure but the analysis is with nucC.

3.3.3.6 H2A.Z is unusually localized at HZAD genes

Because genes contained within the so-called HZAD domains –unlike other H2A.Z regulated genes- are proposed to be regulated by H2A.Z through an anti-silencing mechanism [24], we investigated more closely the distribution of H2A.Z at those genes. First, manual inspection suggested that H2A.Z occupied larger regions at those genes. As shown in **Figure 3.5 A**, the regions occupied by H2A.Z around the *GIT1* (SGD:S000000695) gene (an HZAD gene) extend both upstream as well as downstream within the coding region as compared to a non-HZAD gene such as *SRB8* (see **Figure 3.5 B**). Similar examples can be found in **Figure Sup. 3.3**. In order to investigate this in an unbiased manner, we have calculated the size of the regions localized between the coordinates where the H2A.Z occupancy curve approaches baseline (\log_2 ratio = 0.1). This size provides an estimation of the width of the Z loci. **Figure 3.5 C** shows a plot of that size relative to the distance to chromosome ends. The Figure shows that H2A.Z covers larger uninterrupted regions within the first 20 KB from telomeres compared to telomere distal regions. This is consistent with H2A.Z occupying wider regions around HZAD genes than at other genes in the genome. This conclusion was also reached using a different analysis described in **Figure Sup. 3.4** and **Text Sup. 3.3** and is not dependent on a difference in gene density at telomeres (not shown). This phenomenon is not restricted to HZAD genes, since many telomere-proximal non-HZAD genes also show the same pattern (not

shown). It is therefore possible that the anti-silencing function observed at HZAD genes could actually be extended to more (perhaps most) telomere-proximal genes. Three possibilities could account for this wider distribution: first, H2A.Z could occupy more nucleosomes at those genes. Second, the number of nucleosome(s) occupied by H2A.Z may be the same, but this (those) nucleosome(s) may not always be the same in different cells. Finally, it is also possible that chromatin at those genes is more fluid such that the H2A.Z nucleosome(s) can be found anywhere along a long stretch of DNA.

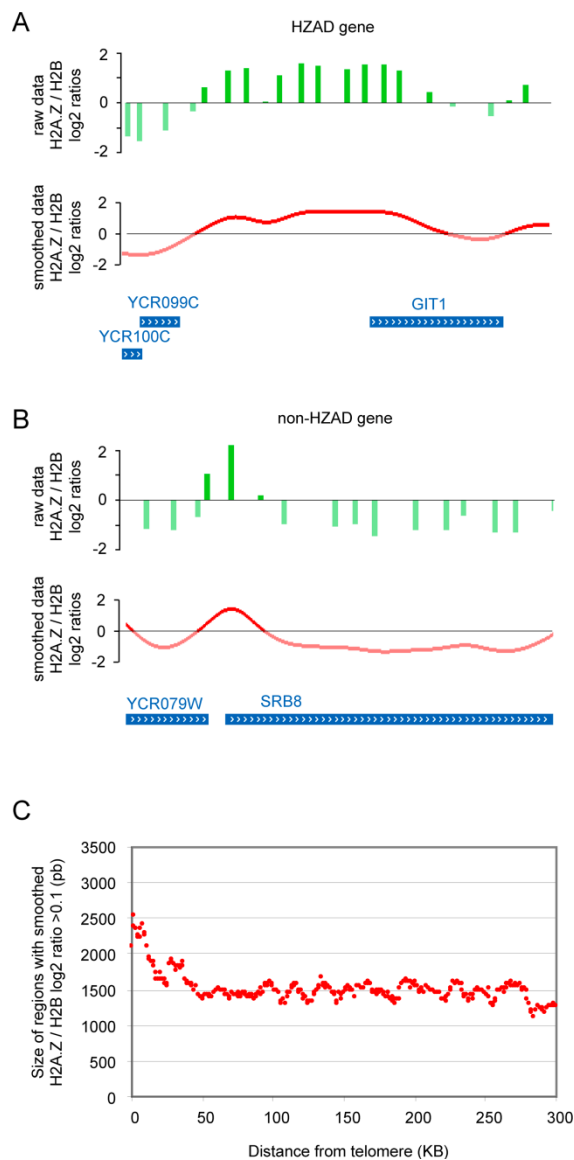


Figure 3.5: H2A.Z is unusually localized in HZAD genes.

(A) The raw (green) and smoothed (red) H2A.Z/H2B log₂ ratios, as determined by our microarray experiment, are shown across a 5 KB region around the *GIT1* genes. The genes present in that region are shown in blue. (B) Same as (A) but for a 5 KB region around the *SRB8* gene. (C) H2A.Z covers wider regions near telomeres. The sizes of the regions covered by H2A.Z (as determined by distance between the coordinates where the smoothed H2A.Z/H2B log₂ ratio reaches “0.1”) are plotted against the distances to telomeres. A moving average (window = 10 KB) was applied to the data.

3.3.4 Discussion

We have investigated the genome-wide distribution of histone variant H2A.Z as compared to the regular S-phase histone H2A. Our results show that H2A.Z can preferentially associate with narrow regions of yeast promoters. Surprisingly, the variant histone is present within the promoters of most nucleosome-containing promoters. Conversely, the promoters of highly transcribed genes are not predominantly enriched in H2A.Z. Furthermore, and much to our surprise, H2A.Z does not appear to be as sharply localized within HZAD genes. The wider distribution of H2A.Z at HZAD genes may be important for preventing Sir-protein spreading in those regions. Since H2A.Z shows a strong preference for inactive gene promoter regions, we also investigated whether the variant histone had an important role in modulating chromatin structure at the *GALI* model gene. Importantly, we find that the replacement of H2A.Z with regular H2A in cells induces a shift in the positioning of a nucleosome located over the transcription start site. In accord with this finding is the fact that nucleosomes at promoter regions bearing Z loci are well positioned as compared to genes not enriched in H2A.Z. Taken together, our results support a model wherein H2A.Z associates with specific nucleosomes in the promoters of inactive genes in order to poise and, perhaps, organize chromatin structure in a fashion that is permissive to transcription initiation. However, it remains to be determined whether the simple incorporation of H2A.Z into specific nucleosomes is, by itself, necessary and sufficient to facilitate gene activity or whether additional factors are required for that action. The first possibility is supported by the fact that *in vitro*-assembled chromatin arrays containing H2A.Z are inherently more resistant to condensation as compared to H2A-containing nucleosomal arrays [51, 52]. On the other hand, work from our laboratory has shown that H2A.Z can interact with regulatory proteins and components of the transcriptional machinery [19], arguing that H2A.Z may regulate gene expression through interactions with such downstream effectors.

Although H2A.Z has been previously shown to be enriched at the *GALI*, *PHO5* (SGD:S000000297), and *PUR5* (SGD:S000001259) genes [21, 22], the fact that H2A.Z globally associates with defined promoter regions of inactive genes is

surprising, since previous studies had suggested that, even though H2A.Z was generally enriched in euchromatic regions, there did not appear to be any correlation between deposition of the variant histone and promoters [24, 30]. However, we note that in those cases where H2A.Z was found more enriched in the ORF as compared to the promoter, the genes are highly transcribed housekeeping genes (*e.g.* *ACT1* (SGD:S000001855) and *ADHI* (SGD:S000005446)) [24, 30]. Recent studies have clearly established that the promoters of such highly active genes are rather depleted of nucleosomes [42-46], and thus the apparent low enrichment of H2A.Z at those promoters could actually reflect low nucleosome occupancy. In support of this conclusion, we find that H2B levels are significantly reduced at the *ACT1* promoter, while the H2A.Z/H2B ratio shows a slight enrichment in the promoter region (see **Figure Sup. 3.3 C**). Because housekeeping genes are, by definition, transcribed in most, if not all, conditions, and since we found that H2A.Z is present at promoters of genes when they are inactive, we favor a model where H2A.Z would be more important for the regulation of non-housekeeping genes that are generally repressed when grown in rich medium but are strongly induced under specific growth conditions (*e.g.* *GAL* genes upon galactose induction, *etc.*). It is interesting to note that this last possibility is in line with published evidence regarding the role of H2A.Z in higher organisms. For example, H2A.Z is not significantly expressed in the non-differentiated inner cell mass tissue of the mouse embryo, whereas it is highly expressed in the differentiating tissues [28]. Furthermore, H2A.Z is required for the proper development of *Xenopus* [13]. These observations may be taken to suggest that H2A.Z is only required when a particular gene expression program is engaged in an organism and not so much for gene expression that pertains to basic cellular metabolism. Since H2A.Z has been shown to be abundantly associated with heterochromatin in higher eukaryotes [28] – which is much more preponderant in metazoans than to yeast [53, 54] – it remains to be determined whether or not the variant histone plays additional functions linked to this localization in metazoans. It is worth emphasizing, however, that this co-localization of H2A.Z with metazoan heterochromatin does not exclude the possibility that it may also be associated with

the promoter regions of inactive genes, as we have shown with yeast. This possibility is currently under investigation.

Our genome-wide location assay has revealed that – as opposed to what is observed at other genes – H2A.Z occupies wider regions at HZAD genes, that spread out well beyond the promoter region and into the ORF. As previously suggested by Meneghini et al. [24], it is conceivable that H2A.Z has a dual role in gene expression, the first being that of poising genes for transcription initiation in euchromatin, and the second being that of protecting euchromatin from silencing. We propose that at most inactive genes, where H2A.Z has a sharp promoter-binding pattern, H2A.Z will simply be incorporated into specific repressive nucleosomes that need to be remodeled quickly prior to gene induction, whereas at HZAD genes, H2A.Z would create nucleated domains that are refractory to gene silencing by heterochromatin. Alternatively to this scenario, it is possible that the broad distribution of H2A.Z at HZAD genes reflects unstable positioning of variant histone-bearing nucleosomes, perhaps due to a chromatin structure that is less well organized at those genes. This last possibility, however, is less likely, since examination of the nucleosome positioning data from Yuan et al. [37] reveals that nucleosomes are well-positioned over the chromosome III-contained HZAD genes. Nevertheless, it is likely that the specific chromatin structure exemplified by this unusual localization of H2A.Z renders the HZAD genes more sensitive to Sir2 silencing.

Our observations that H2A.Z is also important to regulate nucleosome positioning, in particular at the *GALI* promoter, may account for a mechanism by which H2A.Z regulates transcription. We had previously observed that absence of H2A.Z prevents RNA polymerase II as well as TBP (SGD:S000000950) from being efficiently recruited to the *GALI* promoter, whereas the Gal4 (SGD: S000006169) activator was efficiently recruited to the gene [19]. In addition to these findings, recent experiments from our laboratory have shown that Mediator, SAGA, and Swi/Snf also cannot be efficiently recruited to *GALI* in the absence of the variant histone (K. Lemieux and L.G., unpublished data). We thus propose that improper positioning of at least one nucleosome over the *GALI* transcription start site may be sufficient to severely impair the recruitment of the entire transcriptional machinery to

the gene. Our results also emphasize the fact that chromatin architecture is exquisitely regulated, and even subtle changes in its structure may trigger transcription defects. A similar situation was recently reported with Bromodomain factor 1 (Bdf1) (SGD:S000004391) and histone H4-tail acetylation at the *PHO5* gene. It was reported that a shift of 2-3 bp of a positioned nucleosome at a *PHO5-lacZ* reporter construct resulted in a defect in *PHO5* induction in the absence of Bdf1 and H4-acetylation [55]. This is of particular significance since Bdf1 is a component of the Swr1.com that loads H2A.Z into chromatin [29, 30]. It is conceivable that deletion of *BDF1* prevents loading of H2A.Z into chromatin and, therefore, the Bdf1-related defect could be caused by improper loading of H2A.Z at the *PHO5* promoter. Alternatively, it is conceivable that Bdf1 could be a downstream target of H2A.Z important for mediating its function in nucleosome positioning. These possibilities are currently under investigation. Finally, we point out that our data are not mutually exclusive with the possibility that H2A.Z could also be important in transcription elongation, since this histone variant has been shown to interact genetically with the TFIIS elongation factor in yeast [30], and since it appears to be replaced by H2A downstream of the *c-myc* promoter region during elongation [23].

3.3.5 Materials and Methods

3.3.5.1 Yeast strains and genetic methods

S. cerevisiae W303a strains carrying 3xMyc-H2A.Z, 3xMyc-H2A or 3xHA-H2B were generated using PCR epitope tagging as described previously [56]. All *URA3* markers were “popped-out” and strains were verified by PCR and immunoblotting analyses. MAY424 (*htz1Δ*) was described previously [19].

3.3.5.2 Antibodies

Antibodies used in this study were from the following sources: anti-Myc, (9E10; SantaCruz Biotech); anti-HA (12CA5; Roche); anti-TFIIB, kind gift from Richard A. Young.

3.3.5.3 Chromatin immunoprecipitations

ChIP experiments were performed essentially as described previously [19, 22] with some modifications. Briefly, 50 ml of cells were grown in yeast extract-peptone supplemented with 2 % dextrose or 2 % raffinose / 5 % galactose for *GAL*-induced conditions to an OD₆₀₀ of 0.7 and fixed with 1 % formaldehyde as described [19]. 500 µl of whole-cell extract was incubated with the appropriate antibody coupled to magnetic beads (Dynal). Immunoprecipitated DNA was either used for genome-wide location analysis (see below) or Q-PCR analysis. Briefly, immunoprecipitated DNA and input DNA were mixed with the appropriate oligonucleotides and Brilliant SYBR Green Q-PCR Master Mix (Stratagene) in 20 µl final volume. Q-PCR was performed on a Mx3000P Real-Time PCR System (Stratagene). The sequences of the oligonucleotides used for all of the Q-PCR experiments are available in **Table S7**.

3.3.5.4 DNA microarrays

The microarrays used for location analysis were purchased from Agilent Technologies. The microarrays contain a total of 44,290 60-MER probes (including 2,306 controls) covering the entire genome (except for repetitive regions) for an average density of one probe every 270 bp (+/- 100 bp) within the probed regions. The greatest distance between the end of one probe and the beginning of the next is 540 bp and 97 % of the probes are within 350 bp of another one. The probes were mapped to the genome (UCSC) using BLAST. More than 99.9 % of the probes mapped to a single locus whereas 370 probes (<0.1 %) had multiple BLAST outputs and were mapped accordingly. These microarrays were previously used to map histone modifications [57].

3.3.5.5 DNA labeling and hybridization

Labeling was done as described before (Drouin and Robert, in press). Briefly, the immunoprecipitated DNA fragments are blunted with T4 DNA polymerase and ligated to unidirectional linkers. The DNA is then amplified by LM-PCR in the presence of aminoallyl-modified dUTP. The labeling is carried out post-PCR using

monoreactive Cy-dye NHS esters that react specifically with the aminoallyl-modified dUTP. Hybridization was performed as described in Pokholok *et al.* [57] using salmon sperm in place of herring sperm DNA.

3.3.5.6 Data Analysis

The data was normalized and replicates were combined using a weighted average method as described previously [39]. To interpolate between probes and identify Z-loci, a standard gaussian filter (standard deviation = 200 bp) was applied twice to the data. Z-loci were defined as the maxima of this smoothed occupancy curve (excluding maxima smaller than zero).

3.3.5.7 Chromatin analysis

S. cerevisiae nuclei were prepared as previously described [58]. Nuclei were washed once in digestion buffer (15 mM Tris-HCl, pH 8.0, 50 mM NaCl and 1.4 mM CaCl₂) and resuspended in 1.2 ml digestion buffer per gram of dry weight.

For indirect end-labeling experiments, resuspended nuclei were digested with either 0, 1, 3, 5, 25 or 50 U/ml of MNase (USB). DNA was prepared as described [58]. Southern blot analysis was conducted using a *EcoRI-BanII* fragment from the *GALI-10* region as a probe.

For LM-PCR nucleosome boundary mapping, 200 µl aliquots of resuspended nuclei were digested with a final concentration of 100 U/ml of MNase (USB) for 10 minutes at 37°C and stopped with 4 mM EDTA. DNA was then prepared as described [58]. MNase-digested DNA was ligated overnight to unidirectional linkers, and mononucleosome-length fragments were then purified by agarose gel electrophoresis. PCR reactions were conducted with linker-specific and nucleosome-specific oligonucleotides using JumpStart Taq DNA Polymerase (Sigma) using the manufacturer's recommendations. Oligonucleotide sequences are available in the **Table Sup. 3.8**.

3.3.6 Supplementary Documents

Supplementary Tables Sup 3.1-6 are open access text-files containing Raw data and BED from microarrays. Files can be downloaded on the PLoS Biology web site.

Table Sup. 3.1: Raw data for the H2A.Z/H2B, H2A/H2B and H2A.Z/H2A.Z ChIP-chip experiments.

For each experiment, the normalized enrichment ratio is shown for every probe on the microarray.

Table Sup. 3.2: Smoothed data for the H2A.Z/H2B ChIP-chip experiment.

For every coordinate within the genome, the H2A.Z/H2B ratio derived from the raw data by the Gaussian plot analysis is shown. This file is a text file ready for display within the UCSC Genome Browser.

Table Sup. 3.3: Smoothed data for the H2A/H2B ChIP-chip experiment.

For every coordinate within the genome, the H2A/H2B ratio derived from the raw data by the Gaussian plot analysis is shown. This file is a text file ready for display within the UCSC Genome Browser.

Table Sup. 3.4: Smoothed data for the H2A.Z/H2A.Z ChIP-chip experiment.

For every coordinate within the genome, the H2A.Z/H2A.Z ratio derived from the raw data by the Gaussian plot analysis is shown. This file is a text file ready for display within the UCSC Genome Browser.

Table Sup. 3.5: List of Z loci

A table listing the position of each Z locus identified in this study.

Table Sup. 3.6: Genome Browser-ready file.

A BED file ready for displaying the H2A.Z/H2B raw (green) and smoothed (red) data for the entire yeast genome as in **Figure 3.1**.

Table Sup. 3.7: Sequences of the oligonucleotides used to quantify ChIPs by PCR in this study

Loci	Forward oligonucleotide	Reverse oligonucleotide
PRS2 A	AGCATTTCGCGAAACTTCATT	GTGGCTTCTTTAGCGGACAA
PRS2 B	CAACACACTTGGTTCACCGTA	TGGATCTCCATGCTTCACAA
PRS2 C	GCCGTCTTACAAGCGTGAAT	ACGGGGAACATGAAATCAAC
PRS2 D	TCACGAATGCTTTCCCCTAT	TGGGCTTGCTGAGCTTATTT
PRS2 E	CGACAAGATAGTCTTTGGATGATG	GCATCGAGGGATTGCTAAGA
PRS2 F	CGCGATTTGGTAGGAGGATA	GCTCTCTCCGACATTGTTGC
PRS2 G	ATTCCACTGGTCGCTTGAGT	CAATCGATTTTCGGCTTTTTTC
PRS2 H	TCACACTTTAGGAGCCGTA	ATTTGCGAATGCGGTTTTTAG
PRS2 I	CAAAGCAGGCTCACAAGAGA	GGGCGAGCAAGAATATATGG
PRS2 J	ATCACAGACCTGCGGATAC	CATTCTGATAGCCGGTGGTT
PRS2 K	CATCAGACCATCAGAAGAAGACA	CCTGTACAACCTCCGGAAAAA
PRS2 L	AAGGCTGCGAAGGAAAAAGT	TGTTTCAGCGCGTATTCTGTC
PRS2 M	TCCAAGAAAGATGGCAAAGAA	AGGCCAACCAAAAACAAAAA
PRS2 N	TGTATTGACCAAGAAACCTCAA	TTCTGTCTCCCTAGGTTTCACC
PRS2 O	CTTGAGCACGTTGTGTAGCC	GCTGCTAAATGAGGGAGTCG
PRS2 P	GCACCTTTAGCGTGAATTAAGA	GTTGGACTTCATCGGTGGTT
PRS2 Q	TCCAACATACGTCTGAGTGGTT	TTTCCCGGATTTGAAAGATG
PRS2 R	CAGAGCTTGACACCCGTTT	GCAGCAGGCTCCTTATTCTG
PRS2 S	AAATGCCGTACACATCGTCA	CCGGCATTGGAAGAGAATAC
PRS2 T	TGGGACCGTGTTAGAATGA	GAGTCGCCAGACCGTTAAGA
PRS2 U	GGTCCAATGTCATCGCTTCT	TGGCAGAGAAAACTTTGGAAA
PRS2 V	TAATCGTGCATCCCTTCCTT	ACATCGCCCCAGCTATACAC
PRS2 W	CTGGACGCAAGAAAAGGCTA	AGGGGGAAGTTTTGTTGTCA
PRS2 X	TATTGGTTCTCCCTCGTTG	GCTCAAGGCGATACAAGTCC
PRS2 Y	GCACCGCTTAAAAGTTTACTTGA	TATGCCCAGAAGTCGGTAGC
PRS2 Z	TGCGAAACATCTTGACATCC	GTGCTCCGCGAACTTTTAGT
PRS2 a	AGGGACAGTCCGCTAACTT	GACCCTTTCTGCCTAAAGC
PRS2 b	TTCAGCGTGGACATTGAAAC	TACAATACGATGGGCATTGG
PRS2 c	GGAGACTGAACACACCACGA	ACCGCCACAGGTTTTTAGGT
PRS2 d	TGCATCCTGAGCAATGAAAT	AATTGCAGTACCTCCTCAACG
PRS2 e	ATCGATGCAAGGAAAAATGG	CATCGAGATAGAAGGCGAAAG
PRS2 f	CTGGTGCTAAACGTGCTCAA	TTTTGTGTTGGCTGGTTGAC
PRS2 g	TTTCTTGGGCTTCTGAAGGT	TTCATTGTTGGATGGATGGT
PRS2 h	CGGTCAAACCTTGGGTAAGA	CCCCTTTCTTTTTCAGCGTTC
GIT1 A	AAAGAAGGGGAGCGAAAAAG	TTGCAGAACCTTATCGCAGA
GIT1 B	GTTATCAGCCGTCGAAAAGC	TTTGTGGTGCAGACACCATT
GIT1 C	CAGATGTTGCAGTGCGTTTC	TGGCATTTAGGGGATACTGTC
GIT1 D	CGTGATGACGGTAAAGAAA	CGCCGCTATTACTACGTGCT
GIT1 E	CCGTTTGGTGCCACATTTAT	TGGGGGTCCAGTAAGGAAAT
GIT1 F	GGCATTCCGTCCACTGTATC	AAGATAGCCAGATTGAAAACTAGG
GIT1 G	TTGATGAGGGACATCTGCTG	CAACTACATTTCATCGGGGTTT
GIT1 H	CGCAGACGTGAGTTGATCTT	TCCTTCTCATTTACCGATGTGA
GIT1 I	GCAACCCGTAAGTAAACCTC	ACCGGACGCAATGATTGTA
GIT1 J	TGATGGTTACGTAATGGTTCAA	CCGTAATCCATAACAAAAACCTT
GIT1 K	AACGCAGCCCTAGTTGGTATT	TCAGCAGCGATACCCATAAA
GIT1 L	TTGTGGCCACTGCTATCTTG	TCCAAAACATGCCAGGTACA
GIT1 M	GGTGGTATCCTGTTATGGTG	ATGGACCACCGAAGGCTAGT
GIT1 N	GCGATCTGGAGGACTGTTTT	GAACACACTCAATGGCCAGA
GIT1 O	CCTTCCCAAATGGTATTTTCA	TGGTCCTTGATAACGGAAGT
ACT1 A	TCGGATCCTCAAACCCCTTA	GTGGCGGGTAAAGAAGAAAA
ACT1 B	TTTCTTCCACGTCCTCTTGC	AGAGGCGAGTTTGGTTTCAA
ACT1 C	TCACGCTCTCTTTTATCTTCCTT	TGGTTTGAGTAGAAAGGGGAAG

ACT1 D	TTTTTCTTCCCAAGATCGAAA	TTAAATGGGATGGTGCAAGC
ACT1 E	TTTCACTCTCCATAACCTCCT	TTCCAAATTTCAAGCCCCTA
ACT1 F	TTGCTTCATTCTTTTTGTTGC	CAATAACCAAAGCAGCAACCT
ACT1 G	ATGGTCGGTATGGGTCAAAA	TGGATTGAGCTTCATCACCA
ACT1 H	TGTCACCAACTGGGACGATA	TGTTCTTCTGGGGCAACTCT
ACT1 I	ATTGTTTTGGATTCCGGTGA	AACCAGCGTAAATTGGAACG
ACT1 J	TTGGCCGGTAGAGATTTGAC	GAAGTCCAAGGCGACGTAAC
SRB8 A	AAAATGTTGGGCTGAAGCTG	CATACGCATTAGGCCACGAC
SRB8 B	ATTAATAATGATCGGGTCGTG	GGAAATCCAGGCTACTGCAC
SRB8 C	TGTGCACTAGCCTGGATTC	CAAACGTGGACAGCTTATGTTT
SRB8 D	CCGAGGGAAATTAAGGAAA	ACCATTTTGTCCAAGGATCG
SRB8 E	CTGATGAAGGAGGGCGATAG	TGGCAAACCCGTAGAACAAT
SRB8 F	AGACCGTCTGCGATTGTTCT	GGTGCATTCCGGATCATCATA
SRB8 G	TCCGTATGATGATCCGAATG	GTCTTGATCGCAGAGGATGA
SRB8 H	GGTGGGCGACTCTAAGATCA	GTCTGGGTCCAATCGAGAAG
SRB8 I	CTTCTCGATTGGACCCAGAC	CGAGAAAATATCCGGCTCAC
SRB8 J	AGTGAGCCGGATATTTTCTCG	GATGCCGTCGGTGACAAG
SRB8 K	CTTCCTGGCGCTTGTAC	TGCAACAAATTGAGGAGTCTG
SRB8 L	CTGAAAAGGGCGACTCCTCA	GGGCCAGTTCTGTCAACATT
SRB8 M	GGCCCAATGTTGACAGAAT	AAGATGATTATCTAGGACTTGTTC
SRB8 N	CCTGGGTGGAAACAAGTCT	TTTTTCAGAAAGGTGTTTAAATGC
SRB8 O	GCGCATTTTAAACACCTTTC	TGGGGGAGTCAGCAAGTATC
SRB8 P	CGGTTCTGGTCGATACTTGC	GATCTTCTGCTGCGAGTGTGC
SRB8 Q	GAGCACACTGCAGCAGAAGA	GGGAAAGAATGTTGGATTGAG
SRB8 R	CATGAATCATTAGTCACTCAATCCA	GAAAACCCAGGCCGTGTAAT
SRB8 S	TTCCACGACGACATTACACG	CCAATTTTGACCAGCGAAGT
SRB8 T	CTCACACTTCGCTGGTCAAA	TCCATTCGATAAAATAGCAGCA
SRB8 U	AAAGGGCGATTTGGTTGATA	ATCCGTCAGTCTTTAGCA
SRB8 V	CCATGCACCTGCTAAAGGAC	GAAATCTGCCTCTTCCACGTA
SRB8 W	CGTACGTGGAAGAGGCAGAT	TGTAAAGATAATGGCAAAATTCG
SRB8 X	AACTTCGAATTTTCCATTATC	GAAGGGCTCTTTTGAATTGA
SRB8 Y	TTGCGGCTACAATAACATCAA	TCTCGTCATTTATCATTGATTTGC
SRB8 Z	AAATGACGAGAACTACATCATCAA	GATGTGGGAAATATAAACACCTCT

Table Sup. 3.8: Sequences of the oligonucleotides used in the LM-PCR assay to map nucleosomes within the *GAL1* promoter

Name	Sequence
Linker primer-s	GCGGTGACCCGGGAGATCTGAATTC
Linker primer-as	GAATTCAGATC
Gal1NucB-FWD	CGAATCAAATTAACAACCATAGGATGATAATGC
Gal1NucB-REV	GAAGGTTTGTGGGGCCAGGTTACTGC
Gal1NucC-FWD	CTCCTTGACGTTAAAGTATAGAGGTA
Gal1NucC-REV	GTTGATACTTTTATGACATTTGAATAAGAAG

Text Sup. 3.1: Determination of the size of the Z locus within the *SRB8* promoter

The size of the observed domain bound by a given protein, as determined by ChIP assays, is necessarily larger than the actual domain, due to the resolution limit of the technique. The factors that determine the observed size of the domain include the size of the chromatin fragments used in the IP, the size of the PCR fragments used to probe the ChIP reactions, and the distance between these probes. When these parameters are well established, models can be used to calculate the actual size of a domain from the observed data [63]. Alternatively, the size of the domain can be estimated by comparing the observed domain with that of a protein for which the actual domain is known [64] (See **Figure Sup. 3.1 A**). This last strategy was used here to map H2A.Z at the *SRB8* locus. The data is shown in **Figure 3.1 D** and more details are provided below.

In order to map the domain occupied by H2A.Z around the *SRB8* gene we designed 25, ~150 bp-long, overlapping PCR probes covering a ~2.5 KB region around the *SRB8* promoter. These probes were used in Chromatin Immunoprecipitation (ChIP) experiments to map H2A.Z and TFIIB at that region. A schematic representation of the probes is shown in **Figure Sup. 3.1 B** and oligonucleotide sequences can be found in **Table S7**. Since the observed domain occupied by H2A.Z is ~250 bp larger than the one observed for TFIIB (see **Figure 3.1 D**), we can infer that H2A.Z occupies a domain of about 260 bp at this locus. This data is consistent with our ChIP-chip data and provides a higher-resolution map of H2A.Z at the *SRB8* gene.

Text Sup. 3.2: Validation of ChIP-chip data by Q-PCR

Many loci enriched in H2A.Z according to our ChIP-chip assay were validated by Q-PCR. **Figure Sup. 3.2** shows Q-PCR data for these loci. **Figure Sup. 3.2 A** shows that the H2A.Z/H2B ratio is significantly higher in the *PRS2* (SGD:S000000901)-*UBC6* (SGD: S000000902) and *AST2* (SGD: S000000903)-*RPS8B* (SGD: S000000904) intergenic regions than in the ORFs and at the putative terminator region abutting *UBC6* and *AST2*. The figure also shows that the H2A.Z/H2B ratio peaks rather sharply in promoter regions. Interestingly, the promoter of the highly transcribed *RPS8B* gene is devoid of H2A.Z. The case of highly transcribed genes is discussed below. Taken together, these data fully support our conclusion that H2A.Z preferentially targets the promoter regions of inactive yeast genes. **Figure Sup. 3.2 B** shows Q-PCR analyses of ChIP experiments at the *GITI* locus, a gene contained within a HZAD domain. In full support of our ChIP-chip data, the Q-PCR analysis depicted here shows that H2A.Z occupies a wider region at that gene compared with non-HZAD genes like those shown in panel A. **Figure Sup. 3.2 C** shows Q-PCR analysis of the *ACT1* gene. The *ACT1* gene is highly transcribed, and a previous study had implied that H2A.Z levels were higher at that ORF relative to the promoter of that gene. In this same study, however, the level of H2A.Z was not normalized to nucleosome density. Since recent studies have clearly established that the promoters of such highly active genes are depleted of nucleosomes [44, 45, 65-67], the apparent low enrichment of H2A.Z at those promoters could actually reflect low nucleosome occupancy. In support of this conclusion, **Figure Sup. 3.2 C** shows that H2B levels are significantly reduced at the *ACT1* promoter, while H2A.Z/H2B binding shows enrichment in the promoter region relative to the ORF.

Text Sup. 3.3: Z loci have a wider distribution pattern around telomeres

The data presented in **Figure 3.5 C** suggests that Z loci have a wider distribution pattern within the first 20 KB from telomeres compared to telomere-distal regions. The possibility remains, however, that peaks close to telomeres appear wider in this analysis simply because they are taller (tall peaks will necessarily appear wider) or because telomere proximal regions contain a higher density of peak that would result in peak merging (that would look like a single wide peak in the analysis). In order to control for these possibilities, we computed a width index for all Z loci. **Figure Sup. 3.4 A** provides a detailed explanation of this index. In sum, the width index is a measure of the width of the Z loci that is normalized for peak height. We also removed peaks that were close to each other (<1.5 KB on each side) in order to avoid overlapping peaks that may contribute to the observed width of its neighbors. **Figure Sup. 4 B** shows that this width index is larger (meaning wider Z loci) within the first 20 KB from telomeres compared to telomere-distal regions. This is consistent with the data shown in **Figure 3.5 C**, confirms that H2A.Z occupies wider regions around HZAD genes than at other genes in the genome, and shows that this is not due to a higher peak density or to taller peaks.

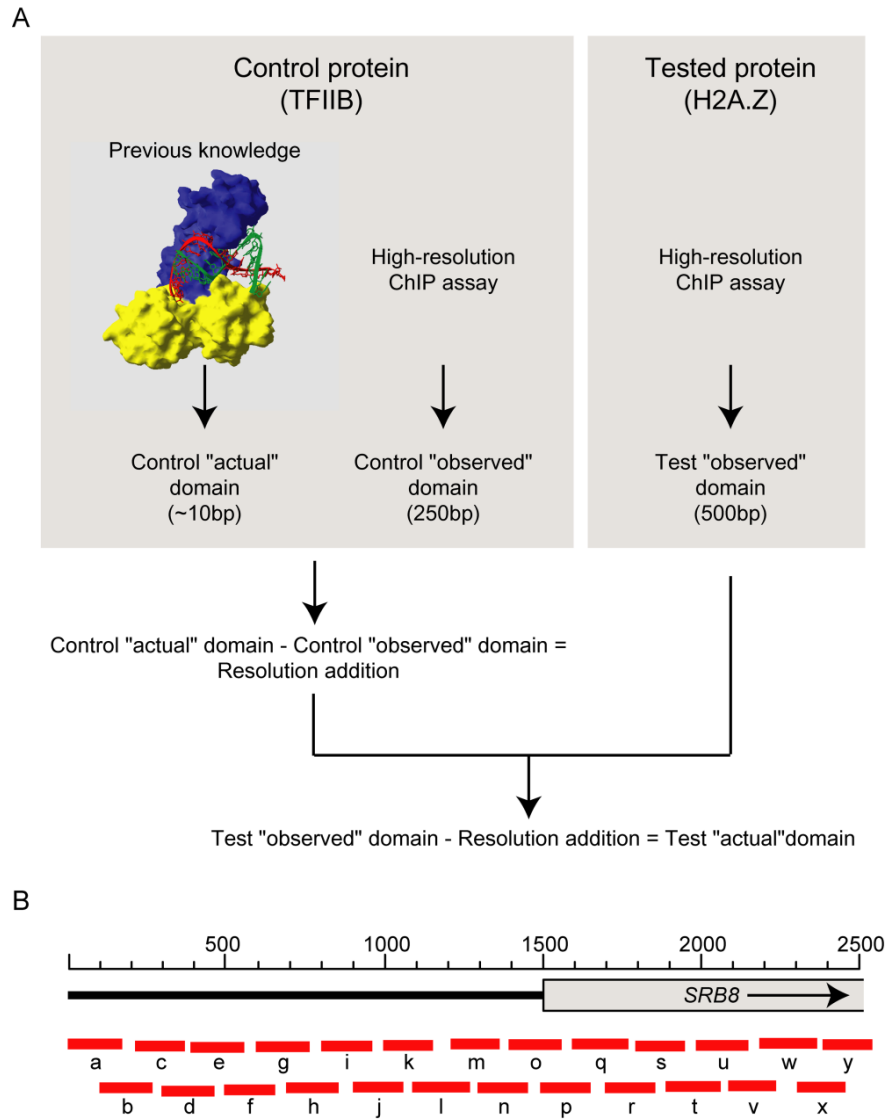


Figure Sup. 3.1: High-resolution mapping of H2A.Z at the SRB8 locus by Q-PCR

(A) Theory for domain size calculation using a control protein. A protein for which previous literature provides the actual size of the domain occupied by the protein on DNA (e.g. TFIIB) can be used to estimate the actual domain of an unknown protein (e.g. H2A.Z). Data from the control protein is used to estimate the amount of DNA that is added due to the limit of the resolution of the ChIP assay. This “Resolution addition” is removed from the “Observed” domain from the tested protein to generate its “Actual” domain. (B) Probes used in our Q-PCR assay. The regions amplified by PCR to probe our ChIP assay are depicted in red.

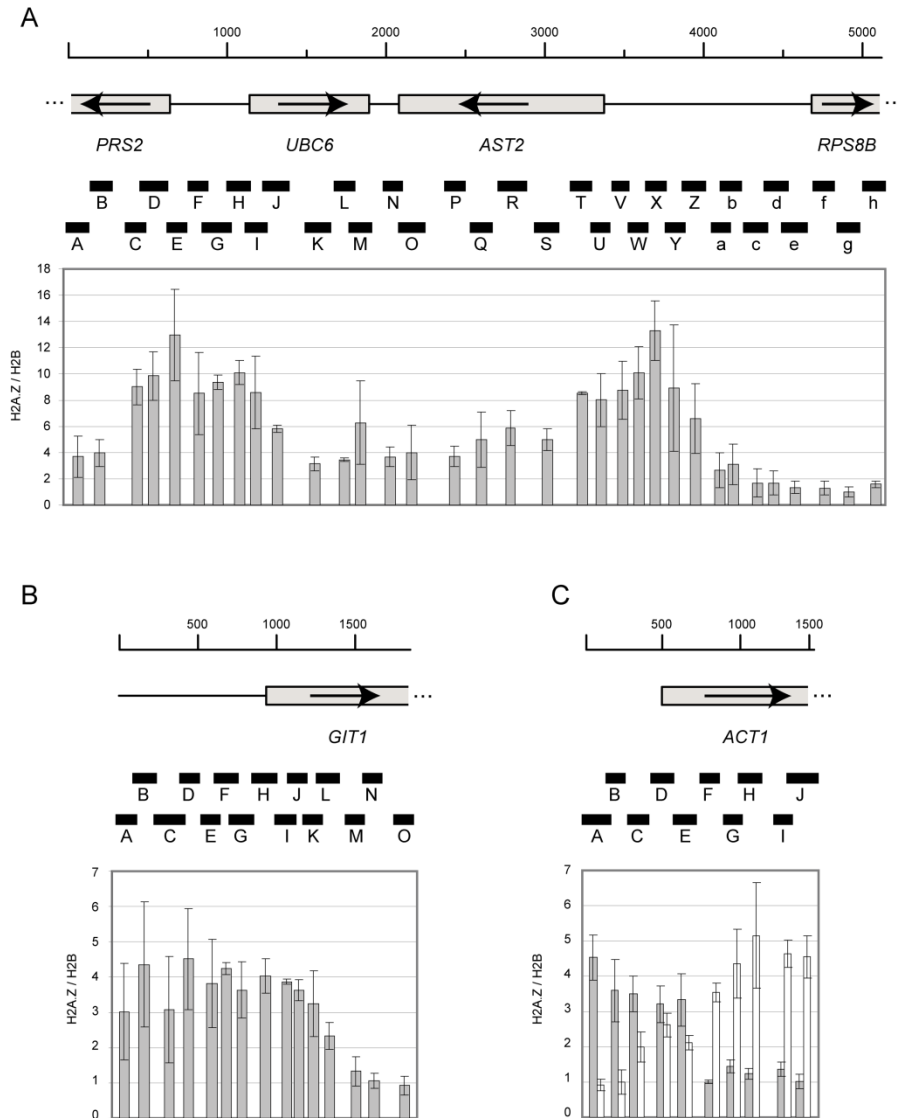


Figure Sup. 3.2: Q-PCR analysis of H2A.Z ChIP experiments.

(A) The region including the *PRS2*, *UBC6*, *AST2* and *RPS8B* genes is drawn to scale, and the positions of the PCR probes used to monitor H2A.Z occupancy are shown as black bars. The H2A.Z/H2B ratios are indicated by grey bars. (B) Same as in (A) but for the *GIT1* locus, a gene contained within a HZAD domain. (C) Same as in (A) and (B) but for the *ACT1* locus. The H2B/Input ratio is also shown as white bars. In each panel, the probe with the lowest ratio was set to “1”.



Figure Sup. 3.3: H2A.Z occupancy at HZAD10.

(A) The raw (green) and smoothed (red) H2A.Z/H2B ratios are shown across a 15 KB region from chromosome IX containing HZAD10 (grey shaded box). The genes in that region are shown in blue. (B) Same as in (A) but for a non-HZAD region from chromosome III.

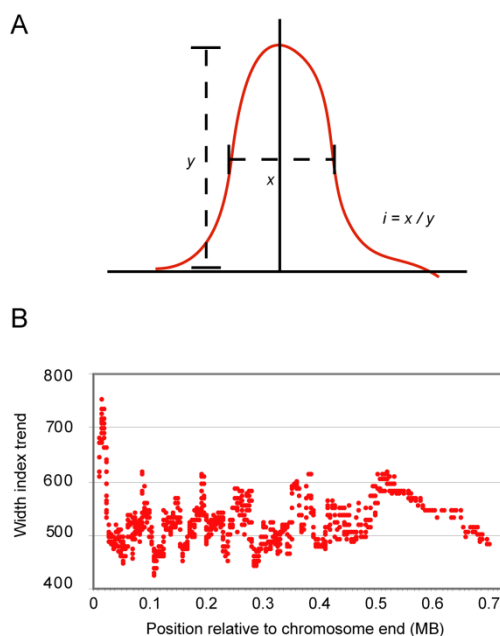


Figure Sup. 3.4: Z loci are wider around telomeres.

(A) Graphical representation of the width index calculation.

In order to look at the spread of the H2A.Z domains across chromosomes we have calculated an index estimating the relative width of these domains. The width index (*i*) is defined as x/y , where *x* is the width observed at $\frac{1}{2} y$, *y* being the height at the maximum of the H2A.Z smooth occupancy curve. The width was divided by the height in order to normalize for the fact that higher peaks will necessarily take more space to reach baseline and will therefore look like wider domains than small peaks. In order to avoid “contamination” by overlapping peaks, maxima localized within less than 1.5 KB from another were not considered in the analysis. **(B)** The width trend was determined by computing a sliding median of the width index (see panel A) across all yeast genes ordered by their distance from a chromosomal end and plotted against that distance.

3.3.7 Acknowledgments

We thank Alain Lavigueur and Benoît Leblanc for critical comments on the manuscript and Svetlana Sadekova for sharing equipment. This work was supported by grants from the CIHR of Canada awarded to L.G. and F.R. and from Genome Québec to M.B. L.G. holds a Canada research chair on mechanisms of gene transcription, and F.R. holds a new investigator award from the CIHR. B.G. is a recipient of a doctoral studentship from NSERC, A.R.B. is a recipient of a doctoral studentship from the IRCM/CIHR cancer research program and N.G. is a recipient of a postdoctoral fellowship from NSERC.

3.4 References

1. Khorasanizadeh, S., *The nucleosome: from genomic organization to genomic regulation*. Cell, 2004. **116**(2): p. 259-72.
2. Vaquero, A., A. Loyola, and D. Reinberg, *The constantly changing face of chromatin*. Sci. Aging Knowledge Environ., 2003. **2003**(14): p. RE4.
3. Felsenfeld, G. and M. Groudine, *Controlling the double helix*. Nature, 2003. **421**(6921): p. 448-53.
4. Flaus, A. and T. Owen-Hughes, *Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer?* Curr. Opin. Genet. Dev., 2004. **14**(2): p. 165-73.
5. Lusser, A. and J.T. Kadonaga, *Chromatin remodeling by ATP-dependent molecular machines*. Bioessays, 2003. **25**(12): p. 1192-200.
6. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**(5532): p. 1074-80.
7. Turner, B.M., *Cellular memory and the histone code*. Cell, 2002. **111**(3): p. 285-91.
8. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* Genes Dev., 2005. **19**(3): p. 295-316.
9. Henikoff, S. and K. Ahmad, *Assembly of Variant Histones into Chromatin*. Annu. Rev. Cell. Dev. Biol., 2005.
10. Thatcher, T.H. and M.A. Gorovsky, *Phylogenetic analysis of the core histones H2A, H2B, H3, and H4*. Nucleic Acids Res., 1994. **22**(2): p. 174-9.
11. Wu, R.S., D. Nishioka, and W.M. Bonner, *Differential conservation of histone 2A variants between mammals and sea urchins*. J. Cell. Biol., 1982. **93**(2): p. 426-31.
12. Liu, X., B. Li, and M.A. Gorovsky, *Essential and nonessential histone H2A variants in Tetrahymena thermophila*. Mol. Cell Biol., 1996. **16**(8): p. 4305-11.

13. Ridgway, P., et al., *Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development*. J. Biol. Chem., 2004. **279**(42): p. 43815-20.
14. Faast, R., et al., *Histone variant H2A.Z is required for early mammalian development*. Curr. Biol., 2001. **11**(15): p. 1183-7.
15. Wu, R.S. and W.M. Bonner, *Separation of basal histone synthesis from S-phase histone synthesis in dividing cells*. Cell, 1981. **27**(2 Pt 1): p. 321-30.
16. Redon, C., et al., *Histone H2A variants H2AX and H2AZ*. Curr. Opin. Genet. Dev., 2002. **12**(2): p. 162-9.
17. Jackson, J.D. and M.A. Gorovsky, *Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants*. Nucleic Acids Res., 2000. **28**(19): p. 3811-6.
18. Clarkson, M.J., et al., *Regions of variant histone His2AvD required for Drosophila development*. Nature, 1999. **399**(6737): p. 694-7.
19. Adam, M., et al., *H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions*. Mol. Cell. Biol., 2001. **21**(18): p. 6270-9.
20. Allis, C.D., et al., *Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila*. Cell, 1980. **20**(3): p. 609-17.
21. Santisteban, M.S., T. Kalashnikova, and M.M. Smith, *Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes*. Cell, 2000. **103**(3): p. 411-22.
22. Larochelle, M. and L. Gaudreau, *H2A.Z has a function reminiscent of an activator required for preferential binding to intergenic DNA*. EMBO J., 2003. **22**(17): p. 4512-22.
23. Farris, S.D., et al., *Transcription-induced Chromatin Remodeling at the c-myc Gene Involves the Local Exchange of Histone H2A.Z*. J. Biol. Chem., 2005. **280**(26): p. 25298-303.

24. Meneghini, M.D., M. Wu, and H.D. Madhani, *Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin*. Cell, 2003. **112**(5): p. 725-36.
25. Swaminathan, J., E.M. Baxter, and V.G. Corces, *The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin*. Genes Dev., 2005. **19**(1): p. 65-76.
26. Dhillon, N. and R.T. Kamakaka, *A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR*. Mol. Cell, 2000. **6**(4): p. 769-80.
27. Fan, J.Y., et al., *H2A.Z Alters the Nucleosome Surface to Promote HP1alpha-Mediated Chromatin Fiber Folding*. Mol. Cell, 2004. **16**(4): p. 655-661.
28. Rangasamy, D., et al., *Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development*. EMBO J., 2003. **22**(7): p. 1599-607.
29. Kobor, M.S., et al., *A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin*. PLoS Biol., 2004. **2**(5): p. E131.
30. Krogan, N.J., et al., *A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1*. Mol. Cell, 2003. **12**(6): p. 1565-76.
31. Mizuguchi, G., et al., *ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex*. Science, 2004. **303**(5656): p. 343-8.
32. Krogan, N.J., et al., *Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4*. Proc. Natl. Acad. Sci. U S A, 2004. **101**(37): p. 13513-8.
33. Rangasamy, D., I. Greaves, and D.J. Tremethick, *RNA interference demonstrates a novel role for H2A.Z in chromosome segregation*. Nat Struct Mol. Biol., 2004. **11**(7): p. 650-5.
34. Leach, T.J., et al., *Histone H2A.Z is widely but nonrandomly distributed in chromosomes of Drosophila melanogaster*. J. Biol. Chem., 2000. **275**(30): p. 23267-72.

35. Suto, R.K., et al., *Crystal structure of a nucleosome core particle containing the variant histone H2A.Z*. Nat. Struct. Biol., 2000. **7**(12): p. 1121-4.
36. Robert, F., et al., *Global position and recruitment of HATs and HDACs in the yeast genome*. Mol. Cell, 2004. **16**(2): p. 199-209.
37. Yuan, G.C., et al., *Genome-scale identification of nucleosome positions in *S. cerevisiae**. Science, 2005. **309**(5734): p. 626-30.
38. Harbison, C.T., et al., *Transcriptional regulatory code of a eukaryotic genome*. Nature, 2004. **431**(7004): p. 99-104.
39. Ren, B., et al., *Genome-wide location and function of DNA binding proteins*. Science, 2000. **290**(5500): p. 2306-9.
40. Ng, H.H., et al., *Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex*. Genes Dev., 2002. **16**(7): p. 806-19.
41. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity*. Mol. Cell, 2003. **11**(3): p. 709-19.
42. Reinke, H. and W. Horz, *Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter*. Mol. Cell, 2003. **11**(6): p. 1599-607.
43. Lee, C.K., et al., *Evidence for nucleosome depletion at active regulatory regions genome-wide*. Nat. Genet., 2004. **36**(8): p. 900-5.
44. Boeger, H., et al., *Nucleosomes unfold completely at a transcriptionally active promoter*. Mol. Cell, 2003. **11**(6): p. 1587-98.
45. Boeger, H., et al., *Removal of promoter nucleosomes by disassembly rather than sliding in vivo*. Mol. Cell, 2004. **14**(5): p. 667-73.
46. Schwabish, M.A. and K. Struhl, *Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II*. Mol. Cell Biol, 2004. **24**(23): p. 10111-7.
47. Fedor, M.J., N.F. Lue, and R.D. Kornberg, *Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast*. J. Mol. Biol., 1988. **204**(1): p. 109-27.

48. Fedor, M.J. and R.D. Kornberg, *Upstream activation sequence-dependent alteration of chromatin structure and transcription activation of the yeast GAL1-GAL10 genes*. Mol. Cell Biol., 1989. **9**(4): p. 1721-32.
49. Fisher-Adams, G. and M. Grunstein, *Yeast histone H4 and H3 N-termini have different effects on the chromatin structure of the GAL1 promoter*. EMBO J., 1995. **14**(7): p. 1468-77.
50. Axelrod, J.D., M.S. Reagan, and J. Majors, *GAL4 disrupts a repressing nucleosome during activation of GAL1 transcription in vivo*. Genes Dev., 1993. **7**(5): p. 857-69.
51. Abbott, D.W., et al., *Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation*. J. Biol. Chem., 2001. **276**(45): p. 41945-9.
52. Fan, J.Y., et al., *The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states*. Nat. Struct. Biol., 2002. **9**(3): p. 172-6.
53. Rusche, L.N., A.L. Kirchmaier, and J. Rine, *The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae*. Annu. Rev. Biochem., 2003. **72**: p. 481-516.
54. Grewal, S.I. and D. Moazed, *Heterochromatin and epigenetic control of gene expression*. Science, 2003. **301**(5634): p. 798-802.
55. Martinez-Campa, C., et al., *Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1*. Mol. Cell, 2004. **15**(1): p. 69-81.
56. Schneider, B.L., et al., *Use of polymerase chain reaction epitope tagging for protein tagging in Saccharomyces cerevisiae*. Yeast, 1995. **11**(13): p. 1265-74.
57. Pokholok, D.K., et al., *Genome-wide Map of Nucleosome Acetylation and Methylation in Yeast*. Cell, 2005. **122**(4): p. 517-27.
58. Svaren, J., U. Venter, and W. Horz, *In vivo analysis of nucleosome structure and transcription factor binding in Saccharomyces cerevisiae*. Methods Mol. Genet., 1995. **6**: p. 153-67.

59. Lieb, J.D., et al., *Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association*. Nat. Genet., 2001. **28**(4): p. 327-34.
60. Kurdistani, S.K., S. Tavazoie, and M. Grunstein, *Mapping global histone acetylation patterns to gene expression*. Cell, 2004. **117**(6): p. 721-33.
61. Wang, B., et al., *53BP1, a mediator of the DNA damage checkpoint*. Science, 2002. **298**(5597): p. 1435-8.
62. Holstege, F.C., et al., *Dissecting the regulatory circuitry of a eukaryotic genome*. Cell, 1998. **95**(5): p. 717-28.
63. Kadosh, D. and K. Struhl, *Targeted recruitment of the Sin3-Rpd3 histone deacetylase complex generates a highly localized domain of repressed chromatin in vivo*. Mol. Cell Biol., 1998. **18**(9): p. 5121-7.
64. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity*. Mol. Cell, 2003. **11**: p. 709-19.
65. Lee, C.-K., et al., *Evidence for nucleosome depletion at active regulatory regions genome-wide*. Nat. Genet., 2004. **36**: p. 900-5.
66. Reinke, H. and W. Hörz, *Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter*. Mol. Cell, 2003. **11**: p. 1599-607.
67. Schwabish, M.A. and K. Struhl, *Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II*. Mol. Cell. Biol., 2004. **24**: p. 10111-7.

4 Le cycle de la l'ARN polymérase II est universel et orchestré le long des gènes par des interactions complexes entre enzymes, kinase, phosphatase et isomérase

4.1 Mise en contexte de l'article

Après l'étude d'un aspect de la chromatine au chapitre précédent, cet article s'intéresse au cycle de phosphorylation du CTD de l'ARN polymérase. Durant plus d'une décennie, les anticorps H5 et H14 ont été utilisés presque exclusivement pour l'étude des marques de phosphorylation des sérines 2 et 5 du CTD respectivement. En 2007, de nouveaux anticorps font leur apparition, permettant de comparer les profils obtenus ainsi que d'obtenir celui de la marque de phosphorylation de la sérine 7 plus récemment identifiée. Publiés à la suite de trois articles sur le CTD de la polymérase, ces travaux apportent cependant de nouveaux éléments, et permettent de clarifier certaines des hypothèses énoncées récemment. La localisation des marques de phosphorylation du CTD ainsi que des études biochimiques nous ont permis de mieux comprendre le rôle des kinases et phosphatases dans le ballet extrêmement chorégraphié qui aboutit à la mise en place des marques de phosphorylation. Cet article met aussi en évidence les limites actuelles des moyens d'investigation.

4.2 Contribution des auteurs

A. Bataille et **C. Jeronimo** ont réalisé la majorité des expériences. **M-È. Fortin**, **L. Laramée**, **A. Forest** et **M. Bergeron** ont réalisé le reste des expériences. **A.R. Bataille** et **P-É. Jacques** ont analysé les résultats de ChIP-chip. **S.D. Hanes** a fourni les souches *ess1* ainsi que la protéine recombinante Ess1 et aidé au développement de l'article. **F. Robert**, **A.R. Bataille** et **C. Jeronimo** ont rédigé l'article.

4.3 Manuscrit

A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays Between Kinase, Phosphatase and Isomerase Enzymes Along Genes.

Alain R. Bataille[©], Célia Jeronimo[©], Pierre-Étienne Jacques, Louise Laramée, Marie-Ève Fortin, Audrey Forest, Maxime Bergeron, Steven D. Hanes and François Robert*

[©]These authors made equal contributions to this work, *Corresponding author

Article publié dans *Molecular Cell* (2012) 45(2):158-70.

Keywords: RNA polymerase II; CTD; Kin28; Ctk1; Bur1; Fcp1; Ssu72; Ess1

4.3.1 Abstract

Transcription by RNA polymerase II (RNAPII) is coupled to mRNA processing and chromatin modifications via the C-terminal domain (CTD) of its largest subunit, consisting of multiple repeats of the heptapeptide YSPTSPS. Pioneering studies showed that CTD serines are differentially phosphorylated along genes in a prescribed pattern during the transcription cycle. Genome-wide analyses challenged this idea, suggesting that this cycle is non-uniform among different genes. Moreover, the respective role of enzymes responsible for CTD modifications remains controversial. Here, we systematically profiled the location of the RNAPII phosphoisoforms in wild type cells and mutants for most CTD modifying enzymes. Together with results of *in vitro* assays, these data reveal a complex interplay between the modifying enzymes, and provide evidence that the CTD cycle is uniform across genes. We also identify Ssu72 as the Ser7 phosphatase and show that proline isomerization is a key regulator of CTD dephosphorylation at the end of genes.

4.3.2 Introduction

Transcription of protein-coding genes is achieved in eukaryotes by RNA polymerase II (RNAPII), a large multisubunit enzyme highly conserved from yeast to humans [1]. Rpb1, the largest subunit of RNAPII, contains a C-terminal domain (CTD) made of repeats of the conserved heptapeptide $Y_1S_2P_3T_4S_5P_6S_7$ [2]. The number of repeats varies from 26 in yeast to 52 in mammals with some degenerate repeats. The CTD is essential for viability and partial truncations or site specific mutations lead to specific growth defects [3]. This unstructured domain, strategically located adjacent to the mRNA exit channel of the enzyme, acts as a scaffold to coordinate the binding of proteins involved in the different phases of transcription and allows for the coupling of transcription with other nuclear processes such as mRNA maturation and the modification of chromatin [4, 5]. The three serines (Ser2, Ser5, Ser7), the threonine and the tyrosine of each repeat can be phosphorylated, and both peptidyl-prolyl bonds can be isomerized by a prolyl isomerase [6-8]. More recently, the CTD has also been shown to be methylated at one degenerate repeat in human cells [9]. The multitude of possible CTD modification states led to the hypothesis of a CTD “code” [10], in which the different combinations of CTD modifications would orchestrate the sequential recruitment of the numerous factors required during the transcription cycle.

RNAPII is recruited to the preinitiation complex (PIC) in an unphosphorylated state [11]. Following its binding to the promoter, Ser5 is rapidly phosphorylated by Kin28 (Cdk7 in human), the kinase subunit of the transcription factor TFIIF. This is thought to facilitate the dissociation of RNAPII from the PIC and, therefore, to promote promoter escape [12]. Concomitantly, Ser5 phosphorylation promotes the recruitment of capping and splicing factors [4] as well as the Set1 methyltransferase complex (COMPASS in yeast, MLL in human) and the Set3C and Rpd3S HDAC complexes [13-16]. During early elongation, P-Ser5 levels drop while Ser2 is phosphorylated by the kinases Ctk1 (Cdk12 in human) [17] and/or Bur1 (Cdk9 in human) [18, 19], leading to the recruitment of the histone methyltransferase Set2 and the activation of the Rpd3S complex, preventing cryptic

transcription initiated within genes [20]. When RNAPII reaches the termination site, mRNA processing factors are recruited to the CTD for cleavage and polyadenylation of the nascent mRNA [21, 22]. This process also requires proper CTD phosphorylation [23]. During or following termination, the CTD is dephosphorylated and the polymerase is recycled for another round of transcription [24]. This final dephosphorylation is thought to be catalyzed by two phosphatases, Ssu72 and Fcp1 [24, 25]. Phosphorylation of Ser7 has also been revealed to be a feature of all RNAPII transcribed genes [7, 26], and Kin28 was demonstrated to be the relevant kinase [27]. While Ser7 phosphorylation has been clearly implicated in transcription termination of non-coding small nuclear RNAs (snRNAs) [28], its function at protein encoding genes has yet to be elucidated.

Despite substantial efforts to dissect this important process, our understanding of how the CTD modifications are added and removed in a timely manner is incomplete. For example, there are two known P-Ser5 phosphatases, Rtr1 and Ssu72. While Rtr1 was recently shown to be responsible for the dephosphorylation of Ser5 at the Ser5-to-Ser2 transition in the 5' region of genes [29], the function of Ssu72 during the CTD cycle remains enigmatic, despite evidence showing that it is a bona fide Ser5 phosphatase. Indeed, Ssu72 is part of the cleavage and polyadenylation factor (CPF) [30, 31], far from the region where most of Ser5 dephosphorylation is thought to occur, a fact difficult to reconcile with its role as a Ser5 phosphatase. The identity of the Ser7 phosphatase is also not certain. It is known that mutations in a proline isomerase, Ess1, lead to an increase in bulk levels of P-Ser5 [32, 33] but it is not certain how exactly this occurs or its impact on the transcription cycle. There is also considerable debate about the relative importance of the two Ser2 kinases (Ctk1 and Bur1) during transcription [34]. While this was until recently thought to be a yeast-specific question, the discovery of a metazoan ortholog of Ctk1 [17] now makes this question relevant to higher eukaryotes. The role of the Ser2 phosphatase, Fcp1, is also under debate. Evidence suggests that it plays a role during elongation [24] while others have shown that it associates with RNAPII in solution [35], implying that Ser2 is dephosphorylated after termination. Finally, and most

importantly, there are two recent reports suggesting that the CTD cycle may not be uniform at all genes, but rather established in a gene-specific manner [36, 37]. This would constitute an important shift in the paradigm, since the CTD cycle was previously regarded as a universal phenomenon. A third study, however, challenged this view by showing that the phospho-serine marks are deposited and removed as a function of the distance from transcription start site (TSS) and termination site, respectively, with no noticeable difference between genes [38].

Here, we sought to resolve these issues by taking a systematic approach to examine the genome-wide distribution of the various CTD modifications using a panel of RNAPII CTD phospho-specific antibodies; both in wild type cells and in mutants for most of the CTD kinases, phosphatases and the isomerase. Surprisingly, the data indicate considerable interplay between the CTD kinases and also between the CTD phosphatases. Our data also identified Ssu72 as the Ser7 phosphatase and showed that the isomerase Ess1 plays a key role in the dephosphorylation of the CTD at the end of genes. Most importantly, our data show that transcription operates in a uniform mode at virtually all genes.

4.3.3 Results

4.3.3.1 Phospho-CTD profiling on active genes using multiple antibodies

To help decipher the CTD cycle, we localized RNAPII and its various phospho-isoforms by performing a systematic ChIP-chip assay using tiling microarrays covering the entire yeast genome [39]. RNAPII occupancy was monitored using antibodies specific for the CTD (8WG16) and the Rpb3 subunit (W0012), as well as a myc antibody (9E10) against a myc-tagged Rpb1 protein. The 8WG16 antibody is known to recognize the unphosphorylated form of the CTD [26] however, when used in ChIP-chip experiments, it gave a strong enrichment all along entire transcription units, following the same pattern as the two other RNAPII antibodies (**Figure 4.1 A**). While this result may seem surprising, it has been observed before [16] and could be explained by the fact that at least some unphosphorylated repeats are likely to be present on most elongating polymerase molecules. Consequently, we used both Rpb3 and 8WG16 antibodies to measure total RNAPII along genes in this study.

Localization of the RNAPII phosphorylation marks was done using phospho-specific antibodies available for each mark (**Figure 4.1 B**). As reported previously by the Eick group and others, CTD antibodies directed against a given P-Ser mark are greatly influenced by phosphate groups on neighboring residues [24, 26]. Moreover, antibody affinities can vary when assayed by different groups and techniques [24, 26, 40]. Consequently, ChIPs for the same mark sometimes give different profiles using different antibodies. For example, the H14 antibody, which is extensively used in the literature for the study of P-Ser5, shows little if no enrichment in the 3' ends of ORFs (**Figure 4.1 B**). This has led to the generally accepted view that Ser5 is dephosphorylated early during elongation. Using a different anti-P-Ser5 antibody (3E8), we and others [38, 41, 42] observed that a significant amount of Ser5 remains phosphorylated along the entire length of the gene. Even more striking is the difference observed using the antibodies specific for P-Ser2. The profile obtained

with H5, an antibody used extensively to study Ser2 phosphorylation, shows a net decrease as RNAPII approaches the polyadenylation site, whereas the 3E10 antibody, directed against the same phosphorylation mark, produced an increase at this location and a sharp decrease after the transcription end (**Figure 4.1 B**). Through this work we have considered the specificity of the CTD antibodies as previously described by Chapman *et al.* [26].

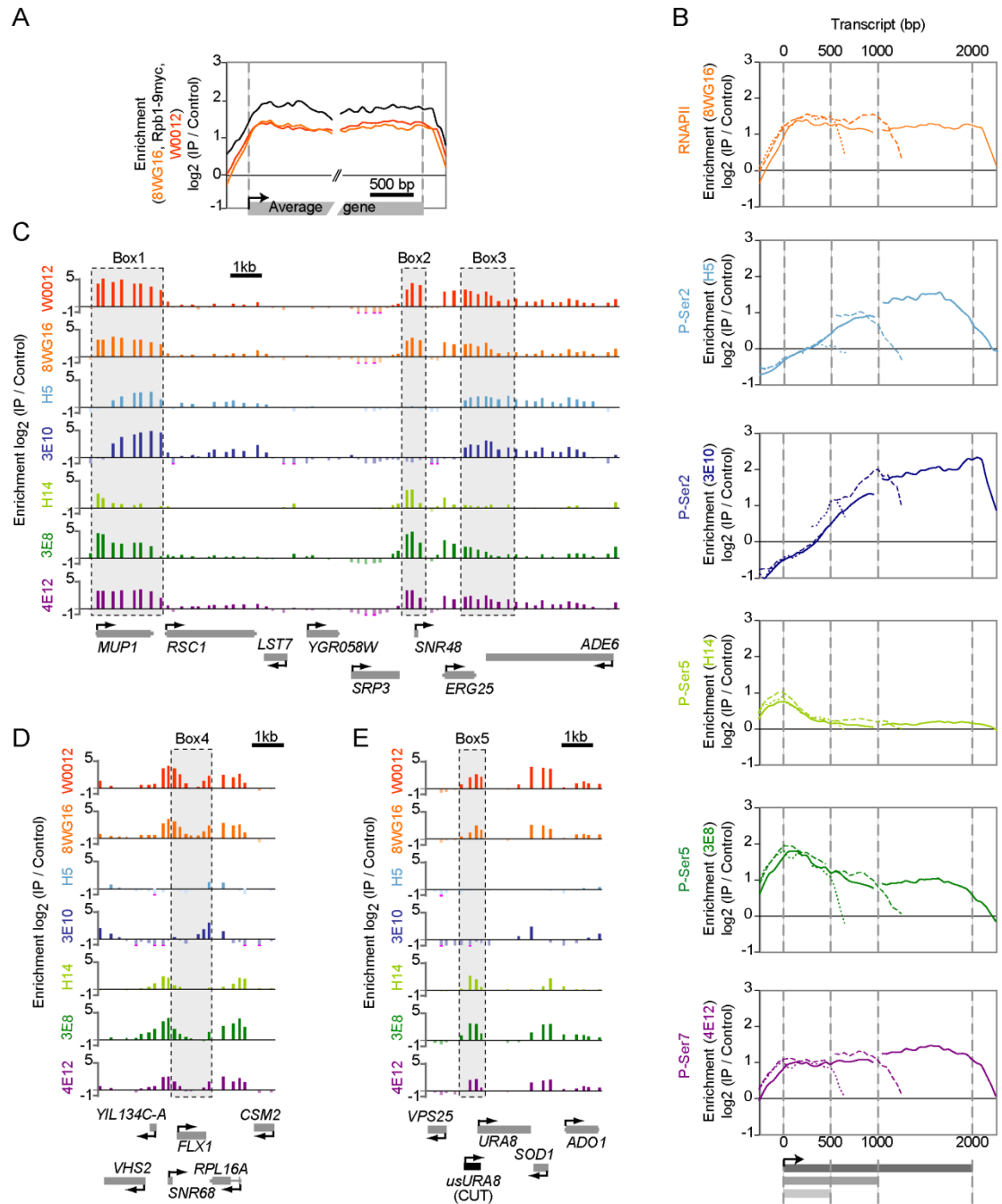


Figure 4.1: A uniform CTD cycle at most transcribed genes.

A) Profile of RNAPII using either anti-CTD (8WG16), anti-myc (9E10; Rpb1-9myc) or anti-Rpb3 (W0012) antibodies on a metagene made with long (1500 bp-2500 bp) and transcribed genes (based on 8WG16 average binding across the ORF). B) Profile of RNAPII (8WG16) and its phosphorylation marks (P-Ser2 [H5 and 3E10], P-

Ser5 [H14 and 3E8], P-Ser7 [4E12]) along metagenes representing three classes of genes based on average gene length (see Experimental Procedures for details). Profiles are shown for long (2000 bp +/- 500 bp; solid line), medium size (1000 bp +/- 250 bp; dashed line) and small (500 bp +/- 100 bp; dotted line) genes. C-E) Genomic regions harboring genes that show canonical (Box1-2) and non-canonical (Box3-5) CTD phospho-serine profiles. Shaded boxes (Box1-5) highlight regions described in the text. Annotated genes are shown in grey and the non-annotated *usURA8* cryptic unstable transcript (CUT) is shown in black. ChIP-chip enrichment ratios are shown for all RNAPII antibodies tested in this study in wild type cells. See also **Figure Sup. 4.1**.

4.3.3.2 A similar CTD cycle operates on genes of different size classes

We examined the CTD phosphorylation patterns of three classes of genes that differ by their average size. It is clear that RNAPII modifications occur with the same kinetics no matter what the gene length (**Figure 4.1 B**). In agreement with recent work by the Cramer group [38], modifications did not scale to gene size, meaning that particular phosphorylation marks along genes depend mostly on the absolute position (in base pairs) from the TSS. Thus, a transcribing RNAPII molecule does not sense whether it has transcribed x or y percent of a gene; instead it only “knows” that it has transcribed x or y base pairs of DNA.

P-Ser5 and P-Ser7 are deposited with the same kinetics within the first few base pairs, after which P-Ser5 (3E8) is partially dephosphorylated within the next 750 base pairs (**Figure 4.1 B**). Both these marks start to decline about 350 base pairs prior to the polyadenylation site. Given the resolution of ChIP assays, this suggests that passing through the polyA signal triggers the dephosphorylation of Ser5 and Ser7. P-Ser2 accumulates with slower kinetics and reaches a plateau after the polymerase has transcribed about one kilobase. The decline of the P-Ser2 signal occurs with kinetics similar to those of P-Ser5 and P-Ser7 when assayed by the H5 antibody but when assayed using the 3E10 antibody declines with total RNAPII, suggesting that it occurs after termination. Because marks are deposited as a function of the distance

from the TSS, short genes do not reach the same P-Ser5 and P-Ser7 plateau and also accumulate significantly lower levels of P-Ser2 (see dotted traces in **Figure 4.1 B** and **Figure 4.1 C**, Box2 for an example (the *SNR48* gene)). This differential phosphorylation pattern on shorter genes likely has implications for the mechanism of transcription termination (see Discussion). However, to obtain a clearer picture of all phosphorylation and dephosphorylation phases, we focused on longer genes to produce metagene profiles.

The average profiles depicted in **Figure 4.1 B** seemed to be the rule for the vast majority, if not the totality, of protein-coding genes. After manually and computationally looking for “exceptional genes”, we came to the conclusion that, in most cases, these can be explained by the genomic landscape (overlapping genes, neighboring genes, anti-sense transcripts, etc) as exemplified in **Figure 4.1 C-E**. For instance, while Boxes 1 and 2 (**Figure 4.1 C**) illustrate long and short genes showing the canonical profile, Box3 highlights a region where two transcribed genes are transcribed toward each other creating the illusion of non-canonical profiles. Indeed, the proximity of *ERG25* and *ADE6* confounds their individual 3’ends profiles. The short *ERG25* gene looks as if RNAPII was reading through (W0012 and 8WG16), while the long *ADE6* shows excessive P-Ser5 (3E8) and P-Ser7 (4E12) in its 3’end. Another example of a gene having a profile skewed by its close neighbors is the *FLX1* gene shown in **Figure 4.1 D** (Box4). This gene appears to have uneven levels of RNAPII (W0012 and 8WG16) and excessive amounts of P-Ser5 (3E8) and P-Ser7 (4E12) at both ends. This gene, however, is not transcribed in these growth conditions [43], so the enrichment observed is a consequence of the proximity of its two highly transcribed neighbors (*SNR68* and *RPL16A*). Finally, what looks like non-canonical profiles can sometimes be explained by the presence of non-annotated transcripts. A very striking example is shown in Box5 (**Figure 4.1 E**). RNAPII and its associated phospho-isoforms are present only in the very 5’end of the *URA8* gene. This creates the impression that RNAPII is paused near the promoter, a process that was never described in yeast. However, there is a cryptic unstable transcript (CUT) in that region known to prevent initiation from the *URA8* promoter [44]. This short transcript

is likely responsible for the observed RNAPII signals. Because unambiguous exceptions are extremely rare, we conclude that the deposition and removal of the different CTD phospho-serines follow the same rules at virtually all genes. This conclusion contrasts with a previous report, as discussed later (see also **Figure Sup. 4.1**).

4.3.3.3 The kinase activity of Kin28 down regulates the activity of Bur1 against P-Ser5 and 7 along the coding region

Our *in vitro* kinase assays using TAP-purified complexes (**Figure Sup. 4.2 A** and **Table Sup. 4.1**) show that CTD kinases have poor specificities *in vitro* (**Figure Sup. 4.2 B**), suggesting that *in vivo* assays are required to understand their function. To start dissecting the role of the CTD kinases, we first used an analog-sensitive *kin28* mutant strain (*kin28-as*) (**Figure 4.2 A**). As expected, the levels of both P-Ser5 (green traces) and P-Ser7 (purple traces) are greatly diminished at the 5' end of genes in that mutant (compare solid with dashed traces). Interestingly, the levels of both P-Ser5 and P-Ser7 increase throughout the ORF to a level far beyond what is normally observed in wild type cells. Consequently, the distribution, but not the overall level of P-Ser5 and P-Ser7, is affected in the absence of Kin28; a result that was also observed by western blot using total cellular extracts (**Figure Sup. 4.2 C**). This result suggests that in the absence of the catalytic activity of Kin28, other kinases phosphorylate Ser5 and Ser7 residues during elongation (see below).

We then investigated the potential role of the kinases Ctk1 and Bur1 in Ser5 and Ser7 phosphorylation using *ctk1Δ* and *bur2Δ* deletions, since analog-sensitive *bur1* and *ctk1* strains are only partially inhibited by ATP analogs [36, 45]. Note that since Bur2 is not the catalytic subunit of the Bur1/2 complex, we first confirmed that Bur1 kinase activity is inhibited in the *bur2Δ* strain by looking at the phosphorylation levels of Spt5, Rpb1 and Rad6 (indirectly by testing H3K4me3) (**Figure Sup. 4.2 D**). In the ChIP-chip analysis, we found that P-Ser5 and P-Ser7 levels are only mildly reduced in the body of genes in the *ctk1Δ* and *bur2Δ* mutants (**Figure 4.2 B**),

suggesting that Ctk1 and Bur1 are not solely responsible for maintaining P-Ser5 and P-Ser7 levels along the gene in wild type cells.

When deleted in a *kin28-as* background, however, *BUR2* leads to a partial but significant reduction of P-Ser5 and P-Ser7 levels (**Figure 4.2 C**). Taken together, the *in vitro* and *in vivo* results suggest that Bur1 is a potent Ser5 and Ser7 kinase, but that its activity is repressed in the presence of a functional Kin28 complex. It should also be noted that none of these kinase mutant strains harbor a complete loss of P-Ser5 or P-Ser7 (**Figure 4.2**), suggesting that other kinases may also be contributing. Ctk1 potentially plays such an opportunistic role in the absence of a functional Kin28. This is supported by the fact that, *in vitro*, Ctk1 complex can phosphorylate the CTD at all three serines, although with different efficiencies (**Figure Sup. 4.2 B**). Because Mediator has been recently proposed to play a role in elongation [46], it is possible that its associated kinase, Srb10 (Cdk8 in human), is involved in this phosphorylation. However, in wild type cells, neither Srb10 nor Ctk1 play any significant role in maintaining P-Ser5 and P-Ser7 levels (see **Figure 4.2 B** for *ctk1Δ* and **Figure Sup. 4.2 E** for *srb10Δ*), suggesting that if these kinases play any role in the phosphorylation of Ser5 and Ser7, they do so mainly in an opportunistic manner, as we showed for the Bur1/2 kinase (**Figure 4.2 C**).

Data from Tietjen et al. (2010), using a *bur1-as* strain, are similar to ours, but the modest decrease of P-Ser7 at the 3' end of genes was attributed to the inefficient inhibition of the analog-sensitive Bur1. Our data using the *bur2Δ* strain, which lacks all detectable kinase activity (**Figure Sup. 4.2 D**), instead suggest that Bur1's contribution to Ser7 phosphorylation is modest due to the presence of active Kin28. We propose that Bur1 and other CTD kinases act opportunistically to phosphorylate Ser5 and Ser7 in the absence of a functional Kin28.

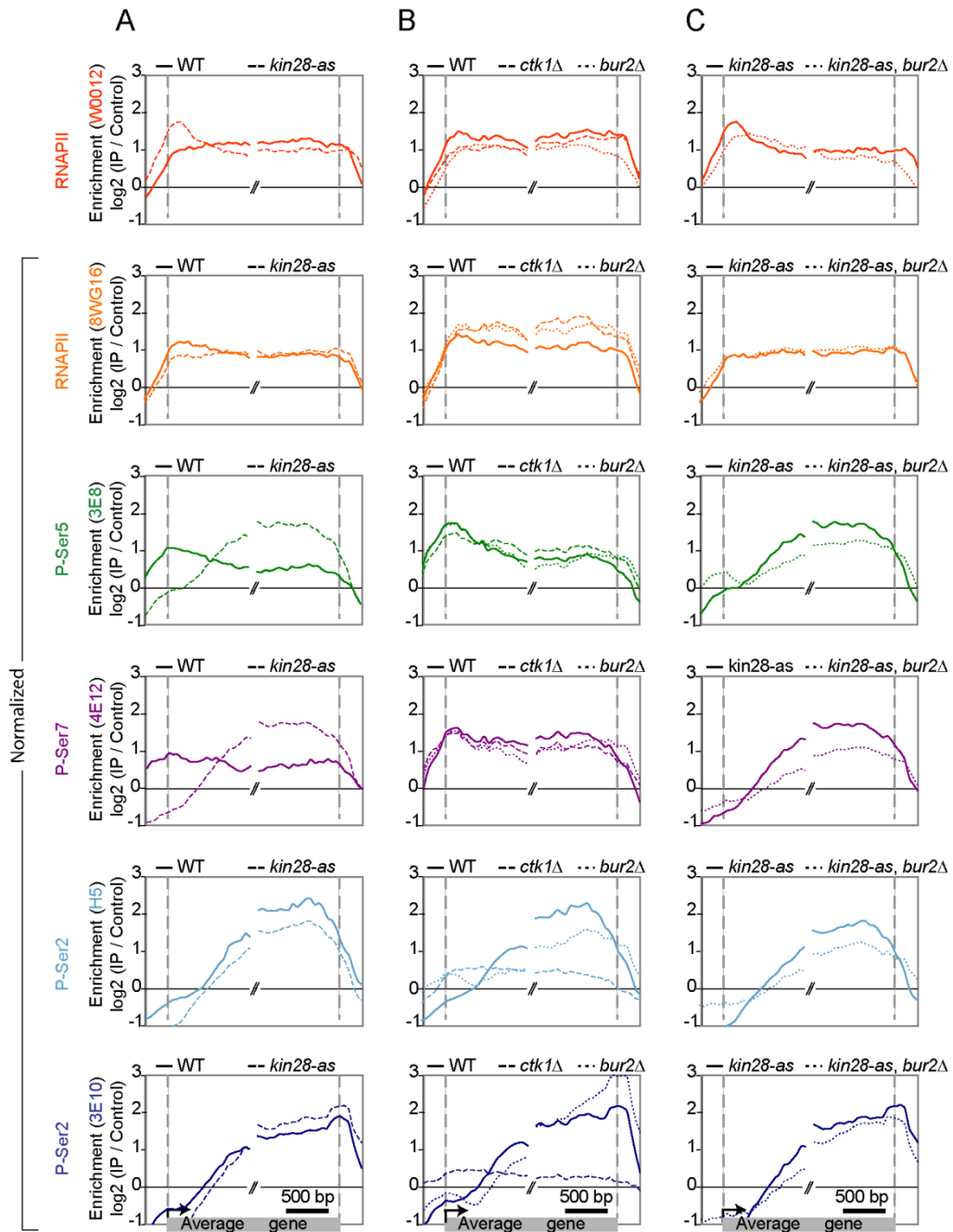


Figure 4.2: Complex interplays between CTD kinases.

Profile of RNAPII and its phospho-isoforms along a metagene corresponding to long and expressed genes in **A**) *kin28-as* (dashed trace) and its isogenic wild type (solid trace); **B**) *ctk1Δ* (dashed trace), *bur2Δ* (dotted trace) and their isogenic wild type (solid trace); **C**) *kin28-as* (solid trace) and *kin28-as/bur2Δ* (dotted trace). The top panel shows total RNAPII (Rpb3, W0012) and the panels below show RNAPII

(8WG16), P-Ser5 (3E8), P-Ser7 (4E12) and P-Ser2 (H5, 3E10), all normalized for the difference in RNAPII (Rpb3, W0012) level in the mutants (as described in Supplemental Experimental Procedures). This allows us to appreciate the difference in phosphorylation levels without being biased by the fact that RNAPII does accumulate at the 5' end of genes in some mutants. See also **Figure Sup. 4.2**.

4.3.3.4 Ctk1 is the major serine 2 kinase during transcription *in vivo*

Ctk1 and Bur1 both phosphorylate Ser2 *in vitro* (**Figure Sup. 4.2 B**); however, there is considerable debate on the extent of Bur1-dependent Ser2 phosphorylation during transcription [34], particularly because Bur1 targets other substrates, notably Spt5 and Rad6 as well as NELF in higher organisms [47]. Our ChIP-chip data show that the bulk of Ser2 phosphorylation on chromatin is dependent on Ctk1 but not Bur1 (**Figure 4.2 B**). It is also noteworthy that some residual P-Ser2 remains, in particular when measured using the H5 antibody, in the *ctk1Δ* mutant (**Figure 4.2 B**). Moreover, deletion of *BUR2* has some partial effect on P-Ser2, especially when assayed by the H5 antibody (**Figure 4.2 B**). Taken together, these results suggest that Ctk1 is the primary Ser2 kinase, while Bur1 contributes mostly to a specific variant of P-Ser2 (the one preferentially recognized by the H5 antibody). This is in agreement with a recent study by the Hinnebusch group which proposed that Bur1 is a minor Ser2 kinase contributing to the phosphorylation of Ser2 in the 5' end of genes [45]. Because the H5 signal depends on both P-Ser5 and P-Ser2, however, we cannot formally exclude the possibility that the contribution of Bur1/2 to the H5 variant of P-Ser2 reflects a partial contribution to P-Ser5, although this appears unlikely, since we do not see any contribution of Bur1/2 on P-Ser5 (**Figure 4.2 B**, green traces). Also noteworthy is the fact that the 3E10 signal does not reach a plateau in the *bur2Δ* strain (**Figure 4.2 B**, lower panel). The signal rather keeps increasing until the very end of the gene, suggesting that Bur1/2 has some inhibitory role on another kinase. Because Ctk1 is the best CTD kinase at generating the 3E10 signal *in vitro* (**Figure Sup. 4.2 B**), one possibility is that the absence of Bur1/2 leaves room for Ctk1 to further phosphorylate the CTD on Ser2 (the 3E10 variant).

Since this effect is not observed for the H5 signal, the data collectively suggest that Bur1/2 is required for keeping the right balance of different variants of P-Ser2 *in vivo*. Interestingly, this excess of 3E10 in the 3' end of genes is eliminated in the *bur2Δ, kin28-as* double mutant (**Figure 4.2 C**, bottom panel), suggesting that Kin28 may be involved in this phenomenon. This, however, is likely to be indirect, since Kin28 occupies promoters (data not shown) and because Kin28 complexes are not very active at generating the 3E10 variant of P-Ser2 *in vitro* (**Figure Sup. 4.2 B**) or in wild type cells *in vivo* (**Figure 4.2 A**). Overall, this data, as well as the data presented in the previous section, show how complex interplays between CTD kinases is key in the establishment of proper CTD phosphorylation marks along genes.

4.3.3.5 Kin28 and Bur1 have antagonistic roles in early steps of transcription

Current models suggest that phosphorylation of Ser5 by Kin28 leads to the dissociation of RNAPII from the PIC, leading to promoter escape [12]. Consistent with this, our *kin28-as* ChIP-chip data revealed a slight accumulation of total RNAPII at the 5' end of genes (**Figure 4.2 A**, top panel), similar to that previously observed [36]. As expected, this defect is not observed in *ctk1Δ* or *bur2Δ* strains (**Figure 4.2 B**). However, accumulation of RNAPII at the 5' end of genes is rescued by the deletion of *BUR2* in the *kin28-as* (**Figure 4.2 C**). These results indicate an opposing role of Kin28 and Bur1 during an early step of transcription, suggesting that the regulation of promoter escape by Kin28 may be more complex than previously envisioned.

4.3.3.6 Ssu72 dephosphorylates both serine 5 and serine 7 prior to termination

Dephosphorylation of Ser5 was proposed to be carried out by one of two RNAPII phosphatases, Ssu72 and/or Rtr1. Rtr1 localizes to the 5' end of genes and contributes to the dephosphorylation of Ser5 in that region [29]. However, the belonging of Ssu72 to the CPF complex was difficult to reconcile with a role of a

Ser5 phosphatase, since Ser5 dephosphorylation was thought to occur at the 5' ends of genes. Our data, in agreement with other recent studies [38, 41, 42], shows that P-Ser5 persists throughout the genes (using the 3E8 antibody; **Figure 4.1 B**), and makes it attractive to think that Ssu72 targets P-Ser5 at the 3' end of genes. Indeed, we found that Ssu72 is physically localized to the 3' end of genes, generating a profile almost identical to Pti1, another member of CPF (**Figure 4.3 A**). Then, using a degron strain for Ssu72 (degron-*ssu72*), we investigated the consequences of Ssu72 depletion on RNAPII phosphorylation along genes by ChIP-chip. Interestingly, we found that P-Ser5 (3E8 antibody) persists further downstream from its usual location in the degron strain (**Figure 4.3 B** and **Figure Sup. 4.3**, green dashed trace). In fact, P-Ser5 decreases with the same kinetics as RNAPII, suggesting that in the absence of a functional Ssu72, transcription terminates with a polymerase that is abnormally phosphorylated on Ser5. This result confirms that Ssu72 is the Ser5 phosphatase responsible for the dephosphorylation of Ser5 just prior to termination.

Because P-Ser7 drops with the same kinetics as P-Ser5 (**Figure 4.1 B**), we hypothesized that Ssu72 may also be the P-Ser7 phosphatase. Indeed, the dephosphorylation of P-Ser7 (4E12) shows a similar overall pattern as P-Ser5 in the degron-*ssu72* strain, with P-Ser7 extending further to the end of the gene and its removal occurring with kinetics similar to RNAPII (**Figure 4.3 B** and **Figure Sup. 4.3**). Very little effect was observed for P-Ser2 (H5, 3E10). We also tested the activity of Ssu72 against P-Ser7 *in vitro*. As shown in **Figure 4.3 C**, GST-Ssu72, but not a catalytic mutant GST-Ssu72 C15S, dephosphorylates both P-Ser5 and P-Ser7. This activity is specific since no activity is detected against P-Ser2. These data show that Ssu72 is a dual phosphatase that targets both Ser5 and Ser7 prior to termination.

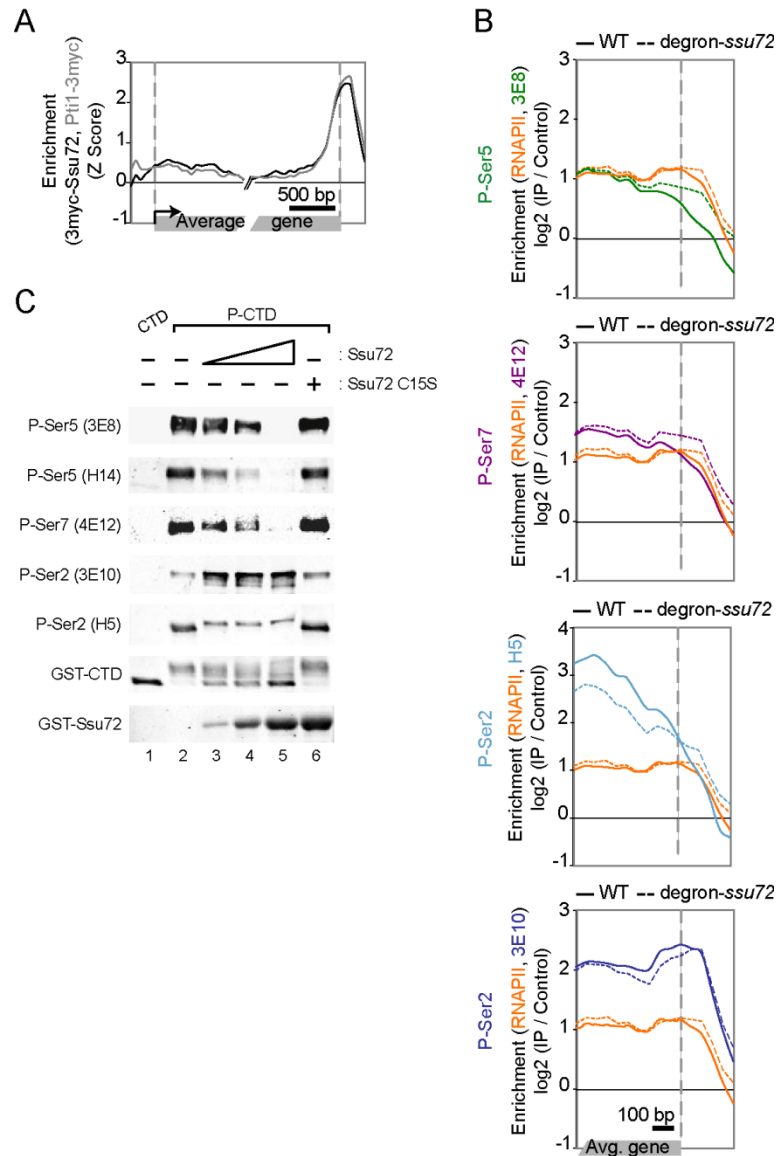


Figure 4.3: Ssu72 dephosphorylates Ser5 and Ser7 at the ends of genes.

A) Profile of Ssu72 (3myc-Ssu72) along a metagenome with the polyadenylation factor Pti1 (Pti1-3myc) as a control. **B)** Profile of RNAPII (8WG16) and its phospho-isoforms (3E8, 4E12, H5 and 3E10) along a metagenome in the degron-ssu72 strain (dashed traces) and its isogenic wild type (solid traces). Only the 3' ends of genes are shown. See **Figure Sup. 4.3** for the complete gene. **C)** *In vitro* CTD phosphatase activity of recombinant Ssu72. GST-CTD substrate was hyperphosphorylated using Kin28- and Ctk1-TAP complexes as described in Experimental Procedures. Increasing amounts (0.5, 2.5 and 5 μ M) of GST-Ssu72 (lanes 3-5) were incubated

with the hyperphosphorylated GST-CTD (P-CTD). An equivalent amount (5 μ M) of a catalytic mutant GST-Ssu72 C15S was used as a negative control (lane 6). Phosphatase reactions were stopped by the addition of gel-loading buffer, resolved on SDS-PAGE gels, and subjected to western blotting with the indicated antibodies. See also **Figure Sup. 4.3**.

4.3.3.7 The prolyl isomerase Ess1 plays a critical role in CTD dephosphorylation by Ssu72

Ess1 was shown to preferentially bind the P-Ser5 over the P-Ser2 form of the CTD [48] and stimulate the dephosphorylation of Ser5 by Ssu72 *in vivo* [32, 33] and *in vitro* [49]. This requires the isomerase activity of Ess1, which catalyzes the *cis/trans* inter-conversion of the peptidyl-proline bond between Ser5-Pro6. Such conformational changes can affect enzyme binding, and the *cis*-isomer form was shown to be the preferred substrate of Ssu72 [49]. We sought to examine the role of Ess1 on both P-Ser5 and P-Ser7 enrichment at the end of genes. For that, we used a *ts*-strain bearing the *ess1*-H164R mutation, which is ~10,000 fold less active than the wild type enzyme at permissive temperature [48]. We found that, in the Ess1 mutant, P-Ser5 (3E8) levels remained high until RNAPII dissociates from DNA (**Figure 4.4**, upper panel and **Figure Sup. 4.4**), as observed with the degenon-*ssu72* strain (**Figure 4.3 B** and **Figure Sup. 4.3**). Interestingly, the same is observed for P-Ser7 (4E12) (**Figure 4.4** and **Figure Sup. 4.4**). To our knowledge, this is the first indication for a role of Ess1 in regulating levels of Ser7 phosphorylation. Very little effect was observed on the levels of P-Ser2 (3E10) (**Figure 4.4**, lower panel and **Figure Sup. 4.4**). These data show that the Ess1 isomerase can specifically stimulate the dephosphorylation of both Ser5 and Ser7 at the 3' end of genes, supporting the idea that Ess1 plays a crucial role in controlling the CTD during the transcription cycle.

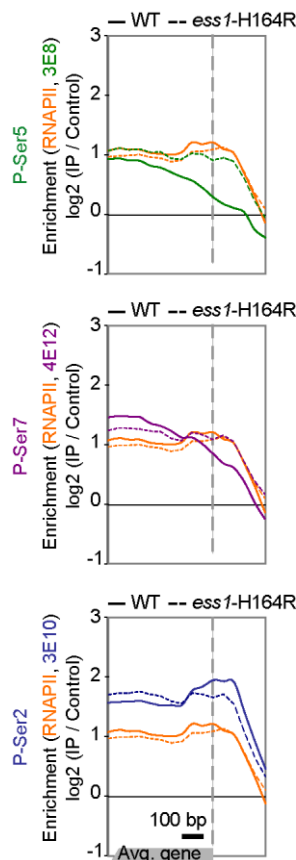


Figure 4.4: Ess1 is required for the dephosphorylation of Ser5 and Ser7 at the ends of genes.

Profile of RNAPII (8WG16) and its phospho-isoforms (3E8, 4E12 and 3E10) along a metagene in the *ess1-H164R* strain (dashed traces) and its isogenic wild type (solid traces). Only the 3' end of genes is shown. See **Figure Sup. 4.4** for the complete gene.

4.3.3.8 Two variants of P-Ser2 are dephosphorylated with different kinetics

P-Ser2 levels decline at or near the 3' end of genes. However, the kinetics of decline differ when assayed using H5 versus 3E10 antibodies; P-Ser2 (H5) declines prior to RNAPII reaching the 3' end, while P-Ser2 (3E10) declines only after or concomitant with termination (**Figure 4.1 B**). Since H5 prefers P-Ser2 with adjacent Ser5 being phosphorylated [26] and because Ser5 is dephosphorylated by Ssu72 in

that region (**Figure 4.3 B** and **Figure Sup. 4.3**), it is possible that the apparent drop in the H5 signal prior to termination is reminiscent of Ser5 dephosphorylation rather than Ser2 dephosphorylation. The H5 signal, however, does not increase at the 3' end of the gene in *degron-ssu72*, a strain where P-Ser5 extends all the way to the end of the gene (see **Figure 4.3 B**), demonstrating that the drop in H5 signal prior to termination cannot be attributed to the decrease of P-Ser5 in that region. Thus, there are at least two variants of P-Ser2, one of which is dephosphorylated prior to termination with similar kinetics to P-Ser5 and P-Ser7 (and preferentially recognized by the H5 antibody) and another that is dephosphorylated after or during termination (preferentially recognized by the 3E10 antibody).

4.3.3.9 Fcp1 partially dephosphorylates Ser2 prior to termination

Fcp1 is well characterized as a Ser2 phosphatase [24, 35]. To test whether Fcp1 dephosphorylates one or both of the P-Ser2 variants, we performed *in vitro* phosphatase assays using TAP-purified Fcp1 (**Figure Sup. 4.2 A**) on a recombinant GST-CTD substrate. Strikingly, this assay shows that, while Fcp1 dephosphorylates P-Ser2 (H5), it has little if any effect on P-Ser2 (3E10) (**Figure 4.5 A**, lanes 1-6). This suggests that Fcp1 is responsible for the dephosphorylation of the H5 variant prior to termination. To test this possibility, we profiled phospho-CTD marks in *fcp1-1*, a strain carrying a thermosensitive allele of this essential gene. Unexpectedly, RNAPII, as assayed by the 8WG16 antibody (recognizing the non-phosphorylated CTD), is dramatically decreased in the body of genes in the *fcp1-1* strain, whereas when assayed using an antibody against Rpb3 (W0012), no significant changes in RNAPII density are observed (**Figure 4.5 B**). Further analyses showed that most antibodies directed against the phospho-CTD showed similar profiles to 8WG16 (**Figure Sup. 4.5 D-F**). Because the *fcp1-1* mutation decreases the signal for both the non-phosphorylated and the phosphorylated CTD, our data suggests that the CTD is somehow “masked” in the *fcp1-1* strain. In order to test for such a possibility, we performed immunoprecipitations from wild type and *fcp1-1* total cellular extracts, followed by western blots to test the ability of our anti-CTD antibodies to interact with their respective epitopes in conditions reminiscent of those present in ChIP

experiments. The results, shown in **Figure Sup. 4.5 A**, demonstrate that in the *fcp1-1* extract, all anti-CTD antibodies (including 8WG16 and the anti-P-Ser2 antibodies) are crippled in their ability to pull down RNAPII. These experiments suggest that the excess of P-Ser2 on the CTD in *fcp1-1* cells causes the CTD to be somehow masked, affecting the ability of anti-CTD antibodies to interact with their epitopes. Quite remarkably, in ChIP experiments, this affects the signal in the 3' end of genes (where P-Ser2 is abundant) but not in the 5' end of genes (where P-Ser2 is absent); see for instance 8WG16 in **Figure 4.5 B**. To correct for this artifact, we normalized all our ChIP data in *fcp1-1* cells to the 8WG16 profile (see **Figure Sup. 4.5 B-I** for all data prior to and after normalization).

After normalization, the results show that P-Ser2 (mostly the H5 variant) is massively increased within the transcribed region in *fcp1-1* cells (**Figure 4.5 B**, lower panel), consistent with previous work showing that Ctk1 and Fcp1 oppose each other during elongation [24]. Note that the 3E10 signal is also increased in these cells, albeit to a lesser extent than the H5 signal. This is perhaps due to the fact that the 3E10 antibody also recognizes the P-Ser5/P-Ser2 epitope, although with less affinity than it recognizes the P-Ser2-only epitope [26]. Most relevant to this study is what occurs at the 3' end of genes, where the P-Ser2 (H5) signal extends further downstream in the *fcp1-1* mutant (**Figure 4.5 B**, lower panel). The signal does not extend all the way to the end of genes, likely because some Fcp1 activity remains at permissive temperature. Together with the *in vitro* assays shown in **Figure 4.5 A**, these data indicate that Fcp1 contributes to the removal of the H5 variant of P-Ser2 prior to termination.

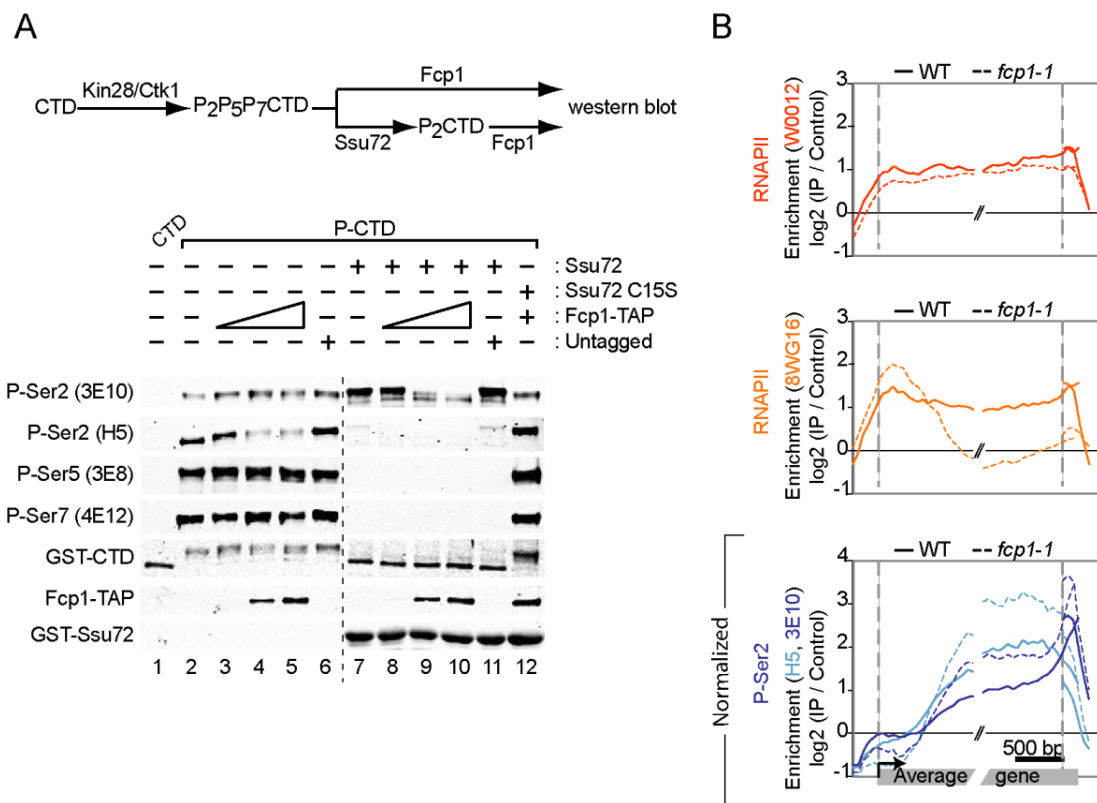


Figure 4.5: Two variants of P-Ser2 are removed in two steps by the concerted action of Fcp1 and Ssu72.

A) *In vitro* CTD phosphatase activity of TAP-purified Fcp1. The experimental diagram is indicated at the top of the panel. GST-CTD substrate was hyperphosphorylated using Kin28- and Ctk1-TAP complexes. Increased amounts of Fcp1-TAP complex (lanes 3-5) were incubated with either fully phosphorylated GST-CTD substrate or with an hyperphosphorylated GST-CTD that was first treated with 5 μ M of GST-Ssu72 (lanes 8-10) or 5 μ M of a catalytic mutant GST-Ssu72 C15S (lane 12). An eluate from an untagged strain was used as a control of the Fcp1-TAP phosphatase activity (lanes 6, 11). Phosphatase reactions were stopped by the addition of gel-loading buffer, resolved on SDS-PAGE gels, and subjected to western blotting with the indicated antibodies. **B)** Profile of RNAPII (W0012, 8WG16) and P-Ser2 (H5, 3E10) along a metagenome in wild type (solid traces) and *fcp1-1* (dashed traces) cells. The gene lists contributing to the metagenes have been selected based on anti-Rpb3 (W0012) occupancy in each strain. The P-Ser2 profile is normalized based on 8WG16 occupancy in order to correct for CTD masking. See also **Figure Sup. 4.5**.

4.3.3.10 Ssu72 enables complete dephosphorylation of Ser2 by Fcp1 after termination

What activity removes the remaining P-Ser2 (3E10) variant after termination? We investigated whether the ability of Fcp1 to remove the 3E10 variant of P-Ser2 is altered during late stages of transcription to allow the complete dephosphorylation of P-Ser2. This could result from prior dephosphorylation of Ser5 (and Ser7) by Ssu72. Therefore we tested if pre-incubation of a fully phosphorylated GST-CTD substrate with Ssu72 would affect further dephosphorylation by Fcp1 *in vitro*. As shown in lanes 7-12 of **Figure 4.5 A**, pre-incubation of the phosphorylated GST-CTD with Ssu72 allows for efficient removal of the 3E10 variant of P-Ser2 by Fcp1. This stimulation requires the catalytic activity of Ssu72 since a catalytic mutant is unable to stimulate Fcp1's activity towards P-Ser2 (3E10). These data suggest a coupled two-step model in which the activity of Ssu72 enables the complete dephosphorylation of Ser2 by Fcp1 at the end of genes or when RNAPII is released (soluble). Because Ssu72 dephosphorylates Ser5, however, it would normally increase the number of P-Ser2-only repeats; the ones recognized preferentially by 3E10 and not efficiently removed by Fcp1 *in vitro*. As mentioned above, however, the 3E10 antibody also recognizes the P-Ser5/P-Ser2 epitope [26]. It is therefore possible that the P-Ser2 variant that requires both Ssu72 and Fcp1 phosphatases is a subset of P-Ser5/P-Ser2 epitopes that are not recognized by Fcp1, unless Ssu72 first dephosphorylates Ser5. Alternatively, this may involve some other modification of the CTD induced by Ssu72. Regardless of the mechanism involved, this interplay between the two phosphatases is likely important for termination and mRNA 3'end processing.

4.3.4 Discussion

4.3.4.1 A universal CTD cycle

While this work was in progress, three different groups reported genome-wide profiles of RNAPII and its phospho-isoforms. Ansari and colleagues proposed that

genes of different functions adopt different CTD phosphorylation profiles [36], while the Cramer group suggested, as we do, that the phospho-marks are placed and removed similarly at most if not all genes [38]. The third report is intermediate since Bentley and colleagues claim that the CTD phosphorylation patterns do not scale with gene size, consistent with our observations and those of the Cramer group, but also propose that promoter structure (defined by their nucleosomal organization) influences CTD phosphorylation profiles [37]. Our re-analyses of the data from Tietjen et al. (2010) shows that the group of genes identified as having non-canonical patterns are heavily biased towards small or large genes as well as genes overlapping with neighbors (**Figure Sup. 4.1**). This, at least in part, explains the difference in the phosphorylation patterns. As for the effect of chromatin structure at promoters, we found that the “occupied proximal nucleosome” genes and the “depleted promoter nucleosome” have similar profiles in our dataset and the datasets from the Cramer and Ansari groups (data not shown). Overall, the data from all four laboratories are quite similar (except for P-Ser7 as discussed below) and the differences are mostly a matter of analysis and interpretation. When accounting for gene size, our data and those of others suggest that the CTD cycle is very similar at all genes. Importantly, this implies that short genes will have higher levels of P-Ser5 and lower levels of P-Ser2 than long genes when they reach the termination site. This most likely explains why short genes, such as snoRNAs and genes coding for small proteins, tend to utilize an alternative mechanism for termination relative to most class II genes [50, 51].

4.3.4.2 Is CTD phosphorylation important for transcription?

A common perception is that phosphorylation of the CTD stimulates transcription. However, the actual evidence on this point is rather weak. An early microarray study using a *kin28-ts* allele had suggested a global role of Kin28 in transcription [43] but later analyses demonstrated this effect was due to a defect in mRNA capping that reduced the half-life of mRNAs, rather than a defect in transcription *per se* [52, 53]. In fact, inhibiting Kin28 did not significantly affect the level of total RNAPII or its P-Ser forms along genes. Similarly, P-Ser2 is thought to

be important for transcription elongation, despite the fact that deletion of *CTK1* has been shown to be dispensable for the recruitment of elongation factors to active genes *in vivo* and also has no detectable effect on the level of RNAPII on genes [22]. Our data extend these studies, since we observe no loss of RNAPII levels on genes in either *kin28-as*, *ctk1Δ* or *bur2Δ* cells. Although inhibition of Kin28 did not affect RNAPII levels along genes, it did, however, affect its distribution: RNAPII accumulates in the 5' region of genes in the absence of a functional Kin28. This may relate to the fact that Kin28, at least in *in vitro* assays, allows for RNAPII to break contact with PIC components, notably Mediator, stimulating promoter escape [12].

We also made the surprising observation that the Kin28 defect can be suppressed by the deletion of Bur1. This suggests that Kin28 may regulate promoter escape at least by regulating a negative effect of Bur1. Regardless of the mechanism involved, our data suggest that, although not required for transcription *in vivo*, the kinase activity of Kin28 is nevertheless involved in fine tuning the transcription reaction and may play a regulatory role. Thus, the work presented here, along with reconsideration of prior studies, indicates that the role of CTD phosphorylation in transcription has been overstated, and that although phospho-serines (in particular P-Ser2) constitute good markers for transcription, they play very minor roles on the efficiency of the transcription reaction *in vivo*. The main function of CTD phosphorylation is more likely to facilitate coupling of transcription to other cellular processes.

4.3.4.3 Phosphorylation of Ser7

Uncertainties remain regarding phosphorylation of Ser7. First, P-Ser7 has been proposed to be deposited in two waves, first by Kin28 and later by Bur1 during elongation [36], but our data clearly show that Bur1 only phosphorylates Ser7 along genes when Kin28 is not functional. Again, although minor, the phosphorylation of Ser7 by Bur1, like that of Ser2, might be functionally important. Second, the P-Ser7 profile itself is not certain, with some groups reporting that P-Ser7 has a similar profile to P-Ser5 (assayed by the 3E8 antibody), with a peak in the 5' end of genes

that is reduced to a plateau during elongation [38]. However, the enrichment observed at the 5' ends may be due to the relatively high concentrations of 4E12 antibody used. Indeed, P-Ser7 profiles, as assayed by ChIP-qPCR on selected genes, have different shapes when low versus high amounts of 4E12 are used [38]. When lower amounts are used (as in our study), P-Ser7 profiles show a rather uniform level along genes. Because the 4E12 antibody cross reacts with the non-phosphorylated CTD under some conditions (data not shown), we elected to use the lower concentrations for our ChIP-chip assays. While we cannot rule out the possibility that we are missing a 5' peak, the profiles we report are specific to P-Ser7, since a strain in which all CTD Ser7 were mutated to alanines shows a complete loss of P-Ser7 signal (**Figure Sup. 4.2 F**). Finally, P-Ser7 was proposed not to correlate with transcription [37]. We looked at correlation between transcription rate and P-Ser7 levels and found Pearson correlations of 0.81, 0.93, 0.88 and 0.45 using data from our study, Mayer et al. [38], Tietjen et al. [36] and Kim et al. [37], respectively. This suggests that in all but the last dataset, P-Ser7 levels do reflect transcription rate, as is the case for the other CTD phospho-serines.

4.3.4.4 Phosphorylation of Ser2 by Bur1/2

Our ChIP-chip data, in agreement with recent work by the Hinnebusch group [45], show that, while Ctk1 is the major Ser2 kinase, Bur1 nevertheless plays a bona fide role in Ser2 phosphorylation *in vivo*. Interestingly, when assessed by western blot using the H5 and 3E10 antibodies, the effect of *BUR2* deletion on P-Ser2 levels in total cell extracts appears much more significant than when assayed by ChIP (**Figure Sup. 4.2 D** and **Figure 4.2 B**). This perhaps suggests that Bur1 may play a role in the phosphorylation of RNAPII outside of the transcription cycle, maybe during mitosis when the CTD is hyperphosphorylated or under certain stress conditions. This, however, would imply that the majority of RNAPII recognized by the 3E10 antibody is not engaged in transcription. Alternatively, this may suggest that the function of the 3E10 variant of Ser2 phosphorylated by Bur1/2 (versus the one phosphorylated by Ctk1) is “masked” in chromatin and, thus, not visible in ChIP experiments. Nevertheless, this discrepancy between the effects of *bur1/2* mutants on ChIP versus

western blot signals may be a clue to some novel roles of this kinase and calls for additional work.

4.3.4.5 Different variants of the phospho-marks

An important concept that emerged from this study is the existence of different variants of the same phospho-serine. At least two observations argue for this complexity. First, P-Ser5 is removed in two waves; initially by Rtr1 and later by Ssu72. If all P-Ser5 were equivalent, Rtr1 would remove them all. The fact that a fraction of the P-Ser5 residues are resistant to Rtr1 suggests that variants exist and that the two P-Ser5 phosphatases have different preferences for different variants. Second, Ser2 is dephosphorylated in two different phases; part of it is removed prior to termination, while the rest is removed after termination. Once again, if all P-Ser2 were equivalent, they would all be removed prior to termination by Fcp1.

What distinguishes the different phospho-serine variants is not clear. However, we envision at least four different possibilities. First, CTD repeats that are proximal to the core of the enzyme have been shown to be functionally distinct from those that are more distal [26, 54, 55]. A second possibility would be that degenerate repeats are functionally distinct from consensus repeats with respect to targeting by the different CTD modifying enzymes. A third possibility is that the *trans* versus *cis* conformations of the Ser-Pro bonds create functionally different phospho-serine variants. Finally, phospho-serine variants may arise from the differential phosphorylation of neighboring serine residues. Future experiments will likely reveal us more about this intriguing level of complexity.

4.3.5 Experimental procedures

4.3.5.1 Yeast strains, antibodies, cell extracts, purification of protein complexes and recombinant proteins

See Supplemental Experimental Procedures.

4.3.5.2 ChIP-chip assay

The ChIP-chip assays were performed as previously described [16]. The microarrays were purchased from Agilent Technologies and contain a total of 44,290 T_m-adjusted 60-mer probes covering the entire yeast genome for an average density of one probe every 275 bp (± 100 bp) within the probed regions (catalog # G4493A). Each microarray was normalized using the Lima Loess method, and replicates were combined using a weighted average method as previously described [56]. More details regarding protocols and data analysis can be found in Supplemental Experimental Procedures.

4.3.5.3 Accession numbers

All raw and processed data have been deposited in the GEO database under accession number GSE29403.

4.3.5.4 *In vitro* CTD phosphatase assay

Recombinant GST-CTD substrate was hyperphosphorylated using Kin28- and Ctk1-TAP complexes as described in Supplemental Experimental Procedures but reactions were incubated at room temperature for 24 hours. Kinase reactions were stopped by removal of the unincorporated ATP using an Illustra MicroSpin G-25 Column (GE Healthcare). Phosphatase reaction mixtures (60 μ L) containing ~200 ng of hyperphosphorylated GST-CTD, 50 mM Tris-HCl pH6.5, 10 mM MgCl₂, 20 mM KCl, 0.1 mM EDTA, 0.025 % Tween 20, 2 % glycerol, 1 mM DTT, and Ssu72 or Fcp1 as specified were incubated for 90 min at 30°C. Reactions were stopped by the addition of SDS-loading buffer, resolved on SDS-PAGE gels, and subjected to western blotting.

4.3.6 Supplemental data

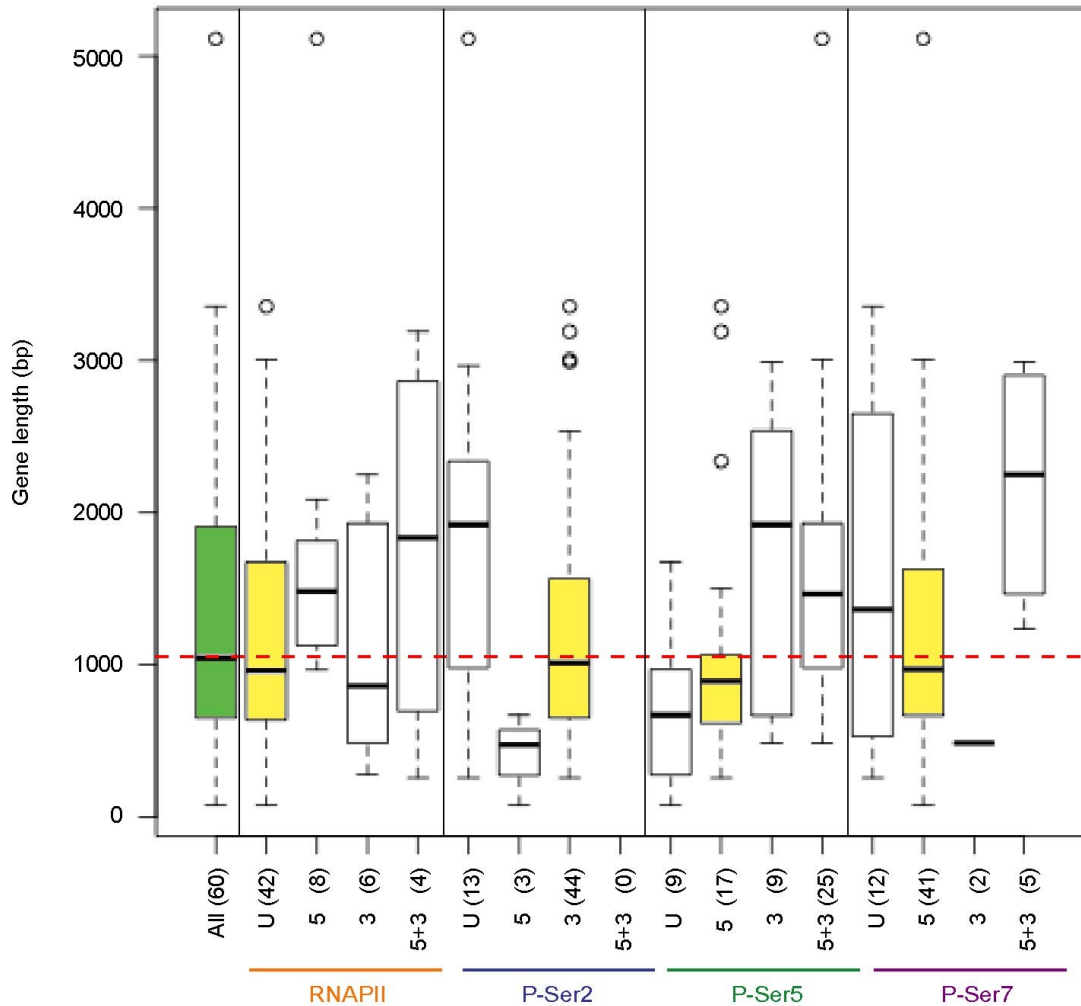


Figure Sup. 4.1: Gene groups previously identified as having non-canonical profiles by Tietjen *et al.*, (2010) [36] are populated by genes with biased gene length.

A companion to **Figure 4.1**. A box plot of the length of the genes included within the clusters identified as having different profiles for RNAPII (orange), P-Ser2 (blue), P-Ser5 (green) and P-Ser7 (purple) by [36]. The gene lists are from the “Set III” in Figure 2 C from [36]. “U” refers to clusters of genes with a uniform level throughout; “5” and “3” refer to genes with higher levels at the 5’ and 3’ end respectively; while

“5+3” refers to genes having high amount at both ends with a dip in the middle. The numbers in parentheses represent the number of genes within each cluster. The clusters representing the expected results have their box labeled in yellow. The green box represents the whole set of genes and the dashed red line represents the average gene size of the whole set. Except in the case of P-Ser5, for each phosphoserine mark, the cluster having the most genes is the expected cluster (yellow clusters). The data show that even if the other clusters were “real”, they would represent a minority of genes. More importantly, the expected clusters are the only ones that are not biased towards gene size. Indeed, and quite strikingly, all the unexpected clusters (white boxes) are biased toward either short or long genes. Because these clusters were identified by parsing the data into an equal number of bins for each gene, short and long genes are somehow distorted, which artificially creates these clusters. For example, long genes will appear to have a more uniform P-Ser2 profile since most of the 5' portion of genes (where P-Ser2 levels increase) will account for only one or two bins. Conversely, short genes have more uniform P-Ser5 levels since P-Ser5 is partially dephosphorylated after about 750 base pairs.

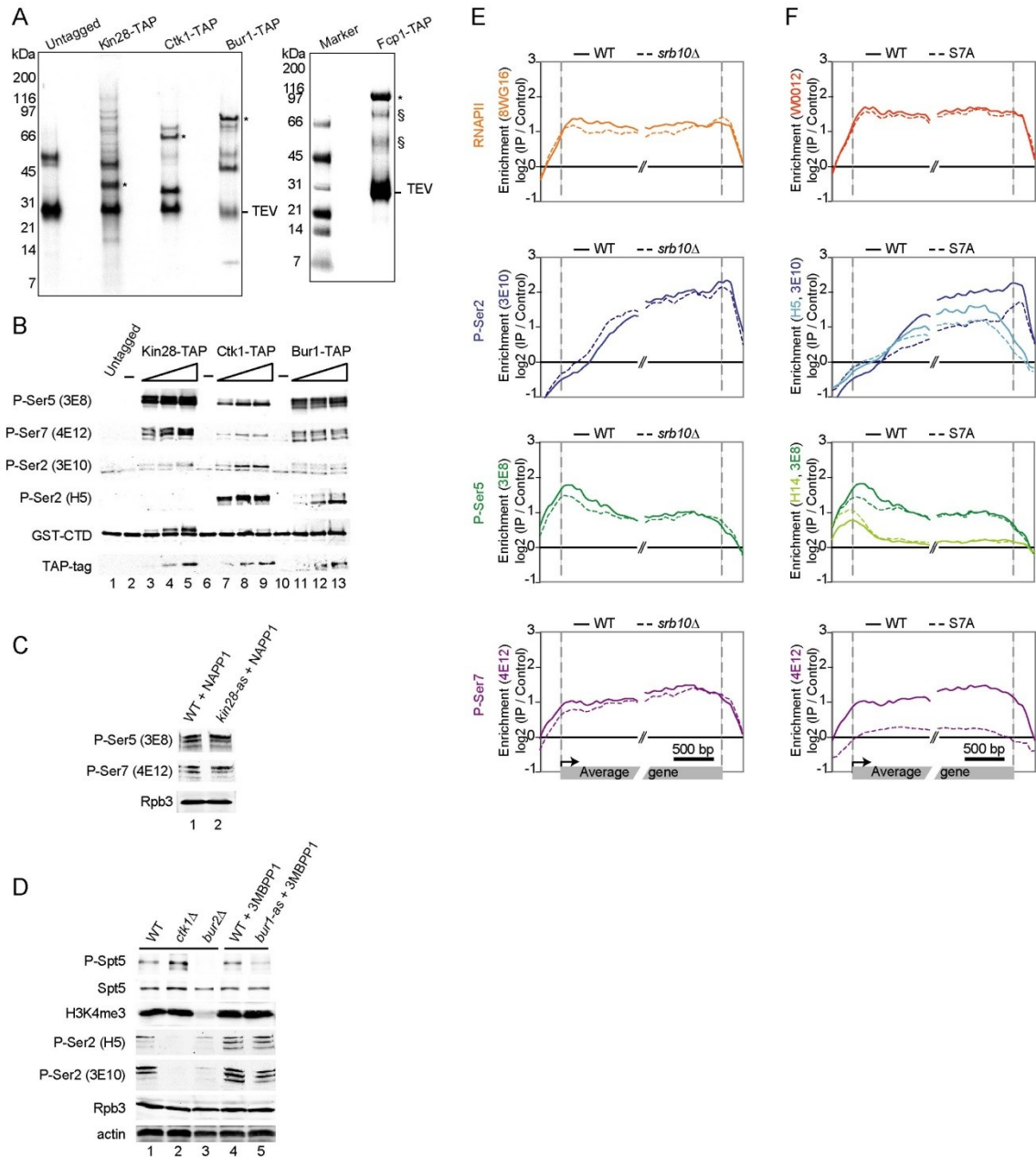


Figure Sup. 4.2: *In vitro* and *in vivo* evidence for Bur1/2 being an opportunistic kinase.

A companion to **Figure 4.2**. **A)** Gels showing the TAP-purified CTD kinase and phosphatase complexes. Kin28-, Ctk1-, Bur1-, and Fcp1-associated complexes were affinity purified via the TAP tag. The asterisk (*) denotes the TAP-tagged (bait) catalytic subunit. An equivalent amount of the eluate from an untagged strain is also shown. Associated proteins were resolved on SDS-PAGE and visualized by silver staining. The band corresponding to the TEV protease used in the TAP procedure is

indicated (TEV). Contaminants identified in the untagged control purification are indicated (§). **B)** *In vitro* CTD kinase activity of TAP-purified Kin28, Ctk1 and Bur1 complexes using recombinant GST-CTD as a substrate. Increasing amounts of Kin28-TAP (lanes 3-5), Ctk1-TAP (lanes 7-9) or Bur1-TAP (lanes 11-13) complexes were titrated into the reaction based on the detection of the TAP-tagged (bait) catalytical subunit. An equivalent amount of the eluate from an untagged strain was used as a control (lane 1). The reactions were stopped by the addition of gel-loading buffer and the reaction products were resolved by SDS-PAGE. The CTD kinase activity was determined by western blot using the indicated antibodies. **C)** A western blot showing total P-Ser5 (3E8), P-Ser7 (4E12) and Rpb3 (W0012) in total cellular extracts prepared from wild type (lane 1) and *kin28-as* (lane 2) cells, both treated with the inhibitor NAPP1. **D)** A western blot showing P-Spt5, Spt5, H3K4me3, P-Ser2 (H5, 3E10), Rpb3 and actin in total cellular extracts prepared from wild type (lane 1), *ctk1Δ* (lane 2) and *bur2Δ* (lane 3) cells as well as *bur1-as* cells (lane 5) and its isogenic wild type (lane 4) both treated with the inhibitor 3MBPP1. All cellular extracts were prepared using a quick extract protocol in order to preserve the phospho mark on Spt5 (see Supplemental Experimental Procedures). **E)** Profile of RNAPII (8WG16) and its phospho-isoforms (P-Ser2 (3E10), P-Ser5 (3E8) and P-Ser7 (4E12)) along a metagene in *srb10Δ* strain (dashed traces) and its isogenic wild type (solid traces). **F)** Profile of RNAPII (W0012) and its phospho-isoforms (P-Ser2 [3E10 and H5], P-Ser5 [3E8 and H14] and P-Ser7 [4E12]) along a metagene in the S7A mutant (dashed traces) and its isogenic wild type (solid traces).

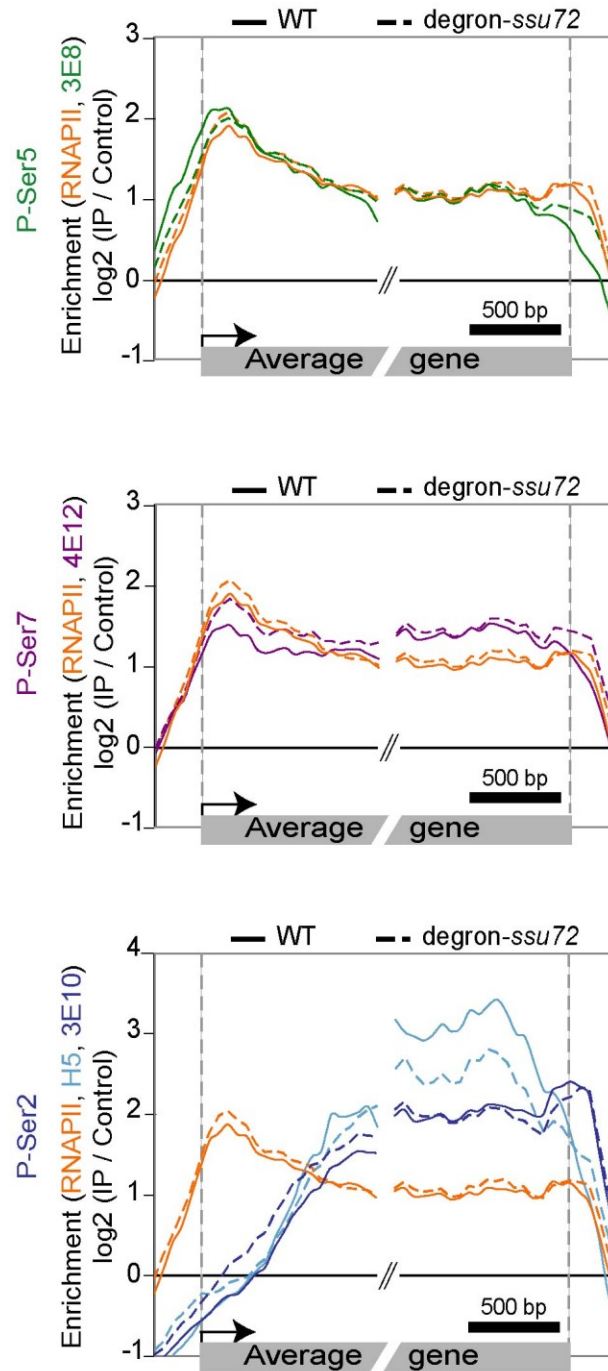


Figure Sup. 4.3: The data from Figure 4.3 is shown for the total length of the metagene.

A companion to **Figure 4.3**

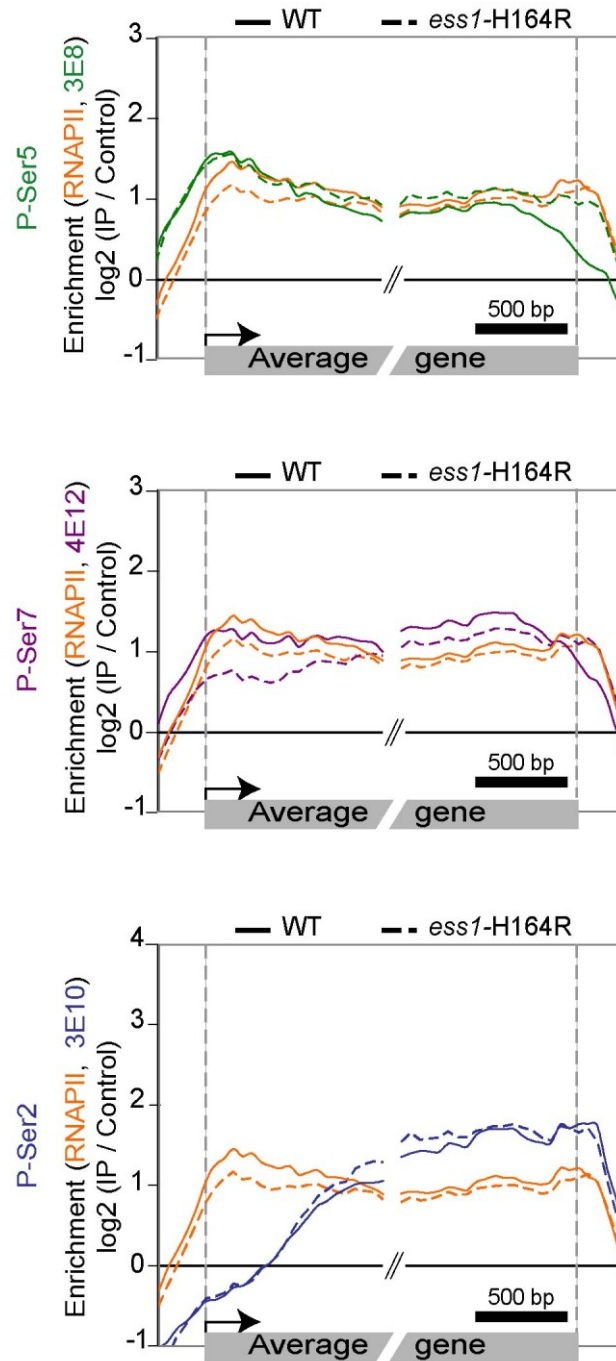


Figure Sup. 4.4: The data from Figure 4.4 is shown for the total length of the metagene.

A companion to **Figure 4.4**.

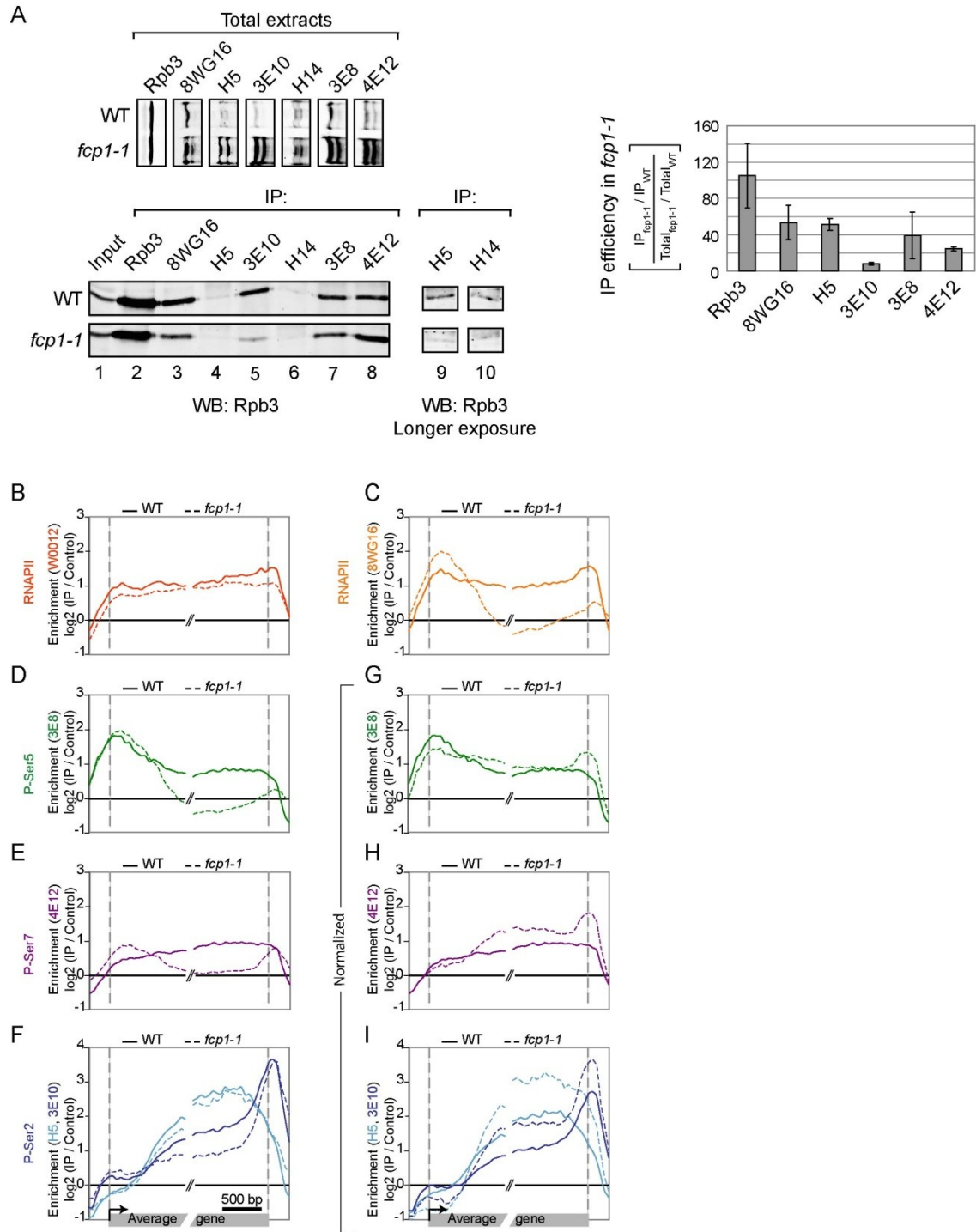


Figure Sup. 4.5: RNAPII CTD is masked in the *fcp1-1* strain.

A companion to **Figure 4.5**. **A)** The top panel shows anti-Rpb3 and anti-CTD (8WG16, H5, 3E10, H14, 3E8 and 4E12) levels detected by western blot in total extracts prepared from wild type and *fcp1-1* cells. The bottom panel shows

immunoprecipitation experiments performed using Rpb3, 8WG16 and phospho-CTD antibodies from the same total extracts used in the top panel. Immunoprecipitations were probed by western blot using anti-Rpb3. The blots were overexposed to reveal the H5 and H14 signals, which are much weaker than the other antibodies. The top panel shows that the signal is stronger for all anti-CTD antibodies in *fcp1-1* cells compared to wild type, while the bottom panel shows that immunoprecipitation of RNAPII using anti-CTD antibodies is equal (H14, 3E8 and 4E12) or less (8WG16, H5 and 3E10) efficient in *fcp1-1* cells compared to wild type. When considering the excess signal in the total extracts (top panel), however, all the anti-phospho-CTD antibodies show reduced efficiency in immunoprecipitating RNAPII from the *fcp1-1* extract (see quantification on the right). These experiments were repeated twice and the quantification of the immunoprecipitation efficiency in *fcp1-1* cells (relative to its wild type) is shown as a histogram. Note that the signals obtained with H14 IPs were too low to generate reliable quantification; thus it has been omitted from the graph.

B-I) Profile of Rpb3 using W0012 (B), RNAPII CTD using 8WG16 (C), P-Ser5 (D, G), P-Ser7 (E, H) and P-Ser2 (F, I), along a metagene in wild type (solid traces) and *fcp1-1* (dashed traces) cells. Panels B-F show raw data while panels G-I show data that have been corrected to remove the effect of CTD masking in the *fcp1-1* strain as in **Figure 4.5 B**. The genes contributing to the metagene have been selected based on anti-Rpb3 occupancy in each strain.

Table Sup. 4.1 : Subunit composition of the TAP-purified complexes shown in **Figure Sup. 4.2 A** based on mass spectrometry data.

Bait	Gene	ORF	Description	MS Score
Kin28	<i>CCL1</i>	YPR025C	Cyclin associated with protein kinase Kin28p	2372
	<i>TFB3</i>	YDR460W	Subunit of TFIIH and nucleotide excision repair factor 3 complexes	2302
	<i>RAD3</i>	YER171W	5' to 3' DNA helicase, involved in nucleotide excision repair and transcription	1845
	<i>KIN28</i>	YDL108W	Serine/threonine protein kinase, subunit of the transcription factor TFIIH	1501
	<i>SSL2</i>	YIL143C	Component of the holoenzyme form of RNA polymerase transcription factor TFIIH	1070
	<i>TFB2</i>	YPL122C	Subunit of TFIIH and nucleotide excision repair factor 3 complexes	1058
	<i>TFB1</i>	YDR311W	Subunit of TFIIH and nucleotide excision repair factor 3 complexes	971
	<i>SSL1</i>	YLR005W	Component of the core form of RNA polymerase transcription factor TFIIH	680
	<i>PET9</i>	YBL030C	Major ADP/ATP carrier of the mitochondrial inner membrane	607
	<i>TFB4</i>	YPR056W	Subunit of TFIIH complex, involved in transcription initiation	413
Ctk1	<i>CTK1</i>	YKL139W	Catalytic (alpha) subunit of C-terminal domain kinase I (CTDK-I)	2787
	<i>CTK2</i>	YJL006C	Beta subunit of C-terminal domain kinase I (CTDK-I)	2163
	<i>CTK3</i>	YML112W	Gamma subunit of C-terminal domain kinase I (CTDK-I)	1513
	<i>GBP2</i> ‡	YCL011C	Poly(A+) RNA-binding protein, involved in the export of mRNAs from the nucleus to the cytoplasm	503
Bur1	<i>SGV1</i>	YPR161C	Cyclin (Bur2p)-dependent protein kinase that functions in transcriptional regulation	4081
	<i>BUR2</i>	YLR226W	Cyclin for the Sgv1p (Bur1p) protein kinase	1977
	<i>MGM101</i> #	YJR144W	Protein involved in mitochondrial genome maintenance	590
Fcp1	<i>FCP1</i>	YMR277W	Carboxy-terminal domain (CTD) phosphatase	3157

Proteins identified in untagged control purifications were removed. Only proteins with a mass spectrometry (MS) score higher than 400 were retained.

‡ As previously described by [57-59].

As previously described by [59].

4.3.7 Supplemental experimental procedures

Yeast growth conditions

Strains were grown to an OD₆₀₀ of 0.6-0.8 (unless otherwise specified) in yeast extract-peptone-dextrose (YPD) at 30°C following standard procedures, except for the *degron-ssu72* strain and its wild type that were transferred to 37°C for 1 hour prior to crosslinking. ATP analog-sensitive *Kin28-as* strains were precultured in yeast nitrogen base (YNB) –URA medium before inoculation into YPD. *Kin28-as* strains and their controls were treated with 6 μM of NAPP1 for 15 min prior to crosslinking. ATP analog sensitive *Bur1-as* strain and its wild type were treated with 6 μM 3MBPP1 for 45 min prior to harvesting.

Construction of the Rpb1 CTD-S7A strain

The Rpb1 CTD-S7A strain was constructed using the break-mediated Delitto perfetto method [60]. Briefly, the delitto CORE cassette was integrated immediately after the 25th repeat of the yeast CTD. Concomitant insertion of a stop codon at this position ensured the viability of the strain during the selection process. Recombination of the CTD and loss of the cassette sequence was obtained by transformation with a PCR product amplifying a minigene encoding a CTD in which all serines at position 7 have been replaced by alanines. The sequence of the minigene, synthesized by IDT (Integrated DNA Technologies), was optimized in order to allow synthesis. The strain was sequenced to verify for the correct replacement and the loss of any unwanted sequence. All sequences are available upon request.

Antibodies

The anti-myc (9E10), anti-Spt5 and anti-P-Spt5 antibodies were kind gifts from Alain Verreault, Grant Hartzog and Steven Hahn, respectively. The 3E8, 3E10 and 4E12 antibodies were purchased from Helmholtz Zentrum in Munich. The H5 (P-Ser2) and H14 (P-Ser5) were from Covance. The anti-Rpb3 (W0012) was purchased from Neoclone Biotechnology. The H3K4me3 (Ab8580), the anti-Actin (Ab8224) and the 8WG16 (Ab817) antibodies were from Abcam. The anti-TAP (directed against the

calmodulin binding peptide region of the TAP tag; 07-482) was from Millipore. The anti-GST antibody was from Cell Signaling (2622).

ChIP-chip assay

The ChIP-chip assays were performed as previously described [16]. For tagged proteins, the ChIP DNA was hybridized in competition with a control ChIP DNA prepared from an isogenic untagged strain, whereas in other cases the ChIP DNA was hybridized in competition with input DNA (non-immunoprecipitated DNA). The following amounts of antibody per immunoprecipitation were used: 3E8 (5 μ L), 3E10 (5 μ L), 4E12 (5 μ L), 8WG16 (2 μ g), H14 (15 μ g), H5 (15 μ g), W0012 (3 μ L) and 9E10 (5 μ g). The antibodies were coupled to Dynabeads (Invitrogen) coated with protein G (3E8, 3E10, 4E12), Pan mouse IgG antibodies (8WG16, H14, W0012 and 9E10) or Sheep anti-mouse IgG (H5). The microarrays were purchased from Agilent Technologies and contain a total of 44,290 Tm-adjusted 60-mer probes covering the entire genome for an average density of one probe every 275 base pairs (\pm 100 base pairs) within the probed regions (catalog # G4493A). All ChIP-chip experiments were done at least in duplicate. Each microarray was normalized using the Lima Loess method and replicates were combined using a weighted average method as previously described [56]. For the localization of Pti1 and Ssu72, the \log_2 ratio of each spot of combined datasets was converted to Z-score.

Metagene analyses

Protein-coding genes (SGD version Feb. 02 2008) were cut in the middle and the first and second halves were aligned on their 5' and 3' boundaries. Transcription unit boundaries were defined as previously reported [61]. As in Rufiange et al. [39], the data were averaged within 50 base pair-windows and a sliding window of 300 base pairs was applied to the ratios. The list of genes contributing to the average profile of a metagene was selected for each strain as follows. First, the 1,500 genes with the highest 8WG16 or Rpd3 average occupancy (as specified) were unbiasedly identified. Second, genes adjacent to intergenic regions smaller than 100 base pairs (defined using all SGD annotations) and genes without two well-defined transcription

boundaries were excluded from the profiles. Average RNAPII occupancy (but not metagene profiling) was calculated using smoothed data obtained as per Guillemette et al. [62]. Last, long (between 2,500 and 1,500 base pairs) genes were selected. Note that **Figure 4.1 B** also shows profiles for “medium” (1,250-750 base pairs) and “small” (600-400 base pairs) genes. Normalization for RNAPII levels (**Figure 4.2 Error! Reference source not found.A, B and C**) or CTD masking (**Figure 4.5 B**) as done as follow. First, at each position i of the metagene profile, a value x_i was computed by subtracting the log₂ value for RNAPII occupancy level (as measured with the Rpb3 (**Figure 4.2**) or 8WG16 (**Figure 4.5**) antibodies) in the mutant from that of the wild type. Second, at each corresponding position i , x_i was added to the phospho-isoform log₂ occupancy value of the mutant. In order to avoid any discrepancy in the comparison of our data with the corresponding data generated by the other laboratories, we remapped all data on the same assembly (sacCer1).

Purification of TAP-tagged protein complexes

TAP-tagged Kin28, Ctk1, Bur1 and Fcp1 protein complexes were purified by one-step affinity purification from extracts of yeast cells (1 L) grown in YPD medium to an OD₆₀₀ of ~ 1.0. Harvesting and cryogenic disruption of cells were conducted as described previously [63]. Frozen cell grindate (5 g) was rapidly thawed into nine volumes of extraction buffer (20 mM Tris-HCl [pH7.5], 500 mM KOAc, 1 mM EDTA [pH8], 0.5 % Triton X-100, 10 % glycerol, Antifoam B (10 µl/50 ml), 1 mM DTT and protease inhibitor mixture) and vortexed for 30 seconds. The resulting extract was clarified by centrifugation at 3,500 rpm for 10 min at 4°C. The tagged complexes were then isolated using 200 µL of pre-washed magnetic Dynabeads (Invitrogen) cross-linked to rabbit IgG (Sigma) as previously described [63]. After binding for 1 hour at 4°C, Dynabeads were collected and washed five times with extraction buffer and three times with TEV protease cleavage buffer (10 mM Tris-HCl [pH8], 500 mM KOAc, 0.5 mM EDTA [pH8], 0.1 % Triton X-100, 10 % glycerol and 1 mM DTT). The isolated protein complex was eluted by incubating the beads overnight at 4°C with 300 U of TEV protease (AcTEV from Invitrogen) in 1 mL of TEV cleavage buffer. After TEV digestion, the eluate was collected,

dialyzed and concentrated in dialysis buffer (10 mM Hepes-KOH [pH7.9], 0.1 mM EDTA [pH8], 100 mM KOAc, 20 % glycerol, 20 % PEG and 1 mM DTT). The concentrated protein complexes were aliquoted and stored at -80°C. Purified proteins were separated by SDS-PAGE on a 4-12 % NuPAGE Novex Bis-tris precast gel (Invitrogen) and visualized by silver staining.

Protein identification by mass spectrometry (LC-MS/MS)

A standard TCA protein precipitation was first performed to remove detergents from the protein samples. Protein extracts were then re-solubilized in 6 M urea buffer followed by reduction and alkylation prior to digestion with trypsin (Promega) at 37°C for 18 hours. The digested peptide mixtures were dried down in vacuum centrifuge and stored at -20°C until LC-MS/MS analysis. Prior to LC-MS/MS, the digested peptide mixtures were re-solubilized in 0.2 % formic acid. Desalting/cleanup of the digested peptides was performed by using C₁₈ ZipTip pipette tips (Millipore). Eluates were dried down in vacuum centrifuge and then re-solubilized in 2 % ACN / 1 % formic acid. The LC column used was a C18 reverse phase column packed with a high-pressure packing cell. A 75 µm i.d. Self-Pack PicoFrit fused silica capillary column 15 cm long (New Objective) was packed with the C18 Jupiter 5 µm 300 Å reverse-phase material (Phenomenex). This column was installed on the Easy-nLC II system (Proxeon Biosystems) and coupled to the LTQ Orbitrap Velos (ThermoFisher Scientific) equipped with a Proxeon nanoelectrospray ion source. Protein database searching was performed with Mascot 2.1 (Matrix Science) against the *S. cerevisiae* NCBI nr protein database (2010-12-14 release). The mass tolerances for precursor and fragment ions were set to 10 ppm and 0.6 Da, respectively. Trypsin was used as the enzyme allowing for up to two missed cleavages. Carbamidomethyl and oxidation of methionine were allowed as variable modifications.

Purification of recombinant proteins

The GST-CTD fusion protein, used in kinase and phosphatase assays, contains the yeast Rpb1 sequence from amino acid 1462 (Ile) to 1733 (Arg), corresponding to the linker (80 amino acids) and complete CTD sequence inserted in the bacterial protein

expression vector pGEX-4T1, generating the plasmid pFR179. GST-Ssu72 (pM712) and catalytic mutant GST-Ssu72 C15S (pM698) plasmids were a gift from Mike Hampsey and previously described in [64]. Purification of recombinant GST-CTD protein from *E.coli* BL21 strain was performed as follows. Cells were grown in 1L of LB plus ampicillin until it reached an OD₆₀₀ of 0.5-0.8. Protein expression was then induced with 1 mM IPTG for 4 hours at room temperature. Cells were pelleted and washed with PBS before flash freezing in liquid nitrogen. Cells were then lysed in ~4 mL/g of lysis buffer (50 mM Tris [pH7.6], 1 M NaCl, 5 mM EDTA, 10 % glycerol, 5 mM DTT and protease inhibitor mixture) using a French press. The cell extract was centrifuged at 35,000 rpm for 45 min at 4°C and the supernatant diluted with an equal volume of no-salt buffer (50 mM Tris [pH7.6], 5 mM EDTA, 10 % glycerol, 5 mM DTT). The GST-tagged protein was isolated using ~2 mL of sepharose Glutathione 4B fast flow beads (GE Healthcare). After binding for 3 hours at 4°C with agitation, beads were collected by centrifugation and washed twice with high salt buffer (50 mM Tris [pH8], 0.5 M NaCl, 5 mM EDTA, 10 % glycerol, 5 mM DTT) and twice with low salt buffer (50 mM Tris [pH8], 0.2 M NaCl, 5 mM EDTA, 10 % glycerol, 5 mM DTT). Proteins were eluted by incubating the beads overnight at 4°C with one volume of glutathione elution buffer (25 mM Tris [pH8], 0.2 M NaCl, 1 mM EDTA, 10 % glycerol, 20 mM reduced glutathione [pH8], 5 mM DTT). Two additional volumes of glutathione elution buffer were added to the beads and immediately collected and pooled with the first. The eluate was concentrated to ~1 mL by dialysis in PEG dialysis buffer (25 mM Tris [pH8], 0.2 M NaCl, 1 mM EDTA, 10 % glycerol, 20 % PEG8000, 0.5 mM DTT) and then dialyzed in No-PEG reaction buffer (10 mM Hepes-KOH [pH7.9], 0.1 M KOAc, 0.1 mM EDTA, 10 % glycerol, 1 mM DTT). The concentrated protein was aliquoted and stored at -80°C. The same protocol was used for GST-Ssu72 and GST-Ssu72 C15S mutant protein, except that bacterial cells were grown at 19°C overnight, and the No-PEG reaction buffer was replaced by the No-PEG dialysis buffer (25 mM Tris [pH8], 0.2 M NaCl, 1 mM EDTA, 10 % glycerol, 0.5 mM DTT).

In vitro CTD kinase assay

TAP-purified Kin28, Ctk1 or Bur1 complexes were used to phosphorylate recombinant GST-CTD. For kinase reaction mixtures (50 μ L), ~900 ng of GST-CTD was incubated with the indicated kinase for 30 min at 30°C in kinase buffer (10 mM Hepes-KOH [pH7.9], 100 mM KOAc, 15 mM MgAc, 10 % glycerol, 1 mM ATP, 1 mM DTT, PhosSTOP phosphatase inhibitor cocktail [Roche] and protease inhibitor mixture). The reactions were stopped by adding SDS-loading buffer and products were resolved on SDS-PAGE gels. Proteins were then transferred to nitrocellulose membranes for western blot analysis as described.

Yeast total cell extracts

For western blot analysis of total cell extracts, yeast extracts were prepared by using a rapid protein extraction procedure as previously described [65]. This method is required to preserve the phosphate group on the Spt5 protein.

Immunoprecipitations followed by western blots

Total extracts from wild type (yFR1003) and *fcp1-1* (yFR1004) cells were prepared as per our TAP-tagged protein purifications except that the frozen cell grindate was resuspended in EB150 (20 mM Tris-HCl [pH7.5], 150 mM KoAc, 1 mM EDTA [pH8], 0.5 % Triton X-100, 10 % glycerol, Antifoam B (10 μ l/50 ml) and 1 mM DTT, supplemented with protease and phosphatase inhibitors) and homogenized for 30 seconds with a Polytron (PT 1200, Kinematic AG) to allow for maximal recovery of chromatin proteins. Five milligrams of wild type and *fcp1-1* mutant extracts were immunoprecipitated overnight at 4°C in EB150 with agitation using various anti-RNAPII antibodies coupled to magnetic beads as per our ChIP protocol. Beads were washed three times with EB150, boiled in SDS-loading buffer and loaded on 8 % SDS-PAGE gel. As the input, 50 μ g of total extract from each strain was also loaded on the gel. Proteins were transferred to nitrocellulose membranes and RNAPII pulldowns were monitored by probing with the anti-Rpb3 antibody (W0012) (**Figure Sup 4.5 A**, lower panel). For each strain, 100 μ g of total extract was also run on SDS-PAGE, transferred to nitrocellulose and probed using all the anti-CTD antibodies and

the anti-Rpb3 antibody as a loading control (**Figure Sup 4.5 A**, upper panel). IP efficiency in *fcp1-1* cells (relative to wild type) is defined for each antibody “x” as:

$$\left[\frac{(\text{IP } x_{fcp1-1} / \text{IP } x_{WT})}{(\text{Input}_{fcp1-1} / \text{Input}_{WT})} \right] / \left[\frac{\text{Total } x_{fcp1-1}}{\text{Total } x_{WT}} \right] \times 100$$

Where:

- IP x_{fcp1-1} is the intensity of the band corresponding to the IP with the antibody “x” in the *fcp1-1* extract.
- IP x_{WT} is the intensity of the band corresponding to the IP with the antibody “x” in the wild type extract.
- Input $_{fcp1-1}$ is the intensity of the band corresponding to the input for the *fcp1-1* extract.
- Input $_{WT}$ is the intensity of the band corresponding to the input for the wild type extract.
- Total x_{fcp1-1} is the intensity of the band obtained when the *fcp1-1* extract was probed with the antibody “x”.
- Total x_{WT} is the intensity of the band obtained when the wild type extract was probed with the antibody “x”.

Table Sup. 4.2: List of strains used in this study.

Strain number	Alias	Description	Background	Source	Genotype
yFR116	Z1256	W303 wild type	W303	[66]	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3
yFR117	N/A	S288C wild type	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0
yFR551	N/A	ctk1Δ	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, ctk1Δ::KanMX4
yFR682	N/A	bur2Δ	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, bur2Δ::KanMX4
yFR999	N/A	srb10Δ	W303	This work	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, srb10 Δ::URA3
yFR842	SHY474	bur1-L149G-as	S288C	[65]	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, BUR1-L149G
yFR763	F1546 (SHY508A)	kin28-L83G-as	S288C	[67]	MATalpha, ade::higG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ0, trp1Δ63, ura3Δ0, kin28::kin28-L83G, [pSH579, ARS CEN URA3 kin28-L83G]
yFR762	F1545 (SHY278, BY4705)	wild type for yFR763	S288C	[67]	MATalpha, ade::higG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ0, trp1Δ63, ura3Δ0
yFR958	YDA502	ess1-H164R	S288C	[68]	MATalpha, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, ESS1::ess1(H164R)+NATMX
yFR957	YDA500	wild type for yFR958	S288C	[68]	MATalpha, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, ESS1::ESS1+NATMX
yFR912	N/A	kin28-L83G-as, bur2Δ	S288C	This work	MATalpha, ade::higG, his3-Δ200, leu2-Δ0, lys2-Δ0, met15-Δ0, trp1-Δ63, ura3-Δ0, kin28::kin28-L83G, bur2Δ::Leu2, [pSH579, ARS CEN URA3 kin28-L83G]
yFR552	XH-24	degron-ssu72	FY23	[69]	MATa, ura3-52, trp1Δ63, leu2Δ1, SSU72-td (Ub-DHFRts-HA-SSU72-URA3)
yFR778	FY23	wild type for yFR552	FY23	[69]	MATa, ura3-52, trp1Δ63, leu2Δ1
yFR1004	YF236 (YMK220)	fcp1-1	W303A	[70]	MATalpha, trp1-1, ura3-1, his3-11,15, leu2-3,112, ade2-1, can1-100, fcp1Δ::LEU2, [p(fcp1-1 ADE2 C/A)]
yFR1003	YF235 (YMK219)	wild type for yFR1004	W303A	[70]	MATalpha, trp1-1, ura3-1, his3-11,15, leu2-3,112, ade2-1, can1-100, fcp1 Δ::LEU2, [p(FCP1 ADE2 C/A)]
yFR787	N/A	S7A	W303	This work	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1::rpb1(CTD-S7A)
yFR487	NA	Rpb1-9myc	W303	This work	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1::RPB1-9myc TRP1
yFR943	N/A	3myc-Ssu72	W303	This work	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, SSU72::3myc-SSU72
yFR945	N/A	Pti1-3myc	W303	This work	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, PTI1::PTI1-3myc-URA3
yFR687	N/A	Kin28-TAP	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, KIN28::KIN28-TAP-HIS3MX6
yFR685	N/A	Ctk1-TAP	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, CTK1::CTK1-TAP-HIS3MX6
yFR686	N/A	Bur1-TAP	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, BUR1::BUR1-TAP-HIS3MX6
yFR758	N/A	Fcp1-TAP	S288C	Open Biosystems	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, FCP1::FCP1-TAP-HIS3MX6

4.3.8 Acknowledgments

We are grateful to members of our laboratory for helpful discussions. Special thanks to Johanne Duhaime for bioinformatics support and Karen Wei for technical assistance in cryogenic grinding of yeast cells. We thank Stephen Buratowski, Alan G. Hinnebusch, Michael Hampsey, Steven Hahn, Alain Verreault and Grant Hartzog for sharing strains and reagents. We also thank David L. Bentley, Patrick Cramer and Aseem Z. Ansari for providing access to datasets. This work was funded by a CIHR grant to FR (MOP-82891). CJ holds a CIHR fellowship.

4.4 References

1. Young, R.A., *RNA polymerase II*. Annu. Rev. Biochem., 1991. **60**: p. 689-715.
2. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD*. Trends Genet., 2008. **24**: p. 289-96.
3. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD*. Genes Dev., 2006. **20**: p. 2922-36.
4. Perales, R. and D. Bentley, *"Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions*. Mol. Cell, 2009. **36**: p. 178-91.
5. Smith, E. and A. Shilatifard, *The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes*. Mol. Cell, 2010. **40**: p. 689-701.
6. Palancade, B. and O. Bensaude, *Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation*. Eur. J. Biochem., 2003. **270**: p. 3859-70.
7. Corden, J.L., *Transcription. Seven ups the code*. Science, 2007. **318**: p. 1735-6.
8. Wu, X., et al., *The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery*. EMBO J., 2000. **19**(14): p. 3727-38.
9. Sims, R.J., et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation*. Science, 2011. **332**: p. 99-103.
10. Buratowski, S., *The CTD code*. Nat. Struct. Biol., 2003. **10**: p. 679-80.
11. Lu, H., et al., *The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex*. Proc. Natl. Acad. Sci. U. S. A., 1991. **88**(22): p. 10004-8.

12. Max, T., M. Sogaard, and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator*. J. Biol. Chem., 2007. **282**: p. 14113-20.
13. Kim, T. and S. Buratowski, *Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions*. Cell, 2009. **137**: p. 259-72.
14. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity*. Mol. Cell, 2003. **11**: p. 709-19.
15. Govind, C.K., et al., *Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes*. Mol. Cell, 2010. **39**: p. 234-46.
16. Drouin, S., et al., *DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes*. PLoS Genet., 2010. **6**: p. e1001173.
17. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1*. Genes Dev., 2010. **24**: p. 2303-16.
18. Komarnitsky, P., E.-J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription*. Genes Dev., 2000. **14**: p. 2452-60.
19. Keogh, M.-C., V. Podolny, and S. Buratowski, *Bur1 kinase is required for efficient transcription elongation by RNA polymerase II*. Mol. Cell. Biol., 2003. **23**: p. 7005-18.
20. Carrozza, M.J., et al., *Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription*. Cell, 2005. **123**: p. 581-92.
21. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1195-1201.

22. Ahn, S.-H., M. Kim, and S. Buratowski, *Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing*. Mol. Cell, 2004. **13**: p. 67-76.
23. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II*. Genes Dev., 1997. **11**: p. 3306-18.
24. Cho, E.-J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain*. Genes Dev., 2001. **15**: p. 3319-29.
25. Krishnamurthy, S., et al., *Ssu72 Is an RNA polymerase II CTD phosphatase*. Mol. Cell, 2004. **14**: p. 387-94.
26. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7*. Science, 2007. **318**: p. 1780-2.
27. Akhtar, M.S., et al., *TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II*. Mol. Cell, 2009. **34**: p. 387-93.
28. Egloff, S., et al., *Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression*. Science, 2007. **318**: p. 1777-9.
29. Mosley, A.L., et al., *Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation*. Mol. Cell, 2009. **34**: p. 168-78.
30. Dichtl, B., et al., *A role for SSU72 in balancing RNA polymerase II transcription elongation and termination*. Mol. Cell, 2002. **10**: p. 1139-50.
31. Steinmetz, E.J. and D.A. Brow, *Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription*. Mol. Cell. Biol., 2003. **23**: p. 6339-49.
32. Krishnamurthy, S., et al., *Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries*. Mol. Cell. Biol., 2009. **29**: p. 2925-34.
33. Singh, N., et al., *The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway*. Mol. Cell, 2009. **36**: p. 255-66.

34. Buratowski, S., *Progression through the RNA polymerase II CTD cycle*. Mol. Cell, 2009. **36**: p. 541-6.
35. Cho, H., et al., *A protein phosphatase functions to recycle RNA polymerase II*. Genes Dev., 1999. **13**: p. 1540-52.
36. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1154-61.
37. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1279-1286.
38. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1272-1278.
39. Rufiange, A., et al., *Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1*. Mol. Cell, 2007. **27**: p. 393-405.
40. Jones, J.C., et al., *C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats*. J. Biol. Chem., 2004. **279**: p. 24957-64.
41. Boehm, A.K., et al., *Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock*. Mol. Cell. Biol., 2003. **23**(21): p. 7628-37.
42. Kim, M., et al., *Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7*. J. Biol. Chem., 2009. **284**: p. 26421-6.
43. Holstege, F.C.P., et al., *Dissecting the regulatory circuitry of a eukaryotic genome*. Cell, 1998. **95**: p. 717-28.
44. Thiebaut, M., et al., *Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae*. Mol. Cell, 2008. **31**(5): p. 671-82.
45. Qiu, H., C. Hu, and A.G. Hinnebusch, *Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters*. Mol. Cell, 2009. **33**: p. 752-62.

46. Donner, A.J., et al., *CDK8 is a positive regulator of transcriptional elongation within the serum response network*. Nat. Struct. Mol. Biol., 2010. **17**: p. 194-201.
47. Lenasi, T. and M. Barboric, *P-TEFb stimulates transcription elongation and pre-mRNA splicing through multilateral mechanisms*. RNA Biol., 2010. **7**(2): p. 145-50.
48. Gemmill, T.R., X. Wu, and S.D. Hanes, *Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in Saccharomyces cerevisiae*. J. Biol. Chem., 2005. **280**: p. 15510-7.
49. Werner-Allen, J.W., et al., *Cis proline-mediated pSer5-dephosphorylation by the RNA polymerase II CTD phosphatase Ssu72*. J. Biol. Chem., 2010.
50. Kim, M., et al., *Distinct pathways for snoRNA and mRNA termination*. Mol. Cell, 2006. **24**: p. 723-34.
51. Lykke-Andersen, S. and T.H. Jensen, *Overlapping pathways dictate termination of RNA polymerase II transcription*. Biochimie, 2007. **89**: p. 1177-82.
52. Kanin, E.I., et al., *Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis*. Proc. Natl. Acad. Sci. U. S. A., 2007. **104**: p. 5812-7.
53. Hong, S.W., et al., *Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome*. Proc. Natl. Acad. Sci. U. S. A., 2009. **106**: p. 14276-80.
54. West, M. and J.L. Corden, *Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations*. Genetics, 1995. **140**: p. 1233.
55. Fong, N. and D.L. Bentley, *Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD*. Genes Dev., 2001. **15**: p. 1783-95.
56. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast*. Cell, 2005. **122**(4): p. 517-27.

57. Ho, Y., et al., *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry*. Nature, 2002. **415**: p. 180-3.
58. Hurt, E., et al., *Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex*. Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(7): p. 1858-62.
59. Breitschütz, A., et al., *A global protein kinase and phosphatase interaction network in yeast*. Science, 2010. **328**(5981): p. 1043-6.
60. Storici, F. and M.A. Resnick, *The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast*. Methods Enzymol., 2006. **409**: p. 329-45.
61. Nagalakshmi, U., et al., *The transcriptional landscape of the yeast genome defined by RNA sequencing*. Science, 2008. **320**: p. 1344-1349.
62. Guillemette, B., et al., *Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning*. PLoS Biol., 2005. **3**: p. e384.
63. Oeffinger, M., et al., *Comprehensive analysis of diverse ribonucleoprotein complexes*. Nat. Methods, 2007. **4**: p. 951-6.
64. Sun, Z.W. and M. Hampsey, *Synthetic enhancement of a TFIIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo*. Mol. Cell. Biol., 1996. **16**: p. 1557-66.
65. Liu, Y., et al., *Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex*. Mol. Cell. Biol., 2009. **29**: p. 4852-63.
66. Ren, B., et al., *Genome-wide location and function of DNA binding proteins*. Science, 2000. **290**: p. 2306-9.
67. Liu, Y., et al., *Two Cyclin-Dependent Kinases Promote RNA Polymerase II Transcription and Formation of the Scaffold Complex*. Mol. Cell. Biol., 2004. **24**: p. 1721-1735.

68. Hanes, S.D., P.R. Shank, and K.A. Bostian, *Sequence and mutational analysis of ESS1, a gene essential for growth in Saccharomyces cerevisiae*. *Yeast*, 1989. **5**: p. 55-72.
69. He, X., et al., *Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1*. *Genes Dev.*, 2003. **17**: p. 1030-42.
70. Kobor, M.S., et al., *An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae*. *Mol. Cell*, 1999. **4**: p. 55-62.

5 Discussion

5.1 H2A.Z

Avec le développement de la chromatine, les eucaryotes ont trouvé le moyen de contrôler le niveau de compaction de l'ADN dans le noyau. Il en résulte une structure extrêmement stable et cependant incroyablement dynamique grâce au développement synchronique de complexes capables de remodeler, modifier ou altérer la composition des nucléosomes. De plus, la chromatine fournit aux cellules le moyen « d'annoter » leur génome, de manière stable ou transitoire, pour de nombreux phénomènes cellulaires. Les variantes d'histones constituent une pièce importante de l'arsenal d'altération de la chromatine. Grâce à sa facilité de manipulation génétique, à la compaction et à la qualité des annotations de son génome, la levure est un organisme idéal pour l'étude des variantes d'histones, malgré une plus grande complexité de la chromatine chez les organismes pluricellulaires. Les études chez la levure ont souvent servi de base aux études chez les eucaryotes supérieurs; cela a d'ailleurs été le cas pour H2A.Z. Chez la levure, seules existent les variantes H2A.Z et CSE4 (CENP-A chez l'humain). Il est cependant intéressant de noter que dans cet organisme la seule histone H3 (exprimée en phase S) est en séquence une histone H3.3 et que la fonction de la variante H2A.X, présente chez les eucaryotes supérieurs, est assumée par l'histone canonique H2A de la levure, grâce à la présence dans la queue C-terminale d'un motif SQ dont la phosphorylation de la sérine est requise pour sa fonction [133].

5.1.1 Localisation de H2A.Z

En 2005, de nombreux groupes s'intéressent aux composants de la chromatine et à leurs modifications ; de plus, les outils de localisation génomique se sont rapidement popularisés et leur développement permet désormais à certains laboratoires, d'obtenir des données avec une résolution de l'ordre du nucléosome [45, 131, 403-407]. H2A.Z, dont l'ancienneté et l'implication dans de multiples phénomènes nucléaires ont favorisé et stimulé l'étude, est certainement l'une des

variantes les plus étudiées à ce jour. Il est donc logique qu'elle ait été la première à être localisée grâce à ces outils. Durant les derniers mois de 2005, quatre articles publieront le profil de localisation génomique de l'histone H2A.Z chez la levure, à plus ou moins grande résolution [50, 157-159]. Ces articles, dont celui présenté au chapitre 3 de cette thèse, aboutissent aux mêmes conclusions générales, mais des différences les distinguent et les opposent. Ainsi, l'anticorrélation à la transcription, observée par la majorité des groupes échappe pourtant à l'équipe de Madhani, sans doute à cause de l'échantillonnage trop restreint des gènes, puisque cette étude, qui se démarque par ailleurs par son degré de résolution supérieur, ne porte que sur une portion du génome, comprenant principalement le chromosome III [50]. C'est aussi le cas de l'emplacement exact et du nombre de nucléosomes occupés par H2A.Z : deux nucléosomes, de part et d'autre du NFR [50] ou un à deux nucléosomes, en aval du NFR [157]. Si les deux conformations existent dans le génome de la levure, la première semble être la règle. La raison de cette divergence n'est pas évidente, mais découle possiblement de la résolution moindre de nos puces. Les deux autres articles ayant une résolution très inférieure n'ont, à l'époque, pas permis de trancher [158, 159].

La localisation génomique de H2A.Z aux promoteurs des gènes a relancé l'intérêt des groupes de recherche pour cette variante d'H2A et de nombreuses études feront suite aux premiers travaux, afin de mieux cerner les fonctions de H2A.Z. Logiquement, la localisation de H2A.Z chez les mammifères a suivi la localisation chez la levure, notamment par Sara Hardy et Pierre-Étienne Jacques dans notre laboratoire [108, 189]. Au cours de l'évolution, la localisation aux promoteurs des gènes a été conservée ; par contre, différence majeure, la présence de H2A.Z est cette fois corrélée à la présence de la polymérase. L'histone H2A.Z est, en effet, recrutée au promoteur « sur demande » lors de l'activation des gènes, où elle augmente le recrutement de la polymérase [108]. Une des observations les plus significatives du travail de Sara Hardy et Pierre-Étienne Jacques est l'observation d'une déplétion de H2A.Z dans les régions codantes [108]. Cette étude suggère que le profil de localisation de H2A.Z est le résultat de trois mécanismes : deux de déposition, un ciblé au promoteur et un plus global aléatoire, et d'un mécanisme d'exclusion dans

les régions codantes. Si le modèle de déposition classique aux promoteurs implique le complexe SWR, nous ne savons actuellement rien du mécanisme qui pourrait permettre l'exclusion de H2A.Z dans la partie codante des gènes. Il apparaît donc particulièrement intéressant de pousser plus loin les travaux afin surtout de mettre en évidence un mécanisme d'exclusion de H2A.Z au niveau de la transcription. Des études sont en cours dans notre laboratoire dans le but de déterminer le rôle des remodeleurs de la chromatine impliqués dans la transcription, pour comprendre le mécanisme de déplétion intragénique de H2A.Z. Ce phénomène étant aussi observable chez *S. cerevisiae* ce projet bénéficie de sa facilité de manipulation [108]. Il est aussi à noter que, chez la levure, dans les domaines HZAD, contrairement au reste du génome, la localisation de H2A.Z s'étend dans la région codante des gènes [157].

Des études, il ressort que H2A.Z assume des localisations et des fonctions plus diverses que la majorité des variantes d'histones, à la fois aux promoteurs des gènes, aux frontières avec l'hétérochromatine (chez la levure), dans l'hétérochromatine (chez les humains) et aux centromères [50, 108, 157, 189].

Les fonctions et la localisation génomique des modifications de H2A.Z restent une voie encore peu exploitée. Nous avons défini dans notre laboratoire que H2A.Z associée à l'hétérochromatine facultative était hypoacétylée [108]. Deux articles ont permis d'associer la monoubiquitination de H2A.Z à l'hétérochromatine facultative du chromosome X-inactif des mammifères, dont H2A.Z est cependant partiellement déplétée [165, 408]. Il est possible que cette modification différencie l'association de H2A.Z à l'euchromatine et à l'hétérochromatine. Un article récent renforce le lien entre la répression des gènes et l'ubiquitination de H2A.Z en démontrant que, chez l'humain, la déubiquitination de H2A.Z est nécessaire à l'activation aux gènes régulés par le récepteur aux androgènes [166].

5.1.2 Implication de H2A.Z dans la mémoire transcriptionnelle

En 2007, le groupe de Brickner implique H2A.Z dans la localisation de Gal1 et INO1 à la membrane nucléaire après induction et propose une fonction dans la mémoire transcriptionnelle [168]. Elle est définie, chez la levure, par une reprise de la transcription plus rapide lors de leur réactivation des gènes suivant une courte période en condition répressive, et confèrerait donc un avantage adaptatif. Cette étude ouvre la voie à de nouvelles hypothèses quant à la fonction de H2A.Z au promoteur des gènes. L'article de Brickner démontre que H2A.Z est nécessaire à une réactivation optimale de *GALI* et *INOI*, en se basant sur des expériences de localisation de chromatine et de quantification d'ARNm, afin de déterminer la position des gènes relatifs à la membrane et leurs niveaux d'induction au cours du temps. [168]. La même année, un second groupe confronte l'hypothèse de Brickner en avançant une origine cytoplasmique à la mémoire de la transcription [409]. En présence de galactose, Gal3 séquestre Gal80, responsable de l'inhibition de l'activateur Gal4. Toutefois, le niveau de transcription relativement faible de Gal3 ne lui permet pas de produire un effet de mémoire à long terme, c'est-à-dire au-delà d'une division cellulaire après laquelle il devient trop dilué. Selon cette équipe, la mémoire transcriptionnelle est due à la présence de Gal1 pouvant agir comme Gal3 pour séquestrer Gal80, bien qu'avec une affinité moindre, grâce à une identité à Gal3 de 70 % (et une similarité de 90 %). L'effet mémoire de Gal1 s'explique par sa stabilité et son induction très importante en présence de galactose, permettant une accumulation suffisante pour procurer un effet sur cinq à sept générations, c'est à dire avant que la dilution ne le ramène en deçà de sa concentration active. Cette hypothèse est renforcée, quelques années plus tard, par la publication de deux articles [410, 411]. Contrairement à l'équipe de Brickner, ces groupes détectent une différence significative du niveau d'induction initial de Gal1 entre les souches *htz1Δ* et sauvage qui se répercute sur le niveau d'accumulation de la protéine Gal1 à l'issue de l'induction initiale. Cette différence dans l'accumulation de la protéine Gal1 durant la première exposition au galactose est, pour eux, la cause de l'effet de mémoire

transcriptionnelle observé par Brickner. Au-delà des différences d'observation dans l'induction initiale de Gal1, difficile à expliquer, ces études soulèvent la question de l'utilisation de Gal1 dans les expériences de mémoire transcriptionnelle. Note intéressante, selon l'équipe de Rine, une partie de la sévérité des phénotypes associés à la délétion de *HTZI* pourrait être la conséquence de l'effet négatif de SWR1 en l'absence de H2A.Z [410]. Ne s'intéressant, cette fois-ci, qu'au gène *INO1*, sans doute dans le but d'éviter toute polémique sur la mémoire transcriptionnelle au gène Gal1, le groupe de Brickner publie, à la même période, un article reliant de nouveau H2A.Z à la mémoire transcriptionnelle [167]. Il différencie alors les mécanismes de ciblage et de rétention à la membrane. Le MRS (*memory recruitment sequence*) identifié dans cet article est un élément de 11 pb responsable de l'incorporation de H2A.Z ainsi que du transfert à la membrane. Contrairement au site Reb1, l'introduction de cet élément entraîne la formation d'un seul nucléosome H2A.Z, sans formation d'un NFR et de façon Nup100 dépendante [167].

Enfin, deux articles récents démontrent une implication du bouclage des gènes dans la mémoire de la transcription, ce qui pose tout de suite la possibilité d'un rôle de H2A.Z dans le bouclage des gènes [412, 413]. L'établissement d'une boucle entre le promoteur et le terminateur des gènes nécessite un premier cycle de transcription [253], mais peut perdurer lors de la répression et pourrait ainsi servir à conserver la mémoire de la transcription [412, 413]. Un mutant *sua7-1* incapable de former des boucles est aussi incapable de mémoire transcriptionnelle [412]. Tan-Wong et ses collègues démontrent l'implication de Mlp1 (*myosine like protein 1*), associée avec le complexe du pore nucléaire, dans le bouclage des gènes et la mémoire de la transcription aux gènes Gal1 et HXK1 (un gène réprimé en glucose, mais activé en présence d'une autre source de carbone). Par contre, cet article, basé sur le temps de réinduction d'*INO1* après le passage en conditions restrictives, remet en cause l'effet de mémoire observé au gène *INO1* par l'équipe de Brickner, et impute l'absence de mémoire à la perte d'interaction avec Mlp1 [412].

Les travaux sur la fonction de H2A.Z dans la mémoire transcriptionnelle restent donc encore très polémiques. Afin d'être pertinents, les travaux sur la mémoire transcriptionnelle au gène GAL1 devront considérer l'accumulation de Gal1

de manière rigoureuse dans leurs essais (en quantifiant la protéine), de manière à dissiper toute ambiguïté quant à l'origine des effets observés. La mémoire transcriptionnelle, la localisation à la membrane nucléaire et le bouclage des gènes semblent intimement liés. Il sera très intéressant de préciser les connexions entre ces phénomènes, et de déterminer la place de H2A.Z. À ce sujet, j'ai obtenu, lors de ma thèse, des résultats préliminaires par l'utilisation de la technique de 3C, suggérant que Htz1 n'est pas requis pour la formation des boucles aux gènes constitutivement actifs, comme c'est le cas de Sua7 (TFIIB) [253]. Il reste à tester, si Htz1 joue un rôle dans le maintien des boucles, à l'induction des gènes, comme Mlp1 [412].

5.1.3 Dissection des domaines de H2A.Z

H2A.Z n'a que 60 % d'homologie avec l'histone canonique H2A. Malgré cela elle peut, comme toutes les variantes d'histones, se substituer à son histone canonique sans causer de changement majeur dans la structure du nucléosome. H2A.Z ne possède pas de domaine qui lui soit spécifique et les zones de divergence entre H2A et H2A.Z sont réparties sur l'ensemble de la protéine, permettant un alignement global de leurs séquences (**Figure 1.5**). De plus, on ne peut clairement établir si H2A.Z possède des modifications qui lui sont réellement propres, comme c'est le cas pour H2A.X ; une modification commune à H2A et H2A.Z pourrait également avoir des fonctions différentes. Malgré ce manque « d'originalité », H2A.Z semble être la variante d'histone accumulant le plus de fonctions, peut-être précisément parce qu'elle n'est hyperspécialisée dans aucune fonction précise. Contrairement à la majorité des variantes d'histones, il ne semble pas facile d'associer une région particulière de H2A.Z à ses fonctions.

Afin de répondre à cette question, deux stratégies ont été tentées pour tester les phénotypes associés à la résistance au stress d'une collection de mutants *htz1*. Durant mon doctorat, j'ai construit une collection de souches chimériques H2A.Z/H2A. L'étendue des régions substituées s'appuie sur trois critères : le degré d'homologie entre H2A.Z et H2A le long de la protéine, la conservation des résidus au cours de l'évolution, et la structure des histones (**Figure 5.1**). Les mutants de cette

collection ont été testés pour six phénotypes de résistance à différents stress par des essais sur plaques. Contrairement à ce que l'on aurait pu attendre d'après les travaux de Larochelle *et al.*, (2003) et Adam *et al.*, (2001) [134, 169] les résultats de cette étude très préliminaire semblent indiquer que seul l'échange d'une majorité du domaine globulaire (et non de la région C-terminale) de H2A.Z pour H2A entraîne un phénotype aussi sévère que la délétion (yFR795-797) (**Figure 5.2**). Aucun des stress ne fait ressortir une souche en particulier. Il n'a donc pas été possible d'associer différentes fonctions de H2A.Z à différentes régions. Il est évident que pour assurer convenablement ses fonctions, H2A.Z doit être correctement localisé, d'où la nécessité d'analyser toutes les mutations à la lumière des expériences de localisation génomique de H2A.Z (et non seulement de son incorporation à la chromatine).

La deuxième stratégie est un criblage par substitution en alanine de tous les résidus de Htz1 (excepté la première méthionine et 15 alanines) réalisé par le groupe de Yamaguchi au Japon [414]. Les résultats de leur essai sur plaques (similaire au nôtre) sont comparés à ceux d'une collection construite à l'identique pour H2A. Cette étude identifie sept mutations ayant un phénotype, mais seulement deux (F32A, I109A) sont spécifiques à Htz1, c'est-à-dire qu'une mutation similaire dans H2A ne reproduit pas le même phénotype (sensibilité au MMS ou au benomyl respectivement). Contrairement à ce que l'on pouvait attendre de ce crible, ces mutations sont localisées dans des régions de divergence modérée entre H2A et H2A.Z; F32 entre la queue N-terminale et la boucle 1, I109 juste après la zone acide. Encore plus inattendu, les résidus isolés sont conservés entre les deux histones (**Figure 5.1**). Les sensibilités au stress causées par les mutations reposent donc peut-être sur des interactions, H2A.Z spécifiques, de ces résidus. Il serait utile de posséder la structure cristallographique d'un nucléosome de levure incorporant H2A.Z; or pour l'instant, elle n'est disponible que pour des histones de métazoaires.

Les résidus identifiés par l'équipe de Yamaguchi se trouvent tous les deux dans des régions de divergence modérée dont l'échange n'a pas d'effet visible dans notre criblage (**Figure 5.1** et **Figure 5.2**). Mais surtout, les résidus sont conservés entre H2A et H2A.Z et donc non mutés dans notre collection. Aucune de ces études n'a réussi à isoler de manière satisfaisante une région responsable des fonctions de

H2A.Z. Ces souches restent des outils particulièrement intéressants qui pourront être utilisés pour tester des phénotypes plus complexes, comme la formation de boucles génomiques par 3C, ou la localisation à la membrane.

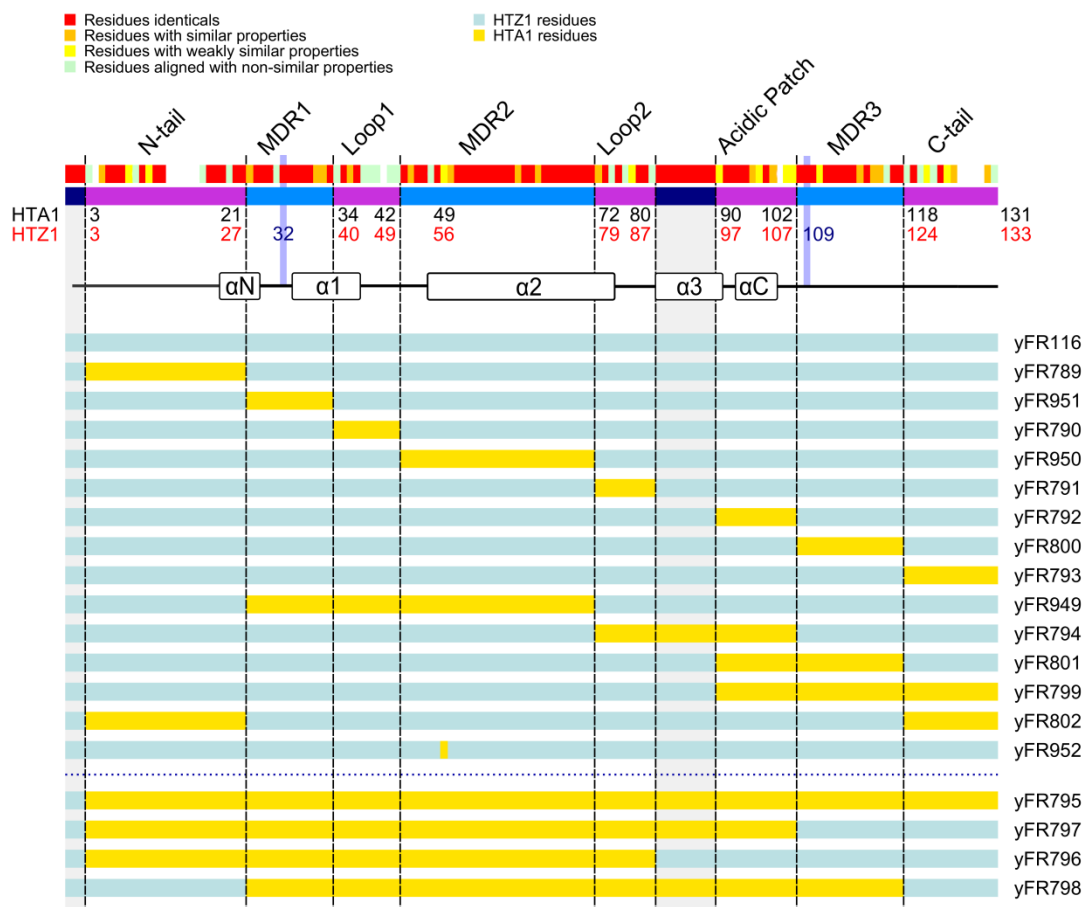


Figure 5.1 : Collection de souches chimériques H2A/H2A.Z

S'appuyant sur l'alignement des séquences H2A/H2A.Z (représenté en haut de la figure à la manière de la **Figure 1.5**), la structure des histones et la conservation au cours de l'évolution, la séquence de HTZ1 a été divisée en 10 segments : deux segments totalement similaires représentés en bleu foncé, cinq segments de divergence prononcée en violet, et trois segments de divergence modérée en bleu. Les résidus aux bornes de ces segments sont indiqués pour HTA1 et HTZ1 en noir et en rouge respectivement. Ces segments ont été remplacés dans HTZ1 (bleu clair), seul ou en combinaison, par la séquence équivalente de HTA1 (jaune) grâce à la méthode de *Dellitto Perfetto* [415]. Au total, 18 souches ont été construites, incluant la substitution d'un seul amino-acide en position 56 de HTZ1 : yFR952. Les résidus identifiés par l'étude de Yamaguchi sont indiqués par un numéro et une bande bleue, afin de pouvoir comparer les deux études [414].

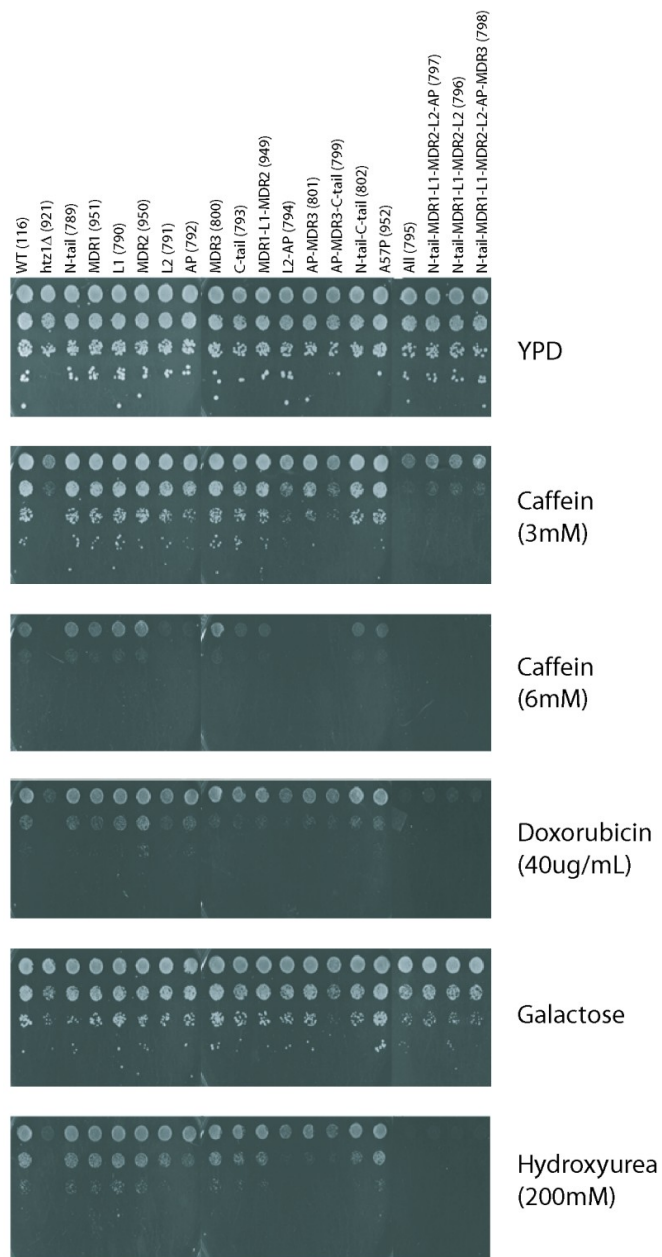


Figure 5.2 : Essai sur plaques de la collection de souches chimériques H2A/H2A.Z

Les souches de la **Figure 5.1** ont été testées par essai sur plaques pour leur résistance au stress. Alors que la croissance des mutants est peu affectée en condition optimale (YPD, 30°C) leur croissance est affectée à différents degrés en fonction du stress chimique appliqué. La souche *htz1Δ* (2^e colonne) est toujours la plus affectée et seulement égalée par les substitutions englobant une majorité de HTZ1 (4 colonnes les plus à droite).

5.2 Transcription

5.2.1 Conclusion

Bien que les mécanismes de transcription soient largement conservés dans l'ensemble du domaine vivant, les eucaryotes ont considérablement complexifié leur machinerie de transcription. Cette évolution reflète, évidemment, l'adoption de la chromatine pour l'empaquetage de l'ADN dans le noyau, mais aussi l'apparition du CTD de l'ARNPolIII (présent chez une majorité des eucaryotes) ayant entraîné une évolution synchronique des kinases et phosphatases dont il est le substrat [416]. La fonction principale du CTD de l'ARNPolIII est d'orchestrer le recrutement de protéines au niveau de la machinerie de transcription, afin de faciliter la maturation de l'ARN messager et de modifier les nucléosomes de la matrice d'ADN. La nature transitoire des marques permet au CTD de l'ARNPolIII d'être modifié dynamiquement en fonction de la progression au cours de la transcription, et forme la base du code du CTD.

Le travail présenté au chapitre 4 de cette thèse suit la publication de trois études génomiques portant sur le cycle de phosphorylation du CTD de l'ARNPolIII chez la levure [342-344]. Cependant, nos résultats clarifient et apportent de nouveaux éléments sur l'établissement des marques du CTD.

Notre travail appuie l'universalité génomique des profils de phosphorylation de l'ARNPolIII, avancée par le groupe de Cramer [344], mais contestée la même année par celui d'Ansari [343]. Remarquablement, les résultats bruts de ces travaux sont similaires, mais les méthodes d'analyse et l'interprétation des résultats sont à l'origine du désaccord. L'étude de nombreux profils de localisation génomique nous a permis de développer l'expertise nécessaire à l'analyse de ce type de données. Quelle que soit l'expérience, aucune technique d'analyse ne peut être appliquée sans évaluer les paramètres qui pourront biaiser les résultats. Il apparaît particulièrement dommageable, chez un organisme dont le génome est aussi compact que celui de la levure, de traiter les métagènes en longueur relative, c'est-à-dire en pourcentage de leur longueur. Cette erreur a pour effet d'agglomérer ensemble des valeurs situées à

des distances extrêmement différentes des frontières des gènes (*transcription start site* TSS et *transcription end site* TES), alors que l'accumulation ou la disparition des marques se produit en fonction de la distance au TSS et au TES, indépendamment de la taille des gènes. Cette remarque s'applique non seulement aux marques de phosphorylation du CTD, mais aussi à toutes les transactions impliquant les histones à la chromatine. Les exceptions à la règle, illustrées au chapitre 4, ne sont dues qu'à la présence d'éléments génétiques fortement exprimés, contaminant les régions voisines. L'universalité des profils de phosphorylation du CTD restera, sans doute, polémique pendant encore quelque temps. La conception d'un algorithme informatique modélisant de manière théorique l'apparition et la disparition des marques du CTD sur chaque gène, pourrait permettre de rechercher les exceptions de façon optimale. La conception d'un tel programme, entreprise dans notre laboratoire par Pierre-Étienne Jacques, s'avère cependant extrêmement complexe, et ne tranchera pas la question dans un avenir proche.

L'établissement du profil de chaque marque de phosphorylation est le résultat de la collaboration de plusieurs enzymes au CTD de la polymérase. Les travaux accumulés durant une décennie sur le code de phosphorylation du CTD ainsi que ceux présentés au chapitre 4 aboutissent à la vision suivante :

La marque de phosphorylation de la sérine 5 est déposée par la Kinase Kin28 lors de l'initiation, partiellement déphosphorylée à proximité du promoteur par la phosphatase Rtr1 qui est recrutée de façon transitoire, pour être complètement retirée en fin de transcription par Ssu72, après isomérisation des ponts peptidyl-proline par Ess1. On peut noter à ce sujet que, malgré nos efforts répétés, il nous a été impossible de reproduire sur l'ensemble du génome les données présentées par Washburn et ses collègues, tant en ce qui concerne la localisation de la phosphatase Rtr1 que le profil de la marque P-Ser5 dans la souche *rtr1Δ* avec l'anticorps H14 [346]. De plus, dans cette souche, nous n'avons observé aucun effet lorsque la détection de la P-Ser5 est réalisée avec l'anticorps 3E8. Il serait donc intéressant de pousser plus loin l'étude de la protéine Rtr1 et de son activité catalytique, afin de confirmer ou d'infirmier le rôle de Rtr1 dans la transcription et de comprendre les résultats obtenus dans chaque laboratoire.

La marque de phosphorylation de la sérine 7 est déposée par la kinase Kin28 au cours de l'initiation. Nos travaux ont identifié Su72 comme la phosphatase de la Ser7 et démontré que l'activité d'isomérisation d'Ess1 jouait aussi un rôle dans le retrait de la marque de phosphorylation de la sérine 7. La Ser5 et la Ser7 partagent donc le même couple kinase/phosphatase aux extrémités des gènes.

La marque de phosphorylation de la sérine 2 est déposée majoritairement par Ctk1, avec la participation de Bur1, partiellement déphosphorylée par Fcp1 dans l'ORF et finalement déphosphorylée après la terminaison par Fcp1. Plus inattendus, les travaux de notre laboratoire suggèrent que la déphosphorylation de la Ser2 nécessite aussi la collaboration de deux enzymes; Fcp1 ne pouvant finir la déphosphorylation de la Ser2 qu'avec le concours de Ssu72. La nature de l'activité de Ssu72 permettant à Fcp1 de compléter la déphosphorylation reste cependant à déterminer.

Nous avons mis en évidence l'activité opportuniste de la kinase Bur1, sur les sérines 5 et 7, en absence de l'activité de Kin28. Cette activité est sans doute partagée par d'autres kinases comme le suggèrent les profils obtenus au chapitre 4. Ces conclusions vont à l'encontre de la théorie avancée par le groupe d'Ansari, selon laquelle la sérine 7 est phosphorylée en deux phases : au promoteur par Kin28, puis par Bur1 dans le corps des gènes, après avoir été déphosphorylée par une phosphatase qui resterait à identifier [343].

Si le code du CTD semble être universel, sa mise en place est incroyablement plus complexe qu'initialement envisagé. Elle est le fruit de la coordination et du recrutement séquentiel d'enzymes au CTD de l'ARNPolIII et repose, finalement, peu sur la spécificité intrinsèque des Kinases. Comme nos résultats le confirment, il est de plus en plus évident que la phosphorylation du CTD n'est pas essentielle au déroulement de la transcription, mais aux réactions qui lui sont associées, puisqu'aucun de nos mutants n'abolit la transcription ou ne modifie le profil de distribution de la polymérase le long de l'unité de transcription.

Basés sur la déphosphorylation en deux étapes des marques P-Ser5 et 2, ces résultats appuient clairement le concept de variantes, pour chacune des marques de phosphorylation. Il est cependant impossible, dans l'état actuel de nos connaissances,

de déterminer clairement ce qui définit une variante. De nombreux paramètres sont cependant à envisager : la présence de marques de phosphorylation sur les sérines adjacentes, la position proximale ou distale le long du CTD, la conformation Cis ou Trans des ponts peptidyl-proline et la présence de résidus dégénérés dans les répétitions. S'il paraît extrêmement intéressant pour le domaine de la transcription de définir le rôle et les interactions des variantes, de telles initiatives sont pour l'instant freinées par les limitations techniques. La structure répétitive du CTD constitue l'obstacle principal à son étude.

À cause de leur dépendance aux anticorps phosphospécifiques, les techniques actuelles pour l'étude de la phosphorylation du CTD ont sans doute atteint leurs limites. Et les techniques d'analyse par spectrométrie de masse sont, pour l'instant, impuissantes à résoudre directement la complexité du CTD.

Le développement d'anticorps discriminant les marques de phosphorylation a fait avancer à grands pas l'étude de la transcription. Les premiers anticorps de ce type, H5 (P-Ser2) et H14 (P-Ser5), ont permis d'établir les premiers profils de phosphorylation, mettant en évidence la nature dynamique de la modification des résidus du CTD [330]. Ces anticorps furent utilisés presque exclusivement pendant près d'une décennie. La mise au point de nouveaux anticorps spécifiques par le groupe de Eick a permis de générer de nouveaux profils et de comparer l'affinité des anticorps. Les résultats démontrent que l'on ne peut ignorer les biais des anticorps, qui apparaissent difficiles à contrevenir. Pour une structure composée de courtes répétitions contenant chacune plusieurs sites de phosphorylation, il paraît improbable de parvenir au développement d'anticorps sans que leur affinité pour une modification soit influencée par la présence d'autres modifications à proximité. Toutefois, il reste possible de caractériser exhaustivement les anticorps à notre disposition de façon à pouvoir intégrer ces connaissances à l'analyse de nos résultats, et corriger au mieux les biais. De tels efforts, entrepris pour certaines combinaisons par le groupe de Eick, se sont déjà révélés très utiles dans l'interprétation de nos résultats, et justifie de pousser plus avant la caractérisation des anticorps [332].

5.2.2 Perspectives

Cette étude a permis à notre laboratoire de développer l'expertise nécessaire à la poursuite des travaux sur le CTD de l'ARNPolIII de façon innovante. Ce nouveau projet s'affranchit de la dépendance aux anticorps pour les phosphoisoformes du CTD et tire avantage des techniques et outils de purification du complexe de la polymérase, optimisés au laboratoire durant notre étude du cycle de phosphorylation du CTD. L'expression ectopique d'une version étiquetée de Rpb1 permet une copurification du complexe de la polymérase et des protéines associées, incluant les partenaires du CTD. Cette stratégie, originalement développée afin d'établir l'inaccessibilité des épitopes du CTD dans une souche *fcp1-1* (chapitre 4), est utilisée en conjonction avec les techniques d'analyse par spectroscopie de masse, dans le but de mesurer l'impact des mutations des kinases et phosphatases sur l'ensemble du réseau d'interactions du CTD. Il est en effet possible de quantifier, de façon relative, les protéines interagissant avec le CTD entre les échantillons d'une souche mutante et d'une souche de type sauvage, grâce à l'abondance des peptides identifiés par spectroscopie de masse. Cette approche permet d'identifier les changements dans la composition des interactants, de manière précise et surtout non biaisée. Il est en outre possible de déterminer l'impact direct de mutations des répétitions par substitution de la séquence codante pour le CTD. Il est, en théorie, possible de tester des mutations qui se révèlent létales lorsqu'introduites dans la copie endogène de Rpb1, comme c'est par exemple le cas pour les mutations de toutes les sérines 2 ou 5 en alanines [223]. Les résultats préliminaires obtenus dans le laboratoire, et déjà confirmés par immunobuvardage par la méthode de Western, valident cette approche et devraient déboucher sur de nouvelles conclusions quant au rôle du CTD dans le recrutement et la coordination des protéines lors de la transcription par l'ARNPolIII.

Ce projet devrait permettre de dévoiler certains mécanismes d'interaction du CTD avec les protéines de la transcription. Il est intéressant de constater qu'un sujet comme le CTD de la polymérase peut encore générer de nouvelles hypothèses et pousser au développement de stratégies innovantes.

6 Conclusion

Au cours des dernières années, l'incroyable développement des techniques d'investigation génomique ont permis d'étudier la chromatine et la transcription d'une manière beaucoup plus globale. Permettant une étude non biaisée, ces travaux ont renforcé le concept d'une communication étroite entre la chromatine et la machinerie de transcription.

Exploitant les facilités de manipulation de la levure *Saccharomyces cerevisiae*, les études incorporées dans cette thèse s'intéressent à deux aspects de la transcription, qui sont spécifiques aux eucaryotes et extrêmement bien conservés au cours de leur évolution : d'une part, la chromatine avec la localisation de la variante d'histone H2A.Z, et, de l'autre, le cycle de phosphorylation de l'ARNPolII.

Dans la lignée des études génomiques, nous avons été parmi les premiers à démontrer une localisation de H2A.Z aux promoteurs de la vaste majorité des gènes de levure et de manière anticorrélée à la transcription. Cette découverte suggère un rôle prépondérant de H2A.Z dans la transcription, bien plus important que celui avancé précédemment. Elle a, en conséquence, relancé l'intérêt pour l'étude de cette variante et incité à sa localisation chez les eucaryotes supérieurs.

Nous avons également travaillé sur la phosphorylation du domaine C-terminal de la sous-unité Rpb1 de l'ARN polymérase II. Cette deuxième étude s'inscrit elle aussi dans une suite de publications portant sur les profils génomiques des marques de phosphorylation du CTD, mais se démarque par l'étendue de ses observations. Ainsi, nous avons pu identifier la phosphatase de la sérine 7 et préciser les relations complexes qui lient kinases et phosphatases. Comme cela fut le cas pour notre étude de la variante d'histone H2A.Z, les observations sur la phosphorylation du CTD pourront servir de base aux travaux sur les eucaryotes supérieurs, généralement plus complexes. Cette thèse souligne également les limites actuelles des techniques de génomique appliquées à l'étude du CTD de l'ARN polymérase II et trace, dans sa discussion, les grandes lignes d'une approche protéomique affranchie de ces limitations, liées à l'utilisation des anticorps. Cette nouvelle perspective devrait permettre de poursuivre l'étude du CTD de l'ARN polymérase II, domaine ayant un

rôle clé dans le processus de maturation des ARN messagers et l'établissement des modifications de la chromatine, résultant de la transcription. Bien qu'intensément étudiée, la complexité des mécanismes de phosphorylation du CTD semble réserver encore de nombreuses surprises.

On peut espérer que des travaux complémentaires à venir, s'appuyant sur cette thèse, ouvriront de nouvelles pistes dans l'étude de la transcription.

7 Bibliographie

1. Paweletz, N., *Walther Flemming: pioneer of mitosis research*. Nat. Rev. Mol. Cell Biol., 2001. **2**: p. 72-5.
2. Trojer, P. and D. Reinberg, *Facultative heterochromatin: is there a distinctive molecular signature?* Mol. Cell, 2007. **28**: p. 1-13.
3. Zacharias, H., *Emil Heitz (1892-1965): chloroplasts, heterochromatin, and polytene chromosomes*. Genetics, 1995. **141**: p. 7-14.
4. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**: p. 737-8.
5. Kornberg, R.D. and J.O. Thomas, *Chromatin structure; oligomers of the histones*. Science, 1974. **184**: p. 865-8.
6. Adkins, N.L., M. Watts, and P.T. Georgel, *To the 30-nm chromatin fiber and beyond*. Biochim. Biophys. Acta, 2004. **1677**: p. 12-23.
7. Fussner, E., R.W. Ching, and D.P. Bazett-Jones, *Living without 30nm chromatin fibers*. Trends Biochem. Sci., 2011. **36**: p. 1-6.
8. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation*. Curr. Opin. Genet. Dev., 2011: p. 1-12.
9. Wu, C., A. Bassett, and A. Travers, *A variable topology for the 30-nm chromatin fibre*. EMBO Rep., 2007. **8**: p. 1129-34.
10. Finch, J.T. and A. Klug, *Solenoidal model for superstructure in chromatin*. Proc. Natl. Acad. Sci. U. S. A., 1976. **73**: p. 1897-901.
11. Robinson, P.J., et al., *EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure*. Proc. Natl. Acad. Sci. U. S. A., 2006. **103**(17): p. 6506-11.
12. Horowitz, R.A., et al., *The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon*. J. Cell Biol., 1994. **125**: p. 1-10.

13. Maeshima, K., S. Hihara, and M. Eltsov, *Chromatin structure: does the 30-nm fibre exist in vivo?* *Curr. Opin. Cell Biol.*, 2010. **22**: p. 291-7.
14. Eltsov, M., et al., *Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ.* *Proc. Natl. Acad. Sci. U. S. A.*, 2008. **105**: p. 19732-7.
15. Meldi, L. and J.H. Brickner, *Compartmentalization of the nucleus.* *Trends Cell Biol.*, 2011: p. 1-8.
16. Cremer, T., et al., *Chromosome territories--a functional nuclear landscape.* *Curr. Opin. Cell Biol.*, 2006. **18**: p. 307-16.
17. Zuleger, N., M.I. Robson, and E.C. Schirmer, *The nuclear envelope as a chromatin organizer.* *Nucleus*, 2011. **2**: p. 1-11.
18. Zimmer, C. and E. Fabre, *Principles of chromosomal organization: lessons from yeast.* *J. Cell Biol.*, 2011. **192**: p. 723-33.
19. Takizawa, T., K.J. Meaburn, and T. Misteli, *The meaning of gene positioning.* *Cell*, 2008. **135**: p. 9-13.
20. Casolari, J.M., et al., *Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization.* *Cell*, 2004. **117**: p. 427-39.
21. Ragoczy, T., et al., *The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation.* *Genes Dev.*, 2006. **20**: p. 1447-57.
22. Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* *Genes Dev.*, 2004. **18**: p. 1119-30.
23. Taddei, A., *Active genes at the nuclear pore complex.* *Curr. Opin. Cell Biol.*, 2007. **19**: p. 305-10.
24. Brickner, J.H., *Transcriptional memory at the nuclear periphery.* *Curr. Opin. Cell Biol.*, 2009. **21**: p. 127-33.
25. O'Sullivan, J.M., et al., *Gene loops juxtapose promoters and terminators in yeast.* *Nat. Genet.*, 2004. **36**: p. 1014-8.

26. Dekker, J., et al., *Capturing chromosome conformation*. *Science*, 2002. **295**: p. 1306-11.
27. Dostie, J., et al., *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements*. *Genome Res.*, 2006. **16**: p. 1299-309.
28. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*. *Science*, 2009. **326**: p. 289-93.
29. Duan, Z., et al., *A three-dimensional model of the yeast genome*. *Nature*, 2010. **465**: p. 363-367.
30. Kornberg, R.D., *Structure of chromatin*. *Annu. Rev. Biochem.*, 1977. **46**: p. 931-54.
31. Verreault, A., et al., *Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4*. *Cell*, 1996. **87**: p. 95-104.
32. Schmid, A., K.D. Fascher, and W. Hörz, *Nucleosome disruption at the yeast PHO5 promoter upon PHO5 induction occurs in the absence of DNA replication*. *Cell*, 1992. **71**: p. 853-64.
33. Svaren, J. and W. Hörz, *Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast*. *Trends Biochem. Sci.*, 1997. **22**: p. 93-7.
34. Lohr, D., *Organization of the GAL1-GAL10 intergenic control region chromatin*. *Nucleic Acids Res.*, 1984. **12**: p. 8457-74.
35. Rainbow, M., J. Lopez, and D. Lohr, *The yeast GAL1-10 UAS region readily accepts nucleosomes in vitro*. *Biochemistry*, 1989. **28**: p. 7486-90.
36. Wu, C., *Analysis of hypersensitive sites in chromatin*. *Methods Enzymol.*, 1989. **170**: p. 269-89.
37. Bellard, M., et al., *Nuclease digestion of transcriptionally active chromatin*. *Methods Enzymol.*, 1989. **170**: p. 317-46.
38. Lutter, L.C., *Digestion of nucleosomes with deoxyribonucleases I and II*. *Methods Enzymol.*, 1989. **170**: p. 264-9.
39. Yuan, G.C., et al., *Genome-scale identification of nucleosome positions in *S. cerevisiae**. *Science*, 2005. **309**(5734): p. 626-30.

40. Schones, D.E., et al., *Dynamic regulation of nucleosome positioning in the human genome*. Cell, 2008. **132**: p. 887-98.
41. Mavrich, T.N., et al., *Nucleosome organization in the Drosophila genome*. Nature, 2008. **453**: p. 358-62.
42. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome*. Cell, 2007. **129**: p. 823-37.
43. Weiner, A., et al., *High-resolution nucleosome mapping reveals transcription-dependent promoter packaging*. Genome Res., 2010. **20**: p. 90-100.
44. Lee, W., et al., *A high-resolution atlas of nucleosome occupancy in yeast*. Nat. Genet., 2007. **39**: p. 1235-44.
45. Liu, C.L., et al., *Single-nucleosome mapping of histone modifications in S. cerevisiae*. PLoS Biol., 2005. **3**: p. e328.
46. Jiang, C. and B.F. Pugh, *Nucleosome positioning and gene regulation: advances through genomics*. Nat. Rev. Genet., 2009. **10**: p. 161-72.
47. Anderson, J.D. and J. Widom, *Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites*. Mol. Cell. Biol., 2001. **21**: p. 3830-9.
48. Rando, O.J. and H.Y. Chang, *Genome-wide views of chromatin structure*. Annu. Rev. Biochem., 2009. **78**: p. 245-71.
49. Rando, O.J., *Global patterns of histone modifications*. Curr. Opin. Genet. Dev., 2007. **17**: p. 94-9.
50. Raisner, R.M., et al., *Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin*. Cell, 2005. **123**: p. 233-48.
51. Dion, M.F., et al., *Dynamics of replication-independent histone turnover in budding yeast*. Science, 2007. **315**: p. 1405-8.
52. Rufiange, A., et al., *Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1*. Mol. Cell, 2007. **27**: p. 393-405.
53. Segal, E., et al., *A genomic code for nucleosome positioning*. Nature, 2006. **442**: p. 772-8.

54. Yuan, G.-C. and J.S. Liu, *Genomic sequence is highly predictive of local nucleosome depletion*. PLoS Comput. Biol., 2008. **4**: p. e13.
55. Kaplan, N., et al., *The DNA-encoded nucleosome organization of a eukaryotic genome*. Nature, 2009. **458**: p. 362-6.
56. Peckham, H.E., et al., *Nucleosome positioning signals in genomic DNA*. Genome Res., 2007. **17**: p. 1170-7.
57. Ramakrishnan, V., *Histone structure and the organization of the nucleosome*. Annu. Rev. Biophys. Biomol. Struct., 1997. **26**: p. 83-112.
58. Kaplan, N., et al., *Nucleosome sequence preferences influence in vivo nucleosome organization*. Nat. Struct. Mol. Biol., 2010. **17**: p. 918-20; author reply 920-2.
59. Zhang, Y., et al., *Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo*. Nat. Struct. Mol. Biol., 2009. **16**: p. 847-52.
60. Mavrich, T.N., et al., *A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome*. Genome Res., 2008. **18**: p. 1073-83.
61. Zhang, Z., et al., *A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome*. Science, 2011. **332**: p. 977-80.
62. Bryant, G.O., et al., *Activator control of nucleosome occupancy in activation and repression of transcription*. PLoS Biol., 2008. **6**: p. 2928-39.
63. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**: p. 251-60.
64. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. Nature, 2000. **403**: p. 41-5.
65. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**: p. 1074-80.
66. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation*. Genes Dev., 2011. **25**: p. 1345-1358.
67. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression*. Nature, 2009. **459**: p. 108-12.

68. Wolffe, A.P., *Histone H1*. The international journal of biochemistry & cell biology, 1997. **29**: p. 1463-6.
69. Vignali, M. and J.L. Workman, *Location and function of linker histones*. Nat. Struct. Biol., 1998. **5**: p. 1025-8.
70. Bustin, M., F. Catez, and J.-H. Lim, *The dynamics of histone H1 function in chromatin*. Mol. Cell, 2005. **17**: p. 617-20.
71. Thomas, J.O., *Histone H1: location and role*. Curr. Opin. Cell Biol., 1999. **11**: p. 312-7.
72. Thoma, F., T. Koller, and A. Klug, *Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin*. J. Cell Biol., 1979. **83**: p. 403-27.
73. Schäfer, G., E.M. Smith, and H.G. Patterson, *The Saccharomyces cerevisiae linker histone Hho1p, with two globular domains, can simultaneously bind to two four-way junction DNA molecules*. Biochemistry, 2005. **44**: p. 16766-75.
74. Ushinsky, S.C., et al., *Histone H1 in Saccharomyces cerevisiae*. Yeast, 1997. **13**: p. 151-61.
75. Downs, J.A., et al., *Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone*. Mol. Cell, 2003. **11**: p. 1685-92.
76. Freidkin, I. and D.J. Katcoff, *Specific distribution of the Saccharomyces cerevisiae linker histone homolog HHO1p in the chromatin*. Nucleic Acids Res., 2001. **29**: p. 4043-51.
77. Richmond, T.J., et al., *Structure of the nucleosome core particle at 7 Å resolution*. Nature, 1984. **311**: p. 532-7.
78. Luger, K. and T.J. Richmond, *DNA binding within the nucleosome core*. Curr. Opin. Struct. Biol., 1998. **8**: p. 33-40.
79. White, C.L., R.K. Suto, and K. Luger, *Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions*. EMBO J., 2001. **20**: p. 5207-18.
80. Kireeva, M.L., et al., *Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription*. Mol. Cell, 2002. **9**: p. 541-52.

81. Luger, K., *Dynamic nucleosomes*. Chromosome Res., 2006. **14**: p. 5-16.
82. Anderson, J.D. and J. Widom, *Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites*. J. Mol. Biol., 2000. **296**: p. 979-87.
83. Gunjan, A., J. Paik, and A. Verreault, *Regulation of histone synthesis and nucleosome assembly*. Biochimie, 2005. **87**: p. 625-35.
84. Osley, M.A. and D. Lycan, *Trans-acting regulatory mutations that alter transcription of Saccharomyces cerevisiae histone genes*. Mol. Cell. Biol., 1987. **7**: p. 4204-10.
85. Marzluff, W.F., et al., *The human and mouse replication-dependent histone genes*. Genomics, 2002. **80**: p. 487-98.
86. Kang, B., et al., *Phosphorylation of H4 Ser 47 promotes HIRA-mediated nucleosome assembly*. Genes Dev., 2011. **25**: p. 1359-64.
87. Marzluff, W.F. and R.J. Duronio, *Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences*. Curr. Opin. Cell Biol., 2002. **14**: p. 692-9.
88. Lycan, D.E., M.A. Osley, and L.M. Hereford, *Role of transcriptional and posttranscriptional regulation in expression of histone genes in Saccharomyces cerevisiae*. Mol. Cell. Biol., 1987. **7**: p. 614-21.
89. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis*. J. Comput. Chem., 2004. **25**: p. 1605-12.
90. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes*. Annu. Rev. Biochem., 2009. **78**: p. 273-304.
91. Mello, J.A. and G. Almouzni, *The ins and outs of nucleosome assembly*. Curr. Opin. Genet. Dev., 2001. **11**: p. 136-41.
92. Loyola, A. and G. Almouzni, *Histone chaperones, a supporting role in the limelight*. Biochim. Biophys. Acta, 2004. **1677**: p. 3-11.
93. Cairns, B.R., *Chromatin remodeling: insights and intrigue from single-molecule studies*. Nat. Struct. Mol. Biol., 2007. **14**: p. 989-96.

94. Tang, L. and E. Nogales, *Structures of SWI/SNF Chromatin Remodeling Complexes and the Mechanistic Implications in Transcription*. Prog. Biophys. Mol. Biol., 2010. **102**: p. 122-128.
95. Owen-Hughes, T., et al., *Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex*. Science, 1996. **273**: p. 513-6.
96. Ito, T., et al., *ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor*. Cell, 1997. **90**: p. 145-55.
97. Yang, J.G., et al., *The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing*. Nat. Struct. Mol. Biol., 2006. **13**: p. 1078-83.
98. Fazio, T.G., et al., *Widespread Collaboration of Isw2 and Sin3-Rpd3 Chromatin Remodeling Complexes in Transcriptional Repression*. Mol. Cell. Biol., 2001. **21**: p. 6450-6460.
99. Morrison, A.J. and X. Shen, *Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes*. Nat. Rev. Mol. Cell Biol., 2009. **10**: p. 373-84.
100. Ebbert, R., A. Birkmann, and H.J. Schüller, *The product of the SNF2/SWI2 paralogue INO80 of Saccharomyces cerevisiae required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex*. Mol. Microbiol., 1999. **32**: p. 741-51.
101. van Attikum, H., O. Fritsch, and S.M. Gasser, *Distinct roles for SWR1 and INO80 chromatin remodeling complexes at chromosomal double-strand breaks*. EMBO J., 2007. **26**: p. 4113-25.
102. Papamichos-Chronakis, M., J.E. Krebs, and C.L. Peterson, *Interplay between Ino80 and Swr1 chromatin remodeling enzymes regulates cell cycle checkpoint adaptation in response to DNA damage*. Genes Dev., 2006. **20**: p. 2437-49.
103. Papamichos-Chronakis, M., et al., *Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity*. Cell, 2011. **144**: p. 200-13.

104. Berger, S.L., *The complex language of chromatin regulation during transcription*. Nature, 2007. **447**: p. 407-12.
105. Suganuma, T. and J.L. Workman, *Crosstalk among Histone Modifications*. Cell, 2008. **135**: p. 604-7.
106. Lee, J.-S., E. Smith, and A. Shilatifard, *The Language of Histone Crosstalk*. Cell, 2010. **142**: p. 682-685.
107. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells*. Cell, 2007. **130**: p. 77-88.
108. Hardy, S., et al., *The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition*. PLoS Genet., 2009. **5**: p. e1000687.
109. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly*. Science, 2001. **292**: p. 110-3.
110. Allis, C.D., et al., *New nomenclature for chromatin-modifying enzymes*. Cell, 2007. **131**: p. 633-6.
111. Roth, S.Y., J.M. Denu, and C.D. Allis, *Histone acetyltransferases*. Annu. Rev. Biochem., 2001. **70**: p. 81-120.
112. Peserico, A. and C. Simone, *Physical and functional HAT/HDAC interplay regulates protein acetylation balance*. J. Biomed. Biotechnol., 2011. **2011**: p. 371832.
113. Verreault, A., et al., *Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase*. Curr. Biol., 1998. **8**: p. 96-108.
114. Parthun, M.R., *Hat1: the emerging cellular roles of a type B histone acetyltransferase*. Oncogene, 2007. **26**: p. 5319-28.
115. Ekwall, K., *Genome-wide analysis of HDAC function*. Trends Genet., 2005. **21**: p. 608-15.
116. Imoberdorf, R.M., I. Topalidou, and M. Strubin, *A role for gcn5-mediated global histone acetylation in transcriptional regulation*. Mol. Cell. Biol., 2006. **26**: p. 1610-6.
117. Shahbazian, M.D. and M. Grunstein, *Functions of site-specific histone acetylation and deacetylation*. Annu. Rev. Biochem., 2007. **76**: p. 75-100.

118. Grant, P.A., et al., *Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex*. Genes Dev., 1997. **11**(13): p. 1640-50.
119. Durant, M. and B.F. Pugh, *Genome-wide relationships between TAF1 and histone acetyltransferases in Saccharomyces cerevisiae*. Mol. Cell. Biol., 2006. **26**: p. 2791-802.
120. Durant, M. and B.F. Pugh, *NuA4-directed chromatin transactions throughout the Saccharomyces cerevisiae genome*. Mol. Cell. Biol., 2007. **27**: p. 5327-35.
121. Allard, S., et al., *NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p*. EMBO J., 1999. **18**(18): p. 5108-19.
122. Clarke, A.S., et al., *Esa1p is an essential histone acetyltransferase required for cell cycle progression*. Mol. Cell. Biol., 1999. **19**(4): p. 2515-26.
123. Boudreault, A.A., et al., *Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin*. Genes Dev., 2003. **17**(11): p. 1415-28.
124. Li, B., et al., *Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin*. Science, 2007. **316**: p. 1050-4.
125. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity*. Mol. Cell, 2003. **11**: p. 709-19.
126. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1*. Cell, 2004. **119**: p. 941-53.
127. Tu, S., et al., *Identification of histone demethylases in Saccharomyces cerevisiae*. J. Biol. Chem., 2007. **282**: p. 14262-71.
128. Pedersen, M.T. and K. Helin, *Histone demethylases in development and disease*. Trends Cell Biol., 2010. **20**: p. 662-71.
129. Kamakaka, R.T. and S. Biggins, *Histone variants: deviants?* Genes Dev., 2005. **19**: p. 295-310.
130. Thatcher, T.H. and M.A. Gorovsky, *Phylogenetic analysis of the core histones H2A, H2B, H3, and H4*. Nucleic Acids Res., 1994. **22**: p. 174-9.

131. Mito, Y., J.G. Henikoff, and S. Henikoff, *Genome-scale profiling of histone H3.3 replacement patterns*. Nat. Genet., 2005. **37**: p. 1090-7.
132. Ahmad, K. and S. Henikoff, *Histone H3 variants specify modes of chromatin assembly*. Proc. Natl. Acad. Sci. U. S. A., 2002. **99 Suppl 4**: p. 16477-84.
133. Szilard, R.K., et al., *Systematic identification of fragile sites via genome-wide location analysis of gamma-H2AX*. Nat. Struct. Mol. Biol., 2010. **17**: p. 299-305.
134. Adam, M., et al., *H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions*. Mol. Cell. Biol., 2001. **21**: p. 6270-9.
135. Santisteban, M.S., T. Kalashnikova, and M.M. Smith, *Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes*. Cell, 2000. **103**: p. 411-22.
136. Faast, R., et al., *Histone variant H2A.Z is required for early mammalian development*. Curr. Biol., 2001. **11**: p. 1183-7.
137. Clarkson, M.J., et al., *Regions of variant histone His2AvD required for Drosophila development*. Nature, 1999. **399**: p. 694-7.
138. Ridgway, P., et al., *Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development*. J. Biol. Chem., 2004. **279**: p. 43815-20.
139. Liu, X., B. Li, and M. Gorovsky, *Essential and nonessential histone H2A variants in Tetrahymena thermophila*. Mol. Cell. Biol., 1996. **16**: p. 4305-11.
140. Allis, C.D., et al., *Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila*. Cell, 1980. **20**: p. 609-17.
141. Zlatanova, J. and A. Thakar, *H2A.Z: view from the top*. Structure, 2008. **16**: p. 166-79.
142. Suto, R.K., et al., *Crystal structure of a nucleosome core particle containing the variant histone H2A.Z*. Nat. Struct. Biol., 2000. **7**: p. 1121-4.

143. Viens, A., et al., *Analysis of human histone H2AZ deposition in vivo argues against its direct role in epigenetic templating mechanisms.* Mol. Cell. Biol., 2006. **26**: p. 5325-35.
144. Weber, C.M., J.G. Henikoff, and S. Henikoff, *H2A.Z nucleosomes enriched over active genes are homotypic.* Nat. Struct. Mol. Biol., 2010. **17**: p. 1500-1507.
145. Chakravarthy, S., et al., *Structural characterization of histone H2A variants.* Cold Spring Harb. Symp. Quant. Biol., 2004. **69**: p. 227-34.
146. Luk, E., et al., *Stepwise histone replacement by SWR1 requires dual activation with histone H2A.Z and canonical nucleosome.* Cell, 2010. **143**: p. 725-36.
147. Jensen, K., et al., *Histone H2A.Z acid patch residues required for deposition and function.* Molecular genetics and genomics : MGG, 2011. **285**: p. 287-96.
148. Abbott, D.W., et al., *Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation.* J. Biol. Chem., 2001. **276**: p. 41945-9.
149. Park, Y.-J., et al., *A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome.* J. Biol. Chem., 2004. **279**: p. 24274-82.
150. Thambirajah, A.A., et al., *H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation.* J. Biol. Chem., 2006. **281**: p. 20036-44.
151. Jin, C. and G. Felsenfeld, *Nucleosome stability mediated by histone variants H3.3 and H2A.Z.* Genes Dev., 2007. **21**: p. 1519-29.
152. Jin, C., et al., *H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.* Nat. Genet., 2009. **41**: p. 941-5.
153. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* J. Mol. Biol., 2000. **302**: p. 205-17.

154. Redon, C., et al., *Histone H2A variants H2AX and H2AZ*. *Curr. Opin. Genet. Dev.*, 2002. **12**: p. 162-9.
155. Meneghini, M.D., M. Wu, and H.D. Madhani, *Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin*. *Cell*, 2003. **112**: p. 725-36.
156. Dhillon, N. and R.T. Kamakaka, *A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR*. *Mol. Cell*, 2000. **6**: p. 769-80.
157. Guillemette, B., et al., *Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning*. *PLoS Biol.*, 2005. **3**: p. e384.
158. Zhang, H., D.N. Roberts, and B.R. Cairns, *Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss*. *Cell*, 2005. **123**: p. 219-31.
159. Li, B., et al., *Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling*. *Proc. Natl. Acad. Sci. U. S. A.*, 2005. **102**: p. 18385-90.
160. Millar, C.B., et al., *Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast*. *Genes Dev.*, 2006. **20**: p. 711-22.
161. Leach, T.J., et al., *Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster**. *J. Biol. Chem.*, 2000. **275**: p. 23267-72.
162. Bruce, K., et al., *The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken*. *Nucleic Acids Res.*, 2005. **33**: p. 5633-9.
163. Buchanan, L., et al., *The *Schizosaccharomyces pombe* JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains*. *PLoS Genet.*, 2009. **5**: p. e1000726.
164. Hardy, S. and F. Robert, *Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism?* *Epigenetics*, 2010. **5**: p. 1-5.

165. Sarcinella, E., et al., *Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin*. Mol. Cell. Biol., 2007. **27**: p. 6457-68.
166. Draker, R., E. Sarcinella, and P. Cheung, *USP10 deubiquitylates the histone variant H2A.Z and both are required for androgen receptor-mediated gene activation*. Nucleic Acids Res., 2011. **39**: p. 3529-42.
167. Light, W.H., et al., *Interaction of a DNA Zip Code with the Nuclear Pore Complex Promotes H2A.Z Incorporation and INO1 Transcriptional Memory*. Mol. Cell, 2010. **40**: p. 112-125.
168. Brickner, D.G., et al., *H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state*. PLoS Biol., 2007. **5**: p. e81.
169. Larochelle, M. and L. Gaudreau, *H2A.Z has a function reminiscent of an activator required for preferential binding to intergenic DNA*. EMBO J., 2003. **22**: p. 4512-22.
170. Krogan, N.J., et al., *Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4*. Proc. Natl. Acad. Sci. U. S. A., 2004. **101**: p. 13513-8.
171. Krogan, N.J., et al., *A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1*. Mol. Cell, 2003. **12**: p. 1565-76.
172. Kobor, M.S., et al., *A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin*. PLoS Biol., 2004. **2**: p. E131.
173. Luk, E., et al., *Chz1, a nuclear chaperone for histone H2AZ*. Mol. Cell, 2007. **25**: p. 357-68.
174. Mizuguchi, G., et al., *ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex*. Science, 2004. **303**: p. 343-8.
175. Zofall, M., et al., *Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs*. Nature, 2009. **461**: p. 419-22.

176. Straube, K., J.S. Blackwell, and L.F. Pemberton, *Nap1 and Chz1 have separate Htz1 nuclear import and assembly functions*. *Traffic*, 2010. **11**: p. 185-97.
177. Huisinga, K.L. and B.F. Pugh, *A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae*. *Mol. Cell*, 2004. **13**: p. 573-85.
178. Altaf, M., et al., *NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWRI complex*. *J. Biol. Chem.*, 2010. **285**: p. 15966-77.
179. Lu, P.Y.T., N. Lévesque, and M.S. Kobor, *NuA4 and SWRI-C: two chromatin-modifying complexes with overlapping functions and components*. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 2009. **87**: p. 799-815.
180. Auger, A., et al., *Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants*. *Mol. Cell. Biol.*, 2008. **28**: p. 2257-70.
181. Doyon, Y. and J. Côté, *The highly conserved and multifunctional NuA4 HAT complex*. *Curr. Opin. Genet. Dev.*, 2004. **14**: p. 147-54.
182. Doyon, Y., et al., *Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans*. *Mol. Cell. Biol.*, 2004. **24**: p. 1884.
183. Billon, P. and J. Côté, *Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance*. *Biochim. Biophys. Acta*, 2011.
184. Keogh, M.-C., et al., *The Saccharomyces cerevisiae histone H2A variant Htz1 is acetylated by NuA4*. *Genes Dev.*, 2006. **20**: p. 660.
185. Babiarz, J.E., J.E. Halley, and J. Rine, *Telomeric heterochromatin boundaries require NuA4-dependent acetylation of histone variant H2A.Z in Saccharomyces cerevisiae*. *Genes Dev.*, 2006. **20**: p. 700-10.
186. Altaf, M., et al., *Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin*. *Mol. Cell*, 2007. **28**(6): p. 1002-14.

187. Shia, W.-J., S.G. Pattenden, and J.L. Workman, *Histone H4 lysine 16 acetylation breaks the genome's silence*. *Genome Biol.*, 2006. **7**: p. 217.
188. Shia, W.-J., B. Li, and J.L. Workman, *SAS-mediated acetylation of histone H4 Lys 16 is required for H2A.Z incorporation at subtelomeric regions in *Saccharomyces cerevisiae**. *Genes Dev.*, 2006. **20**: p. 2507-12.
189. Creyghton, M.P., et al., *H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment*. *Cell*, 2008. **135**: p. 649-61.
190. Hua, S., et al., *Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression*. *Mol. Syst. Biol.*, 2008. **4**: p. 188.
191. Dunican, D., P. McWilliam, and O. Tighe, *expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA*. *Oncogene*, 2002.
192. Rhodes, D.R., et al., *Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression*. *Proc. Natl. Acad. Sci. U. S. A.*, 2004. **101**: p. 9309-14.
193. Zucchi, I., et al., *Gene expression profiles of epithelial cells microscopically isolated from a breast-invasive ductal carcinoma and a nodal metastasis*. *Proc. Natl. Acad. Sci. U. S. A.*, 2004. **101**: p. 18147-52.
194. Gévry, N., et al., *Histone H2A.Z is essential for estrogen receptor signaling*. *Genes Dev.*, 2009. **23**: p. 1522-33.
195. Svtelisl, A., et al., *H2A.Z overexpression promotes cellular proliferation of breast cancer cells*. *Cell Cycle*, 2010. **9**: p. 364-70.
196. Rangasamy, D., *Histone variant H2A.Z can serve as a new target for breast cancer therapy*. *Curr. Med. Chem.*, 2010. **17**: p. 3155-61.
197. Perales, R. and D. Bentley, *"Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions*. *Mol. Cell*, 2009. **36**: p. 178-91.

198. Govind, C.K., et al., *Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes*. Mol. Cell, 2010. **39**: p. 234-46.
199. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1195-1201.
200. Ahn, S.-H., M. Kim, and S. Buratowski, *Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing*. Mol. Cell, 2004. **13**: p. 67-76.
201. Morgan, T., *The relation of genetics to physiology and medicine*. Nobel Lecture, 1934.
202. Morange, M., *The scientific legacy of Jacques Monod*. Res. Microbiol., 2010. **161**: p. 77-81.
203. Ullmann, A., *Jacques Monod, 1910-1976: his life, his work and his commitments*. Res. Microbiol., 2010. **161**: p. 68-73.
204. Monod, J., *From enzymatic adaptation to allosteric transitions*. Science, 1966. **154**: p. 475-83.
205. Hurwitz, J., *The discovery of RNA polymerase*. J. Biol. Chem., 2005. **280**: p. 42477-85.
206. Roeder, R.G. and W.J. Rutter, *Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms*. Nature, 1969. **224**: p. 234-7.
207. Kedinger, C., et al., *Alpha-amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus*. Biochem. Biophys. Res. Commun., 1970. **38**: p. 165-71.
208. Werner, M., P. Thuriaux, and J. Soutourina, *Structure-function analysis of RNA polymerases I and III*. Curr. Opin. Struct. Biol., 2009. **19**: p. 740-5.
209. Young, R.A., *RNA polymerase II*. Annu. Rev. Biochem., 1991. **60**: p. 689-715.
210. Kornberg, R.D., *The molecular basis of eukaryotic transcription*. Proc. Natl. Acad. Sci. U. S. A., 2007. **104**: p. 12955-61.

211. Cramer, P., *Architecture of RNA Polymerase II and Implications for the Transcription Mechanism*. Science, 2000. **288**: p. 640-649.
212. Bushnell, D.A. and R.D. Kornberg, *Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription*. Proc. Natl. Acad. Sci. U. S. A., 2003. **100**: p. 6969-73.
213. Orlicky, S.M., et al., *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., 2001. **276**: p. 10097-102.
214. Feig, M. and Z.F. Burton, *RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation*. Biophys. J., 2010. **99**: p. 2577-86.
215. Gnatt, A.L., et al., *Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution*. Science, 2001. **292**: p. 1876-82.
216. Kireeva, M., M. Kashlev, and Z.F. Burton, *Translocation by multi-subunit RNA polymerases*. Biochim. Biophys. Acta, 2010. **1799**: p. 389-401.
217. Kaplan, C.D., K.-M. Larsson, and R.D. Kornberg, *The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin*. Mol. Cell, 2008. **30**: p. 547-56.
218. Wang, D., et al., *Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis*. Cell, 2006. **127**: p. 941-54.
219. Bushnell, D.A., P. Cramer, and R.D. Kornberg, *Structural basis of transcription: alpha-amanitin-RNA polymerase II cocystal at 2.8 Å resolution*. Proc. Natl. Acad. Sci. U. S. A., 2002. **99**: p. 1218-22.
220. Meinhart, A., et al., *A structural perspective of CTD function*. Genes Dev., 2005. **19**: p. 1401-15.
221. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD*. Trends Genet., 2008. **24**: p. 289-96.
222. Palancade, B. and O. Bensaude, *Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation*. Eur. J. Biochem., 2003. **270**: p. 3859-70.

223. West, M.L. and J.L. Corden, *Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations*. Genetics, 1995. **140**: p. 1223-33.
224. Woychik, N.A. and M. Hampsey, *The RNA polymerase II machinery: structure illuminates function*. Cell, 2002. **108**: p. 453-63.
225. Chapman, R.D., et al., *The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability*. Nucleic Acids Res., 2004. **32**: p. 35-44.
226. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution*. Science, 2001. **292**: p. 1863-76.
227. Corden, J.L., *Transcription. Seven ups the code*. Science, 2007. **318**: p. 1735-6.
228. Wu, X., et al., *The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery*. EMBO J., 2000. **19**: p. 3727-38.
229. Sims, R.J., et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation*. Science, 2011. **332**: p. 99-103.
230. Stiller, J.W. and M.S. Cook, *Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs*. Eukaryot. Cell, 2004. **3**: p. 735-40.
231. Meinhart, A. and P. Cramer, *Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors*. Nature, 2004. **430**: p. 223-6.
232. Fong, N. and D.L. Bentley, *Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD*. Genes Dev., 2001. **15**: p. 1783-95.
233. Hampsey, M., *Molecular genetics of the RNA polymerase II general transcriptional machinery*. Microbiol. Mol. Biol. Rev., 1998. **62**: p. 465-503.
234. Ranish, J.A., N. Yudkovsky, and S. Hahn, *Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme*

- recruitment and a postrecruitment role for the TATA box and TFIIB*. *Genes Dev.*, 1999. **13**: p. 49-63.
235. Orphanides, G., T. Lagrange, and D. Reinberg, *The general transcription factors of RNA polymerase II*. *Genes Dev.*, 1996. **10**: p. 2657-83.
236. Basehoar, A.D., S.J. Zanton, and B.F. Pugh, *Identification and distinct regulation of yeast TATA box-containing genes*. *Cell*, 2004. **116**: p. 699-709.
237. Bhaumik, S.R., *Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID*. *Biochim. Biophys. Acta*, 2010. **1809**: p. 97-108.
238. Conaway, R.C. and J.W. Conaway, *Function and regulation of the Mediator complex*. *Curr. Opin. Genet. Dev.*, 2011. **21**: p. 225-30.
239. Yatherajam, G., et al., *Protein-protein interaction map for yeast TFIID*. *Nucleic Acids Res.*, 2003. **31**: p. 1252-60.
240. Sanders, S.L., K.A. Garbett, and P.A. Weil, *Molecular Characterization of Saccharomyces cerevisiae TFIID*. *Mol. Cell. Biol.*, 2002. **22**(16): p. 6000-6013.
241. Thomas, M.C. and C.-M. Chiang, *The general transcription machinery and general cofactors*. *Crit. Rev. Biochem. Mol. Biol.*, 2006. **41**: p. 105-78.
242. Tora, L., *A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription*. *Genes Dev.*, 2002. **16**: p. 673-5.
243. Verrijzer, C.P., et al., *Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II*. *Cell*, 1995. **81**: p. 1115-25.
244. Sauer, F. and R. Tjian, *Mechanisms of transcriptional activation: differences and similarities between yeast, Drosophila, and man*. *Curr. Opin. Genet. Dev.*, 1997. **7**: p. 176-81.
245. Grant, P.A., et al., *A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation*. *Cell*, 1998. **94**: p. 45-53.

246. Shen, W.-C., et al., *Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly*. EMBO J., 2003. **22**: p. 3395-402.
247. Merino, A., et al., *DNA topoisomerase I is involved in both repression and activation of transcription*. Nature, 1993. **365**(6443): p. 227-32.
248. Ma, D., et al., *Separation of the transcriptional coactivator and antirepression functions of transcription factor IIA*. Proc. Natl. Acad. Sci. U. S. A., 1996. **93**(13): p. 6583-8.
249. Auble, D.T., et al., *Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism*. Genes Dev., 1994. **8**(16): p. 1920-34.
250. Inostroza, J.A., et al., *Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription*. Cell, 1992. **70**(3): p. 477-89.
251. Deng, W. and S.G.E. Roberts, *TFIIB and the regulation of transcription by RNA polymerase II*. Chromosoma, 2007. **116**: p. 417-29.
252. Lagrange, T., et al., *New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB*. Genes Dev., 1998. **12**(1): p. 34-44.
253. Singh, B.N. and M. Hampsey, *A transcription-independent role for TFIIB in gene looping*. Mol. Cell, 2007. **27**: p. 806-16.
254. Flores, O., et al., *The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex*. Proc. Natl. Acad. Sci. U. S. A., 1991. **88**: p. 9999-10003.
255. Robert, F., et al., *Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIIF*. Mol. Cell, 1998. **2**: p. 341-51.
256. Nguyen, B.D., et al., *NMR structure of a complex containing the TFIIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCPI*. Proc. Natl. Acad. Sci. U. S. A., 2003. **100**: p. 5688-93.

257. Zawel, L., K.P. Kumar, and D. Reinberg, *Recycling of the general transcription factors during RNA polymerase II transcription*. Genes Dev., 1995. **9**: p. 1479-1490.
258. Conaway, J.W. and R.C. Conaway, *A multisubunit transcription factor essential for accurate initiation by RNA polymerase II*. J. Biol. Chem., 1989. **264**: p. 2357-62.
259. Kumar, K.P., S. Akoulitchev, and D. Reinberg, *Promoter-proximal stalling results from the inability to recruit transcription factor IIIH to the transcription complex and is a regulated event*. Proc. Natl. Acad. Sci. U. S. A., 1998. **95**: p. 9767-72.
260. Kugel, J.F. and J.A. Goodrich, *Promoter escape limits the rate of RNA polymerase II transcription and is enhanced by TFIIE, TFIIH, and ATP on negatively supercoiled DNA*. Proc. Natl. Acad. Sci. U. S. A., 1998. **95**: p. 9232-7.
261. Flores, O., I. Ha, and D. Reinberg, *Factors involved in specific transcription by mammalian RNA polymerase II. Purification and subunit composition of transcription factor IIF*. J. Biol. Chem., 1990. **265**(10): p. 5629-34.
262. Goodrich, J.A. and R. Tjian, *Transcription factors IIE and IIIH and ATP hydrolysis direct promoter clearance by RNA polymerase II*. Cell, 1994. **77**: p. 145-56.
263. Egly, J.-M. and F. Coin, *A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor*. DNA Repair, 2011.
264. Akhtar, M.S., et al., *TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II*. Mol. Cell, 2009. **34**: p. 387-93.
265. Schaeffer, L., et al., *DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor*. Science, 1993. **260**(5104): p. 58-63.
266. Schaeffer, L., et al., *The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor*. EMBO J., 1994. **13**(10): p. 2388-92.
267. Tirode, F., et al., *Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7*. Mol. Cell, 1999. **3**: p. 87-95.

268. Schultz, P., et al., *Molecular structure of human TFIIH*. Cell, 2000. **102**: p. 599-607.
269. Zurita, M. and C. Merino, *The transcriptional complexity of the TFIIH complex*. Trends Genet., 2003. **19**: p. 578-84.
270. Guzder, S.N., et al., *RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription*. Nature, 1994. **369**(6481): p. 578-81.
271. Guzder, S.N., et al., *DNA repair gene RAD3 of S. cerevisiae is essential for transcription by RNA polymerase II*. Nature, 1994. **367**(6458): p. 91-4.
272. Takagi, Y., et al., *Revised subunit structure of yeast transcription factor IIIH (TFIIH) and reconciliation with human TFIIH*. J. Biol. Chem., 2003. **278**: p. 43897-900.
273. Chen, J., et al., *Xpd/Ercc2 regulates CAK activity and mitotic progression*. Nature, 2003. **424**(6945): p. 228-32.
274. Adamczewski, J.P., et al., *MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH*. EMBO J., 1996. **15**(8): p. 1877-84.
275. Rodríguez-Navarro, S., *Insights into SAGA function during gene expression*. EMBO Rep., 2009. **10**: p. 843-50.
276. Daniel, J.A. and P.A. Grant, *Multi-tasking on chromatin with the SAGA coactivator complexes*. Mutat. Res., 2007. **618**: p. 135-48.
277. Wu, P.-Y.J., et al., *Molecular architecture of the S. cerevisiae SAGA complex*. Mol. Cell, 2004. **15**: p. 199-208.
278. Timmers, H.T.M. and L. Tora, *SAGA unveiled*. Trends Biochem. Sci., 2005. **30**: p. 7-10.
279. Sato, S., et al., *A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology*. Mol. Cell, 2004. **14**: p. 685-91.
280. Chadick, J.Z. and F.J. Asturias, *Structure of eukaryotic Mediator complexes*. Trends Biochem. Sci., 2005. **30**: p. 264-71.
281. Takagi, Y. and R.D. Kornberg, *Mediator as a general transcription factor*. J. Biol. Chem., 2006. **281**: p. 80-9.

282. Thompson, C.M. and R.A. Young, *General requirement for RNA polymerase II holoenzymes in vivo*. Proc. Natl. Acad. Sci. U. S. A., 1995. **92**: p. 4587-90.
283. Takahashi, H., et al., *Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors*. Cell, 2011. **146**: p. 92-104.
284. Galbraith, M.D., A.J. Donner, and J.M. Espinosa, *CDK8: A positive regulator of transcription*. Transcription, 2010. **1**: p. 4-12.
285. Hieb, A.R., et al., *An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation*. EMBO J., 2006. **25**: p. 3100-9.
286. Pokholok, D.K., N.M. Hannett, and R.A. Young, *Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo*. Mol. Cell, 2002. **9**: p. 799-809.
287. Holstege, F.C.P., P.C. van Der Vliet, and H.T. Timmers, *Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH*. EMBO J., 1996. **15**: p. 1666-77.
288. Pal, M., A.S. Ponticelli, and D.S. Luse, *The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II*. Mol. Cell, 2005. **19**: p. 101-10.
289. Nechaev, S. and K. Adelman, *Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation*. Biochim. Biophys. Acta, 2011. **1809**: p. 34-45.
290. Gilchrist, D.A., et al., *NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly*. Genes Dev., 2008. **22**: p. 1921-33.
291. Wu, C.-H., et al., *NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila*. Genes Dev., 2003. **17**: p. 1402-14.
292. Martinez-Rucobo, F.W., et al., *Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity*. EMBO J., 2011: p. 1-9.

293. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs*. *Genes Dev.*, 1998. **12**: p. 343-56.
294. Winston, F., et al., *Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae*. *Genetics*, 1984. **107**: p. 179-97.
295. Winston, F., et al., *Three genes are required for trans-activation of Ty transcription in yeast*. *Genetics*, 1987. **115**: p. 649-56.
296. Fassler, J.S. and F. Winston, *Isolation and analysis of a novel class of suppressor of Ty insertion mutations in Saccharomyces cerevisiae*. *Genetics*, 1988. **118**: p. 203-12.
297. Formosa, T., *FACT and the reorganized nucleosome*. *Mol. Biosyst.*, 2008. **4**: p. 1085-93.
298. Hirtreiter, A., et al., *Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif*. *Nucleic Acids Res.*, 2010. **38**: p. 4040-4051.
299. Liu, Y., et al., *Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex*. *Mol. Cell. Biol.*, 2009. **29**: p. 4852-63.
300. Zhou, K., et al., *Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5*. *Proc. Natl. Acad. Sci. U. S. A.*, 2009. **106**: p. 6956-61.
301. Sims, R.J., R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it*. *Genes Dev.*, 2004. **18**: p. 2437-68.
302. Kettenberger, H., K.-J. Armache, and P. Cramer, *Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage*. *Cell*, 2003. **114**: p. 347-57.
303. Izban, M.G. and D.S. Luse, *Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates*. *J. Biol. Chem.*, 1992. **267**: p. 13647-55.
304. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II*. *Nature*, 1992. **358**(6388): p. 641-5.

305. Williams, S.K. and J.K. Tyler, *Transcriptional regulation by chromatin disassembly and reassembly*. Curr. Opin. Genet. Dev., 2007. **17**: p. 88-93.
306. Orphanides, G., et al., *FACT, a factor that facilitates transcript elongation through nucleosomes*. Cell, 1998. **92**: p. 105-16.
307. Orphanides, G., et al., *The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins*. Nature, 1999. **400**: p. 284-8.
308. Belotserkovskaya, R., et al., *FACT facilitates transcription-dependent nucleosome alteration*. Science, 2003. **301**: p. 1090-3.
309. Ruone, S., A.R. Rhoades, and T. Formosa, *Multiple Nhp6 molecules are required to recruit Spt16-Pob3 to form yFACT complexes and to reorganize nucleosomes*. J. Biol. Chem., 2003. **278**: p. 45288-95.
310. Stillman, D.J., *Nhp6: a small but powerful effector of chromatin structure in Saccharomyces cerevisiae*. Biochim. Biophys. Acta, 2010. **1799**: p. 175-80.
311. Heo, K., et al., *FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16*. Mol. Cell, 2008. **30**: p. 86-97.
312. Schlesinger, M.B. and T. Formosa, *POB3 is required for both transcription and replication in the yeast Saccharomyces cerevisiae*. Genetics, 2000. **155**: p. 1593-606.
313. Ivanovska, I., et al., *Control of chromatin structure by Spt6: different consequences in coding and regulatory regions*. Mol. Cell. Biol., 2010. **31**: p. 531-541.
314. Bortvin, A. and F. Winston, *Evidence that Spt6p controls chromatin structure by a direct interaction with histones*. Science, 1996. **272**: p. 1473-6.
315. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites*. Science, 2003. **301**: p. 1096-9.
316. Ansari, A. and M. Hampsey, *A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping*. Genes Dev., 2005. **19**: p. 2969-78.

317. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model*. Genes Dev., 2006. **20**: p. 954-65.
318. Kim, M., et al., *Transitions in RNA polymerase II elongation complexes at the 3' ends of genes*. EMBO J., 2004. **23**: p. 354-64.
319. Kim, M., et al., *Distinct pathways for snoRNA and mRNA termination*. Mol. Cell, 2006. **24**: p. 723-34.
320. Jiang, Y. and J.D. Gralla, *Uncoupling of initiation and reinitiation rates during HeLa RNA polymerase II transcription in vitro*. Mol. Cell. Biol., 1993. **13**: p. 4572-7.
321. Yudkovsky, N., J.A. Ranish, and S. Hahn, *A transcription reinitiation intermediate that is stabilized by activator*. Nature, 2000. **408**: p. 225-9.
322. Björklund, S. and C.M. Gustafsson, *The yeast Mediator complex and its regulation*. Trends Biochem. Sci., 2005. **30**: p. 240-4.
323. Mapendano, C.K., et al., *Crosstalk between mRNA 3' end processing and transcription initiation*. Mol. Cell, 2010. **40**: p. 410-22.
324. Kim, W.Y. and M.E. Dahmus, *Purification of RNA polymerase IIO from calf thymus*. J. Biol. Chem., 1988. **263**: p. 18880-5.
325. Payne, J.M., P.J. Laybourn, and M.E. Dahmus, *The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit Ila*. J. Biol. Chem., 1989. **264**(33): p. 19621-9.
326. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II*. Proc. Natl. Acad. Sci. U. S. A., 1985. **82**: p. 7934-8.
327. Laybourn, P.J. and M.E. Dahmus, *Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription*. J. Biol. Chem., 1990. **265**: p. 13165-73.
328. O'Brien, T., et al., *Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation*. Nature, 1994. **370**: p. 75-7.

329. Dahmus, M.E., *Phosphorylation of the C-terminal domain of RNA polymerase II*. Biochim. Biophys. Acta, 1995. **1261**: p. 171-82.
330. Komarnitsky, P., E.-J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription*. Genes Dev., 2000. **14**: p. 2452-60.
331. Buratowski, S., *The CTD code*. Nat. Struct. Biol., 2003. **10**: p. 679-80.
332. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7*. Science, 2007. **318**: p. 1780-2.
333. Kim, M., et al., *Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7*. J. Biol. Chem., 2009. **284**: p. 26421-6.
334. Bregman, D.B., et al., *Cytostellin distributes to nuclear regions enriched with splicing factors*. J. Cell Sci., 1994. **107 (Pt 3)**: p. 387-96.
335. Warren, S.L., et al., *Cytostellin: a novel, highly conserved protein that undergoes continuous redistribution during the cell cycle*. J. Cell Sci., 1992. **103 (Pt 2)**: p. 381-8.
336. Cho, E.-J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain*. Genes Dev., 2001. **15**: p. 3319-29.
337. Jones, J.C., et al., *C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats*. J. Biol. Chem., 2004. **279**: p. 24957-64.
338. Lu, H., *The Nonphosphorylated Form of RNA Polymerase II Preferentially Associates with the Preinitiation Complex*. Proceedings of the National Academy of Sciences, 1991. **88**: p. 10004-10008.
339. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex*. Mol. Cell, 2008. **31**: p. 337-46.
340. Max, T., M. Sogaard, and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator*. J. Biol. Chem., 2007. **282**: p. 14113-20.
341. Egloff, S., et al., *Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression*. Science, 2007. **318**: p. 1777-9.

342. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1279-1286.
343. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1154-61.
344. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex*. Nat. Struct. Mol. Biol., 2010. **17**: p. 1272-1278.
345. Buratowski, S., *Progression through the RNA polymerase II CTD cycle*. Mol. Cell, 2009. **36**: p. 541-6.
346. Mosley, A.L., et al., *Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation*. Mol. Cell, 2009. **34**: p. 168-78.
347. Keogh, M.-C., V. Podolny, and S. Buratowski, *Bur1 kinase is required for efficient transcription elongation by RNA polymerase II*. Mol. Cell. Biol., 2003. **23**: p. 7005-18.
348. Prelich, G. and F. Winston, *Mutations that suppress the deletion of an upstream activating sequence in yeast: involvement of a protein kinase and histone H3 in repressing transcription in vivo*. Genetics, 1993. **135**: p. 665-76.
349. Yao, S., A. Neiman, and G. Prelich, *BUR1 and BUR2 encode a divergent cyclin-dependent kinase-cyclin complex important for transcription in vivo*. Mol. Cell. Biol., 2000. **20**: p. 7080-7.
350. Wood, A. and A. Shilatifard, *Bur1/Bur2 and the Ctk complex in yeast: the split personality of mammalian P-TEFb*. Cell Cycle, 2006. **5**: p. 1066-8.
351. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1*. Genes Dev., 2010. **24**: p. 2303-16.
352. Qiu, H., C. Hu, and A.G. Hinnebusch, *Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters*. Mol. Cell, 2009. **33**: p. 752-62.
353. Archambault, J., et al., *An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U. S. A., 1997. **94**: p. 14300-5.

354. Archambault, J., et al., *FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase II*. J. Biol. Chem., 1998. **273**: p. 27593-601.
355. Cho, H., et al., *A protein phosphatase functions to recycle RNA polymerase II*. Genes Dev., 1999. **13**: p. 1540-52.
356. Hausmann, S. and S. Shuman, *Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5*. J. Biol. Chem., 2002. **277**: p. 21213-20.
357. Krishnamurthy, S., et al., *Ssu72 Is an RNA polymerase II CTD phosphatase*. Mol. Cell, 2004. **14**: p. 387-94.
358. Reyes-Reyes, M. and M. Hampsey, *Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation*. Mol. Cell. Biol., 2007. **27**: p. 926-36.
359. Sun, Z.W. and M. Hampsey, *Synthetic enhancement of a TFIIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo*. Mol. Cell. Biol., 1996. **16**: p. 1557-66.
360. Pappas, D.L. and M. Hampsey, *Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in Saccharomyces cerevisiae*. Mol. Cell. Biol., 2000. **20**: p. 8343-51.
361. Dichtl, B., et al., *A role for SSU72 in balancing RNA polymerase II transcription elongation and termination*. Mol. Cell, 2002. **10**: p. 1139-50.
362. Ganem, C., et al., *Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast*. EMBO J., 2003. **22**: p. 1588-98.
363. Gavin, A.-C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes*. Nature, 2002. **415**: p. 141-7.
364. Nedeá, E., et al., *Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends*. J. Biol. Chem., 2003. **278**: p. 33000-10.

365. He, X., et al., *Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1*. *Genes Dev.*, 2003. **17**: p. 1030-42.
366. Krishnamurthy, S., et al., *Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries*. *Mol. Cell. Biol.*, 2009. **29**: p. 2925-34.
367. Werner-Allen, J.W., et al., *Cis proline-mediated pSer5-dephosphorylation by the RNA polymerase II CTD phosphatase Ssu72*. *J. Biol. Chem.*, 2010.
368. Hanes, S.D., P.R. Shank, and K.A. Bostian, *Sequence and mutational analysis of ESS1, a gene essential for growth in Saccharomyces cerevisiae*. *Yeast*, 1989. **5**: p. 55-72.
369. Gemmill, T.R., X. Wu, and S.D. Hanes, *Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in Saccharomyces cerevisiae*. *J. Biol. Chem.*, 2005. **280**: p. 15510-7.
370. Myers, J.K., et al., *Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1*. *Biochemistry*, 2001. **40**: p. 8479-86.
371. Kobor, M.S., et al., *An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae*. *Mol. Cell*, 1999. **4**: p. 55-62.
372. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription*. *Cell*, 2002. **108**: p. 501-12.
373. Egloff, S. and S. Murphy, *Cracking the RNA polymerase II CTD code*. *Trends Genet.*, 2008. **24**: p. 280-8.
374. Fabrega, C., et al., *Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II*. *Mol. Cell*, 2003. **11**: p. 1549-61.
375. Burley, S.K. and N. Sonenberg, *Gimme Phospho-Serine Five! Capping Enzyme Guanylyltransferase Recognition of the RNA Polymerase II CTD*. *Mol. Cell*, 2011. **43**: p. 163-5.

376. Ghosh, A., S. Shuman, and C.D. Lima, *Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code*. Mol. Cell, 2011. **43**: p. 299-310.
377. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II*. Genes Dev., 1997. **11**: p. 3306-18.
378. Schoenberg, D.R. and L.E. Maquat, *Re-capping the message*. Trends Biochem. Sci., 2009. **34**: p. 435-42.
379. Allemand, E., E. Batsché, and C. Muchardt, *Splicing, transcription, and chromatin: a ménage à trois*. Curr. Opin. Genet. Dev., 2008. **18**: p. 145-51.
380. Beyer, A.L. and Y.N. Osheim, *Splice site selection, rate of splicing, and alternative splicing on nascent transcripts*. Genes Dev., 1988. **2**: p. 754-65.
381. Morris, D.P. and A.L. Greenleaf, *The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II*. J. Biol. Chem., 2000. **275**: p. 39935-43.
382. Hirose, Y., R. Tacke, and J.L. Manley, *Phosphorylated RNA polymerase II stimulates pre-mRNA splicing*. Genes Dev., 1999. **13**: p. 1234-9.
383. Millhouse, S. and J.L. Manley, *The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein*. Mol. Cell. Biol., 2005. **25**: p. 533-44.
384. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice*. Nat. Struct. Mol. Biol., 2008. **15**: p. 786-94.
385. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II*. Mol. Cell. Biol., 2003. **23**: p. 4207-18.
386. Hampsey, M. and D. Reinberg, *Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation*. Cell, 2003. **113**: p. 429-32.
387. Ng, H.H., et al., *Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79*. J. Biol. Chem., 2002. **277**: p. 34655-7.

388. Jaehning, J.A., *The Paf1 complex: platform or player in RNA polymerase II transcription?* Biochim. Biophys. Acta, 2010. **1799**: p. 379-88.
389. Drouin, S., et al., *DSIF and RNA Polymerase II CTD Phosphorylation Coordinate the Recruitment of Rpd3S to Actively Transcribed Genes.* PLoS Genet., 2010. **6**: p. e1001173.
390. Dehé, P.-M. and V. Géli, *The multiple faces of Set1.* Biochem. Cell Biol., 2006. **84**: p. 536-48.
391. Sun, Z.-W. and C.D. Allis, *Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast.* Nature, 2002. **418**: p. 104-8.
392. Briggs, S.D., et al., *Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae.* Genes Dev., 2001. **15**: p. 3286-95.
393. Dover, J., et al., *Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6.* J. Biol. Chem., 2002. **277**: p. 28368-71.
394. Laribee, R.N., et al., *BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex.* Curr. Biol., 2005. **15**: p. 1487-93.
395. Sims, R.J., et al., *Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing.* Mol. Cell, 2007. **28**: p. 665-76.
396. Kim, T. and S. Buratowski, *Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions.* Cell, 2009. **137**: p. 259-72.
397. Pijnappel, W.W., et al., *The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program.* Genes Dev., 2001. **15**: p. 2991-3004.
398. Ginsburg, D.S., C.K. Govind, and A.G. Hinnebusch, *NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5.* Mol. Cell. Biol., 2009. **29**: p. 6473-87.

399. Strahl, B.D., et al., *Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression*. Mol. Cell. Biol., 2002. **22**: p. 1298-306.
400. Xiao, T., et al., *Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast*. Genes Dev., 2003. **17**: p. 654-63.
401. Li, B., et al., *The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II*. J. Biol. Chem., 2003. **278**: p. 8897-903.
402. Li, B., et al., *Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription*. J. Biol. Chem., 2009. **284**: p. 7970-6.
403. Bannister, A.J., et al., *Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes*. J. Biol. Chem., 2005. **280**: p. 17732-6.
404. Zhang, L., et al., *Altered nucleosome occupancy and histone H3K4 methylation in response to 'transcriptional stress'*. EMBO J., 2005. **24**: p. 2379-90.
405. Xiao, T., et al., *Histone H2B ubiquitylation is associated with elongating RNA polymerase II*. Mol. Cell. Biol., 2005. **25**: p. 637-51.
406. Wirén, M., et al., *Genomewide analysis of nucleosome density histone acetylation and HDAC function in fission yeast*. EMBO J., 2005. **24**: p. 2906-18.
407. Dion, M.F., et al., *Genomic characterization reveals a simple histone H4 acetylation code*. Proc. Natl. Acad. Sci. U. S. A., 2005. **102**: p. 5501-6.
408. Stock, J.K., et al., *Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells*. Nat. Cell Biol., 2007. **9**: p. 1428-35.
409. Zacharioudakis, I., T. Gligoris, and D. Tzamarias, *A yeast catabolic enzyme controls transcriptional memory*. Curr. Biol., 2007. **17**: p. 2041-6.
410. Halley, J.E., et al., *Roles for H2A.Z and Its Acetylation in GAL1 Transcription and Gene Induction, but Not GAL1-Transcriptional Memory*. PLoS Biol., 2010. **8**: p. e1000401.

411. Kundu, S. and C.L. Peterson, *Dominant role for signal transduction in transcriptional memory of yeast GAL genes*. Mol. Cell. Biol., 2010. **30**: p. 2330-2340.
412. Tan-Wong, S.M., H.D. Wijayatilake, and N.J. Proudfoot, *Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex*. Genes Dev., 2009. **23**: p. 2610-24.
413. Lainé, J.-P., et al., *A physiological role for gene loops in yeast*. Genes Dev., 2009. **23**: p. 2604-9.
414. Kawano, A., et al., *Global analysis for functional residues of histone variant Htz1 using the comprehensive point mutant library*. Genes Cells, 2011. **16**: p. 590-607.
415. Storici, F. and M.A. Resnick, *The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast*. Methods in Enzymology, 2006. **409**: p. 329-45.
416. Guo, Z. and J.W. Stiller, *Comparative genomics of cyclin-dependent kinases suggest co-evolution of the RNAP II C-terminal domain and CTD-directed CDKs*. BMC Genomics, 2004. **5**: p. 69.

