



**HAL**  
open science

**Etude biochimique, structurale et fonctionnelle du  
complexe chaperonne d'histone/facteur d'élongation  
Spt6/Iws1**

Marie-Laure Diebold

► **To cite this version:**

Marie-Laure Diebold. Etude biochimique, structurale et fonctionnelle du complexe chaperonne d'histone/facteur d'élongation Spt6/Iws1. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Strasbourg, 2012. Français. NNT : 2012STRAJ008 . tel-00819618

**HAL Id: tel-00819618**

**<https://theses.hal.science/tel-00819618>**

Submitted on 2 May 2013

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

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



# Université de Strasbourg

Ecole doctorale des Sciences de la Vie et de la Santé

## Thèse

Présentée pour l'obtention du grade de :

**DOCTEUR DE L'UNIVERSITE DE STRASBOURG**

Discipline : Sciences du Vivant

Mention : Aspects Moléculaires et Cellulaires de la Biologie

Par

DIEBOLD Marie-Laure

### **ETUDE BIOCHIMIQUE, STRUCTURALE ET FONCTIONNELLE DU COMPLEXE CHAPERONNE D'HISTONE/FACTEUR D'ELONGATION SPT6/IWS1.**

Soutenue le 26 mars 2012 devant un jury composé de :

Dr. Olivier BENSAUDE	Rapporteur Externe
Dr. Yves BOURNE	Rapporteur Externe
Dr. Pascale ROMBY	Examineur
Dr. Christophe ROMIER	Directeur de thèse

Thèse préparée au sein de l'équipe de Biologie Structurale des Cibles Epigénétiques,  
du Département de Biologie Structurale Intégrative,  
à l'Institut de Génétique et Biologie Moléculaire et Cellulaire



## Remerciements

Je tiens tout d'abord à remercier les Drs. Pascale Romby, Olivier Bensaude et Yves Bourne d'avoir accepté de juger ce travail.

Comment ne pas remercier le Dr. Christophe Romier qui a accepté de me former et de m'encadrer pendant ces quatre années de thèse (plus les stages bien sûr !). Je ne sais pas si un « immense merci » suffit pour te remercier de ta confiance, ton soutien, tes conseils et ta bonne humeur dispensés depuis plusieurs années maintenant. Cela a réellement été un grand plaisir de travailler avec toi !!!

Un grand merci également au Pr. Jean Cavarelli d'avoir accepté de m'accueillir dans son équipe. Merci pour les nombreuses discussions qui nous ont permis de faire avancer ce projet, pour tous vos conseils cristallographiques et merci d'avoir sû, il y a maintenant sept ans, me transmettre votre intérêt pour la cristallographie.

Merci également à nos collaborateurs outre Atlantique sans qui ce projet n'aurait pas eu toute la portée qu'il a maintenant : Le Dr Fred Winston et Erin Loeliger à la Harvard Medical School de Boston. Thank you so much for this nice collaboration, I really enjoyed working with you (even if it was only through mails and post!).

Au cours de ma thèse j'ai eu la chance de pouvoir interagir avec un très grand nombre de personnes qui nous ont permis d'élargir notre champ de recherche et je tiens à les remercier ici : le Dr. Anne-Catherine Dock-Bregeon, le Dr. Eric Ennifar, le Dr. Ali Hamiche et toute son équipe, le Dr. Bruno Kieffer et le Dr. Christian Koehler, le Dr. Bernardo Reina et Ebe Shiavo, le Dr. Pascal Romby et toute son équipe, le Dr. Bertrand Séraphin et Céline Faux, le Dr. Marc Vigneron et le Dr Michel Werner.

Un très grand merci à tous les membres de l'équipe Cavarelli pour la bonne ambiance, les discussions scientifiques et les simples papotages au cours des nombreux goûters que nous avons organisés. J'ai également beaucoup appris de vous. Merci à Vincent, pour les voyages synchrotron, à Nathalie, pour les Western Blot (entre autre choses), à Jean-Marie, à Luc et à tous mes collègues thésards et post-doc : Martin bien sur, Morgane, Magdalena, Justine, Stéphanie et Arnaud. Je n'oublie pas non plus ceux qui sont partis vers d'autres horizons : Michael, avec qui j'ai débuté dans le labo et sur ce projet, Pierre, pour sa constante bonne

humeur, et bien sur Emeline, pour notre amitié (les ateliers cuisine, nos papotages et nos fous rires). Je me rends bien compte d'avoir été un peu trop maniaque parfois, je m'en excuse.

Merci à tous les membres de la plateforme de Biologie Structurale. Merci tout particulièrement à Alastair et Pierre pour leur aide lors de la réalisation de mes nombreux tests de cristallisation. Merci également à Edouard, Loubna et Didier pour les discussions sur les clonages récalcitrants et pour m'avoir laissé utiliser votre robot pour la réalisation des milliers de tests d'expression que j'ai bien dû réaliser ces quatre dernières années.

Merci également à tous les membres du département de Biologie Structurale Intégrative et notamment les personnes en charge de ce département : Patrick Schultz et Dino Moras. Je ne prendrais pas le risque de citer tous les autres de peur d'oublier quelqu'un, mais merci à chacun de vous pour les simples discussions de tous les jours.

Merci aussi à ma famille, sans vous tout ceci n'aurait pas été possible. Merci de tout cœur à mes parents, merci pour vos encouragements et votre soutien, à ma petite sœur, ma grand-mère et tous les (nombreux) autres.

Merci à mon Alex, pour sa patience (et il en faut !), son soutien, ses petits plats et simplement d'être à mes côtés tous les jours. Merci également à sa maman, Christine et à Jean-François pour leur gentillesse.

Mon parcours a commencé par un stage au LCM3B de Nancy (maintenant CRM2A). Là-bas, j'ai rencontré Fred (le Dr. Frédérique Tête-Favier) qui m'a encouragée à faire une thèse et sans qui je ne me serai probablement pas ré-orientée. Merci de m'avoir fait changer d'avis Fred, ça en valait le coup ! Merci également à toutes les personnes qui m'ont poussé, au cours de ma scolarité à choisir cette voie, pas si simple finalement.

# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Table des matières</b>	<b>5</b>
<b>Liste des abréviations</b>	<b>13</b>
<b>Préambule</b>	<b>15</b>
<b>Chapitre 1 Introduction</b>	<b>19</b>
<b>1.1. Production d'un ARN messager fonctionnel : mécanismes de la transcription eucaryote</b>	<b>21</b>
1.1.1. Mécanismes de la transcription eucaryote	21
1.1.1.1. L'initiation de la transcription - Formation du complexe de pré-initiation	22
a) Reconnaissance des éléments régulateurs et du promoteur – implication des activateurs	22
b) Les différents complexes co-activateurs	23
c) Assemblage du complexe de pré-initiation	26
1.1.1.2. L'ARN polymérase II	28
1.1.1.3. Fin de l'initiation de la transcription – Echappée du promoteur	29
1.1.1.4. L'élongation de la transcription	30
a) Régulation de la phase d'élongation : « promoter pausing »	30
b) Entrée en phase d'élongation processive	32
c) Autres complexes impliqués dans la régulation de l'élongation	32
1.1.1.5. Mécanismes de terminaison	33
1.1.2. Implication du CTD de l'ARN polymérase II dans la régulation de la transcription : le code CTD	34
1.1.3. Maturation et export des ARN messagers - Mécanismes co-transcriptionnels	36
1.1.3.1. Maturation des ARN messagers	36
1.1.3.2. Export des ARN messagers	37
<b>1.2. Structure de la chromatine et implication dans la régulation de la transcription</b>	<b>38</b>
1.2.1. Les protéines histones et le nucléosome	39
1.2.1.1. Les protéines histones	40
1.2.1.2. Structure du nucléosome	41
1.2.1.3. Le positionnement des nucléosomes sur l'ADN	43
1.2.2. La régulation épigénétique de la transcription	44
1.2.2.1. La modification des histones	45

1.2.2.1. Les variants d'histones _____	47
1.2.3. Facteurs de remodelage et chaperonnes d'histone _____	49
1.2.3.1. Les facteurs de remodelage de la chromatine _____	49
1.2.3.2. Les protéines chaperonnes d'histones _____	51
<b>1.3. Fonctions des protéines Spt6, Iws1 et de leur complexe _____</b>	<b>55</b>
1.3.1. Identification des protéines Spt6 et Iws1 chez la levure _____	55
1.3.1.1. Les protéines de la famille SPT _____	55
1.3.1.2. Identification de Iws1 et du complexe Spt6/Iws1 _____	57
1.3.2. Rôle de Spt6 et Iws1 dans la modulation de la structure de la chromatine. _____	57
1.3.2.1. Rôle de chaperonne d'histone – interaction avec l'histone H3 _____	57
1.3.2.2. Implication des facteurs de remodelage de la chromatine _____	58
1.3.2.3. Spt6 est impliquée dans la reconstitution des nucléosomes _____	59
1.3.3. Rôle de Spt6 et Iws1 dans l'élongation de la transcription _____	61
1.3.3.1. Colocalisation et interaction de Spt6 et Iws1 avec l'ARN polymérase II hyperphosphorylée _____	61
1.3.3.2. Spt6 et Iws1 participent et stimulent la phase d'élongation de la transcription _____	63
1.3.3.3. Implication des facteurs épigénétiques dans la régulation transcriptionnelle induite par Spt6 et Iws1 _____	64
1.3.4. Le complexe Spt6/Iws1 est impliqué dans les mécanismes co- et post-transcriptionnels _____	65
1.3.5. Autres fonctions de Spt6 _____	67
<b>1.4. But de l'étude – méthodologie employée _____</b>	<b>67</b>
<b>Chapitre 2 Matériel et Méthodes _____</b>	<b>71</b>
<b>2.1. Techniques de biologie moléculaire _____</b>	<b>73</b>
2.1.1. Génération des vecteurs d'expression bactériens _____	73
2.1.1.1. 1 <sup>ère</sup> étape : la PCR et la génération des vecteurs navettes _____	73
a) Réaction de polymérisation en chaîne (PCR) _____	74
b) Insertion dans le vecteur navette _____	74
2.1.1.2. 2 <sup>ème</sup> étape : le sous-clonage en vecteurs d'expression _____	77
a) Les vecteurs d'expression bactériens de la suite pET-MCN _____	77
b) Le vecteur pGexNB _____	78
c) Sous-clonage par restriction/ligation _____	79
d) Sous-clonage par recombinaison homologue _____	79
2.1.2. Concaténation des vecteurs pour l'expression de complexes chez la bactérie _____	80
2.1.3. Mutagénèse dirigée _____	81
2.1.3.1. Mutagénèse par PCR emboîtée (Nested PCR) _____	82
2.1.3.2. Mutagénèse par rolling-circle _____	82

2.1.4. Génération de bacmides pour l'expression en cellules d'insectes	83
2.1.4.1. Principe de l'expression en cellules d'insectes	83
2.1.4.2. Génération des Bacmides grâce au système Bac to Bac®	83
<b>2.2. Tests d'expression, production et purification de protéines et complexes protéiques</b>	<b>85</b>
2.2.1. Le système d'expression bactérien pET et son dérivé, le système pET-MCN	85
2.2.2. Les milieux de culture bactérien	87
2.2.2.1. Milieu complet (2LB)	88
2.2.2.2. Milieu Graffinity®	88
2.2.2.3. Milieu Auto-induit	89
2.2.2.4. Milieu minimum N15-C13	89
2.2.3. Tests d'expression	90
2.2.3.1. Réalisation des tests d'expression en bactéries	90
a) Transformation des bactéries	90
b) Culture des bactéries en milieu liquide	92
c) Réalisation des tests d'expression	93
2.2.3.2. Interprétation des résultats des tests d'expression	94
a) Les tests d'expression d'une protéine seule	94
b) Les tests de co-expression	96
2.2.4. Production des protéines en grand volume dans un système exogène	97
2.2.4.1. Production en bactéries	98
2.2.4.2. Production en cellules d'insectes	99
2.2.5. Purification des protéines	99
2.2.5.1. 1 <sup>ère</sup> étape : la lyse des cellules	100
2.2.5.2. Chromatographie par affinité	100
a) Fixation de la protéine à la résine et lavages	101
b) Elution de la protéine	101
c) Elution par coupure de l'étiquette de purification	102
2.2.5.3. Chromatographie d'exclusion ou filtration sur gel	103
2.2.5.4. Concentration de l'échantillon et estimation de sa concentration	104
<b>2.3. Caractérisation physico-chimique et fonctionnelle de la protéine</b>	<b>105</b>
2.3.1. Caractérisation physico-chimique	105
2.3.1.1. La diffusion dynamique de lumière (DLS)	105
2.3.1.2. Microcalorimétrie (ITC)	106
2.3.1.3. Fluorescence Thermal Shift Assay (Thermofluor)	108
2.3.2. Caractérisations fonctionnelles	109
2.3.2.1. GST pulldown	109



2.3.2.2. Etudes de génétique de levure _____	110
<b>2.4. Etudes structurales par cristallographie aux Rayons X et résonance magnétique nucléaire</b>	<b>111</b>
2.4.1. Résolution de structure par cristallographie aux rayons X _____	111
2.4.1.1. Principe _____	112
2.4.1.2. Cristallisation et collecte des jeux de données _____	113
a) La cristallisation de l'échantillon protéique _____	113
b) Congélation des cristaux et collecte des jeux de données _____	116
2.4.1.3. Traitement des données _____	117
2.4.1.4. Résolution du problème des phases _____	117
a) Le remplacement moléculaire _____	117
b) Phasage par la technique MIRAS _____	118
2.4.1.5. Construction, affinement et validation de la structure _____	120
a) Construction et affinement de la structure _____	120
b) Validation du modèle _____	121
2.4.2. Résolution de structure par résonance magnétique nucléaire _____	122
2.4.2.1. Principe _____	122
a) La résonance magnétique nucléaire _____	122
b) Application à la résolution de structures de protéines _____	124
2.4.2.2. Préparation de l'échantillon et enregistrement des spectres _____	124
2.4.2.3. Stratégie d'assignement _____	126
a) Assignement séquentiel _____	126
b) Assignement des chaînes latérales _____	128
2.4.2.4. Calcul de la structure _____	128
a) Identification des effets NOE et calcul de la structure _____	128
b) Identification des éléments de structure secondaire _____	129
2.4.2.5. Temps de relaxation T1/T2 – dynamique moléculaire _____	129
 <b>Chapitre 3 Etude biochimique, structurale et fonctionnelle du domaine carboxy-terminal de Spt6 et de son recrutement par l'ARN polymérase II _____</b>	 <b>131</b>
<b>3.1. Introduction _____</b>	<b>133</b>
3.1.1. Rôles et architecture des domaines SH2. _____	133
3.1.1.1. Rôle des domaines SH2 dans la signalisation cellulaire _____	133
3.1.1.2. Architecture des domaines SH2 _____	134
3.1.1.3. Interaction des domaines SH2 avec les phosphotyrosines _____	135
3.1.1.4. Poche de spécificité de la reconnaissance _____	137
3.1.2. Le domaine SH2 de Spt6 – intérêt de son étude _____	137

<b>3.2. Brève description des résultats</b>	<b>138</b>
<b>3.3. Publication numéro 1</b>	<b>141</b>
<b>3.4. Etude des interactions du tandem SH2 de la protéine Spt6</b>	<b>152</b>
3.4.1. Tests de l'interaction entre le tandem SH2 et le CTD de l'ARN polymérase II par co-expression	152
3.4.2. Test de l'interaction du tandem SH2 avec des peptides portant des phosphotyrosines	155
3.4.3. Interaction avec la protéine Spt5	155
3.4.3.1. Test de l'interaction Spt6/Spt5 par GST pulldown	156
3.4.3.2. Test de l'interaction Spt6/Spt5 par co-expression	157
<b>3.5. Discussion</b>	<b>160</b>
3.5.1. Comparaison des structures du domaine carboxy-terminal de Spt6 d' <i>A. locustae</i> , de <i>C. glabrata</i> et de <i>S. cerevisiae</i>	160
3.5.2. Interaction des tandems SH2 avec des peptides phosphorylés	162
3.5.3. Modèle de l'interaction entre le tandem SH2 de Spt6 et le CTD de l'ARN polymérase II	164
3.5.4. Structure des domaines SH2 et interaction avec le CTD de l'ARN polymérase - Conclusion	166
3.5.5. Interaction du tandem SH2 de Spt6 avec Spt5	166
<b>Chapitre 4 Etude du complexe Spt6/lws1 impliqué dans l'export des ARN messagers</b>	<b>169</b>
<b>4.1. Introduction</b>	<b>171</b>
<b>4.2. Brève description des résultats</b>	<b>171</b>
<b>4.3. Publication numéro 2</b>	<b>173</b>
<b>4.4. Etude des complexes formés par Spt6 et lws1</b>	<b>187</b>
4.4.1. Complexe Spt6/lws1 – mutants de surface de lws1	187
4.4.2. Interaction de lws1 avec l'ARN polymérase II	189
4.4.2.1. Test de l'interaction avec l'ARN polymérase II par pulldown	190
4.4.2.2. Interaction de lws1 avec des peptides phosphorylés mimant le CTD de l'ARN polymérase II	191
4.4.3. Interaction de lws1 avec l'histone méthyltransférase Set2 et le facteur d'export Yra1	193
<b>4.5. Discussion</b>	<b>194</b>
4.5.1. Comparaison des structures de lws1 et du complexe Spt6/lws1	194
4.5.2. Caractérisation fonctionnelle – expression de mutants des protéines lws1 et Spt6	197
4.5.3. Interaction de lws1 avec l'ARN polymérase II	199
<b>Chapitre 5 Caractérisation biochimique, fonctionnelle et structurale de l'extrémité amino-terminale de Spt6 et des complexes Spt6/TFIIS, Spt8/TFIIS et Med13/TFIIS</b>	<b>201</b>
<b>5.1. Introduction – brève description des résultats</b>	<b>203</b>

<b>5.2. L'extrémité amino-terminale de Spt6 contient plusieurs motifs hydrophobes conservés</b>	<b>204</b>
<b>5.3. Etude des interactions réalisées par l'extrémité amino-terminale de Spt6</b>	<b>206</b>
5.3.1. Iws1 n'interagit qu'avec le motif IF5 de Spt6	206
5.3.2. Interaction avec TFIIIS	208
5.3.2.1. TFIIIS interagit avec deux motifs hydrophobes conservés de Spt6	208
5.3.2.2. Spécificité de l'interaction Spt6/TFIIIS	209
5.3.2.3. Caractérisation biochimique des complexes Spt6/TFIIIS	213
<b>5.4. Structure du complexe Spt8/TFIIIS et implications pour les complexes Spt6/TFIIIS</b>	<b>215</b>
5.4.1. Purification et essais de cristallisation des complexes Spt6 IF3/TFIIIS, Spt6 IF5/TFIIIS et Spt8/TFIIIS	215
5.4.2. Résolution de la structure du complexe Spt8/TFIIIS par RMN	216
<b>5.5. Etude fonctionnelle des motifs hydrophobes conservés de l'extrémité amino-terminale de Spt6</b>	<b>220</b>
<b>5.6. Discussion</b>	<b>223</b>
<b>Chapitre 6 Interaction de Spt6 avec les histones – co-expression de l'octamère d'histone</b>	<b>225</b>
<b>6.1. Introduction</b>	<b>227</b>
<b>6.2. Brève description des résultats</b>	<b>227</b>
<b>6.3. Publication numéro 3</b>	<b>228</b>
<b>6.4. Validation de l'octamère d'histone co-exprimé – reconstitution du nucléosome</b>	<b>240</b>
<b>6.5. Etude de l'interaction entre des chaperonnes et les histones par co-expression</b>	<b>241</b>
6.5.1. Interaction de Asf1 avec la paire H3/H4	241
6.5.2. Interaction de Spt6 avec les histones	243
<b>6.6. Conclusions</b>	<b>244</b>
<b>Chapitre 7 Conclusions générales et perspectives</b>	<b>247</b>
<b>Chapitre 8 Annexes</b>	<b>257</b>
<b>8.1. Protocoles de biologie moléculaire</b>	<b>259</b>
8.1.1. La réaction PCR	259
8.1.2. Clonage en vecteur navette par restriction/ligation	260
8.1.3. Sous clonage en vecteur d'expression	261
8.1.3.1. Sous-clonage par restriction / ligation	261

8.1.3.2. Sous-clonage par recombinaison homologue	261
8.1.4. Les souches bactériennes	262
8.1.5. Les vecteurs de co-expression bactériens	262
8.1.6. Vecteurs baculovirus	263
<b>8.2. Données supplémentaires de la publication 1</b>	<b>264</b>
<b>8.3. Préparation des extraits de levure pour la réalisation d'expériences de pulldown</b>	<b>272</b>
<b>8.4. Données supplémentaires de la publication 2</b>	<b>272</b>
<b>8.5. Purification des complexes de TFIIIS et méthylation des lysines</b>	<b>281</b>
<b>8.6. Données supplémentaires de la publication numéro 3</b>	<b>281</b>
<b>8.7. Reconstitution du nucléosome à partir de l'octamère d'histone co-exprimé</b>	<b>288</b>
8.7.1. Préparation de l'ADN Widom 601	288
8.7.2. Reconstitution des nucléosomes	288
<b>8.8. Publication numéro 4</b>	<b>289</b>
<b>Chapitre 9 Références bibliographiques</b>	<b>295</b>
<b>Chapitre 10 Liste des figures et des tableaux</b>	<b>313</b>
<b>Liste des figures</b>	<b>315</b>
<b>Liste des tableaux</b>	<b>319</b>



## Liste des abréviations

<i>A. locustae</i>	:	<i>Antonospora locustae</i>
AcMNPV	:	Autographa californica multiple nuclear polyhedrosis virus
ADN	:	Acide désoxyribonucléique
AID	:	Activation-induced cytidine deaminase
ARM	:	Armadillo
ARN	:	Acide ribonucléique
ARNnc	:	ARN non codant
ATP	:	Adénosine tri-phosphate
BRE	:	TFIIB Recognition Element
BSA	:	Bovine Serum Albumin
<i>C. glabrata</i>	:	<i>Candida glabrata</i>
cAMP	:	Adénosine monophosphate cyclique
CBP	:	Cap Binding Protein
CCD	:	Charge-Coupled Device
CENP-A	:	Centromere Protein A
CF1A/1B	:	Cleavage Factor
CHD	:	Chromodomain
CID	:	CTD Interacting Domain
CPF	:	Cleavage and Polyadenylation Factor
CTD	:	Carboxy-terminal domain
C-terminal	:	Carboxy-terminal
DLS	:	Dynamic Light Scattering
dNTP	:	Desoxy-ribo nucléotides tri-phosphate
DO	:	Densité optique
DPE	:	Distal Promoter Element
DSIF	:	DRB-sensitivity-inducing factor
<i>E. cuniculi</i>	:	<i>Encephalitozoon cuniculi</i>
FACT	:	Facilitates Chromatin Transcription
FRAP	:	Fluorescence Recovery After Photobleaching
GST	:	Glutathion-S-Transferase
<i>H. Sapiens</i>	:	<i>Homo sapiens</i>
HAT	:	Histone Acetyl Transferase
HAT	:	Histone Acetyltransferase
HDAC	:	Histone Deacetylase
HEXIM1	:	Hexamethylene Inducible mRNA
HFD	:	Histone Fold Domain
HSF	:	Heat shock factor
HU	:	Hydroxyurée
INO80	:	Inosito
INR	:	Initiator

IPTG	:	Isopropyl $\beta$ -D-1-thiogalactopyranoside
ISWI	:	Initiation of Switch
ITC	:	Isothermal Titration Calorimetry
IWS1	:	IWS1
LB	:	Lysogeny Broth
MBP	:	Maltose Binding Protein
MIRAS	:	Multiple Isomorphous Replacement and Anomalous Scattering
NELF	:	Negative Elongation Factor
NOE	:	Nuclear Overhauser Effect
N-terminal	:	Amino-terminale
PAF	:	Polymerase Associated Factor
pb	:	Paires de bases
PCR	:	Polymerase Chain Reaction
PEG	:	Polyéthylène Glycol
PHD	:	Pleckstrin Homology Domain
PIC	:	Pre-initiation Complex
PMSF	:	Phenylmethanesulfonylfluoride
ppm	:	Partie par million
pré-ARNm	:	ARN pré-messager
PRMT	:	Protein Arginine Methyltransferase
PTEF-b	:	Positive elongation factor b
RMN	:	Résonnance magnétique nucléaire
rpm	:	Rotation par minute
RSC	:	Remodels Structure of Chromatin
<i>S. cerevisiae</i>	:	<i>Saccharomyces cerevisiae</i>
SAD/MAD	:	Single/Multiple Anomalous Dispersion
SAGA	:	Spt-Ada-Gcn5-Acetyltransferase
SAXS	:	Small Angle X-Ray Scattering
SH2	:	Src Homology 2
SIR/MIR	:	Single/Multiple Isomorphous Replacement
siRNA	:	RNA Silencing
snRNP	:	Small Nuclear Ribonucleoprotein
SPT	:	Suppressor of Ty
SWI/SNF	:	Mating-Type Switching/Sucrose Non-Fermenting
TAF	:	TBP Associated Factor
TBP	:	TATA-Binding Protein
TF	:	Transcription Factor
TREX	:	Transcription Export
TRX	:	Thioredoxine
TSS	:	Transcription start site

*Les modifications épigénétiques des histones seront décrites en utilisant la syntaxe : HxRy où x décrit l'histone affectée et y la position du résidu R*

## Préambule

La transcription est un mécanisme fondamental de la vie cellulaire qui peut être subdivisé en trois étapes : initiation, élongation et terminaison. Dans le cadre de la transcription eucaryote par l'ARN Polymérase II (ARNPII), différents mécanismes essentiels sont opérés au cours de chacune de ces étapes, et vont permettre d'aboutir à la production d'un ARN messager (ARNm) fonctionnel. Chacun de ces mécanismes nécessite le recrutement par la polymérase de facteurs spécifiques. Au cours de la phase d'élongation de la transcription, ces différents facteurs vont notamment permettre (i) le remodelage de la chromatine, nécessaire au passage de la machinerie transcriptionnelle, (ii) la maturation de l'ARNm avec notamment l'élimination des introns et (iii) le recrutement de facteurs d'export. Bien qu'une vision globale des liens fonctionnels unissant ces différents mécanismes commence à émerger, la compréhension de ces liens au niveau moléculaire reste très souvent à définir.

Le complexe Spt6/Iws1 est l'un des facteurs recruté par l'ARNPII au cours de la phase d'élongation de la transcription. Ce complexe lie le domaine carboxy-terminal (CTD) de l'ARNPII (constitué chez l'homme de 52 répétitions de la séquence  $Y_1S_2P_3T_4S_5P_6S_7$ ) lorsque les sérines en position 2 de ses répétitions sont phosphorylées. Il a été montré que le complexe Spt6/Iws1 était impliqué à la fois dans les mécanismes transcriptionnels, en augmentant les taux de transcription de la polymérase, dans le remodelage de la chromatine, par le recrutement de l'histone méthyltransférase Setd2, ainsi que dans la maturation et l'export des ARNm, par le recrutement du facteur d'export REF1/Aly. De plus, (i) l'interaction entre Spt6 et les histones et (ii) l'implication de Spt6 dans la reconstitution des nucléosomes dans le sillage de l'ARNPII au cours de la transcription, suggèrent un rôle de chaperonne d'histone pour le complexe Spt6/Iws1.

L'ensemble des rôles fonctionnels joués par le complexe Spt6/Iws1 montre clairement l'importance de ce complexe dans le processus de la transcription. Cependant, au début de ces travaux de thèse, le mode d'action du complexe Spt6/Iws1 au niveau moléculaire restait inconnu. Le but de ces travaux a consisté à élucider biochimiquement, structuralement et *in vivo* le rôle de ce complexe au cours de la transcription. Ces travaux ont notamment permis de mieux comprendre (i) comment le complexe Spt6/Iws1 est formé et (ii) comment il est recruté par l'ARNPII. De plus, ces travaux ont permis de mettre en évidence une interaction entre la



protéine Spt6 et le facteur d'élongation de la transcription TFIIIS, le rôle de ce complexe restant à définir.

### Caractérisation du domaine carboxy-terminal de Spt6 – Chapitre 3

Il a été montré que la protéine Spt6 est recrutée via le domaine SH2 de son extrémité carboxy-terminale par l'ARNPII. Les domaines SH2 sont impliqués dans les interactions protéine/protéine et interagissent avec des phosphotyrosines. Étonnamment, le domaine SH2 de la protéine Spt6 est le seul domaine SH2 connu qui reconnaisse des phosphosérines. De plus, les domaines SH2 sont retrouvés majoritairement chez les mammifères, la protéine Spt6 étant en fait la seule protéine de levure contenant ce type de domaine. Afin de tenter de comprendre en termes moléculaires le mode de liaison particulier de ce domaine à l'ARNPII, nous avons résolu la structure du domaine carboxy-terminal entier de la protéine Spt6 du parasite *Antonospora locustae* par diffraction des rayons X.

Cette structure a révélé que le domaine carboxy-terminal de la protéine Spt6 contient en fait deux domaines SH2 organisés en tandem. Alors que le premier domaine présente les caractéristiques classiques de ce type de domaine, le second domaine SH2 possède des caractéristiques hautement non canoniques : l'architecture globale des domaines SH2 est bien retrouvée dans ce deuxième domaine SH2 mais les résidus canoniques de reconnaissance du phosphate sont partiellement manquants. Cependant, l'observation de la présence d'un ion sulfate dans la poche de fixation du phosphate de chacun des domaines SH2 suggère que tous deux ont conservés la capacité de lier des résidus phosphorylés. En accord avec ces observations, nos expériences de pulldown réalisées sur des extraits de levure montrent que le domaine carboxy-terminal entier de Spt6 est nécessaire au recrutement de Spt6 par l'ARNPII. Ces résultats ont été confirmés par des expériences de génétique de levure réalisées par nos collaborateurs américains. Cependant, nous n'avons pas pu mettre en évidence une interaction stable entre le tandem SH2 de Spt6 et de petits fragments phosphorylés de l'ARNPII, montrant que cette interaction est soit très faible, soit qu'elle nécessite de nombreux déterminants.

### Caractérisation du complexe Spt6/Iws1 – Chapitre 4

Il a été montré que la protéine Iws1 interagit avec l'extrémité amino-terminale de Spt6 lorsqu'elle interagit avec l'ARNPII. Le complexe ainsi formé va permettre le recrutement de différents facteurs nécessaires à la production d'un ARNm fonctionnel. Afin de comprendre comment le complexe Spt6/Iws1 est formé, les régions de chacune de ces protéines

impliquées dans l'interaction ont été définies par co-expression de différentes constructions des deux protéines. La structure du complexe entre un fragment amino-terminal de Spt6 et le domaine central de Iws1 d'*Encephalitozoon cuniculi* a été résolue par diffraction des rayons X. Cette structure a permis de mettre en évidence deux principales zones d'interaction, l'une étant formée par un motif hydrophobe (motif IF) de Spt6 qui interagit au niveau d'une poche hydrophobe de Iws1. Afin de comprendre le rôle du complexe Spt6/Iws1, nous avons introduit des mutations ponctuelles dans les deux zones d'interaction de la protéine Spt6. Ces mutations abolissent *in vitro* la formation du complexe, tandis qu'elles conduisent *in vivo* à l'apparition de phénotypes thermosensibles chez la levure. Ces résultats confirment nos données structurales et démontrent ainsi que le complexe Spt6/Iws1 joue un rôle important dans la transcription.

#### Caractérisation des complexes Med13/TFIIS, Spt8/TFIIS et Spt6/TFIIS – Chapitre 5

La résolution de la structure de la protéine Iws1 a permis de mettre en évidence une homologie structurale avec les domaines amino-terminaux des facteurs d'élongation TFIIS et ElonginA, ainsi que de la sous-unité MED26 du co-activateur de la transcription Médiateur. Cette homologie structurale est notamment très forte au niveau de la zone d'interaction avec la protéine Spt6. De plus, le motif de Spt6 reconnu par Iws1 semble également présent dans la séquence de deux interactants de TFIIS : Med13, une autre sous-unité du Médiateur et Spt8, une sous-unité du complexe transcriptionnel SAGA. Nos tests de co-expression ont permis de montrer que les motifs IF des protéines Med13 et Spt8 interagissent avec TFIIS, et que la mutation de ces motifs affaiblit fortement le complexe. Nous avons ainsi pu caractériser biochimiquement les complexes Spt8/TFIIS et Med13/TFIIS qui avaient seulement été décrits fonctionnellement jusqu'à présent.

L'homologie structurale entre Iws1 et TFIIS nous a également poussé à tester par co-expression l'interaction entre Spt6 et TFIIS, dont une interaction génétique avait déjà été décrite précédemment. Une interaction entre ces deux protéines a ainsi pu être mise en évidence. De plus, l'extrémité amino-terminale de Spt6, fortement chargée négativement, contient plusieurs motifs hydrophobes proches du motif IF (le motif IF interagissant avec Iws1 étant le 5<sup>ème</sup> de ces motifs ; IF5). Des tests de co-expression ont révélés qu'un second motif (noté IF3) interagissait avec la protéine TFIIS, ce motif ne retenant pas la protéine Iws1. Ces tests ont également révélés qu'un fragment comprenant les motifs IF3 et IF5 pouvait interagir simultanément avec Iws1 et TFIIS.

Nos tentatives de cristalliser ces complexes ont été infructueuses, aussi la technique de résonance magnétique nucléaire a été utilisée afin de résoudre la structure des complexes formés par TFIIS, et ainsi comprendre leur mode d'interaction. Nos résultats préliminaires sur le complexe Spt8/TFIIS révèlent un mode de liaison spécifique de Spt8 à TFIIS, différent de celui observé dans le cas du complexe Spt6/Iws1. Ces travaux en cours devraient permettre de mieux comprendre l'interaction de TFIIS avec différents effecteurs de la transcription. De plus, ces travaux sont actuellement complétés par des expériences de génétique de levure réalisées par nos collaborateurs qui ont déjà permis de montrer que les différents motifs hydrophobes de l'extrémité amino-terminale de Spt6 participaient fortement à la fonction de Spt6.

#### Interaction entre Spt6 et les histones – Chapitre 6

Le rôle de chaperonne d'histone de la protéine Spt6 a été suggéré par l'observation de l'interaction entre Spt6 et la paire d'histone H3/H4 (Bortvin et Winston, 1996). Afin de caractériser ce complexe, nous avons dans un premier temps mis en place un système qui permet de co-exprimer aussi bien les paires d'histones que l'octamère d'histone. Ce système a ensuite été utilisé afin de tester par co-expression l'interaction entre Spt6 et les histones. Cependant, aucun complexe n'a pu être mis en évidence, suggérant qu'une interaction stable entre Spt6 et les histones ne pourrait se faire que dans le contexte du nucléosome.

L'ensemble de ces travaux de thèse ont permis d'aider à la compréhension au niveau moléculaire du rôle du complexe Spt6/Iws1 dans la transcription. Ces travaux, ainsi que ceux de nos compétiteurs révèlent le rôle essentiel du complexe Spt6/Iws1, notamment par son interaction avec plusieurs protéines impliquées dans la formation d'ARNm fonctionnels. La compréhension du rôle de ces complexes va nécessiter leur caractérisation, nos travaux constituant la base de ces caractérisations, qui permettront la compréhension des mécanismes moléculaires qui permettent la production des ARNm fonctionnels.

# **Chapitre 1**

## **Introduction**



La transcription est un mécanisme central de la vie cellulaire puisqu'elle constitue la première étape de la production des protéines et des ARN fonctionnels. Chez les eucaryotes la transcription réalisée dans le but de la synthèse protéique est catalysée par l'ARN polymérase II, l'ARN polymérase I synthétisant les ARN ribosomiaux (sauf l'ARN 5S) et l'ARN polymérase III synthétisant les ARN de transferts et l'ARN ribosomal 5S. Ces trois enzymes vont utiliser l'ADN cellulaire afin de synthétiser une molécule d'ARN par complémentarité à l'ADN. Dans le cas de l'ARN polymérase II, l'ARN messager produit doit alors être exporté vers le cytoplasme où se trouvent les ribosomes. Cependant, au cours de ce mécanisme, la machinerie transcriptionnelle doit faire face à la structure de la chromatine qui impose des contraintes à la transcription. Ainsi, un très grand nombre de protéines vont être impliquées pour permettre la modulation de la chromatine et, ainsi, réguler la transcription.

## **1.1. Production d'un ARN messager fonctionnel : mécanismes de la transcription eucaryote**

La production d'un ARN messager fonctionnel va impliquer différentes étapes. Tout d'abord la transcription va permettre la synthèse d'un ARN pré-messager (pré-ARNm). Ce pré-ARNm va ensuite être épissé et mûri : les séquences non codantes qu'il contient (introns) vont être éliminées et ses extrémités 5' et 3' vont être modifiées. L'ARN messager mature ainsi produit va être pris en charge par la machinerie d'export.

### **1.1.1. Mécanismes de la transcription eucaryote**

Le mécanisme de la transcription par l'ARN polymérase II est subdivisé en trois étapes principales : (i) l'étape d'initiation qui débute par la formation du complexe de pré-initiation, au cours de laquelle la machinerie transcriptionnelle se met en place et commence à synthétiser un court fragment d'ARN, (ii) une étape d'élongation au cours de laquelle l'ARN messager est synthétisé et finalement (iii) une étape de terminaison au cours de laquelle l'ARN en cours de synthèse va être détaché de la machinerie de transcription par clivage. La maturation de l'ARN est réalisée tout au long de ce processus, l'extrémité 5' de l'ARN messager étant synthétisée en début de phase d'élongation alors que l'extrémité 3' est modifiée au cours de la phase de terminaison. L'épissage est lui réalisé pendant ou après la phase d'élongation.

#### 1.1.1.1. L'initiation de la transcription - Formation du complexe de pré-initiation

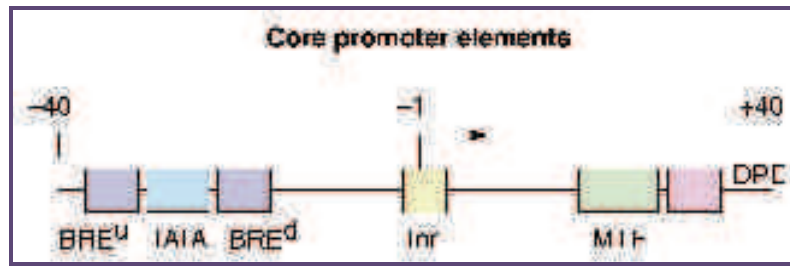
##### *a) Reconnaissance des éléments régulateurs et du promoteur – implication des activateurs*

Les promoteurs sont composés de deux régions : le promoteur distal, qui contient des séquences activatrices et le promoteur proximal, qui contient à la fois des séquences activatrices et des sites de liaison de la machinerie transcriptionnelle générale.

L'activation de la transcription est généralement médiée par une ou plusieurs protéines activatrices, composées d'un domaine d'activation et d'un domaine de liaison à l'ADN permettant aux protéines de fixer les séquences activatrices localisées au niveau des promoteurs. Un activateur peut activer la transcription de plusieurs gènes impliqués dans un même processus. Ce sont, par exemple, les récepteurs hormonaux (RAR, RXR, PPAR $\gamma$  ...) dans les cellules mammifères ou encore, chez la levure, la protéine Pho4 qui régule l'expression du gène *PHO5* en réponse à la quantité de phosphate présent dans le milieu. Ces activateurs peuvent alors recruter, à travers leurs domaines activateurs, directement ou à l'aide de différents complexes protéiques co-activateurs, la machinerie transcriptionnelle au niveau des promoteurs proximaux. Ces activateurs jouent donc un rôle clé dans la régulation de la transcription en réponse à l'environnement, leur mutation et leur dérégulation conduisant à l'apparition de maladies.

Le promoteur proximal comporte aussi différents éléments qui permettent sa reconnaissance par des facteurs de transcriptions qui vont initier la formation du complexe de pré-initiation (PIC). Le PIC va permettre le positionnement de l'ARN polymérase II au niveau du site de début de la transcription et de définir l'orientation du brin à transcrire. Dans le promoteur proximal, différentes séquences ont pu être identifiées qui sont reconnues par les composants du PIC : les TATA box reconnues par la protéine TBP (TATA Binding Protein) qui est un élément central du complexe de pré-initiation, les éléments BRE (TFIIB Recognition Element) reconnus par le facteur TFIIB, ou encore les éléments INR (Initiateur), DPE (Downstream Promoter Element), ou MTE (Motif 10 Element), tous trois reconnus par des protéines TAF (TBP Associated Factor). Certains de ces éléments sont localisés dans le promoteur proximal en amont du site de début de la transcription (éléments TATA et BRE) alors que d'autres sont localisés dans la phase codante du gène (éléments MTE et DPE),

l'élément INR étant localisé au niveau du site d'initiation de la transcription (Figure 1) (Smale et Kadonaga, 2003).



**Figure 1 : Différents éléments constituant les promoteurs des gènes.**

*Le promoteur possède différents éléments qui sont reconnus par différentes sous-unités du complexe de pré-initiation. L'élément TATA est reconnu par la protéine TBP, les éléments BRE sont reconnus par le facteur de transcription TFIIB et les éléments INR, MTE et DPE sont reconnus par différentes protéines TAF (adapté de Juven-Gershon et Kadonaga, 2010).*

Un promoteur est ainsi constitué de la combinaison de différents éléments, qui vont permettre le recrutement de différents facteurs lors de la phase d'activation de la transcription (Juven-Gershon et al., 2008; Juven-Gershon et Kadonaga, 2010). Ainsi, 20% des promoteurs de levure contiennent une TATA-box (Basehoar et al., 2004), contre 40% chez la drosophile (Kutach et Kadonaga, 2000) et 22% chez l'homme (Gershenzon et Ioshikhes, 2004). De plus, la plupart des gènes qui possèdent une TATA box sont des gènes inductibles (Basehoar et al., 2004).

#### *b) Les différents complexes co-activateurs*

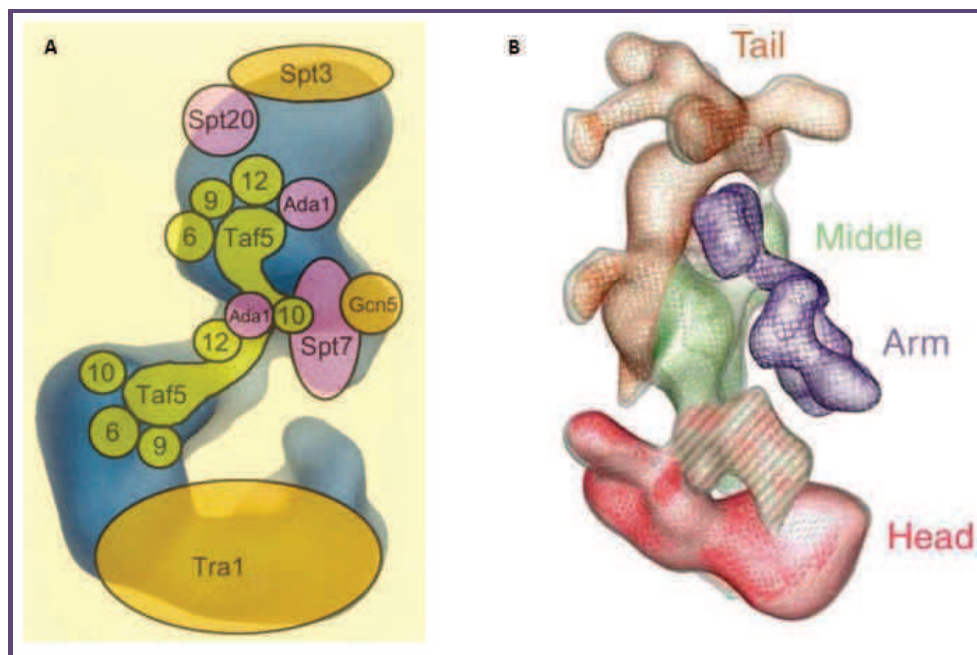
Les co-activateurs sont des complexes protéiques qui permettent de relier les protéines activatrices à la machinerie transcriptionnelle basale. Deux classes de co-activateurs peuvent être définies : les co-activateurs qui interagissent directement avec la TBP (TFIID et SAGA) et les co-activateurs qui inhibent ou favorisent l'interaction de la TBP à l'ARN polymérase et la formation du complexe de pré-initiation (PIC), notamment le complexe Mediateur. Les travaux réalisés au cours de cette thèse ont notamment porté sur des sous-unités des complexes SAGA et Médiateurs qui sont présentés ici.

##### *(i) Le complexe SAGA*

Chez la levure il a été démontré que l'initiation de 10% des gènes est réalisée par le complexe SAGA (Spt-Ada-Gcn5-Acetyltransferase), tous ces gènes étant inductibles et possédant une TATA box dans leur promoteur (Basehoar et al., 2004). Ce complexe est composé de 20 sous-unités chez la levure, alors qu'il existe différents isoformes dans les



cellules mammifères (TFIIIC, PCAF, STAGA). Parmi les sous-unités de SAGA on retrouve les protéines de la famille Spt (Spt3, 7, 8, 20) ainsi que certaines protéines TAF identiques à celles retrouvées dans TFIID (Taf5, 6, 9, 10 et 12) (Figure 2). L'interaction de SAGA au niveau des promoteurs avec la protéine TBP est médiée par les sous-unités Spt3 et Spt8 du complexe (Mohibullah et Hahn, 2008). SAGA participe également à la régulation de la transcription par la modulation de la structure de la chromatine. En effet, plusieurs sous-unités du complexe possèdent des activités enzymatiques : une activité d'acétylation des histones (HAT, Histone Acetyl Transferase) portée par Gcn5 et une activité d'ubiquitylation des histones portée par Ubp8 (Bhaumik, 2011). Des études réalisées chez la levure ont démontrées une interaction entre la sous-unité Spt8 de SAGA et le domaine amino-terminal du facteur d'élongation TFIIIS, cette interaction permettant vraisemblablement de recruter TFIIIS au niveau des promoteurs de certains gènes chez la levure et ainsi de constituer des PIC actifs (Wery et al., 2004; Kim et al., 2007). Le complexe SAGA joue ainsi un rôle important en permettant l'initiation de la transcription par son interaction avec la TBP et la communication de l'état transcriptionnel par la modification des histones.



**Figure 2 : Architecture des complexes co-activateurs de la transcription SAGA et Médiateur.**

*Structure à basse résolution des complexes SAGA (A) et Médiateur (B) résolues par microscopie électronique. Ces complexes sont constitués d'un grand nombre de sous-unités qui forment une structure modulaire impliquée dans la régulation de l'initiation de la transcription (adapté de Wu et al., 2004; Cai et al., 2010)*

### (ii) Le complexe Médiateur

Le Médiateur est un très grand complexe qui permet la communication entre les protéines activatrices et la machinerie transcriptionnelle (Figure 2), mais dont le rôle n'est que peu compris pour l'instant. En effet, ce complexe, retrouvé chez tous les eucaryotes intervient dans différents processus, certains parfois opposés. Ainsi, le médiateur (i) participe à l'activation de la transcription des gènes en favorisant le recrutement de l'ARN polymérase II et de la protéine TBP au niveau des promoteurs, (ii) coordonne la formation du PIC, (iii) stimule la phosphorylation de la sous-unité Rpb1 de l'ARN polymérase II par TFIIH et PTEF-b (Positive Elongation Factor b) et, (iv), réprime l'étape d'initiation de la transcription lorsqu'il est lié à son domaine kinase CDK8. D'autre part, le Médiateur stimule également la transcription réalisée à partir des gènes non-induits. Ainsi, des études montrent que, chez la levure, le Médiateur est retrouvé sur la plupart des promoteurs, suggérant que ce complexe joue un rôle central dans la mise en place de la transcription. La réalisation de ces différentes fonctions va impliquer le recrutement de différentes protéines ou complexes qui vont participer au processus de régulation et notamment des protéines qui modifient les histones. Différentes formes du complexe ont pu être isolées : une forme interagissant avec l'ARN polymérase II, cette interaction étant réalisée à travers les sous-unités Med17 du Médiateur et Rpb3 de l'ARN polymérase II (Soutourina et al., 2011) et une forme en complexe avec un domaine kinase CDK8, ce domaine étant composé des protéines Cdk8, Cycline C, Med12 et Med13.

Des études réalisées chez la levure ont démontré que la sous-unité Med13 du sous-complexe CDK8 du médiateur interagit aussi avec le domaine amino-terminal du facteur d'élongation TFIIS (Wery et al., 2004; Guglielmi et al., 2007; Kim et al., 2007). De la même manière que pour l'interaction Spt8/TFIIS, cette interaction pourrait permettre le recrutement de TFIIS au niveau des promoteurs de certains gènes chez la levure. Ainsi, le complexe Médiateur permet de réguler l'initiation de la transcription par une interaction directe avec l'ARN polymérase II, les différents rôles de ce complexe qui ont été décrits devant encore être replacés dans un contexte global afin de comprendre comment ce complexe agit (Malik et Roeder, 2005; Taatjes, 2010; Conaway et Conaway, 2011).

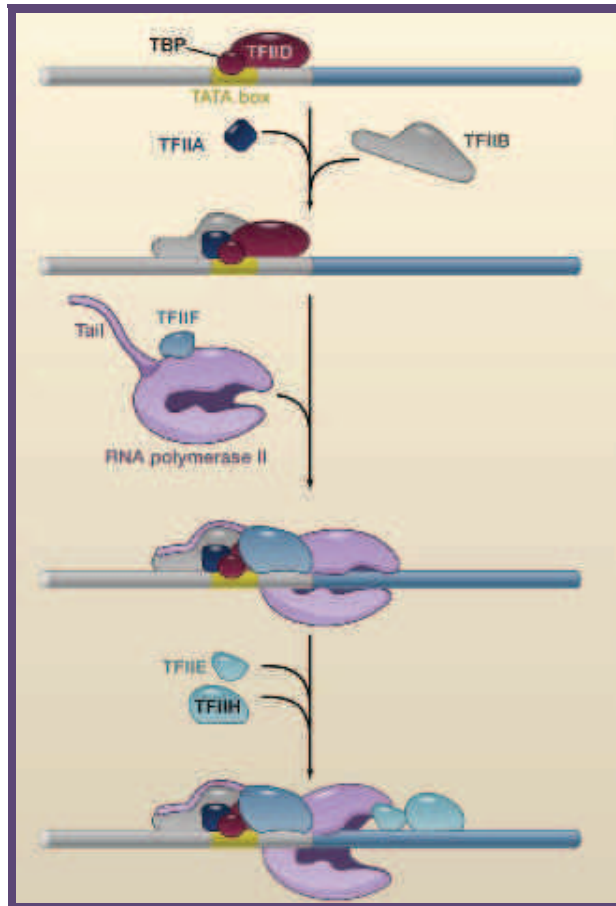
Ainsi, l'activation de la transcription qui conduit au recrutement de la machinerie transcriptionnelle implique des complexes différents du facteur de transcription TFIID, qui avait d'abord été décrit comme l'élément initiateur général de la transcription. Ces complexes

soit remplacent (cas de SAGA), soit assistent TFIID (cas du Médiateur) dans la mise en place du complexe de pré-initiation, leur rôle précis restant encore à définir. Ils illustrent également le couplage entre les mécanismes d'initiation et de répression de la transcription induits par la modification des histones. Parmi les interactions mises en place par ces complexes, l'interaction avec le domaine amino-terminal du facteur d'élongation TFIIS qui a été mise en évidence doit encore être caractérisée biochimiquement et fonctionnellement.

### *c) Assemblage du complexe de pré-initiation*

L'ARN polymérase II seule est capable de réaliser la transcription *in vitro* mais nécessite la présence des facteurs généraux de la transcription (GTF) pour réaliser la transcription dans un environnement chromatinien *in vivo*. Ainsi, une fois la protéine TBP associée au promoteur grâce aux facteurs TFIID ou SAGA, les différents facteurs de transcription qui composent le PIC vont être recrutés séquentiellement. La protéine TBP associée au complexe TFIID va permettre le recrutement des facteurs de transcription TFIIA et TFIIB qui vont stabiliser l'interaction de la protéine TBP avec l'ADN. Le facteur TFIIB reconnaît également les éléments BRE présents dans le promoteur et est nécessaire à l'identification précise du site d'initiation. Le facteur de transcription TFIIF, en complexe avec l'ARN polymérase II hypophosphorylée, va alors interagir avec la plateforme créée par les facteurs TFIID, TFIIA et TFIIB. Les facteurs TFIIE et TFIIH sont alors successivement recrutés (Figure 3).

Dans un autre mécanisme, l'ARN polymérase peut également être recrutée directement sous la forme d'un complexe holoenzyme au niveau du promoteur. Le complexe contient l'ARN polymérase II qui peut être associée à différentes combinaisons de facteurs de transcriptions mais également des co-activateurs, notamment le Médiateur et des complexes de remodelage, le facteur TFIID étant toujours absent de ce complexe. L'initiation est ainsi réalisée en deux temps : (i) le recrutement de TFIID au niveau du promoteur des gènes et (ii) le recrutement de l'holoenzyme contenant l'ARN polymérase II (Orphanides et al., 1996; Thomas et Chiang, 2006; Sikorski et Buratowski, 2009; Baumann et al., 2010; Nechaev et Adelman, 2011)



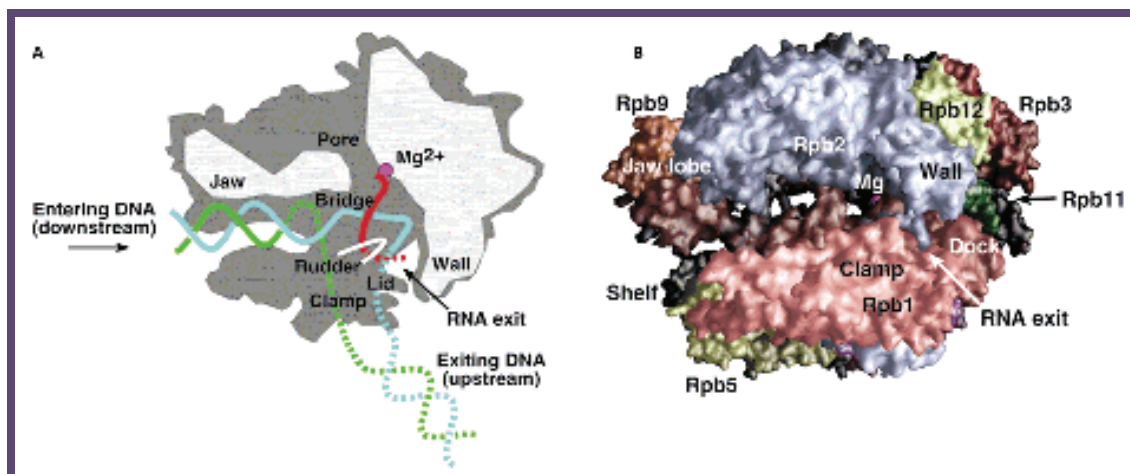
**Figure 3 : Assemblage séquentiel du complexe de pré-initiation de la transcription au niveau des promoteurs**

*Le complexe de pré-initiation de la transcription est formé de plusieurs facteurs de transcriptions dont le recrutement est séquentiel. Ainsi, après la fixation de la protéine TBP au niveau du promoteur, soit par l'intermédiaire de TFIID ou de SAGA, les facteurs de transcription TFIIA et TFIIB sont recrutés. L'ARN polymérase II est alors recrutée en complexe TFIIF puis les facteurs TFIIE et TFIIH (adapté de Levine, 2011).*

Il a par la suite été montré que TFIIS participe également à l'initiation de la transcription en stimulant la formation du PIC et est recruté au niveau des promoteurs de certains gènes chez la levure (Guglielmi et al., 2007; Kim et al., 2007). Cette activité est indépendante du rôle de facteur d'élongation de la protéine et implique le domaine de liaison de la protéine à l'ARN polymérase II (domaine II) et son domaine amino-terminal (domaine I). TFIIS participerait ainsi au recrutement de l'ARN polymérase sur les promoteurs de certains gènes de levure, le recrutement de TFIIS au niveau de ces promoteurs nécessitant les complexes co-activateurs SAGA et Médiateur. Ces données sont notamment compatibles avec l'observation d'une interaction entre le domaine amino-terminal de TFIIS et les sous-unités Spt8 de SAGA et Med13 du Médiateur (Wery et al., 2004).

### 1.1.1.2. L'ARN polymérase II

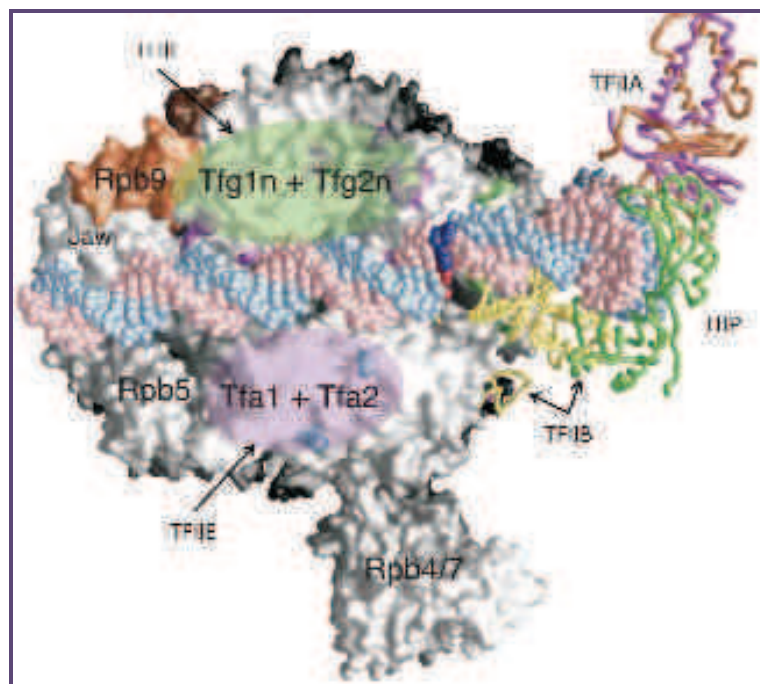
L'ARN polymérase II est un complexe protéique de 0,5MDa, composé de douze protéines, notées Rpb1 à Rpb12 conservées chez tous les eucaryotes. Le complexe est composé de quatre éléments mobiles notamment le cœur et la pince (clamp) (Figure 4B). Le cœur de ce complexe est constitué des protéines Rpb3, 10, 11, 12 et certaines régions des protéines Rpb1 et Rpb2 et représente plus de la moitié du complexe. Au centre de l'enzyme, une poche profonde permet la fixation d'un brin d'ADN après que l'activité hélicase de TFIIH ait permis d'ouvrir le double brin. Dans cette poche, deux ions métalliques vont permettre la catalyse de la réaction d'addition de nucléotides tri-phosphates (Figure 4A). L'entrée de cette poche est masquée par la pince avant l'étape d'initiation et au cours de l'élongation, son déplacement étant nécessaire pour l'entrée d'un simple brin d'ADN lors de l'ouverture du double brin. La résolution de la structure de l'ARN polymérase II (Cramer et al., 2001) a montré également que le point de sortie de l'ARN synthétisé est localisé à proximité du domaine carboxy-terminal (CTD ; non visible dans la structure du fait de sa flexibilité) de la sous-unité Rpb1, ce domaine étant impliqué dans le recrutement de différents facteurs notamment impliqués dans la maturation des ARN messagers (décrit au paragraphe 1.1.2).



**Figure 4 : Structure de l'ARN polymérase II de levure**

*L'ARN polymérase II est composée de douze sous-unités organisées autour des protéines Rpb1 et Rpb2, les sous-unités Rpb4 et Rpb7 étant absentes de ce modèle. Au centre de l'enzyme un canal est formé qui va permettre l'entrée de l'ADN simple brin à transcrire. Ce canal est fermé par un clamp (pince) qui va empêcher l'ADN de ressortir et assurer la processivité du mécanisme de transcription (adapté de Hahn, 2004).*

L'architecture du complexe de pré-initiation a été déterminée par un ensemble de techniques et notamment le pontage chimique et la résolution de la structure de sous-complexes (Figure 5). Ainsi, la protéine TBP et les facteurs de transcription TFIIA et TFIIB interagissent et masquent le site de sortie de l'ARN transcrit alors que le facteur de transcription TFIIF interagit au-dessus du site de fixation de l'ADN et TFIIE en dessous de ce site. Les deux sous-unités Rpb4 et Rpb7 forment un complexe qui interagit sous la pince de l'ARN polymérase II et reste accessible, notamment pour la formation d'interaction avec des protéines ou des molécules d'ADN. Le complexe Rpb4/Rpb7 va notamment permettre de maintenir la pince fermée et le recrutement de phosphatases du CTD (Fcp1) qui vont permettre le recyclage de l'ARN polymérase II (Cramer, 2004).



**Figure 5 : Organisation de l'ARN polymérase II dans le complexe de pré-initiation**

*Les différents facteurs de transcription vont interagir à la surface de l'ARN polymérase II (adapté de Chen et al., 2007).*

### 1.1.1.3. Fin de l'initiation de la transcription – Echappée du promoteur

A la fin de la phase d'initiation, le complexe de pré-initiation est formé au niveau du promoteur du gène. L'association des facteurs TFIIF, TFIIE et TFIIH va permettre l'ouverture du double brin d'ADN grâce à l'activité des sous-unités hélicases XPB et XPD de TFIIH, permettant la formation d'une bulle de transcription dans le site actif de l'ARN polymérase II au niveau du site INR du promoteur. Les premiers ARN courts synthétisés par complémentarité ne sont pas associés de manière stable, la transcription réalisée *in vitro*

aboutissant alors au relargage de fragments de 2 ou 3 nucléotides. De plus, le complexe de pré-initiation doit subir de grands changements conformationnels, notamment au niveau du site de sortie de l'ARN qui est masqué par le facteur TFIIB et qui permet également la stabilisation des transcrits courts dans le site actif (transcrits de 4 paires de bases). Lorsque le transcrit atteint la taille de 8 à 9 paires de bases, la totalité de la bulle de transcription est occupée, elle doit alors être allongée ce qui constitue l'étape d'échappée du promoteur (promoter clearance) et marque la fin de la phase d'initiation. Elle est accompagnée de la phosphorylation du CTD par le facteur TFIIH et le relargage des facteurs de transcription TFIIB, TFIIIE, TFIIIF et TFIIH, ces facteurs pouvant alors être utilisés afin d'initier une nouvelle phase d'initiation (Saunders et al., 2006; Nechaev et Adelman, 2011).

#### 1.1.1.4. L'élongation de la transcription

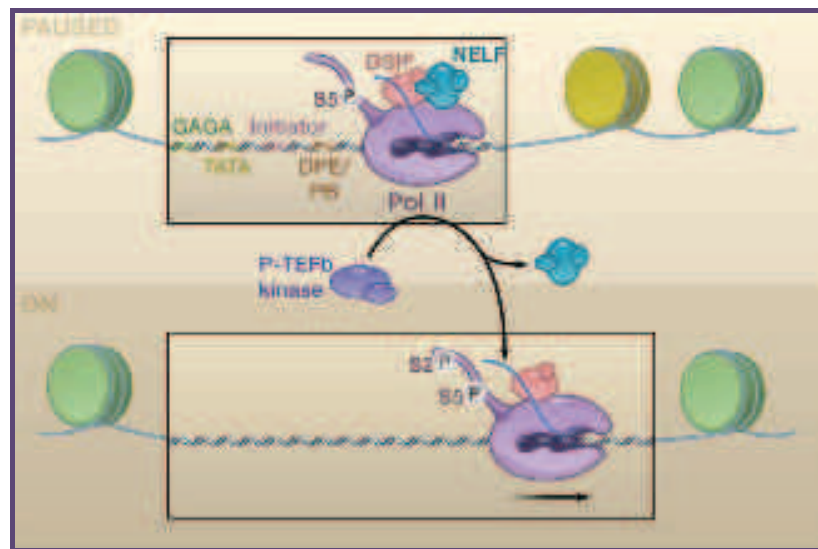
Alors que la forte régulation opérée au moment de l'initiation a été observée depuis longtemps, l'importance de la régulation opérée au cours de l'étape d'élongation n'a que récemment été démontrée. Au cours de cette phase, un grand nombre de facteurs vont être recrutés qui vont permettre la régulation de cette étape et la maturation de l'ARN en cours de synthèse. Certains facteurs vont notamment permettre l'entrée en phase d'élongation processive, d'autres augmentant la processivité de l'ARN polymérase II, alors que d'autres facteurs vont permettre la régulation de cette étape de la transcription.

##### *a) Régulation de la phase d'élongation : « promoter pausing »*

Au début de la phase d'élongation, lorsqu'un fragment d'ARN d'environ 20 paires de bases a été synthétisé, deux facteurs qui inhibent la transcription vont être recrutés : NELF (Negative Elongation Factor) et DSIF (DRB-sensitivity-inducing factor), la présence de ces deux facteurs étant nécessaire à l'induction de la pause transcriptionnelle (Figure 6).

Le facteur DSIF est composé des protéines Spt4 et Spt5, deux protéines de la famille SPT qui ont des liens forts avec Spt6. Spt5 présente une forte homologie avec la protéine bactérienne NusG, impliquée dans la régulation de la transcription et dans le couplage entre la transcription et la traduction protéique bactérienne. Spt5 possède une extrémité aminoterminal fortement chargée négativement et son domaine carboxy-terminal est composé de séquences répétées, proches de celles du CTD de l'ARN polymérase II. La protéine Spt5 interagit directement avec la pince mobile de l'ARN polymérase II, les modèles établis à partir de la structure du complexe formé entre Spt4/Spt5 et la pince de l'ARN polymérase II

suggérant que la région amino-terminale de Spt5 ferme le site actif de la polymérase qui contient l'ARN, augmentant ainsi la processivité de l'enzyme (Martinez-Rucobo et al., 2011). DSIF va permettre le recrutement de NELF à la fin de la phase d'initiation, le complexe restant associé à l'ARN polymérase II au cours de la phase d'élongation et recrutant d'autres facteurs et notamment le complexe PAF, des chaperonnes d'histones, FACT, Spt6, Chd1 et des protéines impliquées dans la maturation des ARN messagers et la surveillance.



**Figure 6 : Passage en phase d'élongation productive de la transcription**

*Le passage en phase d'élongation processive va s'accompagner de la phosphorylation de l'ARN polymérase II et du facteur DSIF par le facteur P-TEFb (adapté de Levine, 2011).*

Le facteur NELF est un complexe de quatre sous-unités qui ne lie pas directement l'ARN polymérase mais plutôt le complexe ARN polymérase/DSIF. La pause induite par l'association de DSIF et NELF à l'ARN polymérase II va aussi permettre le recrutement des facteurs qui vont modifier l'extrémité 5' du transcrit.

Le phénomène de pausing a notamment été mis en évidence au niveau des gènes de choc thermique chez la drosophile (Lis, 1998). Il a été déduit de la présence de molécules d'ARN polymérase II au niveau des promoteurs de ces gènes en l'absence de choc thermique. Il a été démontré que l'ARN polymérase II était localisée au niveau de la région 5' des gènes, un fragment de 20 à 40 nucléotides ayant déjà été synthétisé. Le rôle de ce mécanisme de pausing reste encore à définir mais différentes hypothèses ont été formulées : il pourrait servir de point de contrôle permettant à la machinerie transcriptionnelle de recruter les facteurs de maturation de l'ARN et les facteurs nécessaires à la mise en place d'une élongation processive. Ce mécanisme permet également l'induction rapide de la transcription,



notamment au niveau des gènes de choc thermique chez les mammifères et de tout gène inductible, mais il a également été démontré qu'il était impliqué dans les mécanismes de développement des organismes. Alors que ce mécanisme semble être largement utilisé chez les mammifères, chez la levure son implication reste encore à définir (Sims et al., 2004; Hahn, 2004; Saunders et al., 2006; Li et Gilmour, 2011).

#### *b) Entrée en phase d'élongation processive*

Le passage en phase d'élongation processive va nécessiter le recrutement du facteur PTEF-b : il va phosphoryler le CTD de l'ARN polymérase II mais également le domaine carboxy-terminal de la sous-unité Spt5 du complexe DSIF, conduisant au relargage de NELF et à l'entrée en phase d'élongation (Figure 6). Chez la levure, deux kinases homologues à PTEF-b ont été définies : la kinase Ctk1 et la kinase Bur1. Bur1 va être recrutée dans un premier temps ; elle va phosphoryler la protéine Spt5 et induire le dépôt de marques épigénétiques qui sont nécessaires au recrutement de la kinase Ctk1 qui va alors phosphoryler le CTD de l'ARN polymérase II (Brès et al., 2008). Les facteurs PTEF-b et NELF sont alors relargués et l'ARN polymérase II, en complexe avec DSIF, entre alors en phase d'élongation processive (Sims et al., 2004; Saunders et al., 2006). Le passage en phase d'élongation processive est aussi régulé par la repression de l'activité de PTEF-b par le complexe nucléoprotéique formé par la protéine HEXIM1 (Hemamethylene Inducible mRNA) et l'ARN non codant 7SK, ce mécanisme restant encore mal connu, notamment en termes moléculaires (Diribarne et Bensaude, 2009).

#### *c) Autres complexes impliqués dans la régulation de l'élongation*

Une fois la polymérase entrée en phase d'élongation processive, d'autres facteurs vont être recrutés qui vont favoriser une transcription efficace par l'ARN polymérase II. Parmi ces facteurs les Elongins et les protéines de la famille ELL stimulent l'élongation de la transcription (Sims et al., 2004).

Le facteur TFIIS intervient également au cours de cette phase. En effet, lorsque l'ARN polymérase II est arrêtée du fait de l'incorporation d'un mauvais nucléotide, le domaine carboxy-terminal (domaine III) de TFIIS va entrer dans le site actif de l'ARN polymérase II via le tunnel de sortie de l'ARN et permettre le clivage de l'ARN en cours de synthèse. Ce clivage va permettre un réalignement de l'extrémité 3' du transcrit en cours de synthèse et la reprise de la transcription. Le domaine II (domaine central) de TFIIS va permettre

l'interaction de la protéine avec la surface de la polymérase II au niveau du site de sortie de l'ARN polymérase II au cours de ce mécanisme. Chez la levure, la délétion de TFIIS conduit à l'apparition d'une sensibilité à la 6-azauracile, un composé qui diminue les quantités de GTP et UTP présentes dans le milieu, l'incapacité de ces souches à se développer sur un milieu contenant ce composé reflétant un défaut dans les mécanismes de transcription (Cheung et Cramer, 2011).

Le complexe PAF (Polymerase Associated Factor) intervient également au cours de l'élongation de la transcription. Ce complexe de cinq protéines chez la levure (Paf1, Rtf1, Ctr9, Cdc73 et Leo1, le complexe humain contenant une protéine additionnelle Ski8) est recruté au niveau du complexe ARN polymérase II/DSIF, chez la levure, la phosphorylation de Spt5 par Bur1 permettant de stimuler son recrutement. Le complexe PAF va notamment permettre le dépôt de marques épigénétiques (ubiquitylation de H2B par Rad6/Bre1 et méthylation de H3K4 par Set1), mais également permettre le recrutement de facteurs de remodelage de la chromatine (FACT) et de facteurs de terminaison de la transcription (Jaehning, 2010).

#### 1.1.1.5. Mécanismes de terminaison

Les mécanismes de terminaison sont directement reliés à la maturation de l'extrémité 3' du transcrit. Ce processus nécessite un grand nombre de facteurs, ces facteurs étant différents chez la levure et dans les cellules de mammifères. Après le codon Stop, dans la région 3' UTR (Untranslated Region), des séquences vont permettre le recrutement de facteurs de clivage et de polyadénylation (facteurs CPF chez la levure, pour Cleavage and Polyadenylation Factor) et de différents facteurs de clivage, notamment les complexes CF1A et CF1B (Cleavage Factor) chez la levure, le complexe CF1A contenant la protéine Pcf11. Ces différents facteurs vont diriger la coupure endonucléolityque de l'ARN messager transcrit et permettre la poly-adénylation de l'extrémité 3' ainsi générée. Ce fragment poly-A va être stabilisé et protégé des exonucléases par la liaison de protéines PABP (Poly-A Binding Protein) (Millevoi et Vagner, 2010).

La transcription est donc un mécanisme complexe qui nécessite l'interaction d'un grand nombre de facteurs qui vont à la fois activer et réguler ce processus. Ces différents facteurs vont notamment être recrutés au niveau de l'extrémité carboxy-terminale (CTD) de la sous-unité Rpb1.

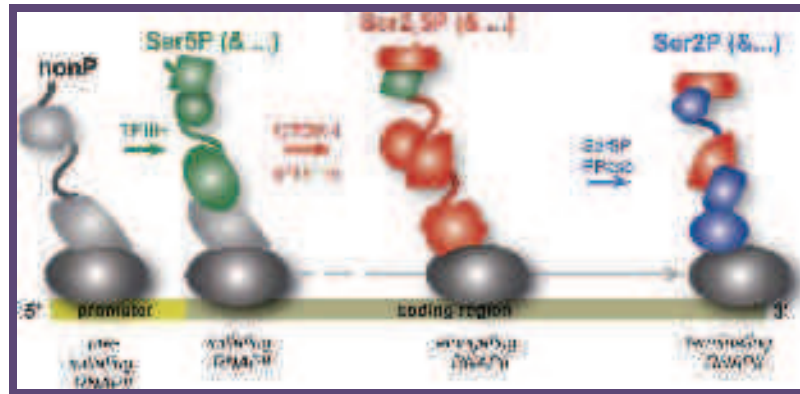
### 1.1.2. Implication du CTD de l'ARN polymérase II dans la régulation de la transcription : le code CTD

La plus grande sous-unité de l'ARN polymérase II, Rpb1, contient à son extrémité carboxy-terminale une séquence qui contient un grand nombre de répétitions de la séquence Tyr<sub>1</sub>-Ser<sub>2</sub>-Pro<sub>3</sub>-Thr<sub>4</sub>-Ser<sub>5</sub>-Pro<sub>6</sub>-Ser<sub>7</sub>, le nombre de répétitions variant selon les espèces. Ainsi, ce motif est répété 26 fois chez la levure, 45 fois chez la drosophile et 52 fois chez l'homme. Au sein de ces répétitions, deux principales zones peuvent être définies : une zone amino-terminale qui contient des répétitions canoniques et une zone à l'extrémité carboxy-terminale du domaine qui contient des motifs divergents, dits non-canoniques. La délétion du CTD de l'ARN polymérase II est létal chez la levure, la drosophile et la souris, montrant l'importance de cette région.

Différentes modifications peuvent avoir lieu sur ce domaine de l'ARN polymérase II au cours de la transcription. Ces modifications participent à la régulation de la transcription. Ainsi, il a été montré que les trois sérines, la thréonine et la tyrosine pouvaient être phosphorylées et les prolines pouvaient être isomérisées. La combinaison de ces différentes modifications permet d'obtenir un nombre considérable de motifs qui vont être reconnus par différents facteurs. Ces différentes observations ont conduit à l'élaboration du « code CTD » (Buratowski, 2003).

Ce domaine va notamment permettre le passage à travers les différentes phases de la transcription (Figure 7). Lors de la formation du complexe de pré-initiation, l'ARN polymérase qui est recrutée ne porte pas de modification. Cette forme non modifiée va notamment être reconnue par le complexe Médiateur. A la fin de la phase d'initiation, la kinase Cdk7 (couplée à la cycline H) du sous-complexe CAK du facteur de transcription TFIIF (Kin28 chez la levure) phosphoryle les sérines en position 5 du CTD, permettant le passage en phase d'élongation. Cette marque va permettre le recrutement des enzymes qui modifient l'extrémité 5' de l'ARN messager en cours de synthèse et de l'histone méthyltransférase Set1 qui va déposer la marque H3K4me3. Le début de la phase d'élongation est marqué par la phosphorylation des sérines en position 2 du CTD, réalisée par la kinase Cdk9 (couplée à la cycline T) du complexe PTEF-b (chez la levure, cette phosphorylation est réalisée par la kinase Ctk1, qui est homologue à Cdk9). La phosphorylation des sérines 2 va permettre le passage de l'ARN polymérase II dans une phase d'élongation processive mais également le recrutement des facteurs d'élongation, d'épissage

et de polyadénylation nécessaires à la production d'un ARN messager fonctionnel, et d'histones méthyltransférases, notamment Set2. Finalement, la déphosphorylation du CTD au cours de la phase de terminaison va notamment permettre le recyclage de l'ARN polymérase II (Phatnani et Greenleaf, 2006; Egloff et Murphy, 2008; Chapman et al., 2008; Buratowski, 2009).



**Figure 7 : Phosphorylation du CTD de l'ARN polymérase II au cours de la transcription**

*Le domaine carboxy-terminal de la plus grande sous-unité de l'ARN polymérase II (Rpb1) contient un grand nombre de résidus qui peuvent être modifiés et qui vont permettre le recrutement de facteurs impliqués dans la transcription (adapté de Phatnani et Greenleaf, 2006).*

L'implication de la phosphorylation de la sérine 7 dans ce mécanisme a été suggérée plus récemment. Chez la levure, cette marque est détectée au niveau des promoteurs et le long de la phase codante des gènes alors que chez les mammifères, elle est détectée sur les promoteurs des gènes codants les petits ARN non codants (snRNA) et à l'extrémité 3' des gènes codants les protéines. Dans les cellules mammifères, cette marque est déposée par la kinase Cdk7 (du complexe TFIIF), alors que chez la levure elle est déposée par la kinase Bur1. Le rôle de cette modification n'est pour l'instant que peu compris dans le cas de la transcription des gènes protéiques mais elle est nécessaire à la transcription des snRNA (Egloff et al., 2007, 2012).

Le CTD de l'ARN polymérase II constitue ainsi une large plateforme d'ancrage des facteurs impliqués dans la transcription. De plus, les différentes modifications de ce domaine vont permettre la discrimination des différents facteurs recrutés notamment dans le cas du recrutement de protéines spécifiques à une phase de la transcription. Parmi ces facteurs, les protéines impliquées dans la maturation et l'export des ARN messagers vont lier le CTD de l'ARN polymérase II.

### 1.1.3. Maturation et export des ARN messagers - Mécanismes co-transcriptionnels

Parallèlement au mécanisme de transcription, le transcrit en cours de synthèse va être maturé (addition de nucléotides modifiés à l'extrémité 5' du transcrit, ajout d'une queue poly-A en 3' et élimination des introns), et des facteurs d'exports vont être recrutés. De plus, le processus de surveillance réalisé par l'exosome va permettre d'éliminer les ARN messagers présentant des défauts de maturation ou des mutations dans leur séquence. Le couplage de ces différents mécanismes (transcription, maturation, export et surveillance) va permettre l'augmentation de l'efficacité et de la précision de ces différents mécanismes, ils sont alors décrits comme étant co-transcriptionnels (Figure 9). En effet, la plupart des facteurs nécessaires à leur réalisation sont directement recrutés par la machinerie d'élongation, notamment le CTD ou les facteurs d'élongation (PTEF-b, complexe PAF).

#### 1.1.3.1. Maturation des ARN messagers

Les transcrits codants pour des protéines chez les eucaryotes, après leur synthèse doivent être maturés. Ce processus de maturation implique l'ajout à l'extrémité 5' du transcrit d'un nucléotide modifié (cap), l'élimination des séquences non codantes, les introns, et l'ajout en 3' d'une séquence poly-A qui va protéger le transcrit des endonucléases présentes dans le milieu.

La maturation de l'extrémité 5' va être réalisée par des enzymes dites de capping qui sont recrutées par le CTD de l'ARN polymérase II phosphorylé sur la sérine 5. Ainsi, dans un premier temps, le phosphate  $\gamma$  du premier nucléotide en cours de synthèse est éliminé, une molécule de GMP (Guanidine MonoPhosphate) est ajoutée par l'enzyme guanylyltransferase, la guanine étant par la suite méthylée. Ce cap va permettre l'initiation de la traduction par le ribosome.

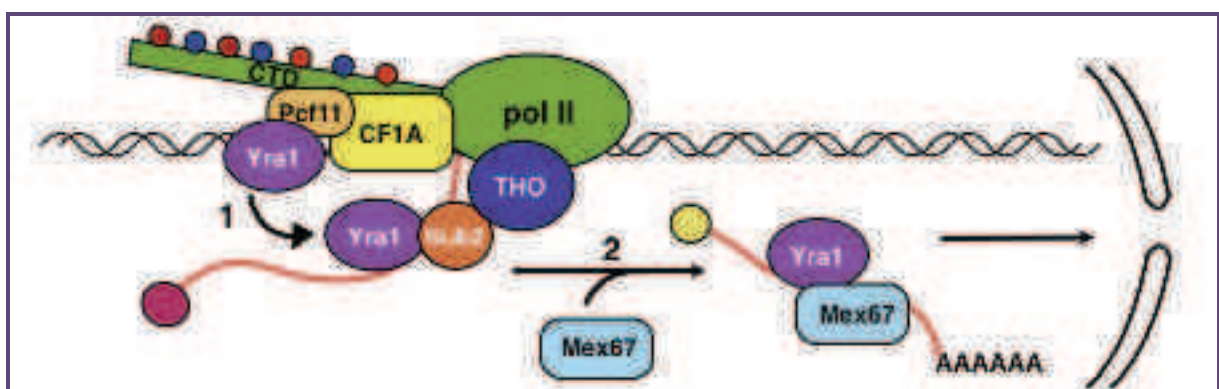
Au cours de la transcription, certains introns vont être éliminés, les ribonucléoprotéines nécessaires à la mise en place de ce mécanisme étant également recrutées par le CTD de l'ARN polymérase, principalement le snRNP U1. Cette particule reconnaît le site d'épissage en 5' de l'intron, l'extrémité 3' de l'intron étant reconnue par la particule U2AF. La particule U2A va alors lier les deux particules, permettant de créer une boucle dans l'ARN en cours de synthèse. Les particules U4-U6/U5 sont alors recrutées et vont cliver et religuer le brin d'ARN. De plus, certaines protéines SR (serine/arginine rich), qui sont

nécessaires à la régulation de ce mécanisme d'épissage et au choix des sites d'épissage alternatifs, interagissent également avec l'ARN polymérase II en cours d'élongation. Ces interactions vont notamment stimuler l'élongation par la polymérase, la vitesse de transcription ayant directement été reliée au choix des sites d'épissage alternatifs.

La maturation de l'extrémité 3' du gène est fortement couplée à l'élongation de la transcription (décrit au paragraphe 1.1.1.5). En effet, le facteur de clivage CF1A de levure est recruté par l'ARN polymérase au début de l'élongation, la sous-unité Pcf11 de ce complexe interagissant avec le CTD de l'ARN polymérase II dont la sérine 2 est phosphorylée.

### 1.1.3.2. Export des ARN messagers

Une fois le transcrit synthétisé et mûré dans le noyau de la cellule, il doit être exporté vers le cytoplasme où est localisée la machinerie traductionnelle. Le recrutement des facteurs d'export a lieu au cours de l'élongation de la transcription par la liaison de protéines adaptatrices qui vont permettre de relier l'ARN messager à la machinerie d'export. En effet, le complexe adaptateur THO/TREX (Transcription Export) est recruté par des protéines qui lient la coiffe (CBP, Cap Binding Protein) ajoutée à l'extrémité 5' du transcrit, cette interaction permettant notamment d'augmenter le taux d'élongation de l'ARN polymérase II. De plus deux sous-unités de ce complexe, Yra1 et Sub2 (appelées respectivement REF1/Aly et UAP52 chez les métazoaires) ont été identifiées comme étant des partenaires des facteurs d'élongation Spt6 et Iws1 chez la souris. L'adaptateur THO/TREX transmet alors l'ARN messager aux facteurs d'export, notamment Mex67 qui va lier la machinerie d'export (Kohler et Hurt, 2007).

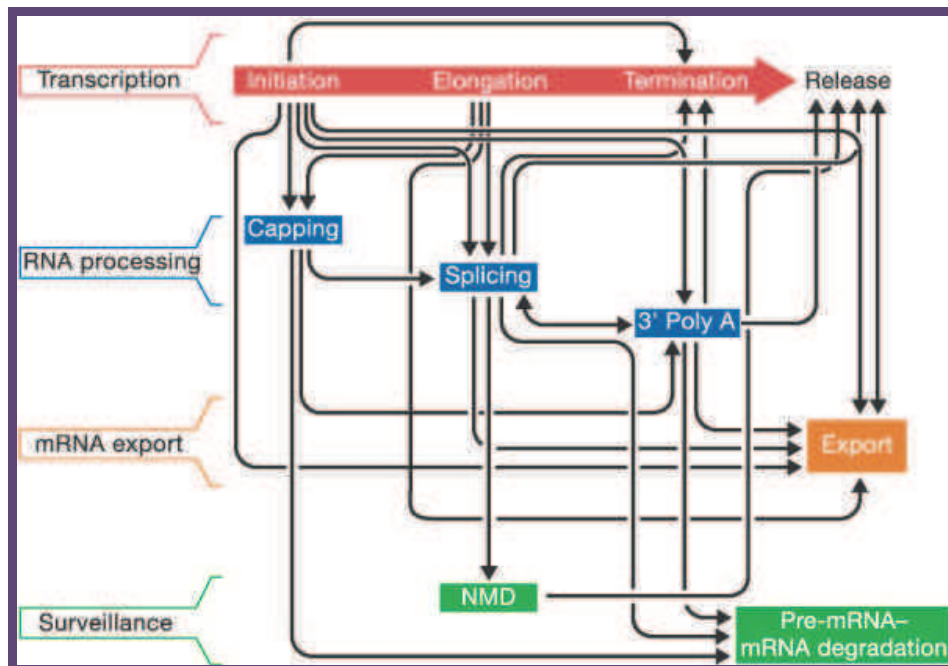


**Figure 8 : Mécanisme d'export des ARN messagers**

*Au cours de l'élongation de la transcription, des protéines adaptatrices des facteurs d'export vont être recrutées, en reconnaissant notamment des facteurs liant le CTD de l'ARN polymérase II phosphorylé (adapté de Perales et Bentley, 2009).*

De plus, il existe un lien direct entre la machinerie d'initiation et d'export. En effet, la sous-unité Sus1 du module de dé-ubiquitylation de SAGA fait également partie du complexe TREX et participe ainsi à la coordination de ces deux événements (García-Oliver et al., 2012).

Finalement, il existe un lien fort entre la machinerie de transcription et les facteurs nécessaires à la maturation et l'export des ARN messagers, ces liens étant notamment illustrés sur la Figure 9.



**Figure 9 : Interconnexion entre les mécanismes de transcription, maturation, export et surveillance des ARN messagers**

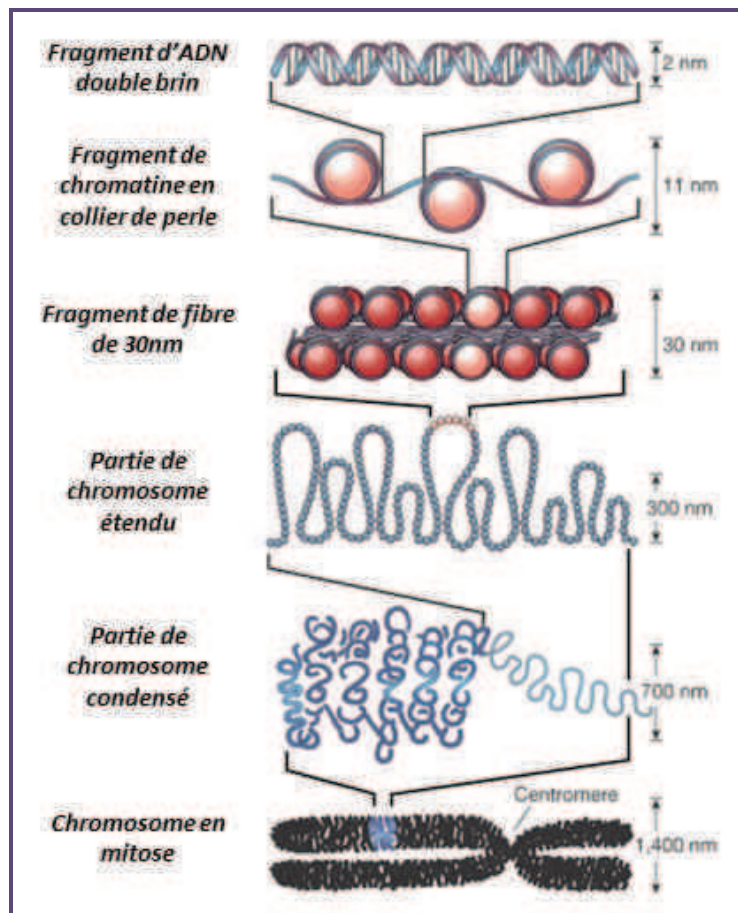
*(Adapté de Maniatis et Reed, 2002)*

Les différents facteurs nécessaires à la mise en place de ces mécanismes vont notamment être recrutés soit directement par le CTD de l'ARN polymérase II, soit par des protéines qui vont lier ce domaine de l'ARN polymérase II comme par exemple le complexe PAF ou les protéines Spt6 et Iws1.

## 1.2. Structure de la chromatine et implication dans la régulation de la transcription

L'ADN cellulaire atteint environ deux mètres et doit être contenu dans le noyau de la cellule qui fait environ un micromètre. Cet ADN doit donc être réduit en une structure compacte, faisant intervenir des protéines, les histones, pour former la chromatine. Différents niveaux de compaction peuvent ainsi être atteints, d'une structure relativement lâche, comme

la structure en collier de perle jusqu'à une structure très condensée tel que le chromosome (Figure 10).



**Figure 10 : Organisation de la chromatine dans le noyau des cellules eucaryotes.**

*L'ADN double brin enroulé autour des protéines histones forme une structure appelée chromatine. La structure en collier de perle peut être condensée en une fibre de 30nm qui peut être encore condensée pour finalement former le chromosome (adapté de Felsenfeld et Groudine, 2003).*

Cependant, l'ADN empaqueté doit rester accessible pour les mécanismes de réplication, réparation, recombinaison et transcription, nécessitant de dérouler cette structure compacte pour laisser l'accès à des machineries complexes. Ainsi, de nombreuses protéines et complexes protéiques vont intervenir au cours de ces processus afin de permettre l'accès à l'ADN, ces différentes protéines et complexes participant également à la régulation de la transcription.

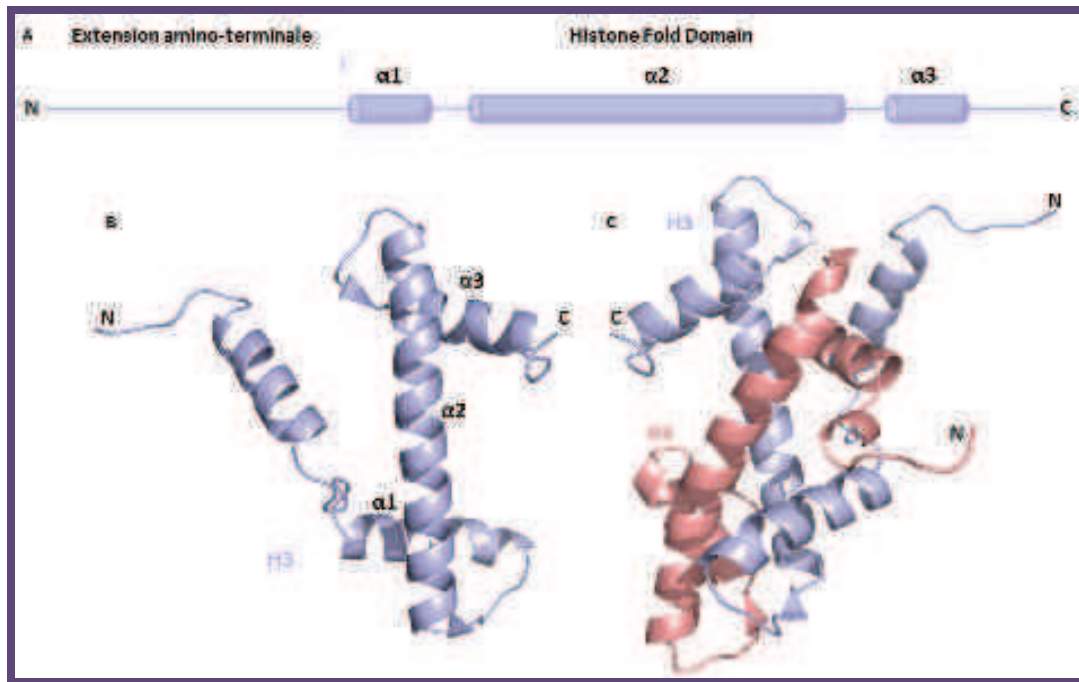
### 1.2.1. Les protéines histones et le nucléosome

La compaction de l'ADN va être réalisée par les protéines histones qui vont former avec l'ADN des structures appelées nucléosomes qui sont l'unité de base de la chromatine.



### 1.2.1.1. Les protéines histones

Les protéines histones H2A, H2B, H3 et H4 sont des protéines courtes (environ 120 résidus) fortement conservées chez tous les eucaryotes. Ces protéines sont très fortement chargées positivement, cette charge leur permettant d'interagir avec des molécules chargées négativement comme l'ADN. Les histones sont toutes composées d'un domaine central, également appelé domaine globulaire, qui adopte un repliement de type « histone fold » (HFD, Histone Fold Domain), chaque protéine histone contenant également une extension amino-terminale variable et non structurée (Figure 11A).



**Figure 11 : Architecture et structure d'une protéine histone, formation de la paire H3/H4.**

*Les protéines histones H2A, H2B, H3 et H4 sont formées d'un domaine central de type « Histone Fold Domain » entouré d'une queue amino-terminale spécifique de chaque histone et carboxy-terminale variable. Les paires d'histones spécifiques H2A/H2B et H3/H4 vont se former par une interaction tête-bêche de leur HFD respectif.*

Les domaines HFD sont constitués d'une hélice  $\alpha$  centrale longue ( $\alpha 2$ ) flanquée de part et d'autre de boucles qui relient deux hélices  $\alpha$  plus courtes ( $\alpha 1$  et  $\alpha 3$ ) (Figure 11B). Ces domaines HFD permettent la mise en place d'interactions protéine-protéine spécifiques. Ils ne sont d'ailleurs pas seulement retrouvés dans les histones. En effet, plusieurs protéines, notamment certaines protéines TAF qui font partie des facteurs co-activateurs de la transcription TFIID et SAGA contiennent également ce type de domaines qui sont impliqués dans la formation de ces complexes (Gangloff et al., 2001). Dans le cas des histones, les

paires H2A/H2B et H3/H4 vont être formées par l'interaction tête-bêche de leurs motifs HFD (Figure 11C).

Une cinquième protéine histone, l'histone H1, va également être impliquée dans la formation du nucléosome (Ramakrishnan, 1997). Cette protéine, qui ne possède pas de domaine HFD va stabiliser l'ADN à l'entrée et à la sortie du nucléosome, stabilisant par la même l'entité nucléosome.

Les gènes codants pour les protéines histones sont présents en plusieurs copies dans les génomes eucaryotes, la levure contenant par exemple deux copies de l'histone H3. De plus, les histones H3 et H2A possèdent des isoformes également appelés variants. Les séquences de ces variants sont très proches des protéines canoniques, les modifications entre un variant et la forme canonique de l'histone étant généralement minimales. Ces variants vont notamment permettre l'annotation de la chromatine lors d'évènements particuliers comme la présence de cassures dans l'ADN ou l'initiation de la transcription et permettent alors le recrutement de facteurs spécifiques. Le remplacement des histones canoniques par son variant va nécessiter des facteurs de remodelage et des chaperonnes d'histones.

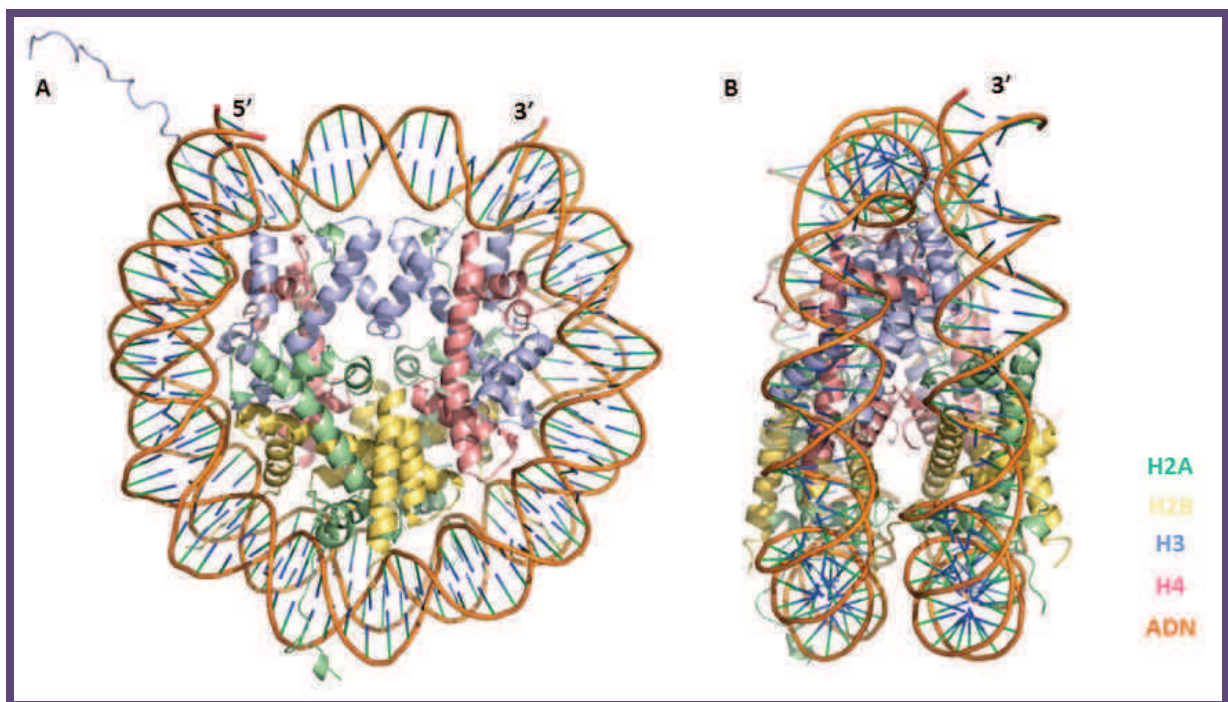
L'expression des protéines histones est fortement régulée au cours du cycle cellulaire, les histones canoniques n'étant produits et déposés sur l'ADN qu'au cours de la phase de synthèse du cycle cellulaire, c'est-à-dire lors de la réplication (phase S). Les variants d'histones sont par contre produits et utilisés indépendamment de la réplication.

#### 1.2.1.2. Structure du nucléosome

Les premières structures de nucléosomes ont été résolues en utilisant un ADN rendu palindromique et issu de séquences répétées retrouvées au niveau des télomères dans le génome humain (satellites  $\alpha$ ) (Luger et al., 1997). Plus récemment, une séquence d'ADN (Widom 601) non palindromique et synthétisée de manière aléatoire a été sélectionnée pour sa forte affinité pour l'octamère d'histone et la structure du nucléosome contenant cet ADN a été résolue (Vasudevan et al., 2010; Makde et al., 2010).

Le nucléosome est formé de l'assemblage d'un octamère des histones H2A, H2B, H3 et H4, qui contient deux copies de chaque protéine, autour desquelles un fragment de 147 paires de bases d'ADN est enroulé, représentant 1,7 tour d'hélice. La structure du nucléosome est sphérique et plane, l'octamère d'histone formant une structure compacte autour de laquelle l'ADN vient s'enrouler (Figure 12). Dans cette structure, seules les queues des histones

émergent mais ne sont pas visibles dans la structure du fait de leur flexibilité. L'octamère d'histone est formé de deux copies de chaque protéine histone, les dimères H2A/H2B s'associant avec le tétramère (H3/H4)<sub>2</sub> pour obtenir l'arrangement (H2A/H2B)(H3/H4)<sub>2</sub>(H2A/H2B). Les deux paires H3/H4 interagissent par une interaction de type « four helix bundle » formée par les deux protéines H3. Ce type d'interaction est principalement hydrophobe et fait intervenir les hélices  $\alpha_2$  et  $\alpha_3$  des deux protéines. Chaque paire H2A/H2B interagit avec le tétramère (H3/H4)<sub>2</sub> à la fois par une seconde interaction de type « four helix bundle » réalisée entre les hélices  $\alpha_2$  et  $\alpha_3$  des histones H4 et H2B et par une interaction de la partie carboxy-terminale de H2A avec les paires H3/H4. Les deux paires H2A/H2B n'interagissent que faiblement ensemble (Figure 12). Lors de sa mise en place sur l'ADN, le nucléosome est formé de manière séquentielle : le tétramère (H3/H4)<sub>2</sub> se lie dans un premier temps à l'ADN, les deux paires H2A/H2B étant alors recrutées pour compléter le nucléosome.



**Figure 12 : Structure tridimensionnelle du nucléosome**

*Le nucléosome est formé des histones qui sont organisés en un octamère central autour duquel l'ADN vient s'enrouler, assurant le premier niveau de compaction de l'ADN.*

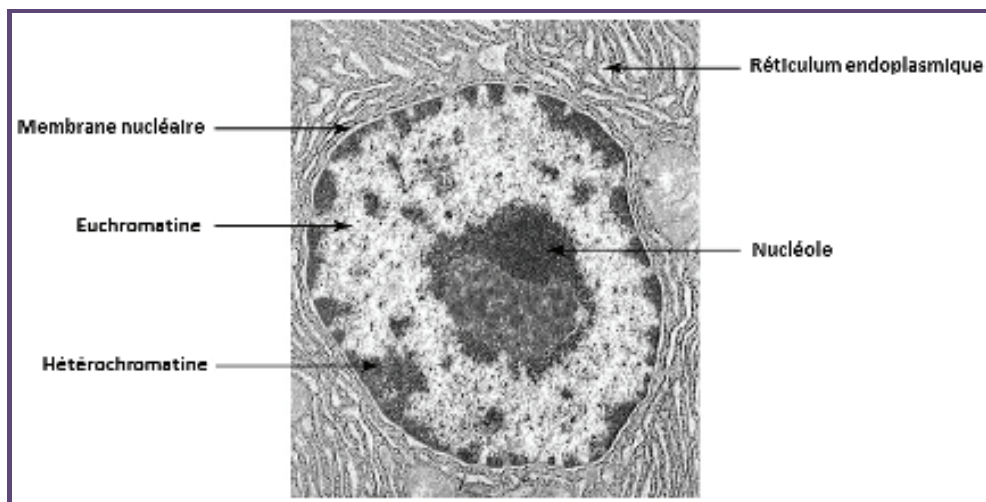
Le tétramère (H3/H4)<sub>2</sub> interagit à la fois avec la région centrale de l'ADN, recouvrant environ 60 paires de bases, et avec les extrémités de l'ADN au niveau de la région aminoterminal non structurée de H3. Chaque dimères H2A/H2B s'associent sur une région d'environ 30 paires de bases avec l'ADN à l'opposé du tétramère (H3/H4)<sub>2</sub>. Toutes les

interactions ADN : protéines réalisées impliquent des résidus des protéines histones et des liaisons phosphodiester du sillon mineur de l'ADN, la plupart de ces interactions étant formées par des liaisons hydrogènes. Cependant, alors que les nucléotides présents dans le sillon mineur ne semblent pas interagir avec les histones, la résolution de la structure du nucléosome contenant l'ADN 601 suggère que la présence de paires adénine – thymine dans le sillon mineur de l'ADN favorise le positionnement correct du nucléosome. Ainsi, les nucléosomes ne semblent pas positionnés aléatoirement sur l'ADN *in vitro*.

L'histone H1 lie l'ADN à la sortie du nucléosome, interagissant notamment avec l'ADN localisé entre deux structures nucléosomales, appelé « linker ». La longueur de ces linker va varier de 10 à 50 paires de bases. Cette protéine va également permettre la compaction de la chromatine en fibre de 30nm.

### 1.2.1.3. Le positionnement des nucléosomes sur l'ADN

Les nucléosomes ne sont pas répartis de manière aléatoire sur l'ADN, leur positionnement va notamment dépendre de la région génomique considérée. En effet, dans le noyau de la cellule, deux types principaux de chromatine peuvent être distingués : l'euchromatine, qui est transcriptionnellement active, et l'hétérochromatine, qui est une forme très condensée de la chromatine qui n'est pas transcrite (Figure 13).



**Figure 13 : Organisation de l'hétéro- et l'euchromatine dans les noyaux des cellules**

*Dans la cellule, observée ici par microscopie électronique, la chromatine peut adopter deux états, l'euchromatine qui est peu condensée et qui est accessible aux machineries cellulaires et l'hétérochromatine qui est fortement condensée et qui forme la chromatine dite silencieuse.*

La position des nucléosomes sur l'ADN est principalement dictée par la séquence nucléotidique, démontré par une forte corrélation entre la position des nucléosomes observée

*in vivo* et celle observée sur un ADN génomique condensé *in vitro* par l'ajout d'histones purifiés (décrit au paragraphe 1.2.1). Cependant, la position des nucléosomes peut également être imposée par des facteurs qui vont déplacer ces structures et notamment les facteurs de remodelage.

Ainsi, chez la levure, des études à grande échelle ont permis de mettre en évidence que plus de 90% des promoteurs des gènes contiennent des régions d'environ 150 paires de bases qui sont dépourvues de nucléosome, le reste du promoteur étant occupé par des nucléosomes qui contiennent des variants d'histone, notamment H2AZ. De plus, chez la levure, le site d'initiation de la transcription est toujours occupé par un nucléosome, appelé nucléosome +1, ce nucléosome étant retrouvé dans les gènes humains, mais légèrement décalé dans la séquence codante. Les nucléosomes entourant ce site particulier sont répartis de manière plus aléatoire.

Le positionnement des nucléosomes au niveau des sites d'initiation de la transcription est également impliqué dans la régulation de l'expression des gènes inductibles chez la levure. Ainsi, en conditions non induites, le promoteur des gènes est condensé par des nucléosomes, seule une courte région de l'ADN étant exposée. Au moment de l'induction, les nucléosomes qui occupent le promoteur sont éliminés, permettant notamment d'exposer les sites de liaison des activateurs et la TATA-box, afin de former le complexe de pré-initiation. Ce mode de régulation est notamment mis en place lors de la régulation de l'expression des gènes *PHO5*, *GALI-10* et *PHO8* (Segal et Widom, 2009; Travers et al., 2009; Bai et Morozov, 2010).

Par contre, la position des nucléosomes sur le corps des gènes au cours de l'élongation de la transcription fait encore débat. En effet, l'élimination complète des nucléosomes au cours de ce mécanisme semble peu plausible. Il semblerait plutôt que les nucléosomes soient dissociés (voire partiellement) devant la machinerie transcriptionnelle puis rapidement reconstitués une fois l'ARN polymérase II passée (Subtil-Rodríguez et Reyes, 2011).

### 1.2.2. La régulation épigénétique de la transcription

La chromatine est une structure dynamique qui peut subir différentes modifications, appelées modifications épigénétiques, qui vont également participer à la régulation de la transcription. Les modifications épigénétiques décrivent les modifications de la chromatine qui ne modifient pas la séquence de l'ADN. Parmi ces modifications trois classes sont principalement distinguées : la modification covalente des protéines histones, l'incorporation

de variants d'histones et la méthylation des séquences ADN riches en nucléotides cytosine et guanidine, également appelés îlots CpG. Cette dernière classe de modification ne sera pas abordée dans cette partie introductive.

### 1.2.2.1. La modification des histones

Les protéines histones peuvent être modifiées par l'ajout de groupements chimiques à l'extrémité des chaînes latérales de certains résidus. Ainsi, plusieurs résidus peuvent être modifiés : (i) les lysines peuvent être acétylées, mono-, di- ou triméthylées, ubiquitinyllées ou sumoylées, (ii) les arginines peuvent être mono- ou diméthylées, (iii) les sérines et les thréonines peuvent être phosphorylées, et (iv) les prolines isomérisées (Figure 14). Ces modifications affectent à la fois les résidus présents aux extrémités amino-terminales des histones, mais également les résidus des domaines globulaires (Kouzarides, 2007; Izzo et Schneider, 2010).



**Figure 14 : Les différents sites de modification des histones canoniques**

*Les différents sites indiqués sont, A : acétylation, M : méthylation, P : phosphorylation, U : ubiquitylation, (adapté de Portela et Esteller, 2010)*

La modification des histones constitue un code complexe, appelé le « code histone » (Strahl et Allis, 2000), qui participe fortement à la régulation des mécanismes cellulaires. En effet, ces modifications vont permettre soit le recrutement de facteurs spécifiques, soit la déstabilisation de la structure de la chromatine, notamment lors de l'acétylation des résidus lysines qui abolit les charges portées par ces résidus. Plusieurs modifications peuvent être

retrouvées sur les histones d'un même nucléosome rendant la compréhension exacte du rôle de ces modifications difficiles. Ainsi, à l'heure actuelle, seule une vision globale du rôle de ces modifications est disponible.

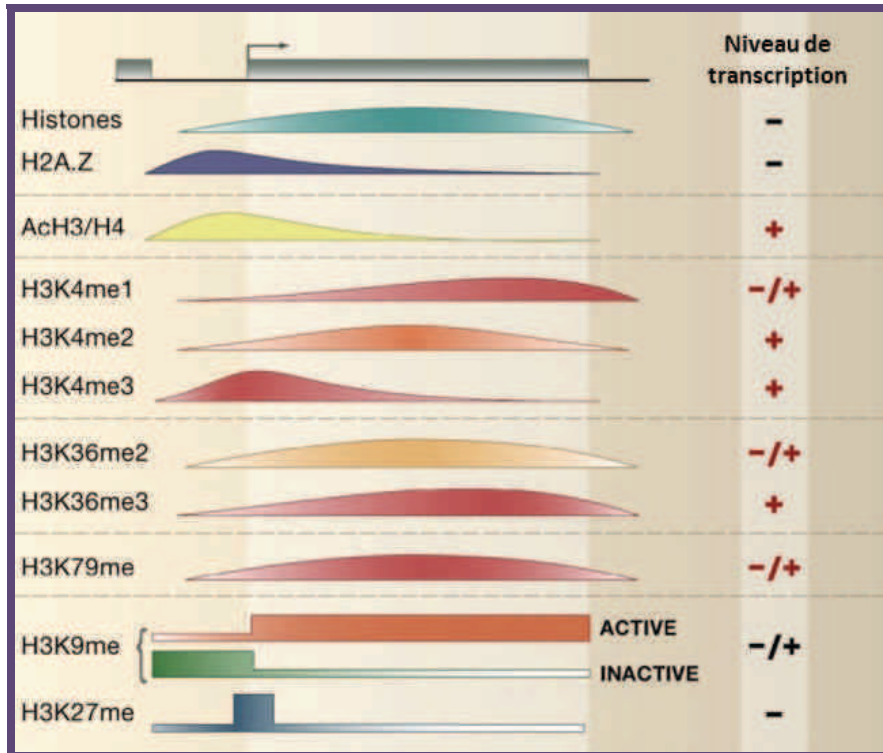
Différents domaines structuraux retrouvés sur les protéines qui interagissent avec ces modifications vont être dédiés à leur reconnaissance. Par exemple, les chromodomains, les domaines WD40 et les domaines PHD (Pleckstrin Homology Domain) vont reconnaître les lysines méthylées alors que les bromodomains reconnaîtront les lysines acétylées.

Ces différentes marques sont déposées par des enzymes spécifiques : la méthylation des lysines va être réalisée par des enzymes contenant des domaines SET, les enzymes de la classe des PRMT (Protein Arginine Methyltransferase) vont méthyler les résidus arginines, et les protéines de la famille des HAT (Histone Acetyl Transferases) vont acétyler les lysines (Taverna et al., 2007). Parallèlement, d'autres enzymes vont être responsables de l'élimination de ces marques, les protéines HDAC (Histone Deacetylase) et Sirtuines (Sir) vont déacétyler les résidus lysine, les lysines déméthylases (responsables de la déméthylation de lysine) étant également nombreuses (Kouzarides, 2007). Le mécanisme de déméthylation des arginines reste encore incertain, mais il semble qu'une lysine déméthylase (JMJD6) puisse également déméthyliser les arginines (Chang et al., 2007).

Ainsi, un très grand réseau d'interconnexions entre ces différentes marques existe, notamment illustré par la présence, dans certaines enzymes de modification des histones d'un ou plusieurs domaines de reconnaissance destinés à reconnaître la ou les modifications particulières d'un autre résidu de la même histone ou des histones adjacents. C'est par exemple le cas de l'histone méthyltransférase G9a qui est responsable de la mono- et diméthylation de la lysine 9 de H3 (H3K9me1,2). Cette protéine reconnaît les lysines 9 des histones H3 méthylées via son domaine ankyrin et va propager cette même marque sur les nucléosomes adjacents conduisant à la répression de la transcription (Collins et al., 2008).

La transcription est fortement régulée par les modifications affectant les histones. La Figure 15 montre une vue simplifiée des interconnexions entre différentes marques épigénétiques et la transcription. Généralement, la modification de certains résidus est synonyme d'activation de la transcription alors que d'autres modifications vont conduire à sa répression. De plus, au niveau du corps de gènes et des promoteurs, ces différentes modifications ne seront pas identiques. Chez la levure, au niveau des promoteurs des gènes, l'activation de la transcription va être reflétée par la triméthylation de la lysine 4 de l'histone

H3 (H3K4me3) et l'acétylation des lysines des histones H3 et H4, alors que la répression de la transcription dans cette même région est assortie de la triméthylation de la lysine 27 de H3 (H3K27me3) et de la lysine 9 (H3K9me3). Au niveau des corps des gènes, c'est principalement la triméthylation de la lysine 36 de H3 (H3K36me3) qui va être retrouvée lors de la transcription (Li et al., 2007).



**Figure 15 : Modifications des histones sur la phase codante des gènes au cours de la transcription**

*Différents degrés de modification des histones sont retrouvés sur le promoteur et la phase codante des gènes transcrits, décrit chez ici chez la levure (adapté de Li et al., 2007)*

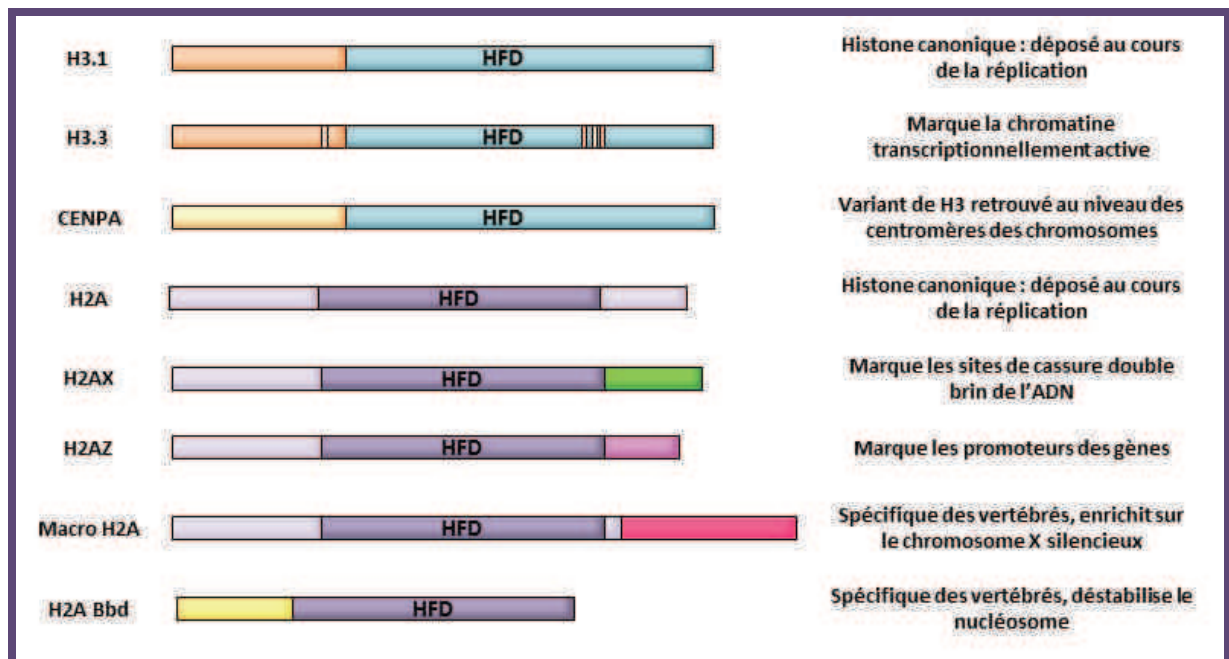
#### 1.2.2.1. Les variants d'histones

Les variants d'histones sont des isoformes des histones dits canoniques. Seules les histones H3 et H2A possèdent de tels variants. Contrairement aux histones canoniques, ces variants sont exprimés en dehors de la phase de réplication du cycle cellulaire et vont être déposés sur la chromatine au cours de processus particuliers.

L'histone H3 possède deux variants principaux (Figure 16) : (i) l'histone H3.3, l'histone canonique étant appelée H3.1 afin de le distinguer de ce variant, et (ii) l'histone CENPA (Centromere Protein A ; Cse4 chez la levure). Le variant H3.3 est intégré aux nucléosomes au cours de la transcription en remplacement des histones H3.1 qui sont éliminés, ils servent ainsi à marquer la chromatine transcriptionnellement active. La séquence



du variant H3.3 est très proche de celle de l'histone canonique, seuls quatre résidus étant remplacés : un résidu de la queue amino-terminale et trois résidus de l'hélice  $\alpha 2$  du HFD (Figure 16). Le variant CENPA est spécifique du centromère des chromosomes et induit une forte condensation de la chromatine dans cette région. Il va notamment permettre l'interaction des kinétochores et des microtubules qui sont nécessaires à la ségrégation des chromosomes dans les cellules divisées (Verdaasdonk et Bloom, 2011). Ce variant d'histone possède un domaine amino-terminal très différent de celui de l'histone canonique, son incorporation au niveau des centromères étant réalisée par la chaperonne d'histone humaine HJURP (Holliday Junction Recognition Protein) en fonction du cycle cellulaire (Shuaib et al., 2010).



**Figure 16 : Architecture globale des variants d'histones**

*Les différents variants d'histone diffèrent de leur histone canonique, au niveau de leur séquence, à différents degrés. Ils sont impliqués dans le marquage de la chromatine lors d'événements particuliers et permettent le recrutement de facteurs spécifiques (adapté de Sarma et Reinberg, 2005)*

L'histone H2A possède quatre isoformes principales : (i) H2AZ, (ii) H2AX, (iii) Macro H2A et (iv) H2A Bbd. L'histone H2AZ (Htz1 chez la levure) est incorporée à la chromatine en dehors de la réplication et des études menées sur le génome de la levure ont montré que H2AZ est réparti ponctuellement dans la chromatine mais présente un enrichissement au niveau des promoteurs des gènes (Zlatanova et Thakar, 2008). Les nucléosomes contenant ce variant sont moins condensés que les nucléosomes canoniques et peuvent glisser sur la chromatine. La chaperonne Chz1 de cette histone a été identifiée chez la levure, le complexe de remodelage permettant l'échange de l'histone canonique par le variant

étant le complexe SWR1. L'histone H2AZ présente 60% d'homologie de séquence avec l'histone H2A canonique.

Le variant H2AX est disséminé à travers la chromatine et est impliqué dans les mécanismes de réparation. En effet, le domaine carboxy-terminal spécifique de cet histone est phosphorylé lors de la formation de cassures doubles brin dans l'ADN et va alors permettre le recrutement des machineries de réparation. Ce variant d'histone possède un motif HFD très proche de celui de l'histone H2A, seule l'extrémité carboxy-terminale de ce variant étant différente. Le variants Macro H2A est quant à lui principalement présent sur le chromosome X silencieux alors que le rôle du variant H2A.Bbd reste encore à définir (Sarma et Reinberg, 2005; Talbert et Henikoff, 2010).

Ainsi, les différents variants d'histones sont impliqués dans l'annotation de la chromatine en réponse aux différents processus cellulaires (transcription, réparation) et permettent le recrutement de machineries spécifiques. Ils sont ainsi fortement impliqués dans la régulation de la transcription.

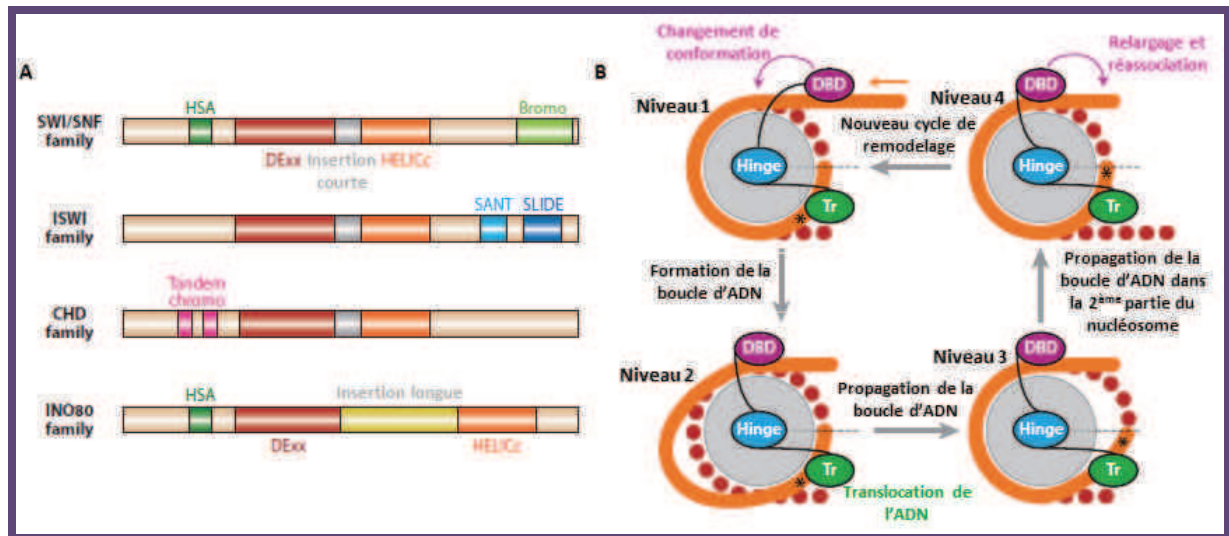
### 1.2.3. Facteurs de remodelage et chaperonnes d'histone

La forte compaction de l'ADN imposée par les nucléosomes va former une barrière lors des mécanismes de réparation, réplication, recombinaison et transcription. Ainsi, des machineries vont être nécessaires lors de la mise en place de ces mécanismes pour déplacer les nucléosomes. Les complexes de remodelage utilisent l'énergie de l'ATP (Adénosine Tri-Phosphate) afin de contrôler la position des nucléosomes. Ces complexes peuvent déplacer les nucléosomes et sont impliqués dans l'échange des histones canoniques par des variants d'histones. Les protéines chaperonnes vont prendre en charge ces histones déplacés par les complexes de remodelage afin d'empêcher leur interaction aspécifique avec l'ADN ou d'autres protéines. Ces deux machineries sont fortement impliquées dans la régulation des processus de réparation, réplication, recombinaison et transcription puisqu'elles sont recrutées par les facteurs impliqués dans ces processus.

#### 1.2.3.1. Les facteurs de remodelage de la chromatine

Les facteurs de remodelages sont de grands complexes protéiques qui contiennent une sous-unité catalytique qui porte une activité ATPase et hélicase et des protéines additionnelles spécifiques. Quatre principales familles de complexes de remodelage ont été définies :

SWI/SNF (Mating-Type Switching/Sucrose Non-Fermenting), ISWI (Imitation of Switch), CHD (Chromodomain) et INO80 (Inosito) (Figure 17A).



**Figure 17 : Architecture des sous-unités catalytiques des quatre familles de complexe de remodelage de la chromatine et mécanisme de remodelage.**

(A) Les complexes de remodelage de la chromatine contiennent tous un domaine ATPase (DExx) qui leur permet d'utiliser l'ATP afin de dissocier les complexes histones-ADN par leur hélicase (HELICc). (B) Mécanisme de remodelage de la chromatine adopté par les complexes de remodelage (adapté de Clapier et Cairns, 2009)

Les complexes de toutes ces familles présentent des caractéristiques communes : (i) une affinité forte pour les nucléosomes, (ii) la reconnaissance spécifique d'histones modifiés covalents, (iii) la présence de domaines ATPase ADN-dépendant (DExx) et hélicase (HELICc) qui permettent la dissociation des interactions histones-ADN, (iv) la présence de domaines ou protéines additionnels qui régulent l'activité ATPase et (v) qui permettent l'interaction avec des facteurs de transcription ou chromatiniens. Ainsi, le bromodomaine présent dans la famille de complexe SWI/SNF va permettre la reconnaissance des histones portant des lysines acétylés, les domaines SANT et SLIDE des complexes ISWI permettant la liaison à l'ADN et à des histones non modifiés, et le chromodomaine des complexes de la famille CHD reconnaissent les histones contenant des lysines méthylées. Il existe ainsi un nombre important de complexes de remodelage de la chromatine qui varient selon leur sous-unité catalytique et les protéines qui leurs sont associées.

Ces complexes de remodelage vont permettre soit de déplacer le nucléosome de quelques centaines de paires de base, soit de le dissocier, ce dernier mécanisme nécessitant la présence de chaperonnes d'histones. Le domaine ATPase de la sous-unité catalytique va

permettre la liaison du complexe à l'ADN du nucléosome, cette interaction pouvant également faire participer des domaines additionnels, notamment les domaines SANT et SLIDE de la famille ISWI. L'ADN enroulé autour du nucléosome va alors être déplacé, induisant la formation d'une boucle d'ADN à la surface du nucléosome qui va être propagée. Dans le cas du complexe SWI/SNF, la boucle et le raccourcissement de l'ADN linker engendrés sur le nucléosome interagissant avec le complexe va déstabiliser le nucléosome adjacent qui sera alors évincé. La résolution de la structure du complexe RSC (Remodels Structure of Chromatin) de levure, qui appartient à cette famille de facteurs de remodelage montre la présence d'une large poche au centre du complexe capable d'interagir avec un nucléosome (Chaban et al., 2008). Ainsi, les complexes de remodelage sont capables d'interagir très largement avec les nucléosomes afin d'exercer leur fonction.

Il a été démontré que le complexe INO80 est impliqué dans le remodelage de la chromatine au cours de la réplication, dans les mécanismes de réparation et l'échange de l'histone canonique H2A par le variant H2AZ au cours de la réplication. Le complexe CHDI intervient quant à lui dans l'élongation de la transcription alors que le complexe ISWI intervient dans le positionnement des nucléosomes au niveau du site d'initiation de la transcription. Ainsi, ces complexes sont impliqués dans diverses fonctions spécifiques, qui sont définies par les différentes protéines qui composent ces complexes (Clapier et Cairns, 2009; Erdel et al., 2011; Kasten et al., 2011)

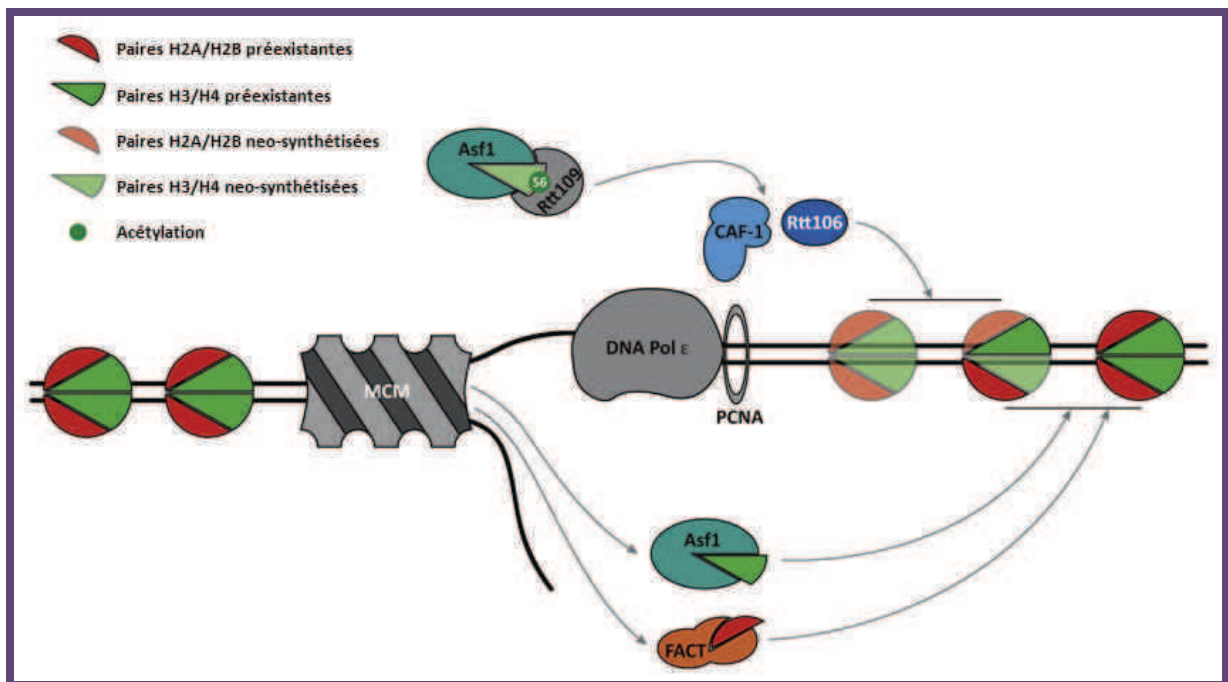
#### 1.2.3.2. Les protéines chaperonnes d'histones

Les protéines histones sont fortement chargées positivement et sont hydrophobes. Afin d'empêcher l'interaction aspécifique de ces protéines *in vivo* avec des molécules chargées négativement, des protéines chaperonnes d'histones vont interagir avec les paires d'histones et les séquestrer. Les protéines chaperonnes d'histones vont ainsi s'associer aux histones depuis leur synthèse dans le cytoplasme de la cellule jusqu'à l'incorporation des histones dans les nucléosomes. Les chaperonnes vont également être impliquées dans la régulation des mécanismes de transcription, réplication, recombinaison et réparation au cours desquelles elles vont participer à la déposition des nucléosomes ou à l'échange des histones canoniques par des histones variants.

Les chaperonnes d'histones peuvent ainsi être séparées en différentes classes selon la paire d'histone avec laquelle elle interagit, chacune pouvant être impliquée dans un ou plusieurs processus cellulaires. Parmi les chaperonnes dont les rôles sont les mieux compris à

l'heure actuelle, le complexe FACT (Facilitates Chromatin Transcription) interagit avec la paire d'histone H2A/H2B et la chaperonne Asf1 interagit avec la paire d'histone H3/H4.

Le complexe FACT est composé de deux sous-unités : Spt16 et Pob3 chez la levure et SPT16 et SSRP1 chez les mammifères, et est impliqué dans la déstabilisation entre les paires H2A/H2B et H3/H4 dans le nucléosome. Ce complexe participe aussi bien aux mécanismes de réplication, de réparation de l'ADN et de transcription. Au cours de la réplication, FACT va interagir avec les paires H2A/H2B qui sont éliminées de la chromatine lors du passage de l'hélicase MCM devant la fourche de réplication (Figure 18).



**Figure 18 : Implication des chaperonnes d'histones FACT et Asf1 au cours de la réplication**

*Les protéines chaperonnes d'histones sont toutes les deux impliquées dans le déplacement des histones au cours de la réplication (adapté de Avvakumov et al., 2011)*

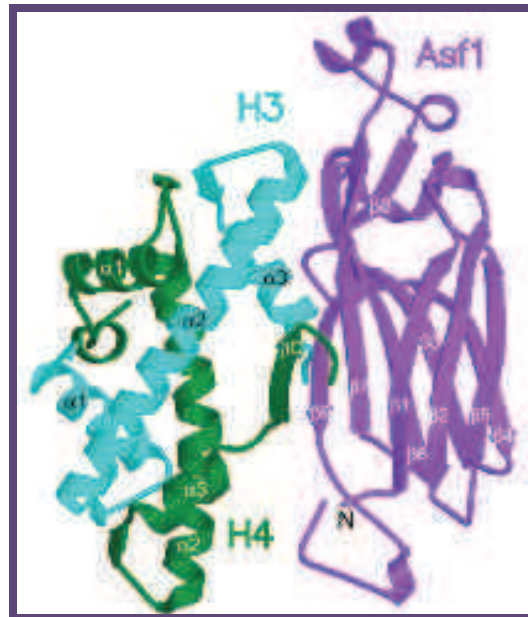
Au cours de la réparation, lorsque des cassures doubles brins sont formées dans l'ADN, les variants d'histones H2AX, disséminés plusieurs milliers de nucléotides autour du site de cassure vont être phosphorylés, constituant un signal pour le recrutement des enzymes de réparation. Une fois la cassure réparée par recombinaison homologue, ces histones phosphorylés doivent être remplacés afin d'éliminer le signal de recrutement. Il a été démontré que FACT facilitait le remplacement de la paire H2AX/H2B phosphorylée par des paires nouvellement synthétisées. Il a été proposé que l'activité de FACT dans ce processus est régulée par la poly-ADP-ribosylation de la sous-unité Spt16 du complexe, l'ajout de groupement poly-ribose inhibant l'activité du complexe. Ainsi, l'histone H2AX phosphorylé

est maintenu dans l'ADN jusqu'à ce que le processus de réparation ait eu lieu (Avvakumov et al., 2011). Au cours de la transcription, FACT participe aussi à l'initiation, ce rôle n'étant que peu compris actuellement, et au processus d'élongation. Ainsi, il a été démontré que FACT facilitait l'initiation de la transcription en participant au désassemblage des nucléosomes présents au niveau des éléments activateurs. FACT participe également au déplacement et à la reconstitution des nucléosomes au cours de l'élongation de la transcription. Ainsi, *in vitro*, FACT est capable de désassembler partiellement les nucléosomes dans un système d'élongation alors que sa mutation chez la levure conduit à l'apparition de transcrits courts qui sont issus d'une transcription initiée à partir de sites d'initiation cryptique dans les gènes. L'implication de FACT dans ce processus semble nécessiter la présence de déterminants additionnels tels que l'ubiquitylation de H2B et la présence du complexe d'élongation PAF. Ainsi, FACT joue un rôle important dans plusieurs processus cellulaires, son mode d'action précis restant encore à définir, notamment car cette chaperonne interagit aussi bien avec H2A/H2B qu'avec H3/H4 *in vitro* (Reinberg et Sims, 2006; Das et al., 2010).

De son côté, la protéine chaperonne Asf1 interagit avec la paire d'histone H3/H4 et participe (i) au mécanisme de réplication, (ii) au dépôt des histones nouvellement synthétisées, et (iii) à l'initiation de la transcription. Asf1 va notamment prendre en charge la paire H3/H4 déplacée par l'hélicase MCM devant la fourche de réplication. Au cours du processus de dépôt des histones nouvellement synthétisées, Asf1 va lier la nouvelle paire H3/H4 et permettre l'acétylation de la lysine 56 de H3 par l'acétyltransférase Rtt109. La paire modifiée va alors être transmise aux chaperonnes CAF-1 et Rtt106 (Figure 18). Au cours de l'initiation de la transcription, Asf1 participe à la formation des nucléosomes au niveau des promoteurs inductibles chez la levure, permettant la régulation de ce mécanisme, ce processus étant dépendant de l'acétylation de la lysine 56 de H3. Les nucléosomes formés qui contiennent la modification H3K56Ac sont moins stables que les nucléosomes non modifiés et permettent une ré-initiation rapide de la transcription au cours des cycles suivants (Das et al., 2010; Avvakumov et al., 2011).

La résolution de la structure de Asf1 a notamment permis de comprendre pourquoi Asf1 interagit avec la paire H3/H4 et non avec le tétramère (Figure 19). En effet, la région de l'histone H3 impliquée dans la formation de l'interaction de type « four-helix bundle » avec le second dimère H3/H4 interagit avec le feuillet central de Asf1, alors que la zone d'interaction

de H4 impliquée dans la fixation de la paire H2A/H2B subit également de forts changements conformationnels (English et al., 2006).



**Figure 19 : Structure du complexe chaperonne d'histone Asf1/H3/H4**

*Asf1 interagit avec la paire d'histone H3/H4 au niveau des sites de formation du tétramère (H3/H4)<sub>2</sub> et d'association de la paire H2A/H2B (adapté de English et al., 2006)*

Ainsi, les mécanismes d'action des différentes chaperonnes d'histone sont divers, une même chaperonne pouvant être impliquée dans plusieurs mécanismes. De même, il existe un grand nombre de chaperonnes d'histones, toutes n'étant pas encore connues à l'heure actuelle. Pour la plupart d'entre elles, leur mode de fonctionnement reste encore à définir, c'est notamment le cas de la chaperonne Spt6 qui est la protéine centrale de ce travail de thèse.

## 1.3. Fonctions des protéines Spt6, Iws1 et de leur complexe

Après la première identification des protéines Spt6 et Iws1 chez la levure, différentes études ont permis de mettre en évidence les différentes fonctions de ces protéines : (i) chaperonne d'histone, (ii) facteur d'élongation de la transcription, (iii) plateforme de recrutement de facteurs d'export et de maturation des ARN messagers et (iv) implication dans la méthylation des histones. Ainsi, ces deux protéines sont fortement impliquées dans la régulation de la transcription en agissant notamment sur la structure de la chromatine.

### 1.3.1. Identification des protéines Spt6 et Iws1 chez la levure

#### 1.3.1.1. Les protéines de la famille SPT

La protéine Spt6 a été identifiée au cours de trois études indépendantes réalisées chez la levure. Dans ces trois études, des mutations de Spt6 ont permis de supprimer l'effet de mutations introduites dans des gènes impliqués dans le métabolisme de la levure. Ainsi, la mutation de Spt6 (Winston et al., 1984), mais également la surexpression de la protéine (Clark-Adams et Winston, 1987), permet de rétablir l'expression du gène *HIS4* dont l'altération est causée par l'insertion d'un élément transposable  $\delta$  (qui dérive des éléments Ty) dans les éléments régulateurs en 5' du gène *HIS4* (phénotype Spt-). La présence des signaux d'initiation de la transcription de l'élément transposable  $\delta$  dans la région 5' non codante du gène *HIS4* va conduire à la production d'un transcrit non fonctionnel, la mutation des facteurs de transcription qui reconnaissent ces éléments initiateurs permettant de rétablir l'expression du gène adjacent (Fassler et Winston, 1988). De nombreux gènes (nommés SPT pour Suppressor of Ty, présentés dans le Tableau 1) ont pu être identifiés par ce type d'études, certains gènes codant pour des protéines essentielles et notamment les protéines histones H2A et H2B ou la protéine TBP.

La mutation de Spt6 permet également (i) de supprimer l'effet de la mutations des gènes *SNF2* et *SNF5*, qui sont nécessaires à l'activation de l'expression du gène *SUC2* (codant pour une invertase) en absence de glucose (Neigeborn et al., 1986) et (ii) permet également l'expression du gène *ADH2* malgré l'absence de glucose et en l'absence de son activateur transcriptionnel ADR1 (Denis, 1984). Ces premières constatations suggèrent déjà



un rôle de régulateur de la transcription de Spt6, la mutation de la protéine permettant de supprimer l'effet de la mutation des régulateurs des gènes considérés.

Gène SPT	Protéine	Gene essentiel	Fonction
Spt1	HIR2	Non	Sous-unité du complexe HIR, complexe qui a un rôle de facteur d'assemblage des nucléosomes impliqué dans la régulation de la transcription des gènes des histones.
Spt2	SIN1, EXA1	Non	Protéine impliquée dans la régulation négative de la transcription. Nécessaire pour la polyadénylation des ARN, interagit de manière régulée avec les histones et le complexe Swi/Snf. Similarité avec la protéine mammifère HMG1.
Spt3	Spt3	Non	Sous-unité du complexe SAGA et du complexe de régulation de la transcription SAGA-like.
Spt4	Spt4	Non	Sous-unité du complexe DSIF. Impliqué dans la régulation de la transcription par l'ARN polymérase I et II, le processing des pre-mRNA, la fonction du kinétochore et le silencing des gènes. Forme un complexe avec Spt5.
Spt5	Spt5	Oui	Sous-unité du complexe DSIF. Agit en complexe avec Spt4.
Spt6	Spt6	Oui	Chaperonne d'histone, activateur de la transcription
Spt7	Spt7	Non	Sous-unité du complexe SAGA
Spt8	Spt8	Non	Sous-unité du complexe SAGA
Spt10	Spt10	Non	Histone acétylase
Spt11/HTA1	H2A	Non (*)	Histone H2A
Spt12/HTA2	H2B	Non (*)	Histone H2B
Spt13	Med15	Non	Sous-unité du médiateur
Spt14	Spt14	Oui	Régulateur de l'expression de plusieurs gènes
Spt15	TBP	Oui	Tata-Binding protein (TBP)
Spt16	Spt16	Oui	Sous-unité du complexe FACT (avec Pob3)
Spt20	Spt20/Ada5	Non	Sous-unité du complexe SAGA
Spt21	Spt21	Non	Régulateur de l'expression de plusieurs gènes

**Tableau 1 : Gènes de la famille SPT (adapté de Yamaguchi et al., 2001).**

*Les gènes de la famille SPT sont tous impliqués dans les processus de régulation de la transcription. (\*) La suppression d'une copie des gènes codants pour les protéines histones n'est pas létale puisqu'il existe plusieurs copies de chaque gène, ce qui n'est pas le cas de la suppression lors de la suppression de toutes les copies de ces gènes (Clark-Adams et al., 1988).*

Ces études réalisées chez *Saccharomyces cerevisiae* montrent également que le gène de la protéine Spt6 est essentiel (Clark-Adams et Winston, 1987), le remplacement du gène codant la protéine par un marqueur d'auxotrophie conduisant à l'apparition d'un phénotype

létal des souches homozygotes. L'analyse de la protéine produite montre qu'il s'agit d'une protéine nucléaire possédant une extrémité amino-terminale acide (Swanson et al., 1990).

#### 1.3.1.2. Identification de Iws1 et du complexe Spt6/Iws1

La protéine Iws1 (Interact with Spt6) a été identifiée par son rôle dans le mécanisme de la transcription lié à la protéine TBP (Fischbeck et al., 2002). En effet, une mutation de Iws1 (K192N) permet de supprimer l'effet d'une mutation particulière de la protéine TBP à cause de laquelle cette protéine TBP mutée est recrutée au niveau du promoteur des gènes mais ne peut plus participer à l'activation de la transcription (Fischbeck et al., 2002). Cette étude a également démontré que Iws1 est une protéine essentielle qui fait partie de la famille des protéines SPT, la délétion du gène *SPNI* étant létal et la mutation K192N permettant de supprimer l'effet de l'insertion d'un élément  $\delta$  dans les gènes *HIS4* et *LYS2*. Dans des cellules de souris, il a été démontré que Iws1 est une protéine nucléaire essentielle dont les taux d'expression sont différents dans tous les tissus (Liu et al., 2007).

L'interaction entre Spt6 et Iws1 a été identifiée dans un premier temps par des études de protéomiques réalisées chez la levure (Krogan et al., 2002; Gavin et al., 2002). La formation du complexe a, par la suite, été confirmée dans les cellules de souris, bien qu'il semble que l'interaction mise en place soit plus faible que celle impliquant les protéines de levure (Yoh et al., 2007, 2008) En effet, ces dernières études montrent que l'interaction de Spt6 avec le CTD de l'ARN polymérase II est nécessaire au recrutement de Iws1 dans ces cellules. De plus, l'implication de l'extrémité amino-terminale de Spt6 et d'une construction du domaine carboxy-terminal de Iws1 dans la formation du complexe a été démontrée à partir des protéines de souris par la réalisation d'expériences de GST pulldown (Yoh et al., 2007).

#### 1.3.2. Rôle de Spt6 et Iws1 dans la modulation de la structure de la chromatine.

Spt6 et Iws1 ont très largement été impliquées dans la modulation de la structure de la chromatine, par la mise en place de différents mécanismes : une activité de chaperonne d'histone pour Spt6 et de recrutement de facteurs épigénétiques et de remodelage de la chromatine pour le complexe Spt6/Iws1.

##### 1.3.2.1. Rôle de chaperonne d'histone – interaction avec l'histone H3

Différentes observations ont conduits très tôt à suggérer un rôle de chaperonne d'histone de Spt6 (Bortvin et Winston, 1996). Tout d'abord, des liens forts ont été démontrés

chez la levure entre Spt6 et plusieurs sous-unité du facteur de remodelage de la chromatine Swi/Snf (Neugeborn et al., 1986; Laurent et al., 1991; Happel et al., 1991), notamment par la suppression des phénotypes causés par la mutation des gènes *SNF2* et *SNF5* qui codent pour deux sous-unités du complexe Swi/Snf (cas du gène *SUC2*, présenté au paragraphe 1.3.1.1). De plus, différents liens ont également été démontrés avec les protéines histones, les phénotypes produits par la suppression des gènes *SPT11* (H2A), *SPT12* (H2B) et *SPT6* étant identiques (Fassler et Winston, 1988). En effet, (i) la mutation des trois gènes conduits à l'apparition de phénotype Spt causée par l'insertion d'un élément  $\delta$  uniquement (Fassler et Winston, 1988), (ii) la surexpression des trois gènes supprime le phénotype Spt- (Clark-Adams et al., 1988), (iii) les trois gènes ont un lien avec les mêmes sous-unités du complexe Swi/Snf (Neugeborn et al., 1986; Hirschhorn et al., 1992) et (iv) la région codante du gène *SUC2* présente le même profil de dégradation par la nucléase micrococcocal lorsque chacun des trois gènes est délété dans une souche portant un gène *SNF5* muté, montrant une modification de la structure de la chromatine (Bortvin et Winston, 1996).

L'interaction entre H3 et Spt6 a par la suite été démontrée dans la levure (i) par la complémentation du phénotype thermosensible d'une souche *spt6-14* par la surexpression de l'histone H3, (ii) par la surexpression d'un mutant du domaine globulaire de H3 ne permettant plus de compléter ce phénotype et (iii) *in vitro* par la réalisation de tests d'interaction et d'expériences de Far Western blot (Bortvin et Winston, 1996). L'implication de Spt6 dans la reconstitution des nucléosomes a été démontrée par la capacité de la protéine à enrouler partiellement un ADN plasmidique autour des protéines histones en présence d'une topoisomérase mais en l'absence d'ATP. Cette étude a ainsi montré que Spt6 participe au contrôle de la structure de la chromatine par une interaction directe avec la paire d'histone H3/H4 et plus particulièrement le domaine globulaire de H3. Cette interaction a par la suite été confirmée par des expériences de pulldown qui ont également permis de suggérer qu'elle impliquait l'extrémité carboxy-terminale de Spt6 (Winkler et al., 2000).

### 1.3.2.2. Implication des facteurs de remodelage de la chromatine

Des facteurs épigénétiques et de remodelage de la chromatine ATP dépendant participent également à l'activité de modulation de la structure de la chromatine du complexe Spt6/Iws1. En effet, en plus des interactions génétiques observées entre *SPT6* et les gènes *SNF2* et *SNF5*, chez la levure, il a été démontré que Iws1 est présente sur le promoteur du gène inductible *CYCI* avant induction (Zhang et al., 2008). Sur ce promoteur, l'ARN

polymérase II dont la sérine 5 est phosphorylée et les protéines nécessaires à la modification de l'extrémité 5' de l'ARN messenger sont présentes. Suite à l'induction, il a été démontré que Spt6 était recrutée au niveau de ce promoteur permettant alors le recrutement du complexe de remodelage de la chromatine ATP dépendant Swi/Snf. Il a également été démontré que Spt6 participait au même processus que le complexe de remodelage INO80 dont le rôle n'est pour l'instant que peu défini (Klopf et al., 2009).

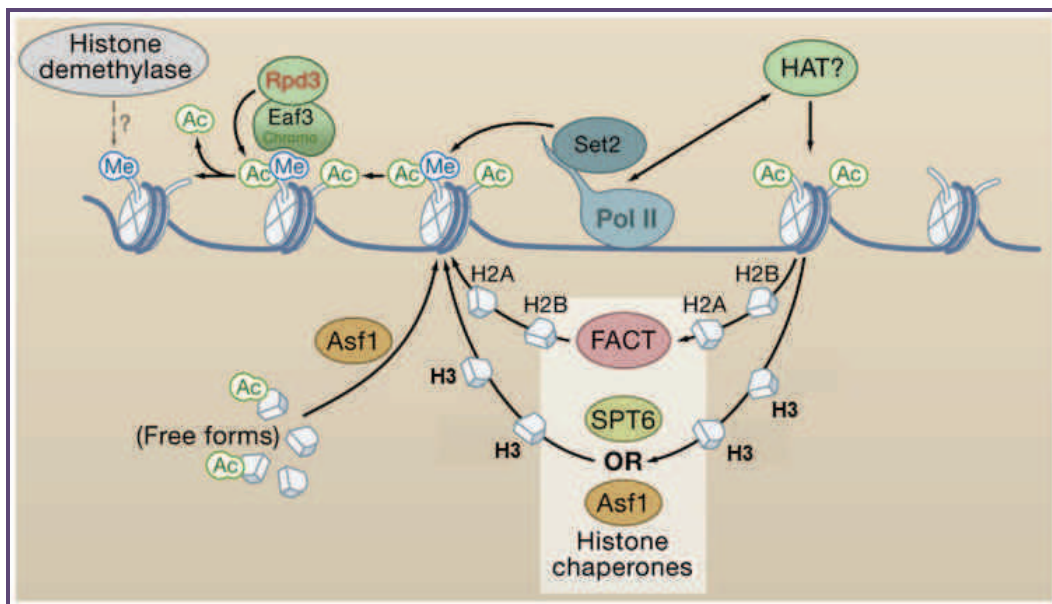
### 1.3.2.3. Spt6 est impliquée dans la reconstitution des nucléosomes

Le caractère essentiel du rôle de chaperonne d'histone de Spt6 a été démontré chez la levure, la mutation de Spt6 conduisant à l'apparition de transcrits courts de certains gènes, notamment des gènes *FLO8* et *RADI*. Ces transcrits résultent d'une initiation réalisée à partir de sites d'initiation cryptiques dans le corps des gènes et reflètent un défaut de reconstitution des nucléosomes dans le sillage de l'ARN polymérase II (Kaplan et al., 2003). Des études réalisées sur tous les gènes de levure ont montré que Spt6 était impliquée dans l'inhibition des sites cryptiques de plus de 1000 gènes (Cheung et al., 2008), l'expression d'une protéine Spt6 mutée montrant une forte diminution de l'occupation des nucléosomes sur un certains nombres de gènes (Ivanovska et al., 2011), montrant ainsi la très large implication de Spt6 dans le mécanisme de reconstitution de la chromatine.

Plusieurs études ont également démontré que Spt6 participait à la régulation de la transcription en modulant la structure de la chromatine, indépendamment de son rôle de facteur d'élongation. En effet, Spt6 participe à la déposition de l'histone H3 sur le gène de choc thermique *HSP104* suite au retour à l'état réprimé du gène, consécutif au choc thermique (Jensen et al., 2008). Chez la levure, il a également été démontré que Spt6 participe à la régulation de la transcription de plusieurs gènes, et notamment du gène *PHO5* en permettant la reconstitution des nucléosomes sur le promoteur de ce gène lors de l'addition de phosphate dans le milieu, conduisant à la répression de la transcription de ce gène (Adkins et Tyler, 2006). Spt6 intervient également dans la répression de la transcription du gène *SER3* induite par un ARN non codant (ARNnc), *SRGI* (Hainer et al., 2011; Thebault et al., 2011; Pruneski et al., 2011). Le gène de cet ARNnc est localisé dans la région intergénique en 5' du gène *SER3*. Lorsque le gène *SER3* est réprimé, la transcription de *SRGI* permet le positionnement d'un nucléosome au niveau du promoteur du gène *SER3*, plusieurs protéines ayant été impliquées dans ce processus et notamment Spt6, toutes ayant un impact sur la reconstitution

du nucléosome présent sur le promoteur du gène *SER3* : les protéines Spt6, Spt2, Spt16 et le complexe PAF.

Ces différentes études ont clairement permis de mettre en évidence l'influence des protéines Spt6 et Iws1 sur la structure de la chromatine (Figure 20). La protéine Spt6 est impliquée directement dans ce processus par son rôle de chaperonne d'histone, les deux protéines participant également par le recrutement de facteurs de remodelage de la chromatine. Il semble ainsi que Spt6 participe à la reconstitution des nucléosomes dans le sillage de l'ARN polymérase II permettant notamment la régulation de la transcription des gènes de chocs thermiques et sur les promoteurs, permettant l'adaptation de l'expression des gènes en réponse à l'environnement. Cependant, ce rôle n'a été observé que dans le cas particulier des gènes inductibles, l'implication de ces deux protéines dans la structure de la chromatine au niveau de tous les gènes restant à définir. De plus, les différents défauts de reconstitution des nucléosomes causés par la perte de Spt6 sont très similaires aux défauts causés par la perte du complexe FACT, suggérant que le mode d'action de Spt6 et de FACT pourrait être similaire.



**Figure 20 : Rôle de chaperonne d'histone de la protéine Spt6**

*La protéine Spt6 participe à la reconstitution des nucléosomes dans le sillage de l'ARN polymérase II (adapté de Li et al., 2007).*

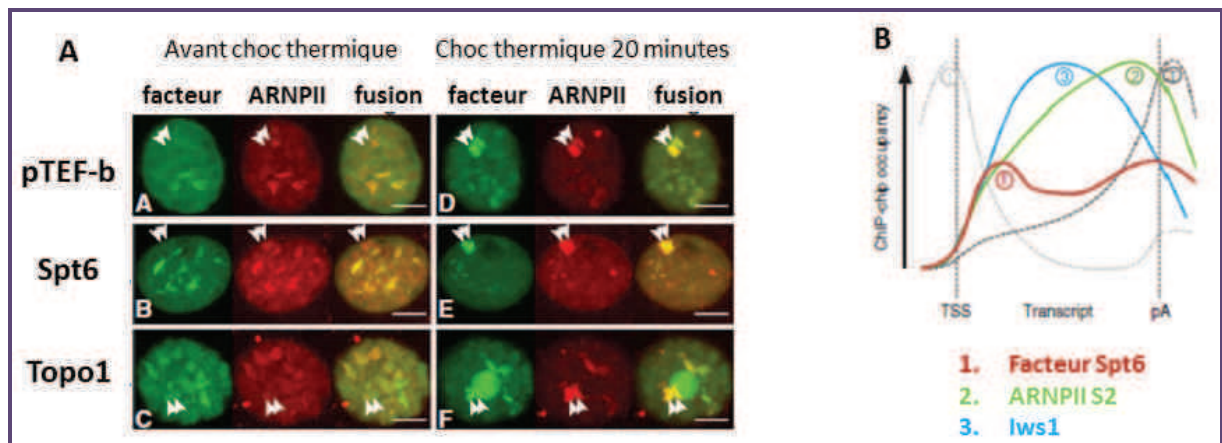
### 1.3.3. Rôle de Spt6 et Iws1 dans l'élongation de la transcription

Les protéines Spt6 et Iws1 participent également à l'élongation de la transcription. Cette activité a été suggérée par (i) l'observation d'une interaction génétique entre les deux protéines et le facteur d'élongation TFIIIS, certaines mutations de Spt6 et Iws1 conduisant à l'apparition de phénotypes thermosensibles lorsqu'elles sont croisées avec une souche *dst1Δ* et (ii) l'apparition dans des souches portant des gènes Spt6 et Iws1 mutés d'une sensibilité à la 6-azauracil, un composé chimique qui réduit les quantités de GTP et UTP présents dans la cellule et induit un arrêt de la transcription (Hartzog et al., 1998; Fischbeck et al., 2002; Zhang et al., 2008).

#### 1.3.3.1. Colocalisation et interaction de Spt6 et Iws1 avec l'ARN polymérase II hyperphosphorylée

La colocalisation de Spt6 avec l'ARN polymérase II hyperphosphorylée a été observée au niveau des chromosomes polytènes chez la drosophile (Kaplan et al., 2000; Andrulis et al., 2000), confirmant l'implication de cette protéine dans la transcription. L'analyse du recrutement de la protéine portant une étiquette fluorescente au niveau des gènes de choc thermique dans les cellules de drosophile a permis d'observer la dynamique de son recrutement (Zobeck et al., 2010). Ces études ont permis de montrer qu'en l'absence de choc thermique, la protéine Spt6 colocalise avec l'ARN polymérase II au niveau des gènes impliqués dans le développement de ces organismes (Figure 21A), la réalisation d'un choc thermique conduisant au recrutement rapide de ces deux protéines au niveau du gène *Hsp70*.

Ainsi, séquentiellement, la protéine HSF est d'abord recrutée (localisation sur le gène 20 secondes après la réalisation du choc thermique), permettant alors le recrutement de l'ARN polymérase II et du facteur PTEF-b (100 secondes après la réalisation du choc thermique) et finalement le facteur Spt6 et la topoisomérase I, qui sont recrutés 20 secondes après l'ARN polymérase II. Le taux de recrutement identique de Spt6 et de l'ARN polymérase II suggère un recrutement parallèle de ces deux protéines.



**Figure 21 : Colocalisation de Spt6 avec l'ARN polymérase II**

(A) L'ARN polymérase II et Spt6 sont colocalisés au niveau des gènes transcrits et recrutés rapidement après un choc thermique de 20 minutes (adapté de Zobeck et al., 2010). (B) Les profils de recrutement de Spt6 et Iws1 sont proches du profil de recrutement de l'ARN polymérase II hyperphosphorylée sur les sérines 2 (ARNPII S2, adapté de Mayer et al., 2010).

Les études d'immunoprécipitation de la chromatine montrent des profils de recrutement de Spt6, Iws1 et de l'ARN polymérase II hyperphosphorylée sur la sérine 2 relativement proches (Figure 21B) (Kim et al., 2004; Mayer et al., 2010; Ivanovska et al., 2011). Le signal des protéines Spt6 et Iws1 apparaît 50pb après le site d'initiation de la transcription (TSS, Transcription Start Site) parallèlement à l'ARN polymérase phosphorylée sur les sérines 2, le taux de d'occupation de ces deux protéines au niveau de la région codante étant différent. En effet, alors que Spt6 atteint un plateau, la quantité de Iws1 augmente tout au long de la phase codante, le taux d'occupation des deux protéines baissant rapidement au-delà des sites de terminaison (Mayer, 2010). Il a par contre été démontré chez la levure que la colocalisation de Spt6 avec l'ARN polymérase II se limitait aux gènes fortement transcrits (Ivanovska et al., 2011).

L'interaction directe de Spt6 avec le CTD de l'ARN polymérase II hyperphosphorylée a été démontrée par co-immunoprécipitation de ces deux facteurs (Yoh et al., 2007) cette interaction ayant été suggérée par l'observation de la colocalisation des deux protéines. L'expression dans des cellules mammifères de différentes constructions de Spt6 suivie de leur purification a permis de mettre en évidence l'implication du domaine carboxy-terminal de Spt6 dans la formation de ce complexe. Ce complexe est aboli lorsque l'arginine 1358 de Spt6 de souris est mutée, démontrant ainsi le rôle direct du domaine SH2 situé dans le domaine carboxy-terminal de Spt6 (Yoh et al., 2007). La spécificité de la reconnaissance de la sérine 2 du CTD a été confirmée par (i) l'abolition de l'interaction lorsque l'ARN polymérase II est

déphosphorylée par la phosphatase Fcp1a, (ii) la formation d'un complexe lorsque les sérines 2 du CTD sont phosphorylées par PTEF-b et (iii) l'absence de complexe lorsque les sérines 5 du CTD sont phosphorylées par le facteur TFIIH (Yoh et al., 2007).

### 1.3.3.2. Spt6 et Iws1 participent et stimulent la phase d'élongation de la transcription

Les études réalisées dans les levures et les cellules de souris ont permis de montrer clairement un rôle de Spt6 dans l'élongation de la transcription, le rôle de Iws1 dans ce mécanisme n'étant pour l'instant que très peu défini, notamment dans les cellules mammifères. Spt6 participe ainsi (i) à la transcription induite par les récepteurs aux œstrogènes (Banahmad et al., 1995), (ii) à la transcription des gènes de HIV1 induite par la protéine Tat (Yoh et al., 2007), (iii) à la transactivation réalisée par la protéine pUL19 du cytomégalo virus humain (Winkler et al., 2000), et (iv) à la transcription des gènes des histones (White et al., 2011). Chez la levure, la protéine Iws1 participe à la transcription des gènes exprimés de manière constitutive et à la régulation de la transcription de certains gènes inductibles (gènes *PHO5*, *HIS3* et *HIS4*) (Fischbeck et al., 2002) et notamment du gène *CYCI* par le recrutement du complexe de remodelage de la chromatine Swi/Snf (Zhang et al., 2008). D'autre part, dans les plantes, Iws1 participe à la régulation de la transcription des gènes en réponse à l'activation du récepteur aux brassinostéroïdes, le facteur de transcription BES1 permettant de recruter Iws1 au niveau du corps des gènes induits (Li et al., 2010). Il a par contre été démontré que dans les cellules de souris, la suppression de l'expression de la protéine Iws1 ne modifiait pas le taux d'expression d'un gène lacZ placé sous le contrôle du promoteur HIV1 en présence de l'activateur Tat (Yoh et al., 2007). Il semblerait ainsi que la protéine Iws1 ne soit pas impliquée directement dans les processus d'élongation de la transcription dans les cellules mammifères.

Il a également été démontré *in vitro* et *in vivo* que Spt6 stimule la phase d'élongation de la transcription. *In vitro*, sur un ADN nu, la déplétion de la protéine d'un extrait nucléaire ralentit fortement l'élongation de la transcription, l'étape d'initiation n'étant pas impactée puisque des transcrits courts (100pb) sont formés même en l'absence de Spt6 (Endoh et al., 2004). L'ajout dans le milieu de la protéine permet de retrouver un niveau de transcription plus élevé, cette complémentation n'étant cependant plus réalisée dans le cas où une protéine Spt6 ne contenant pas d'extrémité carboxy-terminale est ajoutée, confirmant l'implication de ce domaine dans ce mécanisme.



*In vivo*, il a été montré que la suppression de l'expression de Spt6 par l'introduction de siRNA réduisait fortement le taux de transcription du gène de choc thermique *Hsp70* lorsqu'il est induit (Ardehali et al., 2009). La localisation de l'ARN polymérase II au niveau de ce gène en l'absence de Spt6 par des expériences de CHIP révèle une augmentation de l'occupation de l'ARN polymérase II au niveau du site d'initiation de la transcription et une diminution du taux d'occupation au milieu de la phase codante et en 3' du transcrit, reflétant une diminution du nombre de molécules d'ARN polymérase II entrant en phase d'élongation productive. Cette diminution de la vitesse de synthèse de l'ARN polymérase est également détectée par des expériences de FRAP (Fluorescence Recovery After Photobleaching) qui révèlent une diminution du taux de transcription de moitié dans les cellules n'exprimant pas Spt6 (diminuant à 480pb/minutes, contre 1000pb/min dans les cellules contenant Spt6) (Ardehali et al., 2009).

#### 1.3.3.3. Implication des facteurs épigénétiques dans la régulation transcriptionnelle induite par Spt6 et Iws1

Plusieurs études ont également démontré une implication du complexe Spt6/Iws1 dans le dépôt de marques épigénétiques liées à la régulation de la transcription. Chez la levure et dans des cellules de souris, le complexe est ainsi impliqué dans la triméthylation de la lysine 36 de l'histone H3 par les histone méthyltransférases Set2 (chez la levure) et Setd2 chez les eucaryotes supérieurs, cette marque épigénétique reflétant une transcription active (Youdell et al., 2008; Yoh et al., 2008). Chez la souris, plus particulièrement, la suppression de l'expression de Iws1 par siRNA conduit à une diminution spécifique du taux de H3K36me3 au niveau du promoteur et de la phase codante du gène *c-myc*, également observée lors de la suppression de l'expression de la protéine Setd2. Une interaction directe entre Iws1 et Setd2 a pu être détectée *in vivo*, qui implique également l'ARN polymérase II hyperphosphorylée et Spt6.

La mutation de Iws1 et Setd2 conduit également à une augmentation du taux de triméthylation de la lysine 27 de l'histone H3 à l'extrémité 5' du gène *PABPC1* (Yoh et al., 2008). Dans les plantes, il a également été démontré que Iws1 participait à la régulation de la transcription du gène codant pour le transporteur NRT2.1 impliqué dans la morphologie cellulaire en réponse aux quantités d'azote présentes dans le milieu. Cette répression est notamment modulée par le dépôt de marques épigénétiques au niveau du gène de ce

transporteur, le taux de H3K27me3 étant augmenté en présence de grandes quantités d'azote qui conduit à la répression de la transcription de ce gène (Widiez et al., 2011).

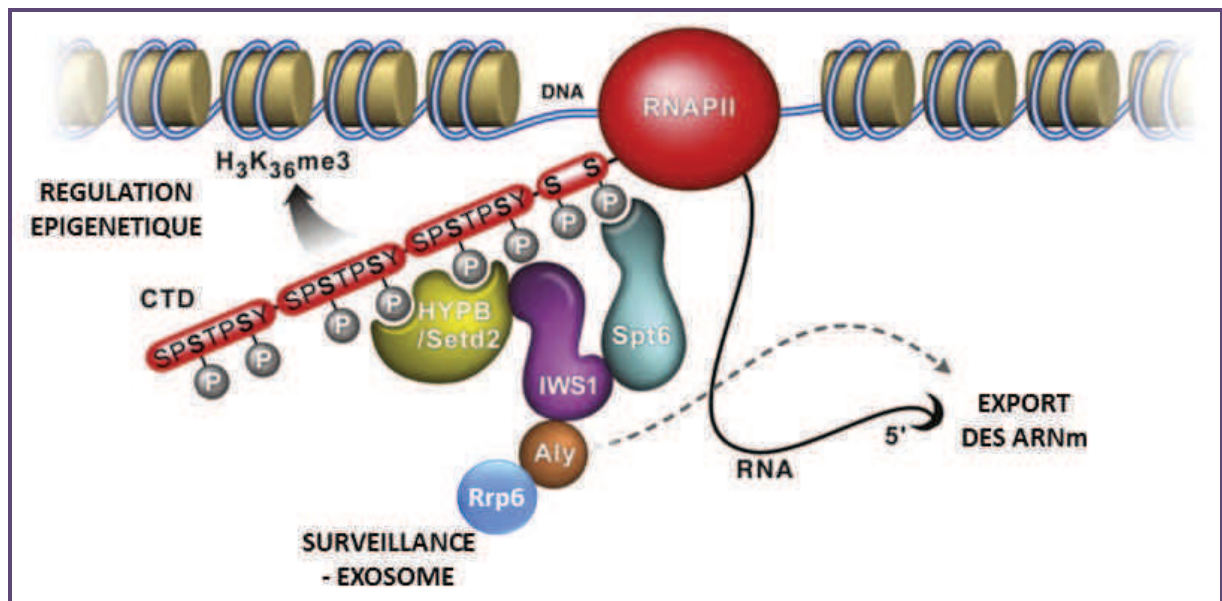
Ainsi, les protéines Spt6 et Iws1 participent également au processus d'élongation de la transcription réalisée par l'ARN polymérase II en activant notamment la transcription de plusieurs gènes et en stimulant la transcription réalisée par l'ARN polymérase. Cependant, de la même manière que dans le cas de l'activité de chaperonne d'histone, l'activité de facteur d'élongation de Spt6 semble être restreinte à un certain nombre de gènes, probablement ceux fortement transcrits. La régulation de la transcription des gènes, médiée par Spt6 et Iws1, implique également le dépôt de marques épigénétiques qui vont refléter une transcription active (cas de H3K36me3) ou permettre le recrutement de facteurs de remodelage qui permettront la recondensation de la chromatine (H3K27me3) suggérant un lien avec le rôle de modulateur de la structure de la chromatine de la protéine Spt6. Cependant, les gènes présentant une forte diminution de l'association des nucléosomes dans une souche *spt6-1004* ne sont pas forcément ceux dont la transcription est fortement dérégulée dans ces souches (Ivanovska et al., 2011). Ces résultats suggèrent que les activités de régulateur de la transcription et de modulateur de la structure de la chromatine de Spt6 pourraient être en partie distinctes.

#### 1.3.4. Le complexe Spt6/Iws1 est impliqué dans les mécanismes co- et post-transcriptionnels

Lorsque les protéines Spt6 et Iws1 interagissent avec l'ARN polymérase II, elles vont aussi recruter des protéines impliquées dans les mécanismes de maturation, d'export et de surveillance des ARN messagers. En effet, chez la levure, une mutation de Spt6 conduit à un défaut de reconnaissance des sites de terminaison à l'extrémité du gène GAL10 couplé au gène GAL7, impliquant la sous-unité Ctr9 du complexe Paf1 (Kaplan et al., 2005). Ce phénomène est notamment expliqué par une accélération de l'ARN polymérase II à l'extrémité 3' des gènes (Ardehali et al., 2009). De plus, *in vivo*, la mutation du domaine SH2 de Spt6, impliqué dans l'interaction avec l'ARN polymérase II, et la suppression de l'expression de Iws1, conduisent à des défauts d'épissage et d'export qui sont reflétés par l'accumulation de transcrits longs et polyadénylés dans le noyau (Yoh et al., 2007). Ces observations ont permis d'identifier dans les cellules de souris une interaction entre Iws1 et le facteur d'export REF1/Aly et la sous-unité Rrp6 de l'exosome, toutes deux impliquant

l'extrémité carboxy-terminal de Iws1. L'interaction de Spt6 avec l'exosome avait été suggérée précédemment chez la drosophile, la réalisation d'expériences de pulldown à partir d'une protéine Spt6 portant une étiquette de purification ayant permis d'identifier les neuf sous-unité de l'exosome dans la fraction retenue (Andrulis et al., 2002).

L'observation de l'implication de Spt6 et Iws1 dans la transcription et la mise en place des mécanismes co- et post-transcriptionnels fait ainsi apparaître un rôle important de ce complexe dans la régulation de ces mécanismes. Ainsi, un modèle a pu être établi qui suggère que le complexe est recruté au niveau du corps des gènes par le CTD de l'ARN polymérase dont les sérines 2 sont phosphorylées. Des interactions sont également mises en place avec d'autres protéines de la machinerie transcriptionnelle et notamment la sous-unité Spt5 du complexe DSIF (Lindstrom et al., 2003) et le complexe PAF (Pruneski et al., 2011), plusieurs interactions génétiques ayant été observées. Le complexe Spt6/Iws1 lié à l'ARN polymérase II va alors permettre le recrutement de facteurs épigénétiques (Setd2), de facteurs d'export (REF1/Aly) et de facteurs impliqués dans les processus de surveillance des ARN messagers (exosome, sous-unité Rrp6) (Figure 22).



**Figure 22 : Implication de Spt6 et Iws1 dans la transcription, la régulation épigénétique, l'export des ARNm et la surveillance**

*Spt6 interagit directement avec l'ARN polymérase II dont les sérines en position 2 sont phosphorylées.*

*Cette interaction va permettre le recrutement de Iws1 mais également d'autres facteurs impliqués dans la modification des histones, l'export et la maturation des ARN messagers (adapté de Yoh et al., 2008).*

Cependant, l'observation que chez la levure, Iws1 peut être recrutée au niveau des promoteurs en l'absence de Spt6 et que le complexe Spt6/Iws1 dans les cellules de souris n'est formé que lorsque Spt6 interagit avec le CTD de l'ARN polymérase II, ce qui n'est pas le cas chez la levure, montre que ce modèle est encore plus complexe.

### 1.3.5. Autres fonctions de Spt6

Différentes fonctions de la protéine Spt6, non liées à celles décrites précédemment ont également été identifiées. Ainsi, Spt6 participe au mécanisme de recombinaison chez la levure, le taux de recombinaison homologue réalisé par Rad51 chez la levure étant augmenté lorsque Spt6 est mutée (Malagón et Aguilera, 1996). Cette fonction pourrait être liée au rôle de chaperonne d'histone de Spt6, la mutation des gènes de H2A et H2B causant le même type de phénotype.

Spt6 intervient également dans la régulation de l'activité de la protéine AID, une cytidine déaminase qui joue un rôle important au cours de la réponse immunitaire. Cette protéine introduit des cassures doubles brins dans l'ADN des lymphocytes B dans les gènes codants pour les anticorps. Ces cassures vont induire (i) l'hypermutation somatique qui augmente l'affinité de l'anticorps pour l'antigène mais aussi (ii) la recombinaison de classe qui a pour but de produire des IgA, E et D à partir des IgM et D produits au début de l'infection. Spt6 a été impliqué dans ce processus par une interaction directe entre son domaine carboxy-terminal et AID, Spt6 participant spécifiquement au mécanisme de recombinaison de classe (Okazaki et al., 2011). Aucun lien n'a pour l'instant été décrit entre ces deux fonctions de Spt6 et l'implication de la protéine dans la transcription et la structure de la chromatine.

## 1.4. But de l'étude – méthodologie employée

Au début de ces travaux de thèse, les différentes fonctions de Spt6 et Iws1 et du complexe Spt6/Iws1 avaient déjà partiellement été décrites. Les rôles de chaperonne d'histone et de modulateur de la structure de la chromatine, de facteur d'élongation de la transcription et de facteur impliqués dans la modification co- et post-transcriptionnelle avait déjà été décrits, le rôle dans la régulation n'ayant alors pas encore été caractérisé. Cependant, les liens unissant ces différentes fonctions n'étaient pas connus, rendant difficile la compréhension du rôle de ces protéines dans ces différents mécanismes. De plus, Spt6, Iws1 et le complexe formé par ces deux protéines n'était que peu caractérisé biochimiquement, aucune structure

n'étant alors disponible. Les travaux de cette thèse ont donc eu pour but de caractériser biochimiquement, structuralement et fonctionnellement ces deux protéines et leur complexe, ainsi que la mise en évidence de leurs liens structure/fonction pour permettre notamment de mieux comprendre le rôle de ces protéines dans la transcription.

Afin de réaliser ces études, nous nous sommes appuyés sur deux approches qui avaient déjà été utilisées dans l'équipe. Premièrement, l'utilisation de la technique de co-expression, qui permet de mettre rapidement en évidence les complexes formés et de les caractériser biochimiquement (Fribourg et al., 2001; Romier et al., 2006). Deuxièmement, l'utilisation des protéines des parasites *Encephalitozoon cuniculi* et *Antonospora locustae*. En effet, l'un des principaux obstacles à la réalisation d'études structurales est la grande flexibilité des protéines de grande taille, comme c'est le cas pour la protéine Spt6 humaine (1726 résidus) et de levure (1451 résidus). Les très petits génomes des deux parasites ont conduits ces organismes (i) à ne conserver que les protéines essentielles à leurs fonctions, certaines étant assurées par l'organisme hôte, et (ii) à réduire la taille de ces protéines afin de conserver uniquement les domaines strictement nécessaires à la réalisation de la fonction de la protéine. Ainsi, les protéines Spt6 et Iws1 sont toutes les deux retrouvées dans ces deux organismes, leur taille étant fortement réduite (environ 900 résidus dans le cas de Spt6). Les protéines de ces parasites avaient déjà été utilisées avec succès lors de l'étude structurale de la sous-unité TAF(II)5 du facteur TFIID (Romier et al., 2007).

Ainsi, en nous appuyant sur les données disponibles sur ces deux protéines et notamment les différents domaines structuraux décrits dans la littérature et identifiés par recherche d'homologie de séquence dans les banques de domaines, l'architecture des protéines Spt6 et Iws1 a pu être définie (Figure 23). Nous nous sommes notamment appuyés sur la protéine bactérienne Tex qui présente une homologie de séquence avec le domaine central de Spt6 (autour de 25% d'homologie de séquence entre les deux protéines) et dont la structure a été résolue par nos compétiteurs au cours de cette thèse (Johnson et al., 2008). Les domaines structuraux de Tex sont retrouvés dans la séquence de Spt6, suggérant que ces deux protéines ont évolué à partir d'un ancêtre commun. La protéine Tex contient notamment différents domaines de liaison aux acides nucléiques (HhH, YqgFc, HtH et S1), le domaine S1 permettant notamment l'interaction de la protéine avec des ARN et ADN double et simple brin.

Ainsi, le champ d'action du complexe Spt6/Iws1 semble être très large mais son rôle exact n'est pour l'instant que peu compris, notamment les liens qui relient ses différentes fonctions. Les travaux de cette thèse ont notamment permis de comprendre en termes moléculaires et fonctionnels (i) comment Spt6 est recrutée par l'ARN polymérase II au cours de la transcription et (ii) comment le complexe Spt6/Iws1 est formé. Ce travail a également permis de suggérer d'autres interactions tel que le complexe Spt6/TFIIS, ouvrant ainsi de nombreuses perspectives scientifiques.

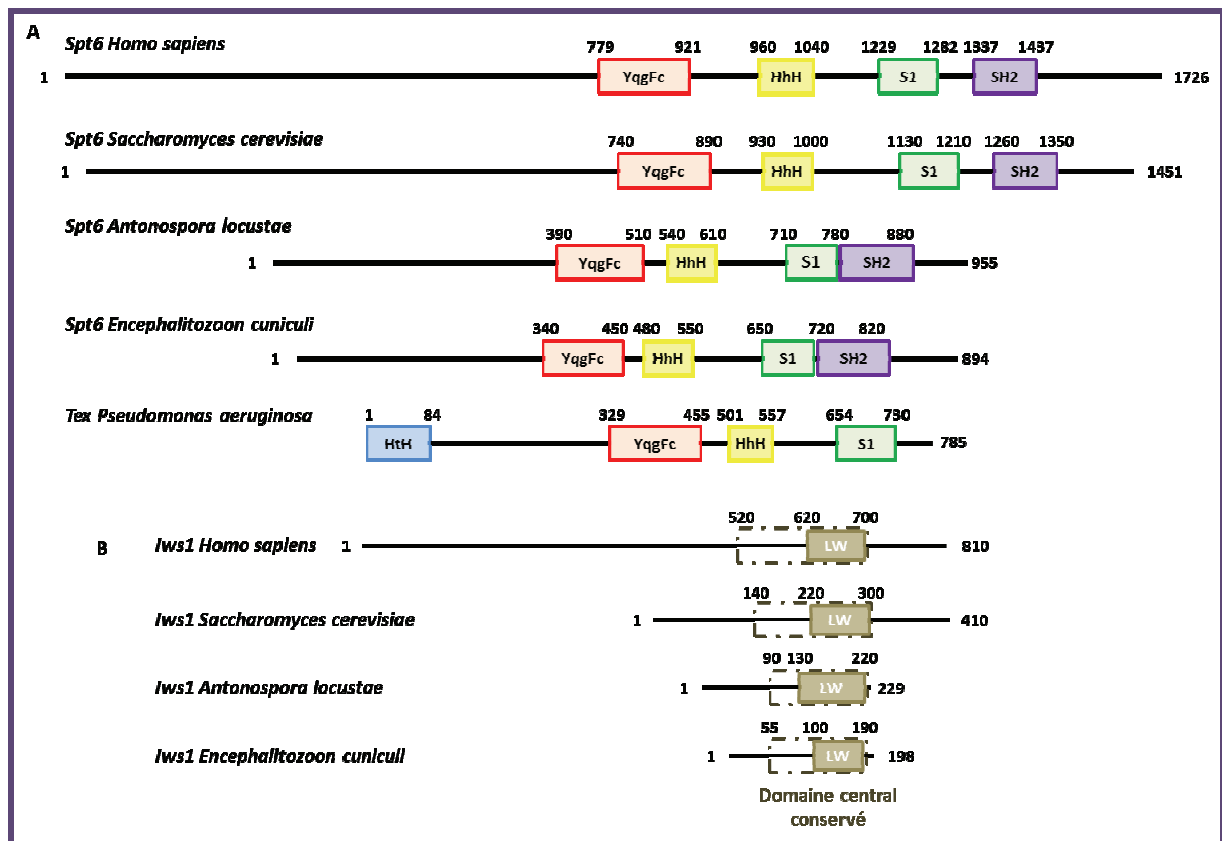


Figure 23 : Architecture des protéines Spt6 et Iws1

Les différents domaines structuraux des protéines Spt6 et Iws1 qui ont pu être identifiés au début de ces travaux de thèse sont indiqués. Leur présence avait été définie par recherche dans des banques de données et par homologie avec la protéine bactérienne Tex.



# **Chapitre 2**

## **Matériel et Méthodes**





Notre but a été de caractériser structurellement et fonctionnellement les différentes protéines et complexes que nous étudions. La réalisation de ces études nécessite la mise en place d'un large panel de techniques recouvrant la biologie moléculaire, la production, la purification, la caractérisation biophysique, la caractérisation fonctionnelle et la résolution de la structure tridimensionnelle des protéines et complexes étudiés. Toutes ces études sont initiées par une étape de biologie moléculaire qui a pour but le clonage des gènes des protéines d'intérêts dans des vecteurs d'expression. Ces étapes de clonages aboutissent à la réalisation de tests d'expression, suivis de la production des protéines ou complexes en grand volume. Une fois la protéine produite, elle est purifiée et caractérisée bio-physiquement et sa structure est résolue. En s'appuyant sur l'observation de la structure des protéines ou complexes, des caractérisations fonctionnelles *in vivo* et *in vitro* sont alors envisagées afin de comprendre leur rôle. Cependant, le processus qui mène à la résolution de la structure des protéines n'est pas linéaire et nécessite généralement un grand nombre d'essais avant d'aboutir à la résolution de la structure. Cette partie décrit les principales techniques qui ont été utilisées au cours de cette thèse.

## 2.1. Techniques de biologie moléculaire

### 2.1.1. Génération des vecteurs d'expression bactériens

L'expression de protéines en bactéries permet d'obtenir, à faible coût, de grandes quantités de protéines utilisables pour des études structurales. De plus, l'utilisation d'hôtes bactériens permet de tester facilement la solubilité de constructions différentes des protéines d'intérêt, notamment dans le cas des complexes protéiques. C'est la raison pour laquelle les premiers tests d'expression sont généralement réalisés dans ce système d'expression.

Pour faciliter les clonages initiaux, les gènes d'intérêt sont d'abord clonés dans des vecteurs navettes qui sont des vecteurs à haut nombre de copies. Une fois les gènes amplifiés séquencés ils sont extraits des vecteurs navettes par digestion et insérés dans les vecteurs d'expression.

#### 2.1.1.1. 1<sup>ère</sup> étape : la PCR et la génération des vecteurs navettes

Nos gènes d'intérêt codant pour différentes protéines provenaient de différentes sources. Dans le cas des gènes codants pour des protéines d'*Encephalitozoon cuniculi*, d'*Antonospora locustae* et de *Saccharomyces cerevisiae*, ceux-ci ont été extraits par PCR à

partir des génomes de ces organismes obtenus auprès de laboratoires les produisant. Les autres gènes ont été achetés auprès de banques.

*a) Réaction de polymérisation en chaîne (PCR)*

La réaction de PCR va permettre d'obtenir de grandes quantités d'un fragment d'ADN contenant le gène d'intérêt. Elle est réalisée par une enzyme polymérase, l'ADN polymérase, qui va catalyser l'ajout de désoxyribonucléotides (dNTP) à l'extrémité 5' d'une amorce par complémentarité à la matrice. Les amorces utilisées vont permettre d'introduire aux extrémités du fragment amplifié : un codon initiateur ATG en 5', un codon stop en 3' ainsi que des séquences reconnues spécifiquement par des enzymes de restriction aux deux extrémités du fragment. Les réactions ont été réalisées avec l'enzyme Phusion dans le cas où le fragment est amplifié dans le but de l'insérer dans un vecteur. L'utilisation de cette enzyme est justifiée par le faible taux d'erreurs introduites. Les conditions de la réaction de PCR sont présentées en annexe 8.1.1 (page 259).

Le produit de la réaction est ensuite déposé sur un gel d'agarose, les bandes correspondantes au produit de l'amplification sont découpées et les fragments sont extraits du gel à l'aide du kit Nucleospin® Extract II (Macherey Nagel).

*b) Insertion dans le vecteur navette*

Différents types de vecteurs navettes ont été utilisés, chacun ayant un mode de clonage différent. Selon la longueur du gène et les sites de restrictions choisis, une technique est privilégiée. Les caractéristiques des principaux vecteurs navettes décrits ci-dessous sont récapitulés dans le Tableau 2.

Le vecteur p207 permet l'insertion de fragments digérés avec les enzymes NdeI en 5' et BamHI ou BglII en 3'. Il porte un gène de résistance à la gentamycine. Ce vecteur est une version modifiée du vecteur pDONOR 207 fourni par Invitrogen : les sites de recombinaison homologues *att* présents de part et d'autres de la cassette de clonage du vecteur commercial ont été utilisés pour introduire des sites de restriction NdeI et BamHI ou NdeI et BglII. Il en résulte qu'une fois le fragment d'ADN cloné dans ce vecteur, la recombinaison homologue (décrite au paragraphe 2.1.1.2.d) peut être utilisée afin de transférer le fragment dans le vecteur d'expression de destination. Pour l'insertion de fragments d'ADN dans ce vecteur, les produits de la réaction PCR sont digérés avec les enzymes de restriction, purifiés sur les colonnes Nucleospin Extract II afin d'éliminer le tampon de l'enzyme de restriction et une

réaction de ligation est réalisée avec la T4 DNA ligase (Epicentre). Ce vecteur est utilisé principalement pour les fragments d'ADN courts (jusqu'à 2000 paires de bases) lorsqu'aucun site de digestion par NdeI et BamHI ou BglII n'est présent dans le gène.

Vecteur navette	Type de clonage	Résistance	Application
<b>p207 (dérive de pDONOR 207, Invitrogen)</b>	Restriction/ligation	Gentamycine	- Gènes courts (<2000 pb) ne possédant pas de sites reconnus par les enzymes de restriction NdeI et BamHI ou BglII, - Sous clonage possible par recombinaison homologue.
<b>PCR Blunt (Invitrogen)</b>	Ligation d'extrémités franches	Kanamycine	- Gènes courts (< 2000 pb) contenant des sites reconnus par les enzymes de restriction NdeI et BamHI ou BglII, - Sous-clonage par restriction / ligation.
<b>pCR2.1 TOPO</b>	TOPO TA cloning	Ampicilline	- Gènes longs - Sous-clonage par restriction / ligation

**Tableau 2: Les différents vecteurs navettes employés et leurs principales caractéristiques.**

Le vecteur PCR Blunt fourni par Invitrogen permet l'insertion de fragments d'ADN possédant des extrémités franches, comme c'est le cas avec les fragments amplifiés par l'enzyme Phusion. Les fragments amplifiés par PCR sont ligués dans le vecteur après leur purification. Les vecteurs ayant intégré le gène d'intérêt vont être sélectionnés grâce au gène ccdB. En effet, le vecteur Blunt est ouvert dans le gène codant pour cet inhibiteur de la gyrase bactérienne. Ainsi, seulement si un ADN est intégré dans le vecteur au niveau de la séquence codante du gène ccdB, la protéine inhibitrice ne sera pas produite et la bactérie pourra se développer. Ce vecteur porte également un gène codant pour une résistance à la Kanamycine. Ce vecteur est utilisé pour les fragments d'ADN courts (jusqu'à 2000 paires de bases) dans le cas où le clonage en vecteur p207 n'est pas réalisable du fait de la présence de sites de digestion par les enzymes de restriction NdeI, BamHI et BglII, il peut alors servir de matrice pour une réaction de mutagenèse dirigée permettant de supprimer ces sites de restriction.

Le vecteur pCR2.1 TOPO (Invitrogen) permet également l'insertion de fragments non digérés. A la fin de la réaction de PCR, 0,5µL de la polymérase de *Thermus aquaticus* (Taq) sont ajoutés au produit de la PCR. Cette enzyme va ajouter de manière non spécifique une adénosine aux extrémités 3' des fragments amplifiés, permettant ainsi de produire des extrémités sortantes qui pourront s'hybrider au vecteur. Le vecteur fourni est quant à lui

linéarisé avec à ses extrémités une séquence sur laquelle l'enzyme Topoisomérase I est fixée sur une thymidine sortante. Lors de la réaction de ligation, l'extrémité adénosine du fragment PCR va s'hybrider à la thymidine du vecteur et la Topoisomérase va réaliser la réaction de ligation.



**Figure 24 : Principe de la réaction de TOPO TA-cloning**

*L'enzyme Topoisomérase I est liée par une liaison phosphodiester à l'extrémité des vecteurs fournis. Cette enzyme va liquer le vecteur ouvert avec le fragment issu de la PCR auquel des adénosines ont été ajoutées par traitement à la polymérase Taq (adapté de Invitrogen).*

Ce vecteur navette est principalement utilisé pour le clonage des très grands gènes (au-delà de 2000 paires de bases).

Une fois les réactions de ligation réalisées, les produits sont transformés dans des cellules DH5α (description des souches bactériennes en annexe 8.1.4, page 262) qui sont ensuite étalées sur un milieu LB-agar contenant l'antibiotique dilué au 1000<sup>ème</sup>, permettant de sélectionner les bactéries ayant intégré le vecteur, et laissées une nuit à 37°C. Le lendemain, des colonies sont repiquées et mises en culture dans un milieu liquide contenant l'antibiotique dilué au 1000<sup>ème</sup>. Au bout de 16 heures de culture, les cellules sont collectées par centrifugation et les plasmides sont extraits à l'aide du kit Nucleospin® Plasmid (Macherey Nagel) puis séquencés.

Différentes techniques sont employées afin de vérifier la présence de l'insert cloné dans le vecteur navette avant le séquençage du vecteur. Tout d'abord, une réaction de PCR sur colonies peut être réalisée, qui consiste à réaliser une réaction de PCR sur la culture bactérienne liquide avec des amorces spécifiques au gène, l'enzyme Taq de *Thermus thermophilus* étant utilisée. Au cours de cette réaction, le premier cycle de dénaturation va durer 3 minutes afin de casser les parois bactériennes, la totalité de l'ADN bactérien (génomique et plasmidique) servant alors de matrice. Ainsi, si le gène d'intérêt est présent dans le vecteur il va être amplifié et sera détecté sur gel d'agarose. De plus, après purification

du plasmide, les enzymes dont les sites de reconnaissances ont été placés en 5' et 3' du gène peuvent être utilisées afin de digérer le vecteur. Le dépôt sur gel d'agarose permettra alors de vérifier la présence d'un insert de bonne taille.

#### 2.1.1.2. 2<sup>ème</sup> étape : le sous-clonage en vecteurs d'expression

Une fois le vecteur navette obtenu et vérifié par séquençage, le gène codant pour la protéine d'intérêt peut en être ré-extrait et inséré dans un vecteur d'expression soit par la technique de restriction/ligation, soit par recombinaison homologue. La majorité des vecteurs d'expression que nous avons utilisés vont notamment permettre d'introduire en amont du gène de la protéine (i) un site de fixation de la polymérase T7 qui va transcrire le gène (ce qui n'est pas le cas du vecteur pGexNB, décrit au paragraphe b), (ii) un site de fixation des ribosomes (*rbs*, « Ribosome Binding Site ») qui vont traduire l'ARN messager produit au cours de la transcription et, en fonction des besoins, (iii) une séquence codant pour une étiquette de purification et d'un site de coupure par une protéase qui permettra la purification par chromatographie d'affinité de la protéine. En aval du gène des codons Stops et un terminateur sont placés qui permettent l'arrêt de la traduction et de la transcription respectivement.

##### a) *Les vecteurs d'expression bactériens de la suite pET-MCN*

Les vecteurs utilisés font partie de la suite de vecteurs construits par le Dr. Romier (décrite sur le site web <http://lbgs.u-strasbg.fr/sbgp/Private/TechRes/CloningProto/Vectors/Vector/SiteWebCRO/Introduction.html>). Cette suite de vecteurs permet de tester différentes étiquettes de purifications placées soit à l'extrémité amino-terminale, soit à l'extrémité carboxy-terminale de la protéine ainsi que des sites de coupure par des protéases différents (Tableau 2). Elle permet également la réalisation de tests de co-expression (décrits au paragraphe 2.2.2, page 87), puisque certains de ces vecteurs permettent l'expression de protéines natives, c'est-à-dire ne portant pas d'étiquette de purification et que chacun de ces vecteurs porte des gènes de résistance à un antibiotique et des origines de répllication différentes.

Les vecteurs de cette suite possèdent (i) des cassettes de clonage identiques facilitant le clonage des gènes d'un vecteur navette vers les vecteurs d'expression et (ii) des sites de restriction spécifiques permettant de concaténer les gènes dans un seul vecteur pour tester des interactions à plusieurs partenaires. La procédure de concaténation est décrite en paragraphe

2.1.2, page 80 et dans la publication 3. Parmi les vecteurs de cette suite, le vecteur pnEA-tH, qui permet l'expression de protéines portant une étiquette de purification (His)<sub>6</sub> suivi d'un site de coupure par la thrombine, et le vecteur pnCS, qui permet l'expression d'une protéine native sont principalement utilisés. Les cartes de ces vecteurs sont présentées sur la Figure 71, en annexe 8.1.5 page 262. Ces vecteurs sont générés par restriction/ligation, sauf dans le cas du vecteur pnEA-tH, pour lequel un vecteur spécifique possédant des sites de recombinaison *att* existe, permettant la réalisation d'une réaction de recombinaison homologue.

Vecteur	Dérive du vecteur	Etiquette de purification	Site protéase	Résistance	Origine de réplication
pnEA-tX	pET-15b	(His) <sub>6</sub> GST His-Trx, Trx His-MBP, MBP His-NusA, NusA Strep	Thrombine TEV Protéase 3C	Ampicilline	pBR322
pnCS	pCDF Duet	Expression de protéines natives	-	Spectinomycine	CDF ori
pnYC	pACYC11b	Expression de protéines natives	-	Chloramphénicol	p15A
pnEK	pET-28b	Expression de protéines natives	-	Kanamycine	pBR322

**Tableau 3 : Les vecteurs de co-expression employés.**

*Les différents vecteurs de la suite pET-MCN ainsi que leurs principales caractéristiques sont présentés. Dans le cas des vecteurs pnEA-tX, X représente n'importe quelle étiquette de purification. Les tests de co-expression vont nécessiter la co-transformation d'un vecteur permettant l'expression d'une protéine native avec un vecteur permettant l'expression d'une protéine portant une étiquette de purification.*

*b) Le vecteur pGexNB*

Dans le cadre de la réalisation des tests de GST pulldown (présentés au paragraphe 2.3.2.1), les protéines ont été produites avec une étiquette de purification GST à leur extrémité amino-terminale. Pour la production de ces protéines, le vecteur pGexNB a été utilisé, qui dérive du vecteur pGEX 4T2 (commercialisé par GE Healthcare) qui a été modifié afin de contenir les sites de restriction NdeI et BamHI, permettant le clonage du gène de la protéine d'intérêt par restriction/ligation. Ce vecteur possède une origine de réplication pBR322, un gène de résistance à l'ampicilline et un promoteur tac. Ce promoteur est un hybride constitué du promoteur trp dans lequel la région -35 de la séquence de Shine-Dalgarno est remplacée par celle du promoteur LacUV5 (De Boer et al., 1983). La transcription à partir de ce

promoteur est réalisée par la polymérase bactérienne, le système pouvant être réprimé par l'inhibiteur LacI et induit par l'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (décrit au paragraphe 2.2.1). Ce vecteur peut être co-transformé avec les vecteurs pnCS, pnYC et pnEK de la suite pET-MCN, mais sa cassette ne peut être concaténée dans ces vecteurs.

#### *c) Sous-clonage par restriction/ligation*

Les sites de coupure par les enzymes de restriction qui ont été placés en 5' et 3' du gène codant pour la protéine lors de la PCR sont utilisés pour extraire les gènes du vecteur navette. Le produit de la digestion est alors déposé sur gel d'agarose, les bandes correspondant à l'insert sont découpées du gel et les fragments d'ADN sont extraits de l'agarose avec le kit Nucleospin® Extract II (Macherey Nagel). Le vecteur est quant à lui également digéré avec les mêmes enzymes, déphosphorylé avec la phosphatase SAP (Shrimp Alkaline Phosphatase, Fermentas) pour éviter qu'il ne se recircularise et également purifié sur gel. Une réaction de ligation est alors réalisée avec la T4 DNA ligase (Epicentre) entre l'insert et le vecteur ouvert en respectant un ratio molaire de 5 : 1 (insert : vecteur). Le produit de ligation est transformé au bout de deux heures dans des cellules DH5 $\alpha$  et des clones sont isolés. La présence des inserts dans les vecteurs obtenus sont vérifiés par digestion. Le protocole de sous-clonage par restriction/ligation est présenté en annexe 8.1.3.1, page 261.

Les enzymes généralement utilisées au cours de ces sous-clonages sont NdeI (digestion en 5' du gène) et BamHI ou BglII (digestion en 3'). Le clonage par restriction/ligation est généralement préféré à la recombinaison homologue puisque seuls trois résidus sont présents dans la protéine d'intérêt après coupure par la protéase, tandis que ce sont neuf résidus qui restent fusionnés à la protéine dans le cas d'une réaction de recombinaison homologue.

#### *d) Sous-clonage par recombinaison homologue*

La recombinaison homologue de type LR est parfois utilisée dans le cas du sous-clonage de gènes dans un vecteur pnEA-tH. Cette réaction est basée sur la reconnaissance entre deux sites attL1 et attL2 présents de part et d'autre du gène dans le vecteur navette et deux sites complémentaires attR1 et attR2 présents dans le vecteur d'expression. L'enzyme LR clonase va catalyser l'échange des deux brins (Figure 25), permettant le transfert du gène dans le vecteur d'expression. Le gène ccdB présent dans la cassette de clonage va permettre



de sélectionner les vecteurs recombinés, ce gène étant échangé avec le gène d'intérêt. Le protocole de sous-clonage par restriction/ligation est présenté en annexe 8.1.3.2, page 261.



**Figure 25 : Principe de la réaction LR d'un clonage Gateway**

*Les sites attL présents des deux côtés du gène dans le vecteur navette ont une séquence complémentaire aux sites attR présents dans le vecteur d'expression qui entourent un gène ccdB. La LR clonase va cliver et échanger les deux brins. A la fin de la réaction, le vecteur d'expression contient le gène codant pour la protéine entouré de sites attB et le vecteur navette contient le gène ccdB entouré de sites attP.*

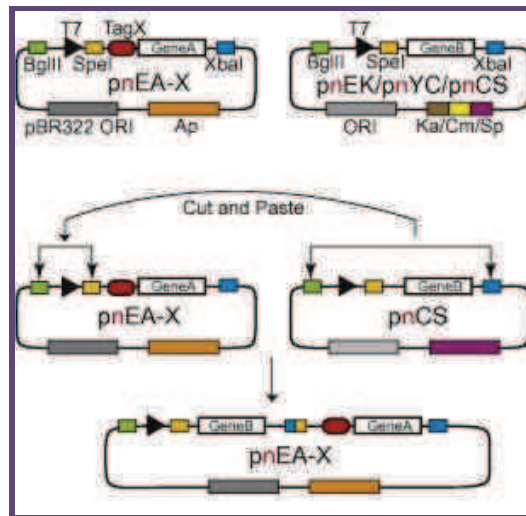
A l'issue de la réaction de recombinaison homologue, le gène codant pour la protéine est placé dans le vecteur en aval de l'étiquette de purification et du site de coupure par la protéase.

### 2.1.2. Concaténation des vecteurs pour l'expression de complexes chez la bactérie

La co-expression par co-transformation permet la production de complexes composés de deux à trois protéines. Au-delà, le nombre de résistance aux antibiotiques nécessaires à la sélection des bactéries portant les plasmides devient trop important et des problèmes de croissance des cellules apparaissent. De plus, dans la plupart des cas, le vecteur pRARE2 est utilisé (décrit au paragraphe 2.2.3.1), qui porte un gène codant pour une résistance au chloramphénicol, empêchant l'utilisation du vecteur pnYC. C'est la raison pour laquelle il est préférable de concaténer les gènes sur un même vecteur lorsque des complexes de plus de deux protéines sont produits.

La concaténation va utiliser les sites de restrictions présents sur les vecteurs d'expression en amont et en aval du gène (Figure 26). Les deux vecteurs codant les protéines du complexe vont être digérés ce qui permettra : (i) l'élimination du promoteur du vecteur dont on conservera l'ossature (digestion BglIII et SpeI) et (ii) l'extraction d'un fragment codant pour le promoteur et le second gène du second vecteur (digestion BglIII et XbaI). La concaténation de ces deux fragments va recréer un vecteur comportant un seul promoteur qui contrôlé l'expression de deux gènes. Pour ce faire, après purification du vecteur et du

fragment, le fragment est ligué dans le vecteur, les sites *SpeI* et *XbaI* étant ligués ensemble du fait de la compatibilité de ligation de ces deux sites.



**Figure 26 : Procédure de concaténation de deux gènes dans un vecteur d’expression.**

*Les sites de restriction BglIII et SpeI sont utilisés pour extraire le promoteur T7 du premier vecteur pnEA-X. Les sites de restrictions BglIII et XbaI sont utilisés pour extraire le promoteur T7 et le gène du second vecteur pnCS. La réaction de ligation est alors réalisée entre le vecteur pnEA-X ouvert et l’insert extrait du vecteur pnCS.*

Le choix des enzymes de restriction utilisées pour la concaténation va dépendre des sites de restrictions présents dans chacun des gènes. Ainsi, les sites *NheI* et *AvrII*, qui sont également compatibles avec le site *SpeI* sont placés en aval du site *XbaI* et peuvent être utilisés de manière alternative dans le cas où un site *XbaI* est présent dans le gène (Figure 71, page 263).

Il est important de noter que ce processus est un processus itératif puisque le fragment gène B – gène A peut-être ré-extrait du vecteur d’expression avec les mêmes enzymes pour être inséré dans un autre vecteur. Trois protéines peuvent ainsi être encodées sur un même vecteur et ainsi de suite (à condition que la taille du vecteur et celle de la partie à transcrire restent raisonnables).

### 2.1.3. Mutagenèse dirigée

La validation des données structurales obtenues sur des complexes protéiques va notamment passer par la mutation des résidus qui ont été identifiés comme participant à la formation du complexe. Les mutations sont introduites par PCR dans les gènes codant pour les protéines et la capacité de la protéine mutée à interagir avec son partenaire est alors testée.

Les mutations ponctuelles sont introduites soit par PCR emboîtée soit par la technique dite de « rolling circle ».

#### 2.1.3.1. Mutagénèse par PCR emboîtée (Nested PCR)

Ce mode d'introduction de mutations ponctuelles nécessite deux étapes de PCR. Pour ces PCR, quatre amorces sont nécessaires : les amorces 5' et 3' du gène et deux amorces qui portent la mutation, l'une complémentaire au brin sens et l'autre complémentaire au brin anti-sens. La première étape va conduire à amplifier deux fragments : (i) un fragment allant de l'extrémité 5' du gène au site de la mutation et (ii) un fragment allant du site de la mutation à l'extrémité 3' du gène. La seconde étape de PCR utilise un mélange des fragments de gènes générés à l'étape précédente afin de reconstituer le gène entier. Cette seconde étape va nécessiter une adaptation du cycle de la PCR, et notamment de la température d'hybridation qui pourra être augmentée afin d'améliorer la spécificité de l'hybridation, notamment au niveau de la zone de recouvrement qui contient la mutation. Le produit de cette troisième réaction de PCR est déposé sur gel d'agarose, les bandes contenant les fragments reconstitués sont découpées et les fragments sont extraits du gel à l'aide du kit Nucleospin Extract II. Les fragments élués sont ensuite insérés dans un vecteur navette afin d'être séquencés.

#### 2.1.3.2. Mutagénèse par rolling-circle

L'introduction d'une mutation peut également être faite en une seule étape par amplification par PCR du vecteur entier. Pour ce faire, un seul couple d'amorces complémentaires portant la mutation est nécessaire, qui va s'hybrider au niveau du site où la mutation doit être introduite. La polymérase utilisée est soit la Phusion soit la Pfu, deux enzymes processives, la matrice étant le vecteur contenant le gène à muter. Le temps d'élongation sera calculé en fonction de la taille du vecteur. A la fin de la réaction de PCR, l'enzyme DpnI est ajoutée qui va dégrader spécifiquement l'ADN méthylé, c'est-à-dire l'ADN du vecteur de départ qui a été méthylé dans la bactérie lors de sa réplication. Le traitement par la DpnI va donc permettre la dégradation des vecteurs ne contenant pas la mutation. Une partie de la réaction est alors transformée dans des bactéries DH5 $\alpha$  et les vecteurs obtenus sont vérifiés par séquençage.

## 2.1.4. Génération de bacmides pour l'expression en cellules d'insectes

Les protéines d'organismes eucaryotes, lorsqu'elles sont produites dans la bactérie sont parfois mal repliées et donc insolubles du fait de l'absence de protéines chaperonnes et de certaines modifications post-traductionnelles. L'expression en cellules d'insecte est une alternative à l'expression en bactérie qui permet de s'affranchir de la plupart de ces problèmes, puisque les cellules d'insectes sont des organismes eucaryotes. Ce système d'expression permet d'exprimer de grandes quantités de protéines à un coût raisonnable, ces protéines étant modifiées post-traductionnellement.

### 2.1.4.1. Principe de l'expression en cellules d'insectes

Dans le cas de la production en cellules d'insectes, le gène codant pour la protéine va être intégré par infection de la cellule avec un virus : le baculovirus qui infecte et se réplique dans les cellules d'insectes. Les cellules d'insecte utilisées pour la production des protéines sont des cellules du papillon *Spodoptera frugiperda* (cellules Sf9 ou Sf21), qui seront infectées par le virus AcMNPV (*Autographa californica* multiple nuclear polyhedrosis virus). L'information génétique de ce virus est stockée sur un ADN double brin circulaire.

Le gène codant pour la protéine d'intérêt va être inséré dans le génome du virus, en remplaçant le gène codant pour la polyhédrine. Cette protéine constitue la matrice protectrice du virus et est fortement exprimée à la fin de son cycle de réplication mais n'est pas essentielle au virus en culture. Le promoteur PolH qui permet l'expression de cette protéine est un promoteur fort et, de ce fait, la protéine d'intérêt sera fortement exprimée.

Le gène de la protéine d'intérêt peut être inséré dans le génome viral de deux manières. Il peut être inséré par co-transformation des cellules Sf9 avec (i) le génome viral linéarisé et (ii) un vecteur adapté contenant le gène de la protéine d'intérêt (ainsi qu'une séquence codant l'étiquette de purification), la recombinaison homologue permettant de recirculariser le génome à l'intérieur des cellules. L'insertion du gène dans le génome viral peut également être réalisée dans la bactérie par un système appelé Bac to Bac®, qui a été utilisé au cours de ce travail, le génome viral est alors appelé Bacmide.

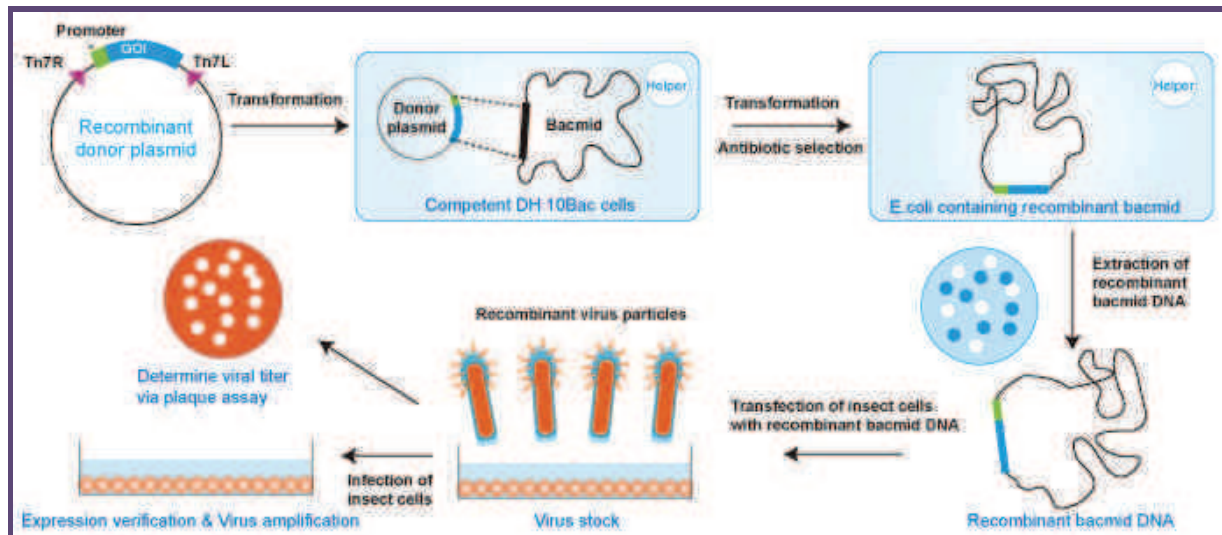
### 2.1.4.2. Génération des Bacmides grâce au système Bac to Bac®

Le gène d'intérêt est cloné par LR dans un vecteur modifié par la plateforme de Biologie Moléculaire de l'IGBMC à partir des vecteurs pFL créée dans l'équipe du Dr. Imre

Berger (Berger et al., 2004; Trowitzsch et al., 2010) et fourni par Redbiotec. Ces vecteurs possèdent une origine de répllication bactérienne ainsi qu'un gène de résistance à un antibiotique (la carte des vecteurs utilisés est présentée sur la Figure 72 page 264). Ils possèdent également des sites de transposition Tn7R et Tn7L, nécessaires à la transposition dans le bacmide, qui entourent le site de clonage. Ce site comprend le promoteur PolH qui est suivi de la séquence codant pour une étiquette de purification puis des sites de recombinaison homologues qui permettent l'insertion du gène de la protéine par recombinaison homologue. Il existe également une version du vecteur qui n'encode pas d'étiquette de purification, permettant la réalisation d'expériences de co-expression (pFC-0GW). Il faut noter que ces vecteurs permettent la concaténation des gènes codant pour les protéines d'un complexe. Cette concaténation est basée sur le même principe que la concaténation de gènes dans un vecteur bactérien (décrit au paragraphe 2.1.2 page 80) et a été adaptée par le service de Biologie Moléculaire de l'IGBMC.

Ces vecteurs sont transformés dans des bactéries DH10Bac (Tableau 18, page 262) dans lesquelles la réaction de transposition va être réalisée. Ces bactéries contiennent le génome du virus qui porte une résistance à la gentamycine et un plasmide helper qui porte une résistance à la kanamycine (Figure 27). Le plasmide helper code pour une transposase qui va réaliser la réaction de transposition entre le vecteur et le génome viral, un gène LacZ placé en aval du promoteur PolH dans le génome viral permet une sélection blanc/bleu des bactéries dans lesquelles la réaction a eu lieu. La présence du gène d'intérêt dans le génome est confirmée par PCR et les bacmides sont extraits des bactéries blanches par extraction à l'isopropanol 100% suivi d'un lavage à l'éthanol.

Le bacmide est alors fourni à la plateforme baculovirus de l'IGBMC, où il sera utilisé pour transformer des cellules Sf9 afin de générer les premiers virus qui seront par la suite utilisés pour produire les protéines (décrit au paragraphe 2.2.4.2).



**Figure 27 : Génération d'un bacmide par la méthode « Bac to bac »**

Les vecteurs baculovirus générés sont transformés dans des bactéries DH10 Bac, qui contiennent le génome du virus. Dans ces bactéries, une partie du vecteur baculovirus est transposée dans le génome du virus. L'évènement de transposition est sélectionné grâce au système blanc/bleu.

## 2.2. Tests d'expression, production et purification de protéines et complexes protéiques

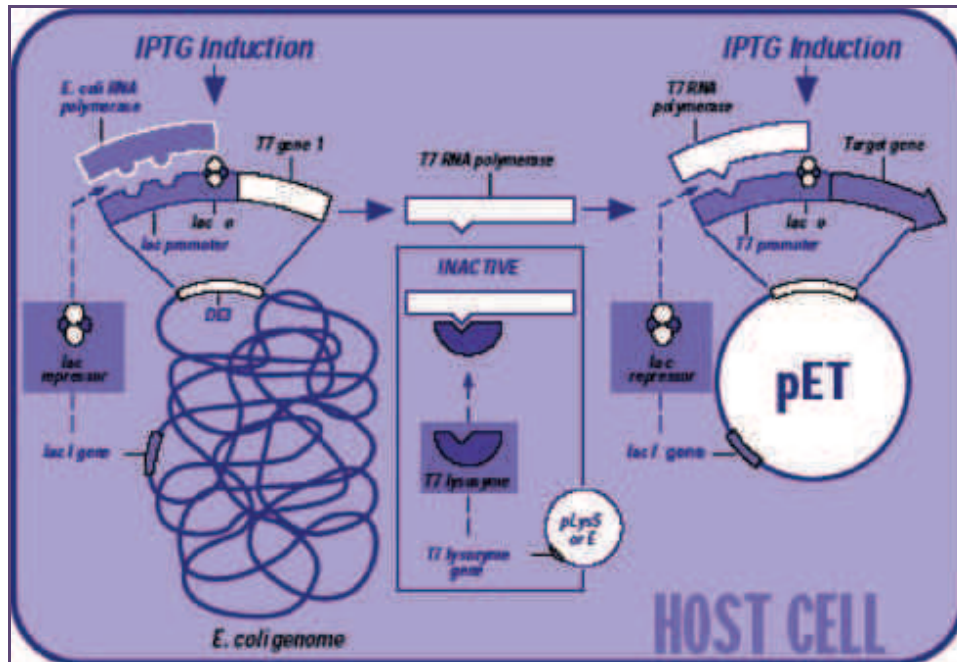
Une fois les vecteurs d'expression obtenus, ils peuvent être utilisés pour produire les protéines et les complexes d'intérêt dans l'hôte choisit. Les tests d'expression vont permettre de définir (i) les conditions de culture qui permettront d'exprimer la protéine de manière optimale (souche bactérienne, milieu de culture et température d'expression) et (ii) les conditions de lyse (principalement le tampon de lyse) qui permettront de purifier les protéines et complexes de manière soluble en grande quantité.

### 2.2.1. Le système d'expression bactérien pET et son dérivé, le système pET-MCN

Les vecteurs d'expression de la suite du Dr. Romier sont basés sur le système d'expression pET, construit par William Studier (Studier et al., 1990) et développé depuis par Novagen®.

Ce système permet un double niveau de régulation lors de la production de la protéine d'intérêt, basé sur la régulation par l'opéron lactose (Figure 28). Ainsi, deux fragments d'ADN ont été insérés dans le génome des bactéries : le segment DE3 du phage  $\lambda$  (les bactéries sont dites lysogènes) ainsi que le gène codant pour l'inhibiteur LacI. Le segment DE3 du phage  $\lambda$  a été inséré dans le gène *int* de la bactérie qui code pour une intégrase,

empêchant ce segment de se recirculariser en l'absence d'un plasmide helper. Il contient le gène qui code pour la polymérase T7, placé sous la dépendance d'un promoteur LacUV5 inducible par ajout de lactose ou de son analogue non hydrolysable l'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG).

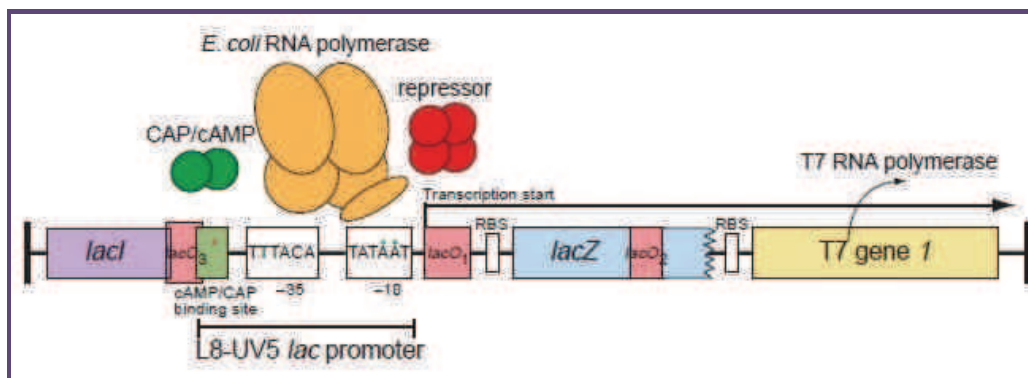


**Figure 28 : Le système d'expression bactérien pET.**

*Le système d'expression bactérien pET est basé sur un double niveau de régulation de l'expression des protéines. En effet, le gène codant pour la polymérase T7 qui est intégré au génome bactérien est placé sous le contrôle du promoteur lac, l'expression de la polymérase étant induite par ajout de lactose ou d'IPTG dans le milieu de culture. Cette régulation est également appliquée pour la transcription du gène de la protéine d'intérêt par la polymérase T7.*

En condition de croissance bactérienne, le répresseur LacI est transcrit par la polymérase bactérienne, bloquant la production de la polymérase T7 et par conséquent la transcription du gène présent dans le vecteur. Au moment de l'induction, l'ajout de lactose ou d'IPTG, va induire un changement de conformation de l'inhibiteur LacI qui ne pourra plus se fixer sur le promoteur LacUV5, permettant la production de la polymérase T7 et par conséquent de la protéine d'intérêt. De plus, le même type de répression s'applique en aval du promoteur présent sur le vecteur d'expression permettant également une régulation directe de l'expression de la protéine d'intérêt. Ce système d'expression permet ainsi une répression de l'expression de la protéine, ce qui est surtout nécessaire lors de l'expression de protéines toxiques. Ce système de répression n'est cependant pas complet puisqu'un faible taux d'ARN polymérase T7 (niveau basal) est détecté en l'absence d'IPTG.

L'utilisation de l'opéron lactose pour la répression de la transcription permet également l'utilisation de glucose pour réprimer plus fortement le promoteur LacUV5 au cours de la phase de croissance (Figure 29). En effet, lorsque la concentration en glucose dans le milieu est forte, la concentration en adénosine monophosphate cyclique (cAMP) est faible. Or le cAMP est nécessaire pour que l'initiation de la transcription soit efficace puisqu'en interagissant avec son récepteur CAP, il permet de stimuler la transcription. Ainsi en présence de glucose, le niveau de transcription de l'ARN polymérase T7 basal est plus faible qu'en l'absence de glucose malgré l'introduction de mutations ponctuelles dans le promoteur qui permettent de réduire l'effet du cAMP sur le taux d'expression.



**Figure 29 : Répression de l'opéron lactose en présence de glucose ou répression des catabolites.**

*Le répresseur LacI se lie au site opérateur lacO<sub>1</sub> puis interagit avec les pseudo-opérateurs lacO<sub>2</sub> et lacO<sub>3</sub> inhibant ainsi la transcription de la polymérase T7. L'ajout d'IPTG va réduire l'affinité du répresseur LacI pour le site opérateur lacO<sub>1</sub> permettant la transcription de la polymérase T7. Quand les concentrations en cAMP sont suffisantes, le complexe CAP/cAMP se lie sur le promoteur afin de stimuler la transcription. Les mutations introduites dans le promoteur afin de réduire l'effet du cAMP sur le niveau de transcription sont indiquées par des points rouges et verts.*

### 2.2.2. Les milieux de culture bactérien

Différents milieux de culture bactériens ont été utilisés au cours de cette thèse. Chacun de ces milieux va présenter un avantage, tous seront donc généralement testés lors de la production en grand volume des protéines. Pour la réalisation des tests d'expression, les milieux 2LB, Graffinity et auto-induits ont principalement été utilisés. La composition détaillée de ces différents milieux ainsi que les conditions d'induction sont présentées dans le Tableau 4.



Milieu	Composition pour 1L	DO <sub>600nm</sub> à l'induction	Induction	Autres
<b>2LB</b>	1L 2LB (20 g bactotryptone 10g d'extrait de levure 20g NaCl)	0,8	750µL IPTG 1M	-
<b>Graffinity</b>	475mL 2LB + 25mL Glucose 10%	1,2	470mL 2LB + 25mL Lactose 12% + 10mL HEPES 1M + 750µL IPTG 1M	-
<b>Auto-induit</b>	1L ZY (10 g bactotryptone 5g d'extrait de levure) + 2mL MgSO <sub>4</sub> (1M) + 20mL 50X52 + 25mM 40XM	0,4	-	<u>50X52 (1L)</u> : 287mL glycérol (87%), 25g glucose, 100g lactose <u>40XM (1L)</u> : 358g Na <sub>2</sub> PO <sub>4</sub> .12H <sub>2</sub> O, 136g KH <sub>2</sub> PO <sub>4</sub> , 107,2g NH <sub>2</sub> Cl, 28,4g Na <sub>2</sub> SO <sub>4</sub>
<b>Milieu minimum</b>	800mL H <sub>2</sub> O + 100mL M9 10X + 100mL NH <sub>4</sub> Cl 1% + 20mL Glucose 10% + sels + vitamines	0,8	750µL IPTG 1M	<u>Sels M9 10X (1L)</u> : 150g Na <sub>2</sub> HPO <sub>4</sub> .12H <sub>2</sub> O + 30g KH <sub>2</sub> PO <sub>4</sub> + 5g NaCl

**Tableau 4 : Composition des différents milieux de culture utilisés.**

La composition des différents milieux de cultures utilisés au cours de ce travail de thèse est présentée dans ce tableau. « sels + vitamines » dans la composition de milieu minimum correspond à : 1mL MgSO<sub>4</sub> 2M, 50µL CaCl<sub>2</sub> 2M, 500µL FeSO<sub>4</sub> 0,1g/L, 1mL Thiamine 10mg/mL et 1mL Biotine 10mg/mL

#### 2.2.2.1. Milieu complet (2LB)

Le milieu 2LB (Lysogeny broth) est un milieu de culture nutritif qui permet d'obtenir à l'issue des cultures une quantité moyenne de bactéries (DO<sub>600nm</sub> ~ 3-4). Ce milieu contient des peptides, des vitamines, des minéraux et des oligo-éléments. Pour la réalisation des tests de co-expression, 10mL de milieu sont utilisés. Pour les productions en grand volume, 1L de milieu est autoclavé dans des erlenmeyer de 5L. Ce milieu est inoculé à une DO<sub>600nm</sub> de 0,05 et placé sous agitation (350rpm) à 37°C. Lorsque la culture atteint une DO<sub>600nm</sub> entre 0,4, la température est baissée et au bout d'une demi-heure induite par ajout de 0,75mM d'IPTG. La culture est laissée sous agitation pendant 16h.

#### 2.2.2.2. Milieu Graffinity®

Le milieu Graffinity a été développé par la société Graffinity (Heidelberg). Il permet de produire de grandes quantités de bactéries (DO<sub>600nm</sub> ~ 5-6 en fin de culture) et peut donc permettre la production de plus grandes quantités de protéines. A un milieu 2LB (typiquement

autour de 500mL pour la production en grands volumes), 0,5% de glucose est ajouté en début de culture permettant une répression plus efficace de la transcription du gène de la polymérase T7 ainsi que la croissance à haute densité des cellules (milieu 1) (décrit paragraphe 2.2.1). Au moment de l'induction, un volume identique à la culture de départ est ajouté, supplémenté de 0,6% de lactose et 0,5mM d'IPTG permettant une induction plus forte de la transcription (milieu 2). La présence d'HEPES (20mM) permet de corriger le pH des cultures qui s'acidifient lorsque les bactéries se multiplient (du fait de la production de déchets) et permet ainsi le retour au pH optimal de croissance bactérienne. Pour la réalisation des tests d'expression, 1,3mL de milieu 1 est inoculé avec les bactéries, 1,3mL de milieu 2 étant ajoutés au moment de l'induction. Pour les productions en grand volume, 475mL de milieu 2LB sont autoclavés en erlenmeyers de 5L auxquels 0,5% de glucose sont ajoutés avant l'inoculation des bactéries puis placés sous agitation à 37°C. Lorsque la  $DO_{600nm}$  atteint 0,6 – 0,8, la température des cultures est baissée à 25 ou 18°C et 470mL de 2LB autoclavé et supplémenté de 0,6% de lactose, 20mM HEPES pH7 et 0,75mM d'IPTG sont ajoutés. Les cultures sont laissées sous agitation pendant 16 h.

#### 2.2.2.3. Milieu Auto-induit

Le milieu auto-induit a été développé par William Studier (Studier F. William, 2005). De la même manière que le milieu Graffinity, de grandes quantités de bactéries peuvent être obtenues en fin de culture ( $DO_{600nm} \sim 5-6$ ) grâce à ce milieu. A un milieu ZY (contenant de l'extrait de levure et de la bactotryptone) un mélange de sel est ajouté (Tableau 4), ainsi que 0,5% de glycérol, 0,05% de glucose et 0,2% de lactose. Les bactéries vont se développer en commençant par consommer le glucose présent dans le milieu, puis, en utilisant le lactose présent qui va alors pouvoir pénétrer dans les cellules, va permettre l'induction de la production des protéines. William Studier a défini de manière empirique que l'induction est réalisée autour d'une  $DO_{600nm}$  de 0,4, moment auquel la température de la culture est diminuée. Les tests d'expression dans ce milieu sont réalisés dans 2,5mL de milieu, ce milieu n'ayant pas été utilisé pour la production des protéines en grand volume.

#### 2.2.2.4. Milieu minimum N15-C13

Ce milieu a uniquement été employé pour la production de protéines incorporant de l'azote 15 et du carbone 13 nécessaires à la réalisation d'études RMN. L'azote marqué est incorporé sous la forme de chlorure d'ammonium ( $NH_4Cl$ ) alors que le carbone 13 est incorporé sous la forme de glucose marqué. Ce milieu est composé des éléments chimiques

strictement nécessaires à la croissance bactérienne, par conséquent la densité optique atteinte en fin de culture est relativement faible ( $DO_{600nm} \sim 1-2$ ). En plus du carbone, apporté par le glucose et de l'azote apporté par le chlorure d'ammonium  $NH_4Cl$ , les sels ajoutés constituent une source de potassium et de phosphore ( $KH_2PO_4$ ), une source de magnésium, de fer et de soufre ( $MgSO_4$  et  $FeSO_4$ ) et une source de calcium ( $CaCl_2$ ) mais aussi des vitamines (Thiamine et Biotine). Pour ces productions en grand volume, de l'eau distillée est autoclavée dans des erlenmeyer de 5L, les solutions de sels étant autoclavées ou filtrées séparément et ajoutées au moment d'inoculer les cultures. Ce milieu étant pauvre, les bactéries se développent moins vite que dans un milieu riche et c'est la raison pour laquelle la culture est inoculée à une  $DO_{600nm}$  de 0,1 avant d'être mise sous agitation à 37°C. Lorsque la  $DO_{600nm}$  atteint 0,6, la température est baissée puis l'expression est induite avec 0,75mM d'IPTG. Elle est laissée sous agitation pendant 12 h.

### 2.2.3. Tests d'expression

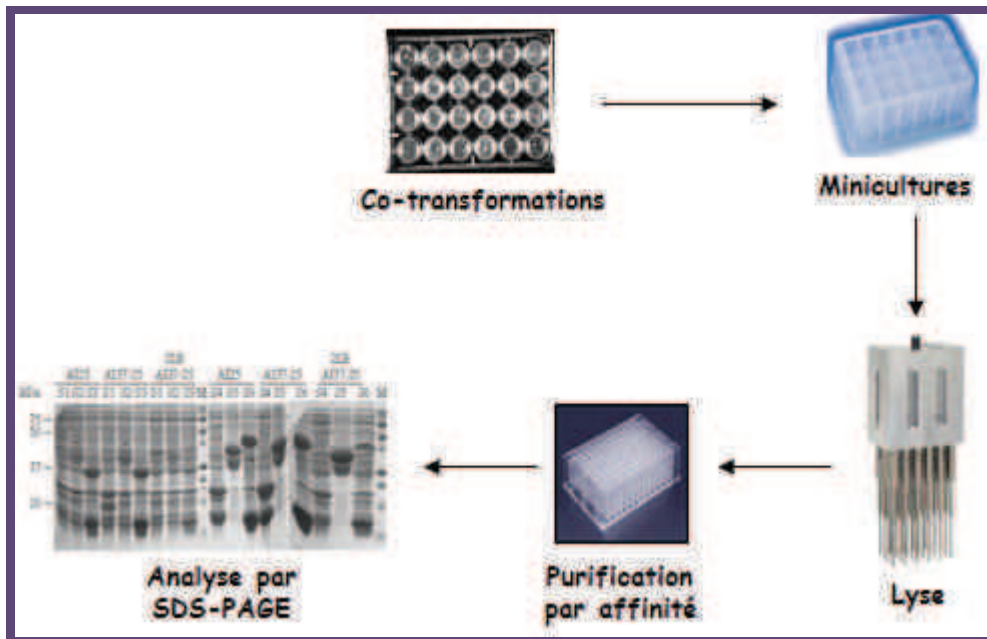
Les tests d'expression vont permettre de répondre à différentes questions tout au long du processus de détermination des conditions d'expression et de purification d'une protéine. Ils vont ainsi permettre (i) d'identifier si une protéine est soluble et quel tampon favorise sa solubilité et (ii) dans le cas de la co-expression d'identifier les zones d'interaction de chaque protéine rapidement. Ces tests peuvent être réalisés aussi bien chez la bactérie que dans des cellules d'insecte, bien que la génération des bacmides pour l'expression dans les cellules d'insecte ralentisse fortement ce type d'analyse. Ces tests d'expression consistent en la réalisation de mini-purifications sur une résine d'affinité. Les protéines liées sont détachées de la résine par l'ajout direct de tampon de charge pour gel d'acrylamide dénaturant et déposées directement sur gel d'acrylamide dénaturant. Les tests effectués au cours de cette thèse ont principalement été réalisés chez la bactérie. C'est la réalisation de ces tests qui est décrite dans la partie suivante.

#### 2.2.3.1. Réalisation des tests d'expression en bactéries

##### *a) Transformation des bactéries*

Dans un premier temps, les vecteurs codant pour les protéines sont transformés dans les bactéries par choc thermique puis étalées sur des boîtes de LB-Agar contenant les antibiotiques (dilués au 1000<sup>ème</sup>) nécessaires à la sélection des bactéries ayant intégré les plasmides (Figure 30, Vijayachandran et al., 2011). L'étalement sur boîte plutôt qu'une

préculture liquide permet une plus forte sélection des colonies ce qui permet d'obtenir une meilleure expression des protéines.



**Figure 30 : Les différentes étapes des tests de co-expression, (adapté de Vijayachandran et al., 2011)**

*Après co-transformation des plasmides dans les bactéries BL21(DE3), les bactéries transformées sont cultivées dans un milieu liquide. Après reprise des cellules dans le tampon de lyse, elles sont lysées par sonication utilisant une tête comportant 4 sondes. L'extrait soluble est alors incubé avec une résine d'affinité, lavée plusieurs fois et la fraction purifiée est déposée sur gel d'acrylamide.*

Pour la réalisation de ces tests, différentes souches bactériennes peuvent être utilisées, les deux principales étant les bactéries BL21(DE3) et BL21(DE3) pRARE2 décrites en annexe (paragraphe 8.1.4 page 262). Ces deux souches permettent de réguler l'expression des protéines d'intérêt via le système pET décrit plus haut (paragraphe 2.2.1, page 85). La souche BL21(DE3) pRARE2 contient un plasmide pRARE2 qui code pour les ARN de transferts rares chez la bactérie. En effet la fréquence d'utilisation des codons chez les bactéries est différente de celle des autres organismes et, par conséquent, les ARN de transfert complémentaires à ces codons sont également moins présents. L'utilisation du vecteur pRARE2 va parfois permettre l'expression ou la solubilisation de protéines dont les gènes contiennent un fort taux de ces codons rares, et notamment dans le cas de l'expression de protéines eucaryotes. Ainsi les tests sont généralement réalisés dans un premier temps dans la souche BL21(DE3) et dans le cas où les protéines ne sont pas exprimées ou exprimées de manière insoluble, ils sont répétés dans la souche BL21(DE3) pRARE2.

### *b) Culture des bactéries en milieu liquide*

Le lendemain, une partie des colonies ayant poussées sur les boîtes d'agar sont resuspendues dans un faible volume de milieu LB et utilisées pour inoculer une culture liquide (antibiotiques dilués au 1000<sup>ème</sup>). Différents milieux ont été utilisés qui sont listés dans le Tableau 5 et décrits plus précisément dans la section 2.2.2 page 87. Selon le milieu dans lequel les tests de culture sont réalisés, le volume de culture réalisé va être adapté. En effet, la densité optique à 600nm atteinte en fin de culture, qui reflète la quantité de cellules en culture va varier en fonction du milieu employé. Ainsi, afin de ne pas introduire un biais lors de l'analyse de la quantité de protéine produite, l'ajustement du volume de culture utilisé va permettre d'équilibrer la quantité de cellules en fin de culture. Le taux d'expression des protéines dans chacun de ces milieux de culture pourra ainsi être comparé directement. Cet ajustement pourrait également être réalisé par la mesure de la densité optique à 600nm des cultures, les tests pouvant alors être réalisés sur une partie seulement des cultures avec des quantités identiques de cellules. Cependant, dans l'optique de la réalisation d'un grand nombre de tests en parallèle, l'ajustement du volume de culture est plus adapté.

Les cultures sont réalisées sous agitation à 37°C pendant 3 à 4 heures permettant aux cellules de se multiplier jusqu'à atteindre la densité optique à 600nm fixée, qui va varier selon le milieu employé (Tableau 5). La température est alors baissée à 25°C afin de ralentir la croissance des bactéries et ainsi favoriser le repliement correct des protéines d'intérêt, puis l'expression des protéines est induite (Tableau 5). Les cultures sont laissées sous agitation à 25°C.

Les cultures sont réalisées soit en falcon de 50mL pour un volume de culture de 10mL, soit en boîtes de culture de 24 puits pour un volume de culture de 2,5mL, ces boîtes étant scellées avec un film protecteur permettant l'oxygénation des cultures et empêchant les contaminations (Figure 30, page 91).

Milieu	Volume de culture initial	DO <sub>600</sub> à l'induction	Induction réalisée avec	Volume de culture après induction
<b>2LB</b>	10mL	0,8	IPTG	10mL
<b>Graffinity</b>	1,3mL	1,2	Lactose + IPTG	2,5mL
<b>Auto-induit</b>	2,5mL	0,4	Lactose	2,5mL

**Tableau 5 : Les différents types de milieu employés pour la réalisation des tests d'expression**

*Le milieu 2LB (Lysogeny broth) est un milieu complet. La culture en milieu Graffinity est réalisée en deux temps : une croissance en 2LB supplémenté en glucose jusqu'à une DO<sub>600nm</sub> de 1,2 puis induction de l'expression des protéines par ajout de 2LB supplémenté en lactose et IPTG. L'induction d'un milieu auto-induit est réalisée sans intervention extérieure puisque le milieu contient un mélange de glucose et de lactose. Lorsque tout le glucose du milieu est consommé, le lactose pénètre dans les cellules et agit comme inducteur de l'expression des protéines.*

### *c) Réalisation des tests d'expression*

La plateforme de Biologie Structurale offre la possibilité de réaliser un grand nombre de tests en parallèle par l'utilisation de sondes de sonication à quatre têtes et de robots de pipetage. Lors de la réalisation de tests d'expression pour vérifier la solubilité d'une protéine ou définir le tampon de lyse qui solubilise le plus efficacement la protéine, différents tampons pourront être testés en parallèle. Typiquement, des tampons contenant 50, 100, 200 et 400 mM de sels et un agent tampon (généralement le Tris) vont être testés. De la même manière, lors de la réalisation de tests de co-expression, (i) l'influence des tampons de lyse et de lavage sur la formation du complexe (ii) la position de l'étiquette de purification (iii) ainsi que différentes constructions de chaque protéine vont également pouvoir être testés rapidement. Il n'est ainsi pas rare que 96 tests d'expressions soient réalisés en parallèle.

Lorsque différents tampons de lyse sont testés, les cultures liquides sont généralement dupliquées et non multipliées afin que les différences dans les taux d'expression qui sont observées reflètent bien une différence de solubilité et non une variation apparue lors de l'expression des protéines. Les cultures, lorsqu'elles sont réalisées en falcon sont transférées dans des boîtes de culture de 24 puits afin de faciliter les étapes de sonication et de transfert sur la résine. Les boîtes de culture sont alors centrifugées (15 minutes à 2500 rpm) afin d'éliminer le milieu de culture et du tampon de lyse est ajouté avant sonication. Un échantillon est alors prélevé, l'extrait total, qui contient toutes les protéines exprimées ce qui

permettra notamment de vérifier que notre protéine d'intérêt est bien exprimée. Les boîtes de culture sont alors centrifugées (20 minutes à 4000 rpm) afin d'éliminer les débris cellulaires ainsi que les protéines insolubles. L'échantillon prélevé après cette étape, l'extrait soluble, va permettre de vérifier que notre protéine d'intérêt est exprimée de manière soluble. Le robot de pipetage de la plateforme est alors utilisé afin de transférer le reste de l'extrait soluble dans une boîte de 96 puits dans laquelle aura été placée la résine d'affinité préalablement équilibrée avec le tampon de lyse (25µL de résine par test). La boîte est placée sous agitation à 4°C pendant une heure puis est centrifugée à faible vitesse (5 minutes à 1000 rpm) afin de faire sédimenter la résine. Le robot de pipetage est utilisé afin d'éliminer l'extrait soluble (qui contient les protéines non retenues sur la résine) et ajouter le tampon de lavage. Une nouvelle étape de lavage est réalisée, puis du tampon de charge (30µL de Laemmli) est ajouté sur la résine sèche afin de décrocher les protéines retenues. Ces échantillons sont alors déposés sur un gel d'acrylamide dénaturant contenant de 15 à 20% d'acrylamide qui est ensuite coloré au bleu de coomassie. Ainsi, pour chaque test, trois échantillons sont obtenus : (i) un extrait total qui contient toutes les protéines produites, (ii) un extrait soluble qui contient les protéines exprimées de manière soluble et (iii) un échantillon de purification qui contient les protéines retenues par affinité à la résine.

### 2.2.3.2. Interprétation des résultats des tests d'expression

#### *a) Les tests d'expression d'une protéine seule*

Les tests d'expression vont permettre d'accéder à un grand nombre d'informations notamment concernant la solubilité d'une protéine ou le milieu qui va permettre une meilleure expression de la protéine. Ils vont notamment être réalisés après les étapes de clonage en vecteur d'expression et vont permettre d'identifier rapidement les protéines exprimées et solubles. Différents cas de figure, récapitulés dans la Figure 31, peuvent se présenter :

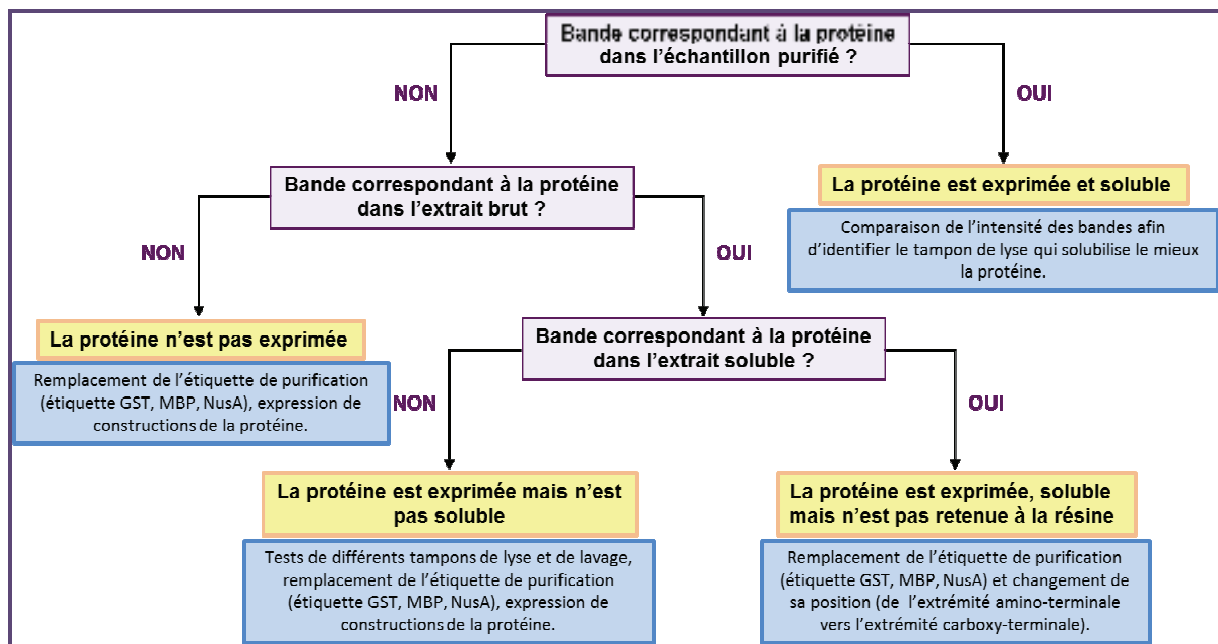
(i) Si la protéine est produite et est soluble, les tests d'expression vont permettre de déterminer dans quel tampon la protéine est la plus soluble par comparaison de l'intensité des bandes observées sur le gel.

(ii) Si aucune bande n'est visible dans les extraits totaux et solubles, alors la protéine n'est pas exprimée dans l'hôte choisi. Dans ce cas, l'ajout à l'étiquette de purification d'une étiquette de solubilisation GST, TRX, MBP ou NusA peut permettre d'exprimer la protéine.

Un changement d'hôte de production peut également permettre d'exprimer la protéine, comme par exemple la production en cellules d'insectes.

(iii) Si une bande est visible dans l'extrait total mais pas dans l'extrait soluble alors la protéine est exprimée mais n'est pas soluble. Dans ce cas, le changement du tampon de lyse peut permettre la solubilisation de la protéine ainsi que l'ajout d'une étiquette de solubilisation. Là encore, le changement de l'organisme dans lequel la protéine est produite peut permettre d'améliorer la solubilité.

(iv) Si des bandes sont visibles dans les extraits totaux et solubles mais pas à l'issue des tests d'expression, dans ce cas la protéine est exprimée et soluble mais ne lie pas la résine car son étiquette de purification n'est pas exposée. Le remplacement de l'étiquette de purification et son positionnement à l'autre extrémité de la protéine peut permettre de retenir la protéine sur la résine d'affinité.



**Figure 31 : Interprétation des résultats des tests d'expression.**

*En fonction des résultats des tests d'expression d'une protéine seule, différentes possibilités vont permettre de poursuivre l'étude menée, décrite ici.*

Ainsi à l'issue de ces tests il peut être décidé de produire la protéine en plus grande quantité si les conditions d'expression et de purification ont pu être définies. Il peut également être décidé de poursuivre les tests d'expression avec, soit de nouveaux tampons de lyse, soit en remplaçant les étiquettes de purification, ce qui nécessite de réaliser de nouveaux sous-clonages ou encore de tester différentes constructions des protéines. Alternativement le



système d'expression peut également être changé notamment pour l'expression en cellules d'insectes. Ainsi, ces tests permettent de savoir rapidement si une protéine ou un complexe protéique peut être produit dans le système d'expression choisi.

*b) Les tests de co-expression*

Les tests de co-expression vont permettre de vérifier la formation d'un complexe et de définir les zones d'interaction en testant différentes régions de chacune des protéines. La co-expression va également présenter de nombreux avantages pour l'expression de complexes protéiques. En effet, dans de nombreux cas, l'expression dans une même bactérie de deux interactants va permettre leur co-solubilisation, cette co-solubilisation étant rendue possible notamment par le repliement correct de chacune des protéines en présence de son partenaire biologique. Ainsi, même les protéines insolubles lors de la réalisation des tests d'expression simples seront testées lors de la réalisation de tests de co-expression. De plus, de la même manière que dans le cas de l'expression de protéines seules, différentes concentrations salines vont également être testées puisque la quantité de sels présente dans le tampon de lyse va également influencer sur la formation des complexes. La formation du complexe va également dépendre de la position de l'étiquette de purification. En effet, la fusion d'un fragment protéique à l'extrémité d'une protéine peut masquer le site d'interaction de son partenaire. Par conséquent, les étiquettes de purification seront toujours placées successivement sur les deux partenaires. Dans le cas des complexes à trois protéines, toutes les combinaisons seront testées (Fribourg et al., 2001; Romier et al., 2006).

La co-expression va donc nécessiter la réalisation d'un grand nombre d'essai, afin de tester ces différents paramètres. Ainsi, (i) un grand nombre de constructions des deux protéines vont être utilisées lors de ces tests, (ii) au moins deux tampons de lyse contenant des quantités de sels différentes seront utilisés et (iii) différentes positions de l'étiquette de purification seront testées. De plus, afin de s'assurer que le complexe observé résulte bien de l'interaction entre les deux protéines et non d'une interaction de la protéine native avec la résine, cette protéine native va être exprimée seule, son incubation avec la résine permettant de tester la possibilité d'une interaction aspécifique. Ce test va également permettre de vérifier la solubilité de la protéine native et son profil de migration, qui sera différent du profil observé pour cette même protéine portant l'étiquette de purification, la présence de l'étiquette augmentant le poids moléculaire de la protéine et donc sa mobilité électrophorétique.

L'interprétation de ces tests sera donc différente des tests de solubilité présentés au paragraphe 2.2.3.2.a. En effet, c'est la présence de deux bandes correspondant à chaque protéine qui sera vérifiée sur le gel d'acrylamide réalisé à la fin des tests, l'intensité des deux bandes devant être équivalente en cas de complexe stoechiométrique. De plus, idéalement, la protéine exprimée sous forme native ne doit pas interagir avec la résine afin de pouvoir estimer de manière claire la stoechiométrie du complexe.

Si une seule bande est observée sur le gel qui correspond à la protéine portant l'étiquette de purification, le test d'expression réalisé va vraisemblablement refléter l'absence d'interaction entre les deux partenaires. Dans ce cas l'inversion de la position de l'étiquette de purification va permettre de s'assurer que l'interaction n'est pas bloquée par cette étiquette qui pourrait masquer le site d'interaction. De même, le test de différentes concentrations salines va permettre de vérifier que le complexe n'est pas formé uniquement en basse ou haute concentration saline.

L'absence de bande correspondant aux protéines peut refléter leur insolubilité. Dans le cas où la protéine portant l'étiquette de purification est insoluble, le complexe n'étant par conséquent pas purifié, la modification de l'étiquette ou l'expression de différentes constructions de la protéine va parfois permettre sa solubilisation. Dans le cas où il s'agit de la protéine native qui est insoluble, l'inversion de l'étiquette de purification peut permettre la solubilisation de la protéine et la formation du complexe.

Les tests de co-expression sont donc un outil essentiel pour définir rapidement les zones de deux ou plusieurs protéines formant un complexe. Dans le cas où des complexes de plus de deux protéines sont étudiés, les gènes codants pour deux des partenaires sont généralement concaténés sur un même vecteur (paragraphe 2.1.2) afin de faciliter les étapes de transformation et de croissance des cellules.

#### **2.2.4. Production des protéines en grand volume dans un système exogène**

La réalisation d'études structurales par cristallographie va nécessiter la production de grandes quantités de protéine, ce qui ne peut être obtenu que par la surexpression dans un organisme hôte. De plus, dans certains cas, les protéines produites doivent être marquées. C'est notamment le cas lors de la réalisation d'études structurales par résonance magnétique nucléaire au cours desquelles de l'azote <sup>15</sup> et/ou du carbone <sup>13</sup> doivent être incorporés à la

protéine ou encore lors de l'introduction de sélénométhionines dans la protéine pour la résolution du problème des phases en cristallographie. Différents milieux de culture vont permettre de cultiver des bactéries qui sont présentées dans cette section.

#### 2.2.4.1. Production en bactéries

Une fois les conditions définies lors des tests d'expression et de co-expression, des cultures bactériennes en grand volume sont réalisées. Ces cultures peuvent être réalisées soit en fioles de 5L, dans lesquelles 1L de culture est placé, soit en fermenteur de plusieurs litres (typiquement plus de 5L de culture). De la même manière que pour les tests d'expression, différents milieux peuvent être utilisés et, là encore, une étape d'optimisation peut s'avérer nécessaire. Les différents milieux utilisés sont présentés dans le paragraphe 2.2.2, page 87. De plus, d'autres paramètres peuvent être modifiés, tels que (i) la densité optique de la culture lors de l'induction, (ii) la température lors de la production, qui est généralement baissée causant un ralentissement de la croissance bactérienne et favorisant le repliement des protéines produites et par conséquent de notre protéine d'intérêt ou (iii) le temps de production, ce qui est notamment le cas lors de la production de protéines toxiques ou de protéines formant rapidement des corps d'inclusion. Ces paramètres vont notamment jouer un rôle sur le taux de production des protéines et sur leur repliement.

La première étape de la production de protéines est la transformation du plasmide codant pour la protéine (ou des plasmides dans le cas de la production de complexes protéiques) dans la bactérie. Cette transformation est réalisée par choc thermique (45 secondes à 42°C) dans des bactéries BL21(DE3) rendues chimiocompétentes par traitement au chlorure de calcium. Cette transformation est ensuite étalée sur LB-Agar contenant les antibiotiques nécessaires dilués au 1000<sup>ème</sup>. Dans le cas où des bactéries contenant un plasmide pRARE2 sont utilisées, le vecteur est déjà présent dans la bactérie (il a été introduit lors de la préparation des bactéries compétentes) afin d'augmenter le taux d'efficacité de la transformation. Le lendemain, quelques colonies sont repiquées et utilisées pour restrier un milieu LB-Agar contenant également les antibiotiques dilués au 1000<sup>ème</sup> afin d'assurer une bonne sélection des bactéries et pouvoir observer les contaminations phagiques qui auraient pu avoir lieu. Le jour de la réalisation de la culture, les bactéries sont remises en suspension dans un milieu LB liquide, leur densité optique à 600nm est mesurée afin de déterminer quel volume utiliser pour inoculer la culture.

Le lendemain, les cultures sont centrifugées pour éliminer le milieu de culture et les bactéries sont resuspendues dans le tampon de lyse adapté à la protéine exprimée (qui a été déterminé au cours des tests d'expression). Si ce tampon n'est pas connu, une fraction de la culture est prélevée et un test d'expression est réalisé avec différents tampons afin de déterminer quel tampon utiliser. Ce type de tests permet également de vérifier que la protéine a bien été produite et d'estimer son taux de production ce qui permettra par la suite d'estimer le volume de résine d'affinité à utiliser lors de la purification. Une fois les cellules resuspendues, elles sont congelées à -20°C.

#### 2.2.4.2. Production en cellules d'insectes

La production dans les cellules d'insectes est réalisée par la plateforme de l'IGBMC. Dans un premier temps, lorsque le bacmide est fourni à la plateforme, il est utilisé pour transformer des cellules Sf9 qui sont des cellules ovariennes du papillon *Spodoptera frugiperda*. Ces cellules vont ainsi produire le virus recombinant qui exprime la protéine d'intérêt. Le virus est alors purifié et utilisé afin de réaliser les premiers tests d'expression qui permettent de vérifier que la protéine est exprimée et soluble. Le protocole de purification à partir de ces tests est identique à celui des tests réalisés à partir de bactéries, les cellules étant cultivées en monocouche dans des grands flacons. Le milieu de culture utilisé est le Grace's Insect Cell Sigma® complétement avec 10% de sérum fœtal bovin et de 28mg/mL de gentamycine les cellules étant cultivées à 27°C.

Si la protéine est exprimée et soluble, de nouveaux tests d'infection sont réalisés qui permettent de définir la quantité de virus nécessaire à l'infection des cellules (MOI, Multiplicity Of Infection) et le temps de culture nécessaire (48h ou 72h) pour que la culture soit optimale. En effet, une plus grande quantité de virus permettra de s'assurer que chaque cellule a été infectée et ainsi d'obtenir une meilleure expression. En fonction de ces tests, des cultures d'un litre sont réalisées en suspension.

Pour ces grandes cultures, le milieu de culture est éliminé après centrifugation à faible vitesse (15 minutes à 1200rpm) car les cellules sont fragiles, elles sont ensuite resuspendues dans le tampon de lyse et congelées à -20°C.

#### 2.2.5. Purification des protéines

La purification des protéines est réalisée à partir des cultures réalisées précédemment. Seule l'étape de lyse diffère entre des protéines exprimées à partir de cultures bactériennes ou

en cellules d'insectes. Différentes techniques permettent de purifier une protéine du milieu cellulaire dans lequel elle a été produite. Les protocoles mis en place comportent généralement une première étape d'affinité qui utilise l'étiquette de purification exprimée en fusion avec la protéine d'intérêt afin de la retenir sur une résine. Vient ensuite une étape de filtration sur gel qui permet d'augmenter la pureté de la protéine et notamment d'éliminer les agrégats et/ou d'échanger le tampon dans lequel elle est purifiée. Cependant, des étapes d'échange d'ions peuvent s'avérer nécessaires, avant l'étape d'affinité, pour éliminer des contaminations par des acides nucléiques ou après l'étape d'affinité pour éliminer des contaminants protéiques qui peuvent être le résultat de la dégradation de la protéine. L'échangeuse d'ion est toujours suivie d'une étape de gel filtration ou de dialyse de manière à transférer la protéine dans un tampon connu. Cependant, dans le cadre de cette thèse cette technique n'a pas été utilisée et ne sera donc pas décrite.

#### 2.2.5.1. 1<sup>ère</sup> étape : la lyse des cellules

1L de culture est resuspendu dans 50mL de tampon de lyse défini au cours des mini-tests d'expression et placé dans un bécher. Le bécher est placé sur la glace et les cellules sont soniquées avec un sonicateur. Les bactéries sont soniquées à une amplitude de 30% pendant un total de 3 minutes avec des impulsions de 5 secondes espacées de 5 secondes afin de permettre un refroidissement du milieu. Les cellules d'insectes sont soniquées à une amplitude de 30% au total 1 minute 30 secondes avec les mêmes impulsions que les bactéries, leurs membranes étant plus fragiles que les parois bactériennes. L'extrait total obtenu est centrifugé à 15000rpm pendant 45 minutes afin de précipiter les débris cellulaires et d'obtenir l'extrait soluble.

#### 2.2.5.2. Chromatographie par affinité

La chromatographie par affinité est basée sur l'affinité particulière de l'étiquette de purification exprimée en fusion avec la protéine pour un ligand qui est couplé à une résine d'agarose ou de Sépharose. Les étiquettes de purification utilisées dans le cadre de cette thèse, leur ligand fixé à la résine et le mode d'élution employé pour chacune est présenté dans le Tableau 6.

Étiquette de purification	Ligand couplé à la résine	Mode d'élution de la résine	Référence des résines
Poly-histidine (His) <sub>6</sub>	Cobalt	Imidazole ou coupure par une protéase	Talon® Superflow de Clontech
Glutathion-S-Transférase (GST)	Glutathion	Glutathion ou coupure par une protéase	Glutathione Sepharose 4B de GE Healthcare

**Tableau 6 : Principales étiquettes de purification utilisées.**

Les chromatographies d'affinité peuvent être réalisées soit sur des systèmes de chromatographie de type Akta, soit en vrac. C'est principalement la technique en vrac qui a été utilisée au cours de cette thèse et qui est donc décrite ici, le principe de ces deux méthodes étant identique.

*a) Fixation de la protéine à la résine et lavages*

Les résines fournies en vrac sont stockées dans de l'éthanol, le ratio d'éthanol et de résine étant de 1 : 1 dans le cas de la résine cobalt. L'éthanol en suspension est éliminé après centrifugation du mélange, la résine d'affinité sédimentant à faible vitesse (maximum 1000rpm). La résine sèche obtenue est ensuite équilibrée avec le tampon de lyse de la protéine puis incubée sous agitation avec l'extrait soluble à 4°C pendant 1h dans des falcon de 50mL. Les tubes sont ensuite centrifugés à basse vitesse (5 minutes à 1000rpm) afin de faire sédimenter la résine. Le surnageant, qui contient les protéines non retenues, est éliminé. La résine est alors reprise dans du tampon de lyse et placée dans une colonne Poly Prep (Bio-Rad). Un lavage de la résine avec du tampon de lyse est alors effectué afin d'éliminer les protéines liées aspécifiquement à cette résine. Les lavages sont poursuivis jusqu'à ce que l'éluat ne colore plus le réactif de Bradford (10µL d'éluat dans 250µL de réactif, le principe de cette technique étant décrit au paragraphe 2.2.5.4). La protéine est alors détachée de la résine soit par élution soit par coupure avec la protéase dont le site est placé entre l'étiquette de purification et la protéine d'intérêt.

*b) Elution de la protéine*

Dans le cas où la protéine est éluee, la résine est lavée dans un premier temps avec des concentrations faibles d'éluant afin de détacher les protéines liées de manière aspécifique à la résine. Dans le cas où l'interaction entre la résine et l'étiquette de purification est faible, la protéine peut déjà être éluee à ces concentrations. Cependant, dans la plupart des cas, la protéine est éluee à des concentrations plus fortes d'éluants. Ainsi, dans le cas des protéines portant une étiquette de purification poly-His retenues par du Cobalt sur la résine, un premier

lavage est généralement réalisé avec un tampon contenant 5mM d'imidazole, puis la protéine est éluée avec 250mM d'imidazole.

*c) Elution par coupure de l'étiquette de purification*

L'éluion de la protéine par coupure de l'étiquette de purification permet d'obtenir une protéine plus pure. En effet lors de l'éluion, les contaminants restants sont élués avec la protéine alors que, lorsqu'une protéase est utilisée, seule la protéine d'intérêt qui est fusionnée avec la séquence reconnue par la protéase est décrochée de la résine. La quantité de protéase ajoutée va être fonction de la quantité de protéine purifiée estimée. Ces protéases sont choisies du fait de leur haut degré de spécificité et de la faible occurrence des séquences qu'elles reconnaissent dans les séquences protéiques. Différentes protéases peuvent être utilisées : la thrombine, la TEV qui est une protéase issue du virus de la mosaïque du tabac (*Tobacco Etch Virus*), le facteur Xa ou l'enzyme Précission (P3C).

C'est principalement la thrombine qui a été utilisée au cours de ce travail de thèse. Cette protéase reconnaît spécifiquement la séquence protéique Leu-Val-Pro-Arg-Gly-Ser et coupe entre l'arginine et la glycine. Ce site de coupure est notamment codé par les vecteurs du Dr. Romier qui ont été utilisés (vecteur pnEA-tH). La protéase est ajoutée à la résine après l'étape de chromatographie d'affinité, la résine étant alors resuspendue dans le tampon de lyse et placée dans un falcon. Le volume dans lequel la résine est resuspendue doit être suffisamment important pour empêcher une précipitation des protéines du fait d'une trop forte concentration dans le tampon après coupure (généralement 3mL). Pour 1mg de protéine purifiée, 1U de thrombine à 1U/ $\mu$ L est ajoutée (MP Biomedicals). Les tubes sont placés sous agitation à 4°C pendant la nuit. Le lendemain la résine est placée dans les colonnes Poly-prep afin de récupérer l'éluat et de laver la résine jusqu'à ce que toute la protéine soit éluée. Un échantillon de résine après coupure permet généralement de vérifier que la coupure de l'étiquette de purification a bien été totale et que toute la protéine a pu être éluée.

La pureté des échantillons obtenus à l'issue d'une étape de chromatographie d'affinité n'est pas toujours suffisante pour la réalisation d'études structurales, c'est la raison pour laquelle au moins une étape de filtration sur gel est ajoutée. Par contre, la protéine obtenue peut être utilisée pour la réalisation de tests fonctionnels, et notamment d'expériences de pulldown, qui ont été réalisés au cours de cette thèse.

### 2.2.5.3. Chromatographie d'exclusion ou filtration sur gel

La filtration sur gel permet de séparer les molécules en fonction de leur taille et de leur encombrement par passage au travers d'un tamis. Ce tamis est constitué de particules sphériques qui sont poreuses, la taille des pores dans une colonne étant variable mais comprise dans une certaine gamme. Trois types de molécules sont distinguées : (i) les molécules très encombrées qui sont trop grosses pour passer à travers les pores et passeront à l'extérieur du tamis et qui sortiront les premières de la colonne, (ii) les molécules d'encombrement faible qui passeront à travers tous les pores et seront donc fortement retardées et finalement (iii) les molécules d'encombrement moyens qui passeront à travers certains pores et qui seront donc séparées en fonction de leur encombrement. La chromatographie d'exclusion a été utilisée au cours de cette thèse pour la purification des protéines mais également pour le désalage d'échantillons protéiques. Le désalage est réalisé sur un tamis ayant un volume d'exclusion très bas, qui va retenir uniquement le sel et les petites molécules mais pas les protéines.

La réalisation d'une chromatographie d'exclusion pour la purification de protéines nécessite l'utilisation d'un système de chromatographie de type Akta (Purifier ou Express). Dans un premier temps, la colonne est équilibrée avec 1,2 volume (de colonne) de tampon. L'échantillon est ensuite injecté. Afin d'obtenir la meilleure résolution possible des pics observés, le volume injecté ne doit pas dépasser un millième du volume de la colonne dans le cas des colonnes préparatives (S75 ou S200, GE Healthcare). Un système de détection UV permet d'identifier rapidement les fractions contenant des protéines, un gel d'acrylamide permettant de localiser la protéine d'intérêt dans les fractions éluées.

Dans le cas des colonnes de désalage (qui ne nécessitent pas de systèmes de chromatographie) le volume d'échantillon injecté peut représenter jusqu'à 20% du volume de la colonne.

La chromatographie d'exclusion peut également être utilisée pour estimer le poids moléculaire apparent d'une protéine (poids qui tient compte de l'encombrement de la protéine). Cette estimation est réalisée en comparant le volume d'exclusion de la protéine d'intérêt au volume d'exclusion de protéines de poids moléculaire connu. Cette observation peut notamment permettre de définir si la protéine est purifiée sous la forme d'un monomère ou d'un dimère.



#### 2.2.5.4. Concentration de l'échantillon et estimation de sa concentration

L'échantillon protéique obtenu à l'issue des étapes de chromatographie est généralement dilué, il doit donc être concentré avant la réalisation d'études structurales. Les échantillons sont concentrés sur des concentrateurs à membrane Amicon® Ultra. La membrane est poreuse et permet le passage des molécules en fonction de leur poids moléculaire et de leur forme. Les petites molécules d'eau, de sels et de tampon vont pouvoir passer au travers de la membrane alors que les protéines sont retenues au-dessus. Le volume de l'échantillon va ainsi être réduit sans perdre la protéine, conduisant donc à sa concentration. La membrane de ces systèmes de concentration va dans un premier temps être rincée avec le tampon de lyse de la protéine afin d'éliminer les traces de glycérol présentes. L'échantillon est ensuite chargé sur la membrane et le tube est centrifugé en plusieurs étapes de quelques minutes entre lesquelles la protéine est resuspendue. Ces étapes courtes permettent d'éviter que la protéine ne précipite sous l'effet du gradient de concentration qui se met en place au cours de la centrifugation. Lorsque la concentration voulue est atteinte, l'échantillon est prélevé sur la membrane puis utilisé pour les études structurales.

La concentration de l'échantillon peut être estimée, soit par la mesure de son absorbance à 280nm, soit en utilisant le réactif de Bradford.

La mesure de l'absorbance de l'échantillon à 280nm va permettre d'estimer la concentration de l'échantillon par l'utilisation de la loi de Beer Lambert. Cette loi relie l'absorbance mesurée à la concentration protéique en prenant en compte l'épaisseur du matériau traversé (en centimètres) et un coefficient d'extinction molaire  $\epsilon$  (en  $L \cdot mol^{-1} \cdot cm^{-1}$ ). Ce sont le tryptophane et la tyrosine qui présentent un pic d'absorption à 280nm, le coefficient d'extinction molaire va donc être fonction du nombre de ces résidus présents dans la séquence protéique. Cependant ce coefficient ne sera qu'une estimation puisque le repliement de la protéine et le pH va également influencer sur la capacité d'absorption des tryptophanes et donc sur la valeur du coefficient d'extinction molaire. Une détermination expérimentale de ce coefficient permet par conséquent le calcul exact de la concentration de la protéine. L'absorbance de l'échantillon protéique à 280nm va être mesurée sur le Nanodrop ND1000 dont est équipé le laboratoire.

Le réactif de Bradford contient du bleu de coomassie qui se lie aux acides aminés aromatiques présents dans la protéine. La fixation du bleu de coomassie sur ces résidus va induire un changement d'absorbance, la mesure étant réalisée à 595nm. Le changement

d'absorbance observé va être proportionnel à la quantité de réactif lié à la protéine, permettant de déduire la quantité de protéines présentes dans le mélange. Cette concentration protéique va être déduite d'une courbe étalon généralement réalisée avec la protéine BSA (Bovine Serum Albumin). Le réactif de Bradford est dilué au 5<sup>ème</sup> dans de l'eau et entre 1 et 20 $\mu$ L de l'échantillon est ajouté à 1mL de réactif dilué placé dans une cuve de spectrophotomètre en plastique. L'absorbance à 595nm est alors mesurée sur le spectrophotomètre du laboratoire. La variation du volume de protéines mélangées va permettre de se placer dans la gamme de linéarité du réactif, comprise entre une absorbance de 0,2 et une absorbance de 0,8.

## 2.3. Caractérisation physico-chimique et fonctionnelle de la protéine

Différentes paramètres physico-chimiques vont permettre de caractériser l'échantillon protéique avant de réaliser les premiers tests de cristallisation ou les mesures par RMN. Ces techniques vont notamment permettre d'estimer l'homogénéité de l'échantillon (DLS) ou sa stabilité. De plus, il est important d'estimer le repliement correct des protéines, c'est la raison pour laquelle des tests d'activité sont réalisés dans le cas d'enzymes ou d'interaction dans le cas de complexes.

### 2.3.1. Caractérisation physico-chimique

La caractérisation physico-chimique inclut toutes les mesures qui peuvent être réalisées et qui permettent de déterminer les propriétés physiques et chimiques de l'échantillon.

#### 2.3.1.1. La diffusion dynamique de lumière (DLS)

La diffusion dynamique de lumière (ou diffusion quasi-élastique de lumière) va permettre d'estimer la distribution de la taille des particules en solution. Cette distribution est représentée par un pourcentage de polydispersité qui va permettre de déduire l'homogénéité de l'échantillon protéique.

Au cours d'une expérience de DLS, l'intensité de lumière diffusée par les molécules en solution va être mesurée en fonction du temps. Cette diffusion de lumière est basée sur l'interaction entre la lumière laser (qui est un rayonnement monochromatique cohérent) et les molécules en solution. C'est le mouvement brownien de ces molécules en solution qui va causer les variations dans l'intensité de la lumière diffusée, l'intensité étant proportionnel à la

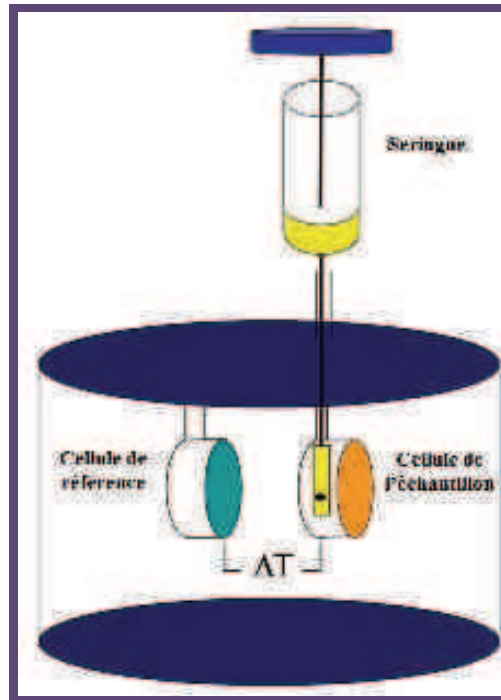
taille des particules. Le traitement mathématique de l'intensité mesurée en fonction du temps va nous renseigner sur la taille des particules en solution et la polydispersité de l'échantillon (profil de la fonction d'autocorrélation). Ainsi, à partir de l'intensité diffusée, un coefficient de diffusion  $D$  va pouvoir être déduit. A son tour il va permettre d'estimer le rayon hydrodynamique des molécules en solution par l'équation de Stokes-Einstein. Le nombre d'espèces présentes dans l'échantillon et leur rayon hydrodynamique vont ainsi pouvoir être déterminé.

Les mesures de DLS sont réalisées sur l'appareil Dynapro-MS disponible au laboratoire. Ces mesures sont réalisées sur l'échantillon concentré avant la réalisation de tests de cristallisation afin d'observer l'effet de la concentration sur la protéine. Après une étape de centrifugation de l'échantillon (10 minutes à 10000rpm) réalisée afin d'éliminer les poussières et les contaminants, 20 $\mu$ L de l'échantillon sont déposés dans la cuve en quartz dans laquelle la mesure est réalisée.

#### 2.3.1.2. Microcalorimétrie (ITC)

La microcalorimétrie ou ITC (Isothermal Titration Calorimetry) permet de caractériser la liaison d'une protéine et de son interactant (qui peut être une protéine, un acide nucléique, ou une petite molécule). Parmi ces paramètres, la constante d'affinité, la stoechiométrie du complexe ainsi que les valeurs d'entropie ( $\Delta S$ ) et d'enthalpie ( $\Delta H$ ) peuvent être mesurées.

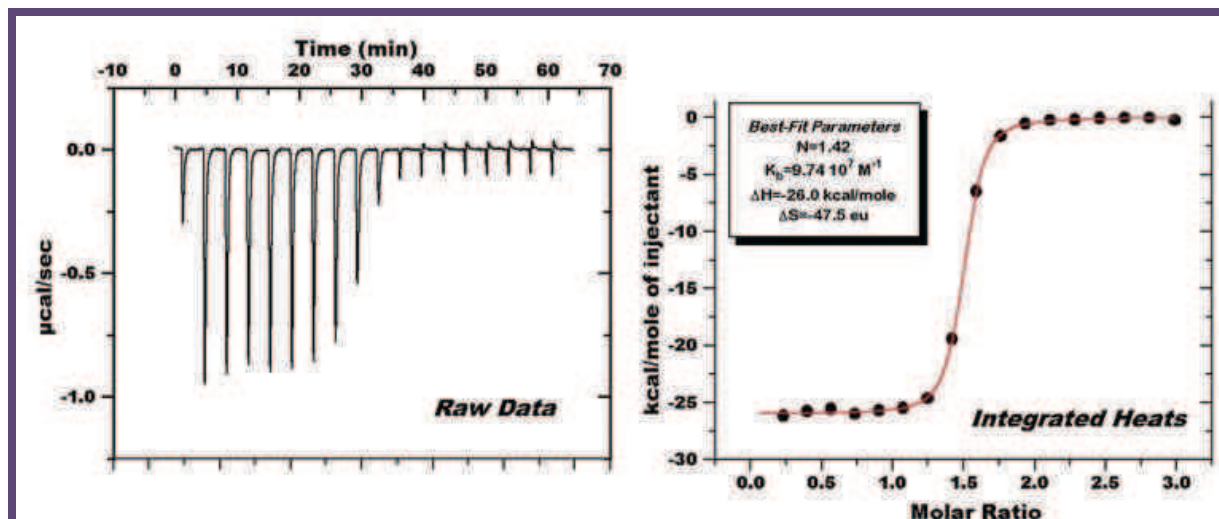
Le dispositif de l'ITC est composé d'une enceinte adiabatique contenant deux cellules : une cellule contenant l'échantillon et une cellule de référence qui va contenir le tampon de la protéine et qui tiendra lieu de standard interne (Figure 32). Le ligand est quant à lui placé dans la seringue.



**Figure 32 : Montage d'une expérience de microcalorimétrie.**

*Une enceinte adiabatique contient deux cellules qui vont contenir pour l'une le tampon de la protéine et pour l'autre la protéine dans laquelle le ligand va être injecté. C'est la quantité de chaleur  $\Delta T$  ( $\mu\text{cal/sec}$ ) à appliquer sur la cellule de l'échantillon pour que sa température reste identique à la chaleur de la cellule de référence qui va être mesurée (adapté de Chung, 2007).*

L'expérience a pour but de quantifier le dégagement ou l'absorption de chaleur provoquée par l'injection de l'interactant dans l'échantillon protéique en fonction du temps (Figure 33). La saturation des sites de liaison sur la protéine va conduire à l'apparition d'un plateau. Ces valeurs sont ensuite ramenées au ratio molaire de ligand injecté par rapport à la quantité de protéines dans la cellule permettant d'obtenir une courbe isotherme. A partir de cette courbe, les constantes d'affinité, la stœchiométrie et les valeurs d'enthalpie et d'entropie du complexe vont pouvoir être déduites. Différents paramètres vont pouvoir être testés et notamment la température ou la concentration en sel du tampon.



**Figure 33 : Résultats typiques d'une expérience de microcalorimétrie.**

*A chaque injection de ligand dans la cellule de l'échantillon, de la chaleur est dégagée (réaction exothermique) ou absorbée (réaction endothermique) par la mise en place de l'interaction. A mesure que les sites de la protéine sont occupés par le ligand, un phénomène de saturation de la protéine apparaît. Ces données sont ensuite intégrées en fonction des ratios protéines/ligand afin de définir les valeurs de  $K_d$ ,  $\Delta H$ ,  $\Delta S$  et la stœchiométrie du complexe.*

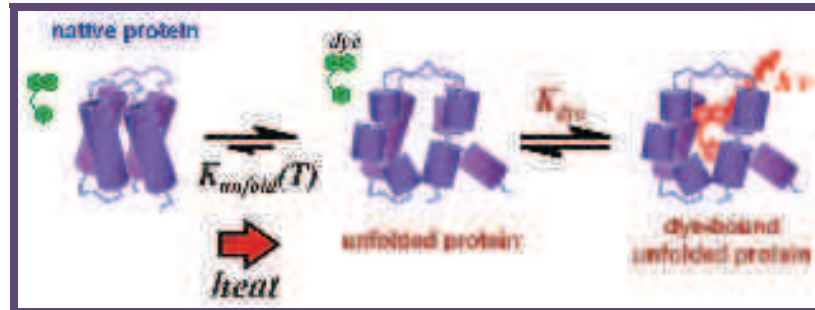
Les mesures ont été réalisées sur l'appareil ITC-200 de GE Healthcare de l'Institut de Biologie Moléculaire et Cellulaire (IBMC) de Strasbourg. 300µL de l'échantillon protéique sont placés dans la cellule de l'échantillon, différentes concentrations étant testées (autour de 20µM). Les peptides sont solubilisés dans le même tampon que la protéine, leurs concentrations étant 10 fois supérieure à la constante de dissociation estimée du complexe. Les quantités de ligand injectées dans la cellule vont dépendre de la force de l'interaction mesurée. Typiquement, 4µL de peptides sont injectés toutes les 120 secondes.

### 2.3.1.3. Fluorescence Thermal Shift Assay (Thermofluor)

La technologie du Thermofluor permet d'étudier la stabilité d'une protéine et sa liaison à un ligand. Cette technique permet d'estimer la température de fusion  $T_m$  d'une protéine qui est directement liée à sa stabilité. Il s'agira donc de mesurer la température de fusion d'une protéine en présence de son interactant ou de différents tampons, une augmentation du  $T_m$  reflétant un effet de stabilisation de la protéine. Cette technique a été utilisée au cours de cette thèse afin de vérifier l'interaction de protéines avec des peptides et d'observer l'effet de mutations ponctuelles.

L'expérience de Thermofluor consiste à mélanger la protéine (avec ou sans ligand selon les paramètres à définir) avec un marqueur fluorescent : le Sypro Orange. Il absorbe les

longueurs d'onde à 490nm et les réémets à 580nm. La fluorescence de ce marqueur est fortement augmentée en présence de patches hydrophobes exposés par des protéines dénaturées du fait d'une diminution de la constante diélectrique du milieu. Ainsi la fluorescence émise par le Sypro orange va augmenter avec la dénaturation de la protéine (Figure 34).



**Figure 34 : Principe de l'expérience de Thermofluor.**

*Au fur à mesure que le mélange réactionnel est chauffé, la protéine est dénaturée et le colorant va se fixer sur les patches hydrophobes exposés. C'est lorsque le colorant est fixé à la protéine qu'il fluoresce. La stabilisation de la protéine par sa liaison à un ligand ou l'utilisation d'un autre tampon va augmenter la température de fusion de la protéine (adapté de Pantoliano et al., 2001).*

Les mesures ont été réalisées sur l'appareil MiniOpticon Real Time PCR Detection System de BioRad, disponible au laboratoire. La protéine utilisée est concentrée à 50 $\mu$ M. Dans le cas où l'expérience est réalisée pour vérifier la formation d'un complexe, deux concentrations de ligand vont être testées : 100 et 200 $\mu$ M. Le marqueur Sypro Orange est ajouté à une concentration finale de 25X. Typiquement, la température est augmentée par paliers de 1°C toutes les minutes, entre 20 et 90°C.

## 2.3.2. Caractérisations fonctionnelles

### 2.3.2.1. GST pulldown

Les expériences de GST pulldown ont pour but d'identifier dans un extrait cellulaire les protéines interagissant directement ou indirectement avec la protéine d'intérêt. En effet, une protéine purifiée retenue sur une résine d'affinité va pouvoir interagir avec ses partenaires présents dans l'extrait cellulaire total. Ces interactants vont pouvoir être identifiés soit par Western Blot dans le cas où ils sont connus soit par spectrométrie de masse après dépôt sur gel d'acrylamide dénaturant et découpage des bandes présentes. Au cours de cette thèse, des expériences de GST pulldown ont été réalisées afin d'observer l'effet de mutations ponctuelles ou de la suppression de domaines structuraux sur la formation d'un complexe.

Dans un premier temps la protéine d'intérêt couplée à une étiquette de purification est produite. L'étiquette de purification choisie est généralement la GST puisqu'elle permet une fixation très spécifique de la protéine à la résine. Les cellules sont ensuite lysées et une étape de chromatographie est réalisée sur une résine Glutathion Sépharose 4B (GE Healthcare), comme décrit dans le paragraphe 2.2.5.2, sans que la protéine soit éluée. Un aliquot de résine est alors déposé sur gel d'acrylamide afin d'estimer la quantité de protéine purifiée. Une partie de la résine est alors incubée pendant 2 heures et 30 minutes à 4°C avec l'extrait cellulaire total préparé précédemment. Dans le cadre de cette thèse, des extraits cellulaires de levure ont été utilisés. Ils ont été obtenus à partir de souche sauvages par disruption mécanique des cellules (vortex) comme décrit dans la publication de Seraphin et Rosbash (1989), le protocole étant présenté en annexe 8.2 page 264. La résine est sédimentée par centrifugation, lavée à trois reprises puis reprise dans du tampon de charge (Laemmli). 10µL sont déposés sur gel d'acrylamide qui est soit coloré en bleu de coomassie afin de voir toutes les protéines retenues, soit utilisé pour la réalisation d'un western blot afin de détecter spécifiquement la présence d'une protéine. Les anticorps utilisés pour la réalisation de ces western blot sont soit commerciaux (cas des anticorps H5 et H14) soit obtenus auprès de collaborateurs (l'anticorps dirigé contre la protéine Spt5 a été obtenu auprès de Grant Hartzog, Santa Cruz, Université de Californie).

#### 2.3.2.2. Etudes de génétique de levure

La génétique de levure est un outil couramment utilisé pour la compréhension du rôle des protéines. En effet, la levure est un organisme eucaryote qui met en œuvre des mécanismes proches de ceux des eucaryotes supérieurs tout en étant facilement manipulable génétiquement. Ainsi, des mutations peuvent être introduites dans les gènes de levure et leur effet être observé au niveau cellulaire par l'apparition de phénotypes. Ce sont ces phénotypes qui vont nous renseigner sur la fonction de la protéine *in vivo*.

L'analyse des structures cristallographiques permet de définir les résidus importants pour la fonction de la protéine. Ce sont ces résidus qui vont être mutés *in vivo* afin d'observer leur importance dans la fonction de la protéine. Ces mutations ponctuelles sont introduites dans le gène codant pour la protéine d'intérêt entière, ce gène étant placé dans un vecteur de levure possédant un marqueur d'auxotrophie. Ce vecteur est ensuite utilisé pour transformer des souches de levure dont le gène natif codant pour la protéine d'intérêt a été déplété. Ces

levures sont ensuite étalées sur différents milieux contenant des agents chimiques induisant des stress.

Ces études de génétique de levure ont été réalisées par nos collaborateurs de la Harvard Medical School : le Pr. Fred Winston et Erin Loeliger. Les méthodes utilisées pour la réalisation de ces études sont décrites précisément dans les parties Matériel et Méthodes des deux publications présentées dans les chapitres suivants.

## 2.4. Etudes structurales par cristallographie aux Rayons X et résonnance magnétique nucléaire

La résolution de la structure d'une macromolécule va permettre de mieux comprendre au niveau moléculaire son mode d'action. Différentes méthodes vont permettre d'obtenir des informations à différentes résolutions. Les techniques de diffusion des rayons X aux petits angles (SAXS) et de microscopie électroniques vont permettre d'observer l'enveloppe de molécules alors que les résolutions atteintes grâce aux techniques de cristallographie aux rayons X et résonnance magnétique nucléaire (RMN) permettent d'obtenir des informations à l'échelle atomique. Cependant, ces deux dernières méthodes présentent chacune des limitations spécifiques : la résonance magnétique nucléaire est limitée aux petites molécules dont le poids moléculaire est inférieur à 30kDa alors que la cristallographie requiert l'obtention de cristaux de bonne qualité. Au cours de cette thèse, la cristallographie aux rayons X a été utilisée afin de résoudre la structure du domaine carboxy-terminal de Spt6 d'*A. locustae* et de plusieurs complexes formés de constructions des protéines Spt6 et Iws1 d'*E. cuniculi* alors que la RMN a été utilisée pour résoudre la structure d'un complexe contenant des constructions des protéines TFIIS (facteur d'élongation de la transcription) et Spt8 (sous-unité du complexe SAGA).

### 2.4.1. Résolution de structure par cristallographie aux rayons X

Suite aux étapes de clonage, purification et caractérisation de la protéine, le processus de résolution de structure par cristallographie aux rayons X est composé de cinq étapes principales : la cristallisation de l'échantillon, la collecte des jeux de données, la résolution du problème de phase, la construction du modèle, l'affinement et la validation de la structure.



### 2.4.1.1. Principe

La cristallographie utilise l'interaction élastique cohérente entre les rayons X et les électrons des atomes. Cette interaction sur un échantillon quelconque va donner naissance au phénomène de diffusion. Sur un cristal, qui est un élément périodique ordonné, la diffusion va se faire uniquement dans certaines directions privilégiées, c'est-à-dire qu'il y a diffraction. A chaque onde diffractée dans une direction particulière est associé un facteur de structure  $F_{hkl}$  qui est fonction de trois indices entiers  $h$ ,  $k$  et  $l$ .

La connaissance de chaque facteur de structure  $F_{hkl}$  va permettre d'accéder à la fonction de densité électronique  $\rho(x, y, z)$  :

$$\rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_{hkl} e^{-i2\pi(hx+ky+lz)}$$

qui représente une distribution continue d'électrons dans la maille cristalline. Cette fonction va être calculée pour chaque position  $(x, y, z)$  de la maille cristalline. Le facteur de structure  $F_{hkl}$  va être calculé pour chaque réflexion  $h, k, l$  :

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}}$$

où  $|F_{hkl}|$  est le module de ce facteur de structure et  $\phi_{hkl}$  est la phase de la réflexion considérée. Le module du facteur de structure va pouvoir être déterminé directement à partir des clichés de diffraction puisque l'intensité de chacune de ces tâches va être proportionnelle au carré du module du facteur de structure selon l'équation :

$$I_{hkl} \sim (|F_{hkl}|)^2$$

où  $I_{hkl}$  est l'intensité de la tâche de réflexion de coordonnées  $h, k$  et  $l$ . L'obtention de la phase associée à chaque réflexion va par contre être plus complexe, puisque cette information est perdue lors de la collecte des données, principalement du fait de l'absence de refocalisation des faisceaux diffractés. Pour déterminer ces phases, deux familles d'approches principales peuvent être utilisées :

- les méthodes qui utilisent le remplacement isomorphe (SIR et MIR) et la diffusion anormale (SAD et MAD) par l'introduction de différences d'intensité dans les facteurs de structure,
- la méthode de remplacement moléculaire par l'utilisation d'un modèle, c'est-à-dire une protéine de repliement proche et dont la structure est connue.

#### 2.4.1.2. Cristallisation et collecte des jeux de données

La première étape de la résolution de la structure d'une protéine par cristallographie va consister à cristalliser l'échantillon protéique.

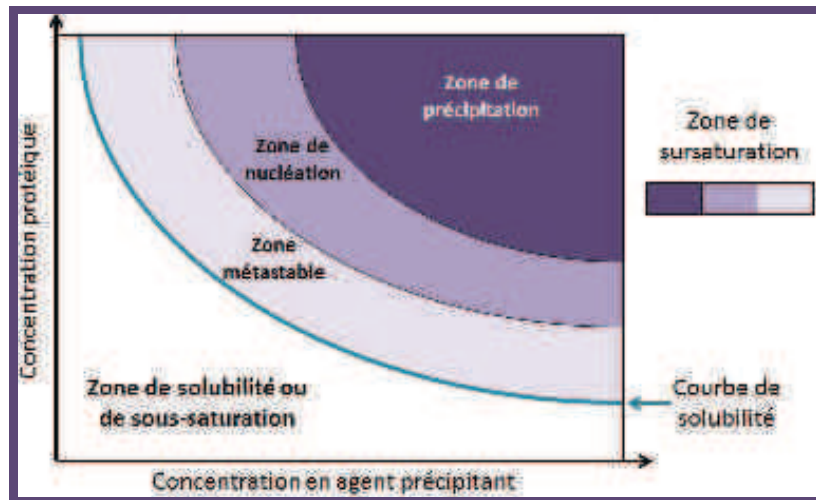
##### *a) La cristallisation de l'échantillon protéique*

La cristallisation de la protéine a pour but de passer d'un état liquide à un état solide ordonné. L'obtention de cet état solide est induite par l'ajout d'agents de cristallisation dont le rôle est de diminuer la solubilité de la protéine. Les agents de cristallisation peuvent être des sels, des polymères ou encore des solvants organiques. La cristallisation va également être influencée par d'autres paramètres tels que la concentration de la protéine, son pH, la température, la présence d'additifs (molécules chimiques ou ligand de la protéine) et la pureté et l'homogénéité de l'échantillon.

L'obtention des cristaux va d'abord consister en une phase de recherche des conditions de cristallisation qui vont permettre d'obtenir les premières pistes cristallines. En principe, les différents agents de cristallisation susceptibles de permettre la cristallisation sont testés, de même que les différents paramètres qui peuvent favoriser la cristallisation de l'échantillon. Cependant, vu le nombre important de conditions qui résulte de cette recherche systématique, des kits de cristallisation qui contiennent (i) soit un nombre restreint de conditions réparties de manière statistique dans l'espace de recherche à N dimensions, (ii) soit des conditions donnant souvent des cristaux, sont généralement utilisés (Classics, JCSG, Peg/Ion/pH, ...). Une fois les premières pistes de cristallisation définies, une étape d'optimisation est le plus souvent nécessaire afin d'améliorer la qualité et la taille des cristaux.

Le processus de passage de l'état liquide à l'état solide ordonné va se faire selon le diagramme de phase, qui comprend autant de dimension que de paramètres testés. Le diagramme représenté sur la Figure 35 est une simplification du diagramme de phase où seules les concentrations en agent de cristallisation et en protéine sont représentées. Ce diagramme est principalement composé de quatre zones : (i) une zone de sous-saturation où la protéine reste en solution, (ii) une zone métastable dans laquelle aucun cristal n'apparaît mais où un cristal peut croître notamment après ensemencement, (iii) une zone de nucléation dans laquelle les germes cristallins apparaissent et (iv) une zone de précipitation dans laquelle la protéine précipite. Des cristaux peuvent être obtenus lorsque la zone de nucléation est atteinte. La diminution de la concentration de protéine en solution va permettre d'atteindre la zone

métastable dans laquelle le cristal va croître jusqu'à ce que la concentration de protéine en solution atteigne la limite de solubilité.



**Figure 35 : Diagramme de phase à deux dimensions**

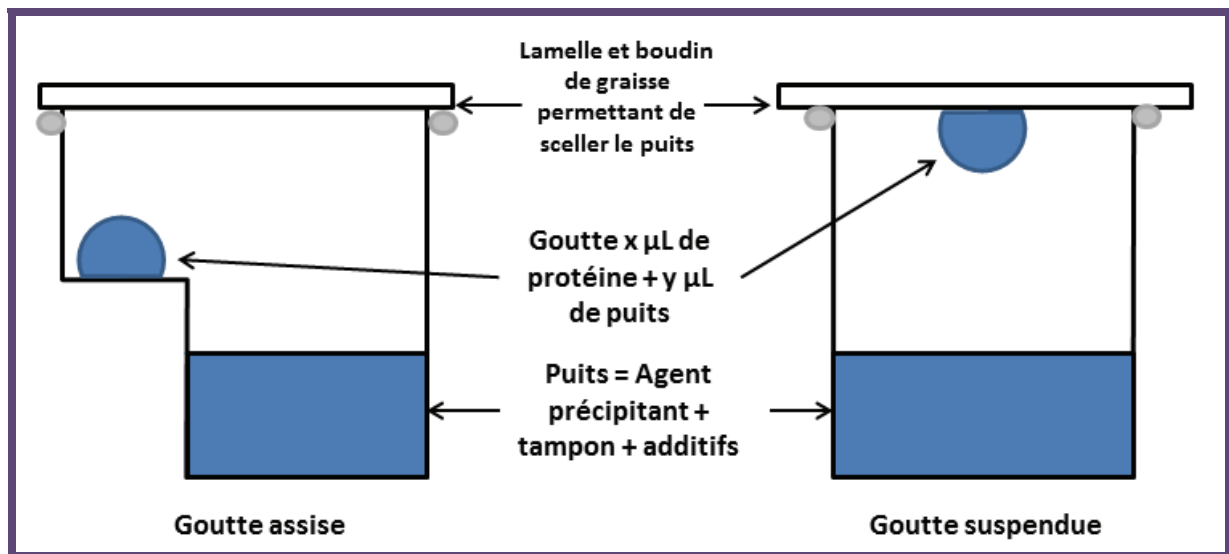
*Dans la zone de sous-saturation la concentration de la protéine est inférieure à sa solubilité, aucun cristal ne peut donc se former. Par contre, dans les zones de sursaturation, la concentration de la protéine est supérieure à sa solubilité et par conséquent des cristaux peuvent être formés.*

C'est la technique de diffusion de vapeur qui a été utilisée au cours cette thèse, notamment par la mise en place des méthodes de goutte assise et goutte suspendue (représentées schématiquement sur la Figure 36).

La cristallisation va être réalisée dans une enceinte close. Une goutte est formée avec un volume de protéine et un volume de puits, conduisant à une dilution de l'agent précipitant par rapport à sa concentration dans le puits. Il y aura donc diffusion des espèces volatiles (eau, alcool ...) de la goutte vers le puits pour rétablir cet équilibre et, par conséquent, concentration de l'agent précipitant et de la protéine dans la goutte. Il y aura ainsi passage de la zone de solubilité vers les zones de sursaturation. Ce processus peut être rapide mais également très lent selon la cinétique de la diffusion. Ainsi, des cristaux peuvent apparaître dans les boîtes de cristallisation plusieurs mois après qu'elles aient été faites.

La technique de la goutte assise a été utilisée lors de la recherche des conditions de cristallisation (Figure 36). La définition de ces conditions de cristallisation s'appuie sur l'utilisation de kits de cristallisation généralistes, chacun regroupant 96 conditions différentes. Ces kits recouvrent un grand nombre de combinaisons d'agents précipitants (sels, PEG ...), de tampons et d'additifs. En plus de tester plusieurs kits, différentes températures et concentrations protéiques sont également testées. Il en résulte que la définition des conditions

de cristallisation va généralement nécessiter de tester plusieurs centaines de conditions de cristallisation. C'est la raison pour laquelle ces tests sont réalisés dans des boîtes de 96 puits à l'aide de robots disponibles sur la plateforme. Le robot Tecan Temo 96-head est utilisé pour le remplissage des boîtes de cristallisation et le robot Cartesian Honeybee 8+1 est utilisé pour la formation des gouttes. Ces robots permettent notamment de n'utiliser que de faibles quantités de protéines puisque des gouttes de  $0,2\mu\text{L}$  de protéine et  $0,2\mu\text{L}$  d'agent de cristallisation peuvent être réalisées. Une fois la protéine et la solution d'agent précipitant dispensés, la boîte est scellée à l'aide d'un film en plastique permettant de créer une enceinte close où l'équilibre est mis en place.



**Figure 36 : Schéma des dispositifs de goutte assise et de goutte suspendue**

*Une goutte est formée entre un volume de protéine et un volume de solution du puits. La différence de concentration en agent précipitant ainsi induite va conduire à de la diffusion d'eau de la goutte vers le puits et ainsi, une augmentation progressive de la concentration en agent de cristallisation.*

La technique de la goutte suspendue a été utilisée pour l'optimisation des conditions de cristallisation. Les cristaux obtenus à l'issue des tests de cristallisation sont parfois trop petits pour permettre la collecte de jeux de données, bien que ce problème soit réduit par la mise en place de lignes de lumière microfocuss dans la plupart des synchrotrons. Pour optimiser les cristaux, des gouttes sont réalisées à la main,  $2\mu\text{L}$  de protéine et  $2\mu\text{L}$  de réservoir étant utilisés pour réaliser les gouttes. Cependant les conditions déterminées lors des premiers tests de cristallisation ne sont pas toujours parfaitement reproductibles lors de l'étape d'optimisation. Très souvent, il est nécessaire de varier encore le pH, les concentrations en agents précipitants et en protéine et la température afin d'obtenir des cristaux utilisables. Ces

étapes d'optimisation sont réalisées dans des boîtes contenant 24 puits de type Greiner. L'enceinte est close par le dépôt d'une lamelle silanisée sur un boudin de graisse (vaseline).

*b) Congélation des cristaux et collecte des jeux de données*

Les sources de lumières utilisées au synchrotron lors de la collecte des jeux de données sont très intenses et vont conduire à l'endommagement des cristaux par la génération de radicaux libres. Afin de minimiser la dégradation des cristaux, ils sont généralement congelés dans l'azote liquide avant d'être soumis à ces faisceaux, les collectes étant alors réalisées entre 100 et 140K. Cependant les cristaux de protéines contiennent une forte proportion de solvant qui forme de la glace cristalline lors du processus de congélation, détruisant le cristal. C'est pourquoi le cristal est préalablement trempé dans une solution de cryoprotectant avant sa congélation. Ce cryoprotectant va remplacer l'eau dans les canaux de solvant du cristal et lors de la congélation va former de la glace amorphe qui n'endommagera pas le cristal. En général, plusieurs cryoprotectants sont testés pour mettre en évidence celui ou ceux qui n'endommagent pas les cristaux.

Les cryoprotectants les plus couramment utilisés sont les PEG de bas poids moléculaire (200 et 400), le glycérol, les sucres (de types sucrose, glucose ou xylitol) ou l'éthylène glycol. Entre 20 et 30 % de ces composés sont incorporés à la solution de cristallisation utilisée. Lors de sa congélation le cristal est prélevé à l'aide d'une boucle dans la goutte et est trempé rapidement dans la solution de cryoprotectant avant d'être plongé dans l'azote liquide.

Les jeux de données peuvent être enregistrés soit au laboratoire soit au synchrotron. Les détecteurs utilisés sont des détecteurs CCD (Charge-Coupled Device), qui permettent une acquisition très rapide, voire instantanée des données. Classiquement, lors de la collecte des données de diffraction, le cristal est tourné perpendiculairement au faisceau des rayons X avec des angles de faible amplitude (entre 0,2 et 1°). Cet angle est fonction des paramètres de maille et de l'orientation du cristal. De plus, le nombre de clichés de diffraction nécessaires à l'obtention d'un jeu de données complet va dépendre de la géométrie cristalline du cristal. C'est pourquoi, dans un premier temps, deux ou trois clichés sont pris, chacun espacés de 45 à 90°. Ces premiers clichés vont permettre d'établir une stratégie de collecte qui sera un compromis entre la collecte du meilleur jeu de données possible et la durée de vie cristal. Cette stratégie est basée sur une estimation de la symétrie cristalline, des paramètres de la maille du cristal et de la diffraction maximale du cristal.

### 2.4.1.3. Traitement des données

Une fois les jeux de données collectés, ils sont traités à l'aide de programmes informatiques qui permettent d'intégrer l'ensemble des données. Dans notre cas nous avons utilisé le programme HKL2000 (Otwinowski et Minor, 1997). Dans un premier temps, le système cristallin va être déterminé à partir du premier cliché de diffraction (indexation). Le jeu de données complet va ensuite être intégré : l'intensité de chaque réflexion, définie par ses coordonnées  $h$ ,  $k$  et  $l$  va être déterminée. Une fois l'intégration réalisée, un fichier est généré qui contient les paramètres de maille et pour chaque réflexion, son intensité ( $I_{hkl}$ ) et la valeur du bruit de fond aux alentours de cette réflexion ( $\sigma_{hkl}$ ). La qualité du jeu de données peut être estimée à travers un certain nombre de paramètres tels que : (i) le  $R_{sym}$  (qui reflète la différence d'intensité entre réflexions équivalentes), (ii) le facteur  $I/\sigma I$  (qui reflète l'intensité moyenne des données collectées rapportées au bruit de fond), (iii) la redondance (nombre de fois où une réflexion et ses symétriques sont mesurées) et (iv) la complétude (pourcentage de réflexions mesurées par rapport au nombre de réflexions attendues). Ces valeurs sont généralement utilisées pour avoir une estimation plus réaliste de la résolution effective du jeu de données.

### 2.4.1.4. Résolution du problème des phases

Le problème des phases, issue de la perte de cette information au cours de la collecte des données, peut être résolue par deux approches principales : le remplacement moléculaire, qui est basé sur la connaissance de la structure d'une protéine de repliement proche, ou l'introduction de différences d'intensité dans les facteurs de structure par les méthodes de remplacement isomorphe (SIR et MIR) et de diffusion anormale (SAD et MAD).

#### *a) Le remplacement moléculaire*

La technique du remplacement moléculaire a principalement été utilisée pour résoudre la structure des complexes Spt6/Iws1.

Le remplacement moléculaire est utilisable lorsque la structure d'une molécule dont le repliement est proche de celui attendu pour la protéine d'intérêt est connue. Cette similitude est généralement détectée par l'observation d'une similarité de séquence entre les deux protéines qui est détectée par la réalisation d'alignements de séquence.

Le phasage consiste alors à prendre la structure de la molécule connue et de trouver son orientation et sa position dans la maille cristalline. Cette recherche est réalisée le plus

souvent par la recherche de la meilleure superposition des fonctions de Patterson expérimentales et du modèle. La fonction de Patterson va représenter sous forme vectorielle les distances entre les atomes dans la molécule, deux protéines de structures proches ayant des fonctions de Patterson similaires. La superposition des fonctions de Patterson peut être subdivisée en deux étapes : la recherche de l'orientation de la molécule dans la maille par l'application d'une fonction de rotation et la recherche de la position dans la maille par l'application d'une fonction de translation. Une fois l'orientation trouvée, les phases du modèle vont être utilisées comme phases initiales pour déterminer la structure de la molécule d'intérêt.

Dans le cas des complexes Spt6/Iws1, c'est la structure de la protéine Iws1, résolue au laboratoire, qui est utilisée comme modèle. Le programme Phaser (McCoy et al., 2007) de CCP4 a été utilisé pour résoudre ces structures.

#### *b) Phasage par la technique MIRAS*

Le phasage par MIRAS (Multiple Isomorphous Replacement and Anomalous Scattering) a été utilisée pour résoudre la structure du domaine carboxy-terminal de la protéine Spt6. Cette technique de phasage est basée sur une combinaison des méthodes de remplacement isomorphe et de diffusion anormale et utilise plusieurs dérivés lourds dont l'un au moins possède une contribution anormale significative, afin de résoudre le problème des phases

Le phasage par remplacement isomorphe va consister à introduire des dérivés lourds dans le cristal par trempage, les paramètres de maille et le groupe d'espace de ce cristal devant être quasiment identiques pour éviter les problèmes de non-isomorphisme. La méthode de la série isomorphe va considérer les jeux de données obtenus à partir d'un cristal natif et de un (méthode SIR) ou plusieurs (méthodes MIR) jeux de données contenant un dérivé lourd. Les changements induits dans la densité électronique de la molécule native résultent uniquement du signal des atomes supplémentaires introduits. Ainsi, en connaissant la position des atomes lourds dans la maille, grâce à la fonction de Patterson différence, l'obtention d'un seul jeu de données contenant un atome lourd va permettre de réduire à deux valeurs possibles la valeur de la phase pour une réflexion donnée. Un deuxième jeu de données permettra de lever complètement l'incertitude sur les phases. Cependant, dans le cas du phasage par la méthode SIR, si le dérivé lourd introduit possède une contribution anormale, cette contribution peut être utilisée afin de lever totalement l'indétermination sur les phases.

La méthode de la diffusion anormale va tenir compte de la présence d'atomes dans le cristal qui ont une contribution anormale. Ces atomes ont un seuil d'absorption proche de la longueur d'onde incidente et vont donc entrer en résonance, conduisant à un changement de phase de l'onde diffractée en fonction de la longueur d'onde incidente (appelé phénomène de diffusion anormale). Ce changement de longueur d'onde est caractérisé par un facteur de diffusion  $f$ , qui modélise l'interaction des rayons X avec un atome donné n'est alors plus égal à un nombre entier et doit être décomposé selon

$$f = {}^0f + {}^\lambda f' + i {}^\lambda f''$$

où  ${}^0f$  correspond à la diffusion « normale » (absence de diffusion anormale) et  $f'$  et  $f''$  sont deux paramètres correctifs qui dépendent de la longueur d'onde. Dans le cas particulier de ces atomes, la loi de Friedel n'est plus respectée, cette propriété pouvant être utilisée pour lever le problème des phases.

Différents métaux lourds ont été testés pour résoudre la structure du domaine carboxy-terminal de Spt6. Les cristaux natifs ont été trempés pendant 12 heures dans la solution de cristallisation contenant 1mM de métaux lourds d'or et plomb. Ils sont alors congelés dans une solution de cristallisation contenant 25% de xylitol qui est ici utilisée comme cryoprotectant. Ces cristaux ont été utilisés au synchrotron. Trois jeux de données ont été collectés : (i) un jeu obtenu à partir d'un cristal natif (ne contenant pas d'atomes lourds), (ii) un jeu obtenu à partir d'un cristal trempé dans une solution à 1mM de potassium dicyanoaurate ( $\text{KAu}(\text{CN})_2$ ) et (iii) un jeu obtenu à partir d'un cristal trempé dans une solution à 1mM et de triméthyl-acétate de plomb ( $((\text{CH}_3)_3\text{PbAc})$ ). Ces jeux de données ont été traités séparément avec HKL2000 et ont été utilisés pour la résolution de structure par le programme auto-Sharp (Vonnrhein et al., 2007). Le phasage par MIR réalisé n'ayant pas permis d'obtenir une carte de densité de qualité suffisante, les contributions anormales de l'or et du plomb ont été utilisées.

De plus, la modification de la densité par nivellement de solvant (réalisée par auto-Sharp à l'aide du programme Solomon) a permis d'améliorer les cartes initiales. Cette technique de modification de densité tient compte de la présence de fortes proportions de solvant dans le cristal qui ne participent pas à la diffraction, ces zones de solvant présentant une densité faible et relativement plate. La connaissance du pourcentage de solvant dans la maille cristalline va être nécessaire afin d'appliquer un masque qui aura une valeur de 1 dans les zones de protéine et 0 dans les zones de solvant. L'application de ce masque et le calcul de



nouvelles phases va permettre l'amélioration de la carte de densité initiale par l'amélioration des phases.

#### 2.4.1.5. Construction, affinement et validation de la structure

##### *a) Construction et affinement de la structure*

Une première carte de densité (carte expérimentale) est obtenue à l'issue de l'étape de phasage. Cette carte est utilisée pour construire le premier modèle avec le programme COOT (Crystallographic Object-Oriented Toolkit, Emsley et Cowtan, 2004). Il est ensuite optimisé à l'aide des programmes Refmac5 de Ccp4 (Collaborative, 1994; Murshudov et al., 1997) ou Phenix Refine (Adams et al., 2002), afin d'améliorer le modèle initial et, par conséquent, les phases initiales du modèle. L'affinement cristallographique réalisé a pour but de minimiser l'écart entre le modèle et les données de diffraction, par la minimisation d'une fonction qui tient compte des paramètres expérimentaux et du modèle. L'incertitude sur les phases est notamment une source d'erreur importante puisque leur dans le calcul de la densité électronique est majeur. Les calculs de minimisation vont permettre de réduire l'erreur sur les phases, ce qui améliorera fortement la carte de densité et par conséquent le modèle. Les paramètres qui vont être optimisés grâce à ces calculs vont être la position des atomes (paramètres x, y et z), les facteurs de déplacement atomique (facteur d'agitation thermique B et facteurs d'occupation). Cependant, le nombre de données à optimiser issues du modèle est bien plus grand que le nombre d'observations réalisées. Ainsi, afin de réduire le nombre de paramètres à optimiser, certains paramètres vont être ajoutés par l'incorporation de contraintes et de restrictions par exemple sur la longueur des liaisons et des angles. Différents algorithmes de minimisation peuvent être utilisés tels que : la méthode des moindres carrés ou la méthode du maximum de vraisemblance.

Dans le cas des complexes Spt6/Iws1, une étape d'affinement par TLS a été réalisée qui consiste à considérer des sous-structures du modèle comme étant des corps rigides. En effet, les facteurs de déplacements atomiques anisotropes ne peuvent être modélisés du fait de la limite de diffraction (par exemple à basse résolution). Les sous-structures définies peuvent alors être modélisées comme des corps rigides, la contribution de leurs facteurs de déplacement atomiques pouvant alors être considérée comme étant proche, permettant de réduire le nombre de paramètres à minimiser. L'affinement par TLS va considérer trois types de mouvement de ces corps rigides : un mouvement de translation (T, translation), un mouvement de vibration (L, libration) et un mouvement de rotation translatrice (S, screw).

Ainsi, lorsque plusieurs sous-structures sont définies, elles vont bouger l'une par rapport à l'autre mais aussi dans la maille cristalline par rapport au solvant. Cette technique d'affinement va notamment améliorer les facteurs d'accord cristallographiques

A l'issue de l'étape de minimisation, différentes cartes de densités vont être calculées qui ont pour but de mettre en évidence les différences et erreurs qui subsistent dans le modèle par rapport aux données expérimentales, permettant ainsi une correction manuelle. Ces cartes, appelées « Fourier différence », vont tenir compte des phases calculées et de leur pondération, ainsi que de la différence entre les facteurs de structures observés et calculés, auxquels des coefficients sont appliqués. Deux cartes sont principalement utilisées : (i) la carte de type «  $2F_o - F_c$  », qui permettra de construire le modèle et (ii) la carte de type «  $F_o - F_c$  » qui présentera une densité positive autour des atomes mal positionnés et négative là où un atome devrait être présent.

L'affinement du modèle de la structure est un processus itératif constitué de plusieurs cycles de minimisation suivis d'une étape de reconstruction manuelle, les améliorations du modèle étant suivies grâce aux facteurs d'accord. Ces facteurs d'accord  $R$  et  $R_{\text{free}}$  vont permettre d'estimer l'adéquation des facteurs de structure du modèle avec les facteurs de structure expérimentaux. Le facteur  $R$  va porter sur la majorité des facteurs de structure alors que le facteur  $R_{\text{free}}$  ne va porter que sur une petite fraction (typiquement 5%) des données qui ne servent pas au processus d'affinement. Ce processus de minimisation suivi de reconstruction manuelle va être répété jusqu'à ce que les facteurs d'accord cristallographiques soient stabilisés à des valeurs acceptables (typiquement entre 15 et 30% selon la résolution pour les macromolécules) et que les cartes de Fourier différence présentent peu de densités positives et négatives après les cycles de minimisation.

#### *b) Validation du modèle*

La dernière étape du processus de résolution d'une structure est une étape de validation et de contrôle-qualité du modèle qui va avoir pour but de comparer les paramètres du modèle construit aux données expérimentales afin de s'assurer que le modèle est le meilleur étant donné les observations et les logiciels utilisés, notamment grâce aux facteurs d'accord cristallographiques  $R$  et  $R_{\text{free}}$ , mais également d'un point de vue de ses caractéristiques géométriques.

Parmi les caractéristiques géométriques inspectées, le diagramme de Ramachandran permet de vérifier la bonne géométrie de sa chaîne principale. L'inspection des rotamères des

différents résidus, la longueur moyenne et les angles moyens des liaisons entre les atomes ainsi que les encombrements stériques vont être vérifiés. Ces données peuvent être obtenues grâce à différents serveurs tels que Molprobit (Davis et al., 2007). Si besoin, le type de domaine structural de la protéine va être identifié grâce au serveur DALI (Holm et Rosenström, 2010) qui va comparer la structure du modèle aux structures disponibles dans la PDB.

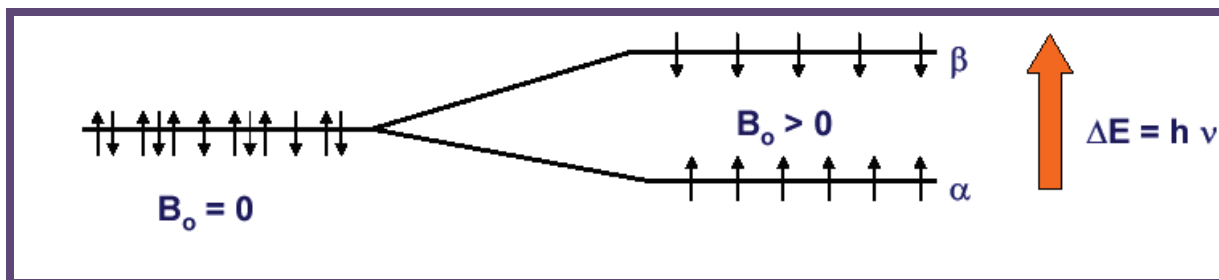
## 2.4.2. Résolution de structure par résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) est un phénomène physique très largement utilisé, notamment en imagerie médicale (IRM) où la résonance des protons de l'eau est utilisée pour visualiser les tissus mous. Dans le cas de la RMN des protéines, ce sont également les atomes d'hydrogènes qui sont étudiés, ainsi que les atomes de carbone et d'azote. La résolution de la structure d'une protéine par RMN va nécessiter la mise en place de quatre étapes : (i) la production d'un échantillon marqué au carbone 13 et à l'azote 15, (ii) l'enregistrement des spectres di- et tridimensionnels, (iii) l'assignement des résonances de chaque atome de la protéine et (iv) le calcul de la structure de la protéine. Au cours de cette thèse, la structure du complexe protéique TFIIS/Spt8 a été résolue par RMN avec l'aide du Dr. Christian Koehler de l'équipe de RMN biomoléculaire du Dr. Bruno Kieffer.

### 2.4.2.1. Principe

#### *a) La résonance magnétique nucléaire*

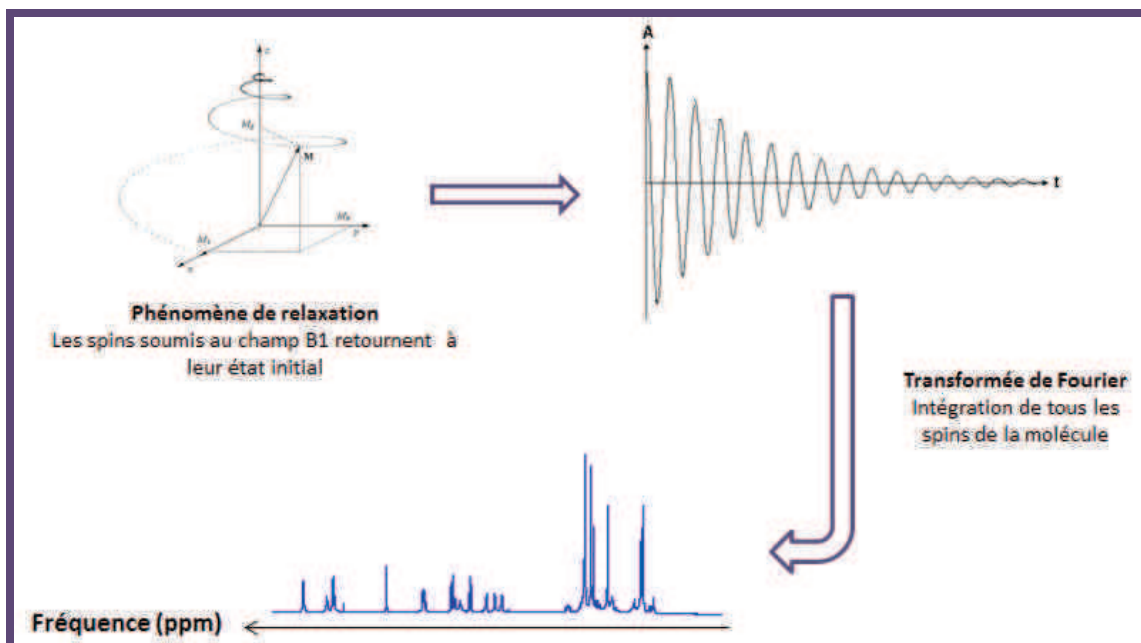
La résonance magnétique nucléaire utilise une propriété des atomes possédant un spin égal à  $\frac{1}{2}$ , ces spins étant répartis en deux états de spin  $+\frac{1}{2}$  et  $-\frac{1}{2}$  sous l'effet d'un champ électromagnétique (Figure 37). Différents atomes peuvent être considérés pour réaliser ces mesures et notamment l'hydrogène  $^1\text{H}$ , le carbone  $^{13}\text{C}$  et l'azote  $^{15}\text{N}$ . Lorsque ces atomes sont placés dans un champ magnétique  $B_0$  chaque état de spin va correspondre à un niveau d'énergie (Figure 37), la fréquence ou moment cinétique de chacun de ces spins est alors fonction du champ  $B_0$ .



**Figure 37 : Etat des spins nucléaires en absence et en présence d'un champ magnétique.**

Un atome de spin  $1/2$  (cas de  $^1H$ ,  $^{13}C$  et  $^{15}N$ ) possède deux états de spin :  $+1/2$  et  $-1/2$  (cas  $B_0=0$ ) en l'absence de champ magnétique. Lorsque cet atome est placé dans un champ magnétique (cas  $B_0>0$ ), chaque état de spin va adopter un niveau d'énergie, l'état de plus faible énergie présentant un léger excès de population. L'impulsion nécessaire au passage des spins des niveaux de basse énergie vers les niveaux de haute énergie est égale à  $\Delta E=h\nu$  et est fonction de l'atome considéré.

Au cours d'une expérience de RMN, l'échantillon soumis au champ initial  $B_0$  va être soumis par impulsion à un second champ magnétique  $B_1$  perpendiculaire au champ  $B_0$ . L'énergie  $\Delta E$  de cette impulsion est égale à la différence entre les deux niveaux d'énergie. Cette seconde impulsion va rompre l'équilibre qui existe entre les spins peuplant les deux niveaux d'énergie, induisant une augmentation de la population du niveau d'énergie le plus élevé. Le phénomène de relaxation décrit le retour des spins de basse énergie à leur état initial (Figure 38).



**Figure 38 : Mesure et exploitation du phénomène de relaxation**

Le retour du spin à son niveau initial va induire des fluctuations de la magnétisation dans le temps qui vont être enregistrés. La transformée de Fourier va permettre de transformer le signal afin qu'il soit exploitable (exprimé en en partie par million ou ppm).

C'est la fréquence de résonance des spins au cours du phénomène de relaxation qui est mesurée au cours de l'expérience de RMN (mesurée en Hertz). Le déplacement chimique  $\delta$  (en partie par million, ppm) va être déduit de la fréquence de résonance du noyau considéré. L'environnement chimique de l'atome considéré, va fortement influencer ce déplacement. Le signal enregistré va donc contenir les informations de tous les noyaux excités présents dans la molécule. La Transformée de Fourier va être utilisée pour extraire le signal de chaque proton du signal sigmoïdal obtenu (Figure 38).

#### *b) Application à la résolution de structures de protéines*

La détermination de la structure d'une protéine par RMN va nécessiter la réalisation d'expériences hétéronucléaires. En effet, le grand nombre d'atomes présents dans la molécule biologique, notamment les protons, conduirait dans le cas d'expériences homonucléaires à la superposition des pics rendant les spectres inexploitable. L'enregistrement des spectres va alors nécessiter l'application de séquences d'impulsions complexes. Ces impulsions vont permettre le transfert de l'aimantation d'un atome d'hydrogène vers les atomes d'azote ou de carbone présents à proximité. Ces transferts peuvent être réalisés soit (i) le long des liaisons atomiques, donnant une information sur les atomes directement reliés à l'atome considéré soit (ii) directement aux atomes présents dans l'environnement proche de l'atome ayant subi l'excitation (moins de 5Å) et constitue l'effet Overhauser (effet NOE). Ainsi pour chaque atome d'hydrogène, de carbone et d'azote, une valeur de déplacement chimique  $\delta$  en ppm va pouvoir être définie. La connaissance de ces valeurs de déplacement chimique et l'observation des effets Overhauser va permettre de définir des contraintes de distance entre les atomes de la protéine. Plusieurs structures qui répondent à ces contraintes ainsi qu'aux contraintes topologiques des protéines (longueur de liaison, angles dièdres, stéréochimie ...) vont ainsi pouvoir être calculées.

#### 2.4.2.2. Préparation de l'échantillon et enregistrement des spectres

Le complexe marqué avec du  $^{13}\text{C}$  et du  $^{15}\text{N}$  est produit puis purifié selon les protocoles décrits précédemment. Le tampon de lyse utilisé contient 50mM NaCl et 10mM d'un tampon phosphate (couple  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ ) à pH7. La mesure de spectre RMN va nécessiter 300 $\mu\text{L}$  d'échantillon à environ 200 $\mu\text{M}$  auquel 10% d'eau deutérée ( $\text{D}_2\text{O}$ ) est ajouté. Après centrifugation, l'échantillon est placé dans un tube fin en verre de type Shigemi, l'échantillon devant avoir une hauteur de 5cm dans ce tube. Le tube est alors placé dans un spectromètre

Brucker 700MHz. Les spectres sont enregistrés à 300K afin de favoriser les transferts d'aimantation.

Différents types de spectres sont enregistrés :

- Des spectres à deux dimensions  $^1\text{H}$ - $^{15}\text{N}$  et  $^1\text{H}$ - $^{13}\text{C}$  qui vont permettre d'observer la corrélation entre les atomes d'azote et d'hydrogène et de carbone et d'hydrogène respectivement. Dans le spectre  $^1\text{H}$ - $^{15}\text{N}$ , un pic de corrélation est observé pour chaque groupement amide de la liaison peptidique (un pic par résidu) ainsi que pour les chaînes latérales des résidus possédant des groupements amides (asparagine, glutamine, tryptophane). Ce spectre constitue la signature d'une protéine.
- Des spectres à trois dimensions vont permettre de réaliser l'attribution séquentielle des déplacements chimiques. Ces spectres ainsi que les corrélations observées sont présentés dans le Tableau 7. Les spectres utilisés pour l'attribution séquentielle sont colorés, ceux colorés de manière identique étant comparés ensemble.
- Deux spectres à trois dimensions NOESY ( $^{15}\text{N}$  et  $^{13}\text{C}$ ) vont permettre l'observation des effets NOE, qui donnent une information sur les atomes proches dans l'espace de l'atome considéré

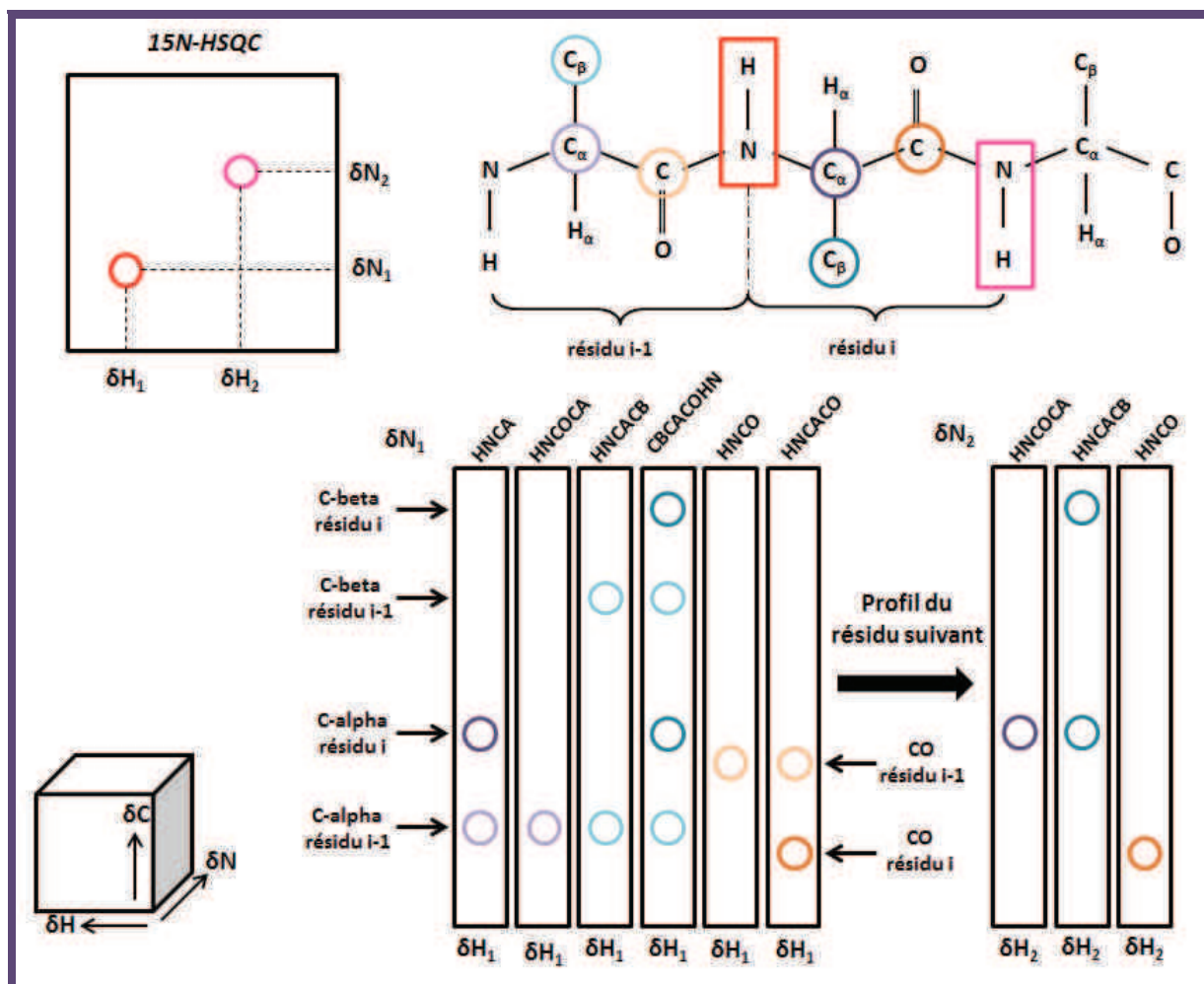
Spectre	N dimensions	Corrélation
<sup>15</sup> N-HSQC	2D	Azote et hydrogène liés covalents
<sup>13</sup> C-HSQC	2D	Carbones et hydrogènes liés covalents
HNCA	3D	Hydrogène de la chaîne principale avec les C-alpha des résidus suivants et précédents
HNCOCA	3D	Hydrogène de la chaîne principale avec les C-alpha des résidus précédents
HNCACB	3D	Hydrogène de la chaîne principale avec les C-alpha et C-béta des résidus suivants et précédents
CBCACONH	3D	Hydrogène de la chaîne principale avec les C-alpha et C-béta des résidus précédents
HNCO	3D	Hydrogène de la chaîne principale avec le carbone du groupement carbonyle des résidus précédents
HNCACO	3D	Hydrogène de la chaîne principale avec le carbone du groupement carbonyle des résidus suivants et précédents
HBHACONH	3D	Hydrogène de la chaîne principale et H-alpha et H-béta de la chaîne latérale du résidu précédent
COSY	3D	Hydrogènes de la chaîne latérale avec hydrogènes de rang +1 et -1
TOCSY	3D	Tous les hydrogènes des chaînes latérales
<sup>15</sup> N-NOESY	3D	A partir de l'azote de la chaîne principale ou des chaînes latérales, tous les hydrogènes reliés et effet NOE.
<sup>13</sup> C-NOESY	3D	A partir du carbone de la chaîne principale ou des chaînes latérales, tous les hydrogènes reliés et effet NOE.

**Tableau 7 : Spectres tridimensionnelles mesurés pour l'attribution tridimensionnelle des spectres RMN**

### 2.4.2.3. Stratégie d'assignement

Dans un premier temps, les déplacements chimiques de tous les atomes d'hydrogène, de carbone et d'azote de la protéine sont déterminés. La réalisation de cet assignement est généralement divisée en deux temps : (i) l'assignement séquentiel qui a pour but de trouver les systèmes de spins qui se suivent dans la séquence et (ii) l'assignement des atomes des chaînes latérales qui permet d'associer le résidu à un système de spin. Ces assignements ont été réalisés avec le logiciel Sparky développé par Goddard et Kneller.

#### *a) Assignement séquentiel*



**Figure 39 : Stratégie d'assignement séquentiel**

Le spectre  $^{15}\text{N}$ -HSQC va permettre de définir la fréquence des atomes d'azote et d'hydrogène de la chaîne principale. En naviguant dans les spectres 3D, la fréquence (i) des carbones des groupements carbonyles (ii) des C-alpha et (iii) des C-beta du résidu considéré et du résidu précédent vont être déterminés. La connaissance de la fréquence des carbones du résidu précédent vont alors permettre de trouver la fréquence de l'azote et de l'hydrogène du résidu précédent, permettant ainsi d'évoluer le long de la chaîne peptidique.

L'attribution séquentielle a pour but de trouver les valeurs de déplacement chimique des atomes de la chaîne principale de la protéine et de relier les atomes de la chaîne principale. Seuls les atomes N, H, C, C-alpha et C-beta de la chaîne principale seront considérés. Le spectre  $^{15}\text{N}$ -HSQC permet dans un premier temps d'obtenir les déplacements chimiques des atomes d'hydrogène et d'azote de la chaîne principale. Les spectres HNCA/HNCOCA, HNCACB/CBCACONH et HNCO/HNCACO vont permettre de déterminer les valeurs des déplacements chimiques des atomes C-alpha, C-beta et CO (carbonyle) du résidu propre i et du résidu précédent i-1 (Figure 39). Le spectre HBHACONH



permet quant à lui de définir les déplacements chimiques des atomes H-alpha et H-béta du résidu précédent i-1.

La connaissance des valeurs de déplacement chimique des atomes C-alpha, C-beta et CO du résidu suivant va permettre de rechercher dans les spectres HNCOCA, HNCACB et HNCO les systèmes de spin présentant ces valeurs. Les valeurs des déplacements chimiques des atomes H et N de la liaison peptidique du résidu suivant i va ainsi pouvoir être retrouvée.

#### *b) Assignement des chaînes latérales*

Une fois les systèmes de spins successifs définis, les spectres COSY et TOCSY sont utilisés pour attribuer ces systèmes à la séquence protéique. Les valeurs de déplacement chimiques des atomes C-alphas et C-beta sont des indications du type d'acide aminé. En effet, les valeurs de déplacement chimiques des atomes des chaînes latérales sont caractéristiques des différents résidus. Ainsi, l'identification des chaînes latérales de plusieurs acides aminés va permettre de replacer la séquence protéique dans la succession des systèmes de spins déterminés lors de l'attribution séquentielle.

### 2.4.2.4. Calcul de la structure

#### *a) Identification des effets NOE et calcul de la structure*

Grâce à l'assignement des déplacements chimiques des atomes, il est ensuite possible d'assigner les signaux NOE dont l'intensité est proportionnelle à la distance entre atomes. Le calcul de la structure est réalisé avec la suite logicielle UNIO (Guerry et Herrmann, 2012). Dans un premier temps, le logiciel va utiliser les spectres NOESY 3D et les valeurs de déplacements chimiques déterminées précédemment afin de définir les effets Overhauser. En effet, les spectres NOESY 3D ( $^{15}\text{N}$  et  $^{13}\text{C}$ ) permettent d'observer les corrélations entre un atome d'hydrogène de la chaîne principale ou latérale et les atomes d'hydrogènes reliés soit covalents, soit spatialement par l'effet Overhauser par transfert de la magnétisation à travers un atome d'azote ( $^{15}\text{N}$  3D NOESY) ou de carbone ( $^{13}\text{C}$  3D NOESY). Ainsi pour un atome d'hydrogène donné, certains pics de corrélation vont correspondre aux atomes d'hydrogènes des chaînes principales et latérales ( $\text{H}\alpha$ ,  $\text{H}\beta$  ...) pour lesquels les valeurs de déplacements chimiques sont connus, alors que d'autres pics seront issus de la corrélation entre l'atome d'hydrogène considéré et un atome d'hydrogène spatialement proche. L'intensité du pic de corrélation va être proportionnelle à la distance entre les deux atomes d'hydrogènes et la valeur du déplacement chimique de l'atome en interaction va permettre de

définir de quel atome et de quel résidu il s'agit. Le programme va ainsi identifier dans les spectres NOESY les corrélations intra- et inter-résidus et déduire les distances séparant deux atomes spatialement proches.

#### *b) Identification des éléments de structure secondaire*

Les déplacements chimiques des atomes de la chaîne principale (C-alpha, C, N et H) et du résidu C-beta sont généralement caractéristiques des éléments de structure secondaire dans lesquels ils sont localisés (hélice, feuillet ou boucle). Ainsi, ces valeurs de déplacement chimique sont comparées aux valeurs de déplacements chimiques déterminées pour des structures résolues par RMN, permettant de déduire des valeurs d'angle phi et psi pour le résidu considéré. Le programme Talos+ (Shen et al., 2009) va utiliser la séquence de la protéine et les valeurs de déplacements chimiques des atomes N, H, C, C-alpha et C-beta afin de définir les éléments de structure secondaire de la protéine. Ces éléments de structure secondaire vont permettre d'introduire des contraintes lors du calcul de la structure.

De plus, les contacts NOE observés vont également donner des informations concernant la structure secondaire de la protéine. En effet, des contacts NOE pourront être observés entre les résidus  $i$  et  $i+3/i+4$  qui sont localisés dans une hélice  $\alpha$ , la présence de feuillets  $\beta$  induisant des contacts NOE entre les résidus face à face dans ces éléments de structure secondaire.

La liste de contrainte de distance et la connaissance des éléments de structure secondaire de la protéine vont ainsi permettre la réalisation de calculs de dynamique moléculaire. Ces calculs sont réalisés avec le programme CYANA qui réalise des dynamiques sur les angles de torsion de la molécule (Güntert, 2004).

#### **2.4.2.5. Temps de relaxation T1/T2 – dynamique moléculaire**

Le retour d'un spin à son état d'origine après application d'un champ B1 peut être décomposé en deux phénomènes : (i) un phénomène de disparition du plan de l'état excité et (ii) un phénomène de réapparition dans le plan de l'état d'équilibre. A chacun de ces phénomènes, une vitesse de relaxation est associée. Ainsi, T1 (relaxation longitudinale) va correspondre au temps de retour des spins à leur état initial et T2 (relaxation transversale) va correspondre au temps de disparition du spin de l'état excité. Ces temps de relaxation sont fonction de l'environnement du spin considéré.

Au cours d'une expérience de relaxation, des spectres  $^{15}\text{N}$ -HSQC sont enregistrés après différents temps  $T_1$  et  $T_2$ . L'évolution de l'intensité des pics de corrélation observés en fonction des temps  $T_1$  et  $T_2$  de relaxation imposés va alors refléter la dynamique de la molécule. Dans ces expériences, seuls les couplages H-N de la chaîne principale sont considérés. Ainsi, si le rapport  $T_1/T_2$  est élevé (entre 4 et 6), le résidu considéré est localisé dans une zone rigide alors que si ce rapport est faible (entre 1 et 3), le résidu est localisé dans une région flexible.

**Chapitre 3**

**Etude biochimique, structurale et  
fonctionnelle du domaine carboxy-terminal  
de Spt6 et de son recrutement par l'ARN  
polymérase II**



## 3.1. Introduction

La protéine Spt6 est recrutée par l'ARN polymérase II au niveau du corps des gènes par son extrémité carboxy-terminale (Yoh et al., 2007). L'analyse de la séquence de la protéine Spt6 de levure a permis de définir la présence dans cette région d'un domaine SH2 (Src Homology 2) (John MacLennan et Shaw, 1993). La présence de ce domaine structural a par la suite été confirmée dans les protéines humaines et de souris (Chiang et al., 1996). Dans ce chapitre, les domaines SH2 et les particularités du domaine SH2 de la protéine Spt6 seront décrits. Les résultats obtenus lors de cette étude seront ensuite présentés sous la forme d'une brève description suivie de la publication parue dans la revue « The Journal of Biological Chemistry » (Diebold et al., 2010b). Les résultats additionnels non publiés seront également présentés. Pour terminer, la discussion permettra de replacer ces résultats dans un contexte biologique et de comparer nos données aux données publiées par d'autres équipes (Dengl et al., 2009; Sun et al., 2010; Close et al., 2011; Liu et al., 2011).

### 3.1.1. Rôles et architecture des domaines SH2.

#### 3.1.1.1. Rôle des domaines SH2 dans la signalisation cellulaire

La différenciation et la prolifération cellulaire ainsi que le contrôle du cycle cellulaire dépendent de signaux externes à la cellule. Ces signaux sont notamment propagés par les récepteurs à activité tyrosine kinase, dont l'activation par un ligand va induire la phosphorylation de résidus tyrosines dans leur partie cytoplasmique (Lemmon et Schlessinger, 2010). La phosphorylation de ces tyrosines va induire le recrutement de protéines contenant notamment des domaines SH2, leur rôle étant de transmettre ce signal dans la cellule. Ces protéines peuvent être des kinases, des protéines adaptatrices ou encore des facteurs de transcription.

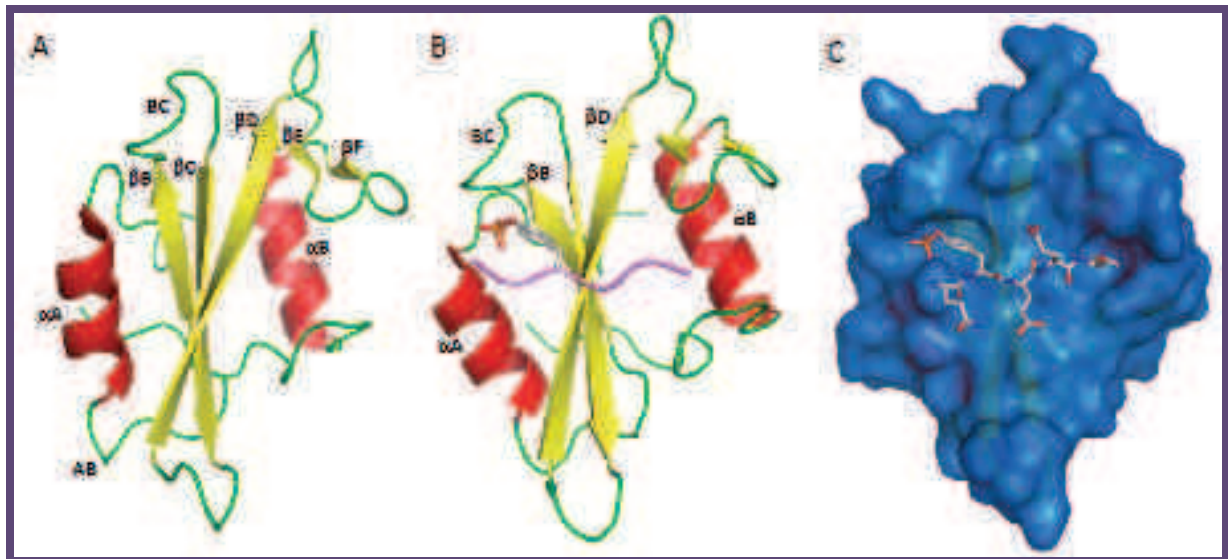
Dans les cellules humaines, 120 domaines SH2 ont pu être identifiés, montrant leur très large implication dans la signalisation cellulaire chez les mammifères (Liu et al., 2010, 2006). Les domaines SH2 jouent ainsi un rôle important dans le développement cellulaire et des mutations qui interviennent dans ces domaines sont parfois à l'origine de cancers, notamment certaines formes de leucémies (Liu et al., 2006).

Les domaines SH2 vont permettre la mise en place d'interactions modérées, la constante de dissociation de l'interaction entre le domaine et sa protéine cible étant de l'ordre

du micromolaire. Cette interaction peut être renforcée par la présence de domaines additionnels tels que les domaines SH3, qui lient les motifs proline-X-X-proline, ou par la présence de deux domaines SH2 successifs (Machida et Mayer, 2005).

### 3.1.1.2. Architecture des domaines SH2

Les domaines SH2 interagissent spécifiquement avec des protéines portant des résidus phosphotyrosine, les résidus en position +3 par rapport à cette phosphotyrosine permettant la reconnaissance spécifique d'un domaine pour sa cible (Huang et al., 2008) (Figure 40). Ces domaines d'une centaine de résidus sont très conservés au niveau de leur séquence et de leur repliement global, la présence d'éléments additionnels permettant de distinguer deux classes de domaines (présentées dans le Tableau 8).



**Figure 40 : Repliement du domaine SH2 de la protéine Src et interaction au peptide phosphorylé.**

*(A) Repliement global des domaines SH2, le domaine SH2 représenté étant celui de la protéine Src (B) Interaction d'un fragment portant un résidu phosphotyrosine avec le domaine SH2 de la protéine Src (C) la surface du domaine formant deux poches (code PDB : 1SPS).*

Type de domaines SH2	Eléments additionnels
<b>Src</b>	Présence d'un petit feuillet additionnel entre le feuillet $\beta$ D et l'hélice $\alpha$ B, constitué des brins $\beta$ E et $\beta$ F. La boucle entre ces deux brins (boucle EF) participe à la reconnaissance spécifique de la cible (résidu en position +3).
<b>STAT</b>	La dernière hélice $\alpha$ B est plus longue que dans les domaines de type Src, le feuillet $\beta$ E – $\beta$ F est absent et est parfois remplacé par une courte hélice $\alpha$ B' (Gao et al., 2004).

**Tableau 8 : Différentes familles de domaines SH2 et leurs caractéristiques.**

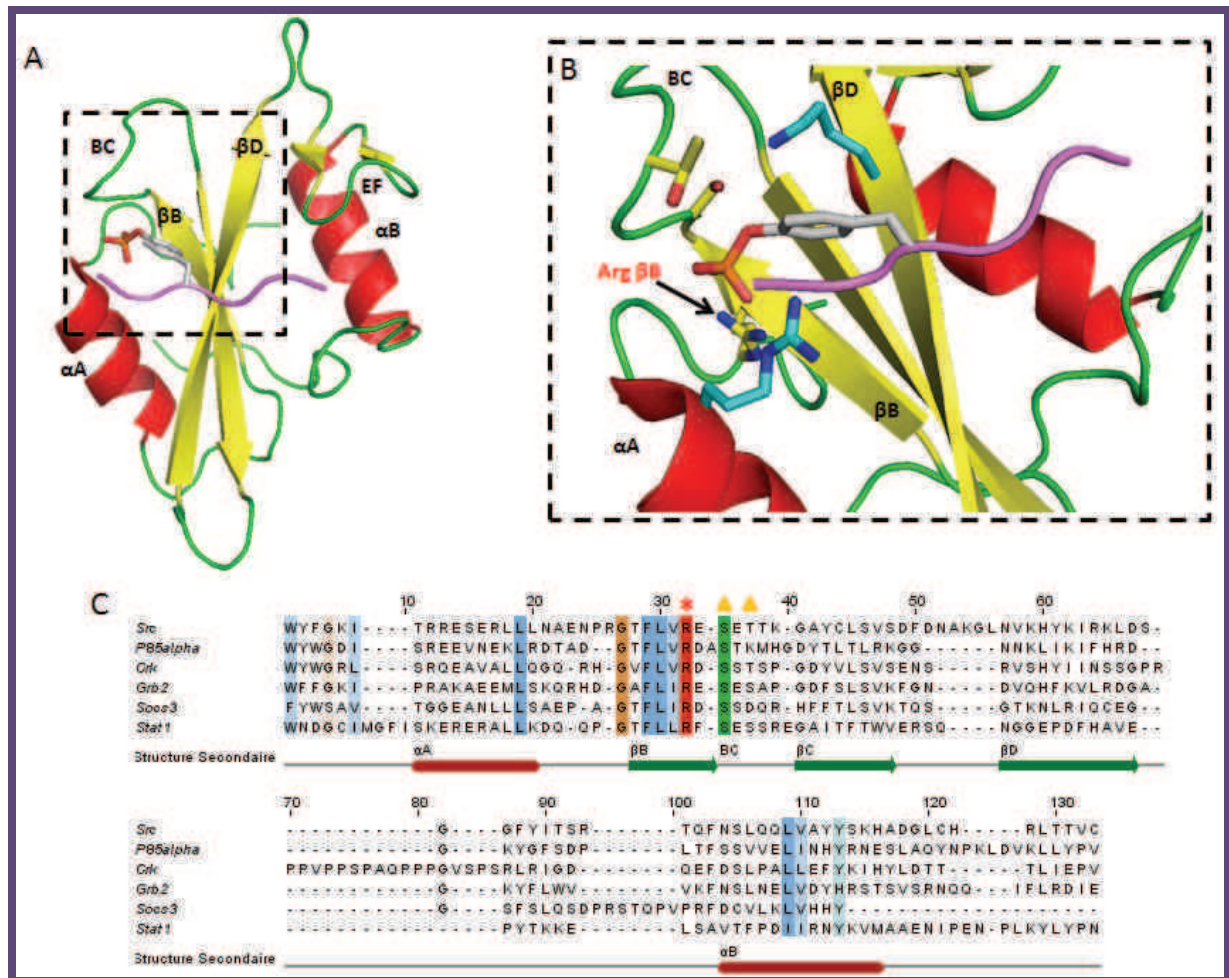
Les domaines SH2 sont constitués d'un feuillet  $\beta$  antiparallèle central constitué de trois brins ( $\beta$ B,  $\beta$ C et  $\beta$ D) pris en sandwich par deux hélices  $\alpha$  ( $\alpha$ A et  $\alpha$ B) (Figure 40A) (Kuriyan et Cowburn, 1997). Les protéines contenant des phosphotyrosines interagissent à la surface du domaine, perpendiculairement au feuillet  $\beta$  central (Figure 40B). La phosphotyrosine interagit dans une poche délimitée par l'hélice  $\alpha$ A et le feuillet  $\beta$ D. La poche de spécificité, située de l'autre côté du feuillet  $\beta$  central entre le brin  $\beta$ D et l'hélice  $\alpha$ B, va permettre la reconnaissance spécifique de la cible (Figure 40C). Ce mode d'interaction au niveau de deux poches profondes et distinctes par un fragment protéique étendu à la surface du domaine est appelé « two-pronged » (Machida et Mayer, 2005).

### 3.1.1.3. Interaction des domaines SH2 avec les phosphotyrosines

Au niveau de la poche d'interaction de la phosphotyrosine seul un résidu arginine du feuillet  $\beta$ B est conservé dans tous les domaines SH2, noté Arg  $\beta$ B (Figure 41), sa mutation abolissant l'interaction avec leurs protéines cibles. L'importance de ce résidu dans la formation des complexes s'explique au niveau structural par la réalisation d'interactions ioniques entre le groupement guanidinium de cette arginine et le groupement phosphate de la phosphotyrosine (Figure 41B). Au niveau de la boucle BC qui relie les brins  $\beta$ B et  $\beta$ C, les chaînes latérales et principales de certains résidus participent également à l'interaction avec le groupement phosphate de la phosphotyrosine. Aucune conservation de séquence n'est observée au niveau de cette boucle BC, bien qu'elle contienne un grand nombre de résidus sérine et thréonine (Figure 41C). De la même manière que la mutation de l'arginine  $\beta$ B abolit l'interaction entre les domaines SH2 et leurs cibles phosphorylées, un peptide non



phosphorylé ne peut interagir avec un domaine SH2. Cette poche de fixation de la phosphotyrosine contribue pour moitié à la force de l'interaction (Kuriyan et Cowburn, 1997).



**Figure 41 : Interaction entre la phosphotyrosine et la poche de fixation du domaine SH2 de la protéine Src.**

(A) le groupement phosphate de la phosphotyrosine interagit au niveau de la poche formée par l'hélice  $\alpha A$ , le feuillet  $\beta B$  et la boucle BC. (B) Agrandissement de la poche de fixation de la phosphotyrosine. Les résidus réalisant des interactions de type électrostatique avec le groupement phosphate sont colorés en jaune, les résidus formant l'interaction de type amino-aromatique sont colorés en bleu. (C) alignement des séquences de certains domaines SH2 humains. Les éléments de structure secondaire indiqués sont ceux communs aux deux classes de domaines SH2, (\*) indique la position de l'arginine  $\beta B$  strictement conservée, ( $\Delta$ ) indique la position des résidus de la boucle BC participant à l'interaction avec la phosphotyrosine.

Le complexe est renforcé par la formation d'interactions de type amino-aromatique entre des résidus de l'hélice  $\alpha A$  et du feuillet  $\beta D$  avec le cycle aromatique de la phosphotyrosine (Figure 41B). Ce type d'interaction fait intervenir des résidus portant un groupement amine (résidus de type lysines, arginines, asparagines ou glutamines) qui réalise une interaction de type Van der Waals avec les électrons  $\pi$  du cycle aromatique d'un résidu

phénylalanine, tyrosine ou tryptophane (Burley et Petsko, 1986). Cependant, ce type d'interaction n'est pas retrouvé dans tous les domaines SH2, l'absence du résidu de l'hélice  $\alpha A$  qui participe à cette interaction modifiant l'orientation du groupement et du cycle aromatique de la phosphotyrosine (cas du domaine SH2 de la phosphatase SH-PTP2) (Kuriyan et Cowburn, 1997).

#### 3.1.1.4. Poche de spécificité de la reconnaissance

La seconde poche retrouvée à la surface des domaines SH2 est bien moins conservée, les domaines SH2 des différentes protéines contenant une poche de spécificité différente. En effet, elle va permettre la reconnaissance spécifique de la protéine cible phosphorylée par son interaction avec le résidu en position +3 par rapport à la phosphotyrosine. Cette seconde poche est formée de résidus de la boucle entre les éléments  $\beta D$  et  $\alpha B$  et notamment la boucle EF qui relie les brins  $\beta E$  et  $\beta F$  dans le cas des domaines SH2 de type Src et de résidus de l'hélice  $\alpha B$ . La chaîne latérale du résidu en position +3 de la phosphotyrosine va entrer dans cette poche, les modes de liaison adoptés étant divers, et vont notamment dépendre du type de motifs reconnus. Dans le cas de la protéine Src, cette seconde poche met en place une interaction hydrophobe avec un résidu isoleucine du peptide YEEL.

En général, les domaines SH2 sont donc constitués de deux poches, l'une fortement conservée qui permet la reconnaissance de la phosphotyrosine et une seconde poche plus variable, qui va être différente selon la protéine cible reconnue. Il existe cependant des domaines SH2 qui mettent en place des modes de liaison différents de ceux décrits ici pour la reconnaissance spécifique de leurs protéines cibles (Kuriyan et Cowburn, 1997).

#### 3.1.2. Le domaine SH2 de Spt6 – intérêt de son étude

Lorsque ce projet a été initié, le rôle du domaine SH2 de Spt6 avait été mis en évidence par des expériences réalisées dans des cellules de souris. Des expériences de pulldown ont montré qu'une construction du domaine carboxy-terminal de Spt6 qui contient le domaine SH2 retient l'ARN polymérase II dont les sérines en position 2 sont phosphorylées (forme hyperphosphorylée, ARNP<sub>IIo</sub>) (Yoh et al., 2007). La mutation de l'arginine  $\beta B$  strictement conservée dans ces domaines abolit l'interaction avec l'ARN polymérase II, confirmant l'implication du domaine SH2 dans cette interaction. De plus (i) la déphosphorylation des sérines en position 2 de l'ARN polymérase II par la phosphatase Fcp1a abolit l'interaction, alors que (ii) l'incubation de l'ARN polymérase II déphosphorylée avec le

facteur PTEF-b permet de reformer le complexe. D'autre part, la phosphorylation des sérines en position 5 par TFIIH ou des tyrosines en position 1 par Abl1 ne permet pas de former le complexe (Yoh et al., 2007).

L'interaction directe entre Spt6 et l'ARN polymérase II permet notamment d'expliquer la colocalisation de ces deux protéines lors de la transcription sur le corps des gènes (Kaplan et al., 2000) et la détection de Spt6 au niveau des promoteurs des gènes par précipitation de chromatine (Chromatin Immunoprecipitation ou ChIP) (Kim et al., 2004). Cette interaction avait de plus déjà été suggérée par la réalisation d'expériences d'immunoprécipitations, sans que le domaine de Spt6 responsable de cette interaction ait été identifié (Lindstrom et al., 2003). Il apparaît donc que le domaine SH2 de Spt6 intervient dans le recrutement de cette protéine par l'ARN polymérase II hyperphosphorylée.

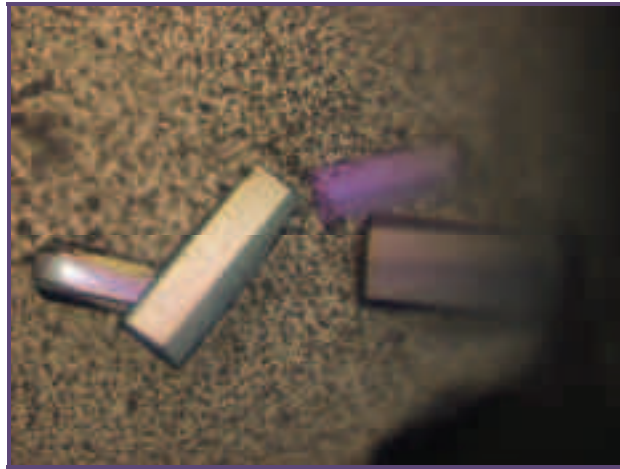
Ainsi, le domaine SH2 de Spt6 apparaît comme non canonique puisqu'il reconnaît une sérine phosphorylée, contrairement aux autres domaines SH2. Des changements structuraux au niveau de la poche de fixation du résidu phosphorylé sont donc attendus qui peuvent être révélés par la résolution de la structure de ce domaine. De plus, bien qu'un grand nombre de domaines SH2 soient connus dans les cellules humaines, le domaine SH2 de la protéine Spt6 est le seul qui ait été détecté chez la levure et les eucaryotes inférieurs en général (John Maclennan et Shaw, 1993). Ces observations confirment donc le rôle fonctionnel important de ce domaine de la protéine Spt6, suggérant qu'il puisse être l'ancêtre des domaines SH2 retrouvés chez les eucaryotes supérieurs.

### 3.2. Brève description des résultats

Cette partie a pour but de présenter succinctement les résultats décrits dans la publication numéro 1 (paragraphe 3.3, page 141).

Les bornes d'un fragment codant pour l'extrémité carboxy-terminale et contenant le domaine SH2 des protéines Spt6 d'*Encephalitozoon cuniculi* et *Antonospora locustae* ont été déterminées à partir d'un alignement multiple de séquences. Les gènes codant pour ces fragments ont ensuite été clonés dans les vecteurs pnEA-tH, permettant d'exprimer les protéines en fusion avec une étiquette de purification poly-(His) à son extrémité amino-terminale. Ces protéines ont été produites dans la bactérie, purifiées et utilisées pour des essais de cristallisation. Des cristaux ont été obtenus à partir de la construction de la protéine

d'*Antonospora locustae* (Figure 42), permettant de résoudre la structure de la protéine par la méthode MIRAS.



**Figure 42 : Cristaux du domaine carboxy-terminal de la protéine Spt6 d'*Antonospora locustae***  
*Cristaux de la construction (783-957) de la protéine Spt6 d'*Antonospora locustae* (groupe d'espace P2<sub>1</sub>2<sub>1</sub>2, résolution 2,2Å)*

Cette structure a révélé que le domaine carboxy-terminal de Spt6 contient en fait deux domaines SH2 de type Src (Tableau 8) organisés en tandem. Le premier de ces domaines SH2 (SH2<sub>1</sub>) possède les principaux éléments caractéristiques de ces domaines. Cependant, la poche de fixation de la phosphotyrosine de ce domaine est moins profonde et donc serait plus adaptée à la reconnaissance d'un résidu phosphosérine. La poche de spécificité de ce domaine est formée en partie par des résidus de la boucle BC du second domaine SH2 (SH2<sub>2</sub>) et de l'hélice  $\alpha$ B du domaine SH2<sub>1</sub> qui est plus longue que dans les domaines SH2 standards. Le second domaine SH2 (SH2<sub>2</sub>) est par contre non canonique. L'arginine  $\beta$ B est remplacée par une tyrosine qui ne pointe pas vers la poche de fixation de la phosphotyrosine alors que l'arginine  $\alpha$ A et la lysine  $\beta$ D qui réalisent les interactions de type amino-aromatique sont remplacés par une alanine et une tyrosine respectivement. L'orientation de la boucle BC et la présence d'une sérine dans cette boucle indique que des interactions électrostatiques peuvent être formées avec un groupement phosphate. De plus, la poche de spécificité ne peut être détectée à la surface de ce domaine. Ainsi, la très forte divergence de ce domaine SH2<sub>2</sub> explique pourquoi ce domaine n'a pas été détecté au niveau de la séquence de la protéine et amène la question de son rôle dans la fonction de Spt6.

Dans chacune des poches de fixation des phosphotyrosines, un ion sulfate peut être placé dans la densité électronique indiquant que les deux domaines SH2 ont conservé leur capacité à fixer des ions phosphates. Dans le premier domaine SH2, la fixation de cet ion

sulfate est compatible avec le mode de fixation du groupement phosphate des domaines SH2 standards. Cet ion sulfate est ancré moins profondément dans le domaine puisque la poche de fixation est moins profonde. Dans le second domaine SH2, par rapport aux domaines SH2 standards, seule la boucle BC participe à la fixation de l'ion. Cet ion réalise également des interactions électrostatiques avec un résidu du brin  $\beta$ C et la tyrosine du brin  $\beta$ D qui remplacent la lysine réalisant normalement l'interaction de type amino-aromatique, ce mode de liaison étant tout à fait atypique. Ainsi, bien que très divergent des domaines SH2 dits standard, ce second domaine SH2 semble avoir conservé la capacité à reconnaître des groupements phosphates.

Afin de tester l'implication de chacun de ces domaines dans la liaison à l'ARN polymérase II, des expériences de GST pulldown ont été réalisées sur des extraits de levure. Les constructions des domaines carboxy-terminal des protéines Spt6 d'*A. locustae*, *E. cuniculi*, *S. cerevisiae* et de la protéine humaine ont été clonées dans un vecteur pGEX, permettant l'expression de ces domaines en fusion avec une étiquette de purification GST. De même, des mutations ont été introduites au niveau des résidus impliqués dans la fixation de l'ion sulfate dans le domaine carboxy-terminal de la protéine Spt6 de levure. La présence de l'ARN polymérase II dans la fraction retenue au cours des expériences de pulldown est révélée par Western Blot en utilisant des anticorps dirigés contre le domaine carboxy-terminal de la plus grande sous unité de l'ARN polymérase II, cet anticorps reconnaissant indifféremment les formes hypo- et hyperphosphorylées (anticorps 7C2) (Besse et al., 1995). Ces expériences de pulldown ont révélé que les domaines carboxy-terminaux, contenant les deux domaines SH2 des protéines Spt6 des différents organismes testés permettaient de retenir l'ARN polymérase II. De plus, l'absence de l'un de ces domaines abolit l'interaction alors que la mutation des résidus impliqués dans la fixation de l'ion sulfate ne fait que l'affaiblir. La réalisation de western blots avec des anticorps reconnaissant spécifiquement l'état hyperphosphorylé sur les sérines 2 (anticorps H5) ou sur les sérines 5 (anticorps H14) indique que le domaine carboxy-terminal de la protéine Spt6 de levure reconnaît préférentiellement le CTD de l'ARN polymérase II phosphorylé sur ces sérines 2.

Afin de comprendre le rôle de ces deux domaines SH2 dans un contexte biologique, des études de génétique de levure ont été réalisées par nos collaborateurs de l'équipe de Fred Winston à la Harvard Medical School de Boston. L'expression dans la levure de versions de la protéine dans lesquelles le domaine SH2<sub>2</sub> ou les deux domaines SH2 ont été délétés conduit

à l'apparition de phénotypes thermosensibles, de phénotype Spt- et à l'initiation de la transcription à partir de sites cryptiques. L'apparition de ces phénotypes montre le rôle très important du domaine carboxy-terminal de Spt6 et confirme l'importance fonctionnelle du second domaine SH2 dont la divergence suggérait un rôle fonctionnel éventuellement moins important.

La connaissance de la structure du domaine carboxy-terminal de Spt6 a permis d'établir un modèle de l'interaction entre le tandem SH2 de Spt6 et le CTD phosphorylé de l'ARN polymérase II. Afin de valider ce modèle, des peptides contenant deux répétitions du motif Y-S-P-T-S-P-S phosphorylés sur les différentes sérines ont été synthétisés. Leur capacité à lier le domaine carboxy-terminal de Spt6 a été testée par microcalorimétrie et Thermofluor. Cependant aucune interaction n'a pu être détectée par aucune de ces techniques. L'ensemble de nos résultats montrent donc que la reconnaissance de l'ARN polymérase II par Spt6 fait appel à des mécanismes compliqués qu'il reste à mettre en évidence.

### 3.3. Publication numéro 1

Les données supplémentaires relatives à cette publication sont présentées en Annexe 8.2, page 264

# Noncanonical Tandem SH2 Enables Interaction of Elongation Factor Spt6 with RNA Polymerase II<sup>\*[S]</sup>

Received for publication, May 20, 2010, and in revised form, August 25, 2010. Published, JBC Papers in Press, October 6, 2010, DOI 10.1074/jbc.M110.146696

Marie-Laure Diebold<sup>‡</sup>, Erin Loeliger<sup>§</sup>, Michael Koch<sup>‡1</sup>, Fred Winston<sup>§</sup>, Jean Cavarelli<sup>‡</sup>, and Christophe Romier<sup>‡2</sup>

From the <sup>‡</sup>Département de Biologie et Génomique Structurales, Institut de Génétique et Biologie Moléculaire et Cellulaire, Université de Strasbourg, CNRS, INSERM, 1 rue Laurent Fries, B.P. 10142, 67404 Illkirch Cedex, France and the <sup>§</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115

Src homology 2 (SH2) domains are mostly found in multicellular organisms where they recognize phosphotyrosine-containing signaling proteins. Spt6, a conserved transcription factor and putative histone chaperone, contains a C-terminal SH2 domain conserved from yeast to human. In mammals, this SH2 domain recognizes phosphoserines rather than phosphotyrosines and is essential for the recruitment of Spt6 by elongating RNA polymerase II (RNAPII), enabling Spt6 to participate in the coupling of transcription elongation, chromatin modulation, and mRNA export. We have determined the structure of the entire Spt6 C-terminal region from *Antonospora locustae*, revealing the presence of two highly conserved tandem SH2 domains rather than a single SH2 domain. Although the first SH2 domain has a canonical organization, the second SH2 domain is highly noncanonical and appears to be unique in the SH2 family. However, both SH2 domains have phosphate-binding determinants. Our biochemical and genetic data demonstrate that the complete tandem, but not the individual SH2 domains, are necessary and sufficient for the interaction of Spt6 with RNAPII and are important for Spt6 function *in vivo*. Furthermore, our data suggest that binding of RNAPII to the Spt6 tandem SH2 is more extensive than the mere recognition of a doubly phosphorylated C-terminal domain peptide by the tandem SH2. Taken together, our results show that Spt6 interaction with RNAPII via a novel arrangement of canonical and noncanonical SH2 domains is crucial for Spt6 function *in vivo*.

Transcription elongation by the eukaryotic RNA polymerase II (RNAPII)<sup>3</sup> is tightly linked to chromatin modulation and mRNA processing and export (1–3). The various factors involved in these mechanisms are often recruited directly by RNAPII through the docking platform provided by the C-terminal domain (CTD) of its largest subunit, Rpb1. The CTD is composed of a large number of heptad repeats which are

generally composed of the canonical motif Tyr<sup>1</sup>Ser<sup>2</sup>Pro<sup>3</sup>Thr<sup>4</sup>Ser<sup>5</sup>Pro<sup>6</sup>Ser<sup>7</sup>, although some repeats may diverge from this consensus sequence (4).

During the transcription cycle, the CTD repeats undergo a large number of phosphorylation and dephosphorylation events, notably at their serine residues (Ser<sup>2</sup>, Ser<sup>5</sup>, and Ser<sup>7</sup>). These changes in the CTD phosphorylation pattern alter its binding surface as transcription proceeds and enable the recruitment by RNAPII of different sets of factors at each transcriptional stage (5–8). The ability of the CTD to recruit multiple activities is essential for the coupling of transcription with chromatin modulation and mRNA processing and export. One such activity recruited by the CTD that plays essential roles is the Spt6/Iws1 complex.

The essential protein Spt6 is a putative histone chaperone that interacts with histone H3 and promotes the reassembly of nucleosomes on promoters and on the body of genes in the wake of the RNAPII (9–13). Depletion of Spt6 causes sustained transcription in repressing conditions and production of aberrant transcripts due to initiation from cryptic sites within the body of genes (13, 14). Spt6 also belongs to the family of elongation factors, enhancing the elongation rates of RNAPII both *in vitro* and *in vivo* (15–17). Yet, this latter role appears independent of Spt6 chaperone activity because elongation enhancement occurs also on naked DNA templates (16, 17).

The association of Spt6 with the essential protein Iws1/Spn1 (17–22) provides an additional link between Spt6, chromatin modulation, and mRNA export. Indeed, mammalian Iws1 is required for the recruitment of the lysine methyltransferase HYPB/Setd2 that trimethylates H3K36 across transcribed regions. In addition, Iws1 also interacts directly with the export factor REF1/Aly, with the depletion of Iws1 leading to splicing defects and nuclear retention of bulk poly(A)<sup>+</sup> mRNAs (17, 23).

Several lines of evidence indicate that the Spt6/Iws1 complex travels with RNAPII. In yeast, both Spt6 and Iws1 have been shown by ChIP analyses to be present on the promoter, the body, and even past the polyadenylation signal of several genes (24). In agreement with these results, *Drosophila* Spt6 colocalizes with hyperphosphorylated RNAPII at transcribed genes (25–28). More recently, a direct interaction of mammalian Spt6 with the RNAPII CTD has been characterized which requires the SH2 motif found in the Spt6 C-terminal region (17, 23). A mutation in this motif leads to the loss of the inter-

\* This work was supported, in whole or in part, by National Institutes of Health Grant GM32967. This work was also supported by the CNRS, INSERM, the Université de Strasbourg, and the European Commission SPINE2-Complexes Contract LSHG-CT-2006-031220.

[S] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables 1 and 2 and Figs. 1–5.

<sup>1</sup> Present address: Institut für Biochemie, Universität zu Köln, Otto-Fischer-Str. 12–14, 50674 Köln, Germany.

<sup>2</sup> To whom correspondence should be addressed. Tel.: 33-3-88-65-57-98; Fax: 33-3-88-65-32-79; E-mail: romier@igbmc.fr.

<sup>3</sup> The abbreviations used are: RNAPII, RNA polymerase II; al, *A. locustae*; cg, *C. glabrata*; CTD, C-terminal domain; ec, *E. coli*; hs, *Homo sapiens*; sc, *Saccharomyces cerevisiae*; TAP, tandem affinity purification.

## Noncanonical Tandem SH2 in Spt6 Interacts with RNAPII

action of Spt6 with the polymerase and causes splicing defects (17).

SH2 motifs are generally found in signaling proteins from higher eukaryotes where they are involved in the specific recognition of phosphotyrosine-containing proteins (29). Phosphate recognition is carried out by several residues, including an invariant arginine on strand  $\beta$ B of the SH2 motif as well as various residues of the BC loop connecting the strands  $\beta$ B and  $\beta$ C of the SH2 motif (30). Yet, phosphotyrosine binding is not sufficient for specific target recognition which also requires interaction with flanking residues (31, 32). Specificity can be reinforced by contacts of the SH2 domain with other regions of the target protein (33) or the use of tandem SH2 domains (34, 35).

Surprisingly, the Spt6-RNAPII interaction requires the hyperphosphorylation of the CTD repeats on a serine (Ser<sup>2</sup>) rather than a tyrosine residue, suggesting that the Spt6 SH2 motif represents an ancient SH2 domain recognizing phosphoserines rather than phosphotyrosines (17, 23). The conservation of this SH2 domain throughout evolution highlights its functional importance (36). Spt6 appears to be the only protein in yeast harboring an SH2 domain.

The recent structure of the SH2 domain of Spt6 from *Candida glabrata* has revealed a rather canonical organization for this domain (37). Notably, the conservation of the invariant arginine on the  $\beta$ B strand and the presence of residues in the BC loop that might participate in phosphate recognition are in agreement with phosphoresidue binding by this SH2 domain. Yet, the absence of residues being able to form aminoaromatic interactions with an incoming phosphotyrosine as well as a shallower phosphoresidue-binding pocket suggested a favored recognition of phosphoserines over phosphotyrosines. However, binding assays *in vitro* showed that the yeast and human Spt6 SH2 domains are unable to interact with a single or tandem CTD repeats phosphorylated on Ser<sup>2</sup> (37).

These results suggest that the entire conserved C-terminal region of Spt6 rather than the single SH2 motif is required for binding to the RNAPII CTD. We have solved the structure of the entire C-terminal region of Spt6 from the yeast-related organism *Antonospora locustae* at 2.2 Å resolution. Unexpectedly, the structure reveals that this region is formed by two SH2 domains arranged in tandem. Whereas the first SH2 domain shows a canonical organization, the second SH2 domain displays many atypical features at the sequence and structural levels. Yet both SH2 domains have retained determinants for phosphate binding. By performing GST pulldown assays on yeast extract, we show that the tandem, but not the individual SH2 domains, is able to interact with RNAPII. Furthermore, deletion of either the two SH2 domains or only the second SH2 domain causes severe mutant phenotypes *in vivo*, confirming the importance of the tandem SH2 domain. In addition, mutation of putative phospho-binding residues in both SH2 domains decreases but does not fully abolish interaction with RNAPII, suggesting an extensive interaction between the Spt6 tandem SH2 and RNAPII. In agreement, short phosphorylated CTD fragments are unable to bind to the tandem SH2. Altogether, our data reveal the presence of an un-

expected, atypical tandem SH2 within Spt6, emphasizing the importance of both the canonical and the noncanonical SH2 domains for Spt6 function *in vivo*.

### EXPERIMENTAL PROCEDURES

**Cloning and Purification**—All constructs used in this study were amplified by PCR techniques and cloned using NdeI and BamHI restriction sites into either the pEA-tH expression vector from the pET-MCN series (38), which encodes an N-terminal His<sub>6</sub> tag and a thrombin cleavage site, or a modified pGEX-4T2 vector containing compatible restriction sites. All proteins were overexpressed in *Escherichia coli* BL21[DE3] in 2LB medium in the presence of the pRARE2 vector (Novagen). When the cultures reached an  $A_{600\text{ nm}}$  of 0.4 they were cooled to 25 °C, and gene expression was induced for 16 h at 25 °C with 0.7 mM isopropyl- $\beta$ -D-thiogalactopyranoside (Euromedex). Cells were then harvested and resuspended in a lysis buffer containing 10 mM Tris (pH 8.0) and either 50 or 400 mM NaCl.

For structural studies, the Spt6 C-terminal regions from *Encephalitozoon cuniculi* (ec; residues 724–894) and *A. locustae* (al; residues 783–957) were used. The cells were lysed by sonication in a lysis buffer containing 400 mM NaCl. The soluble fraction was incubated for 1 h with 500  $\mu$ l of Talon affinity resin (Clontech). The beads were then washed and resuspended in lysis buffer, and bovine thrombin was added overnight for cleaving off the histidine tag. The proteins were then injected onto a gel filtration column Hiload 16/60 Superdex 75 (GE Healthcare) equilibrated with lysis buffer supplemented with 2 mM dithiothreitol (Euromedex). The purified proteins were concentrated on the Amicon 10 K system (Millipore) to a concentration of 20 mg/ml.

**Crystallization and Data Collection**—Crystallization trials were performed using the hanging-drop vapor-diffusion method at 24 °C by mixing equal amounts of protein and crystallization buffer. Crystals for the C-terminal region of *A. locustae* Spt6 were obtained in presence of 0.1 M MES (pH 6.0), 1.5 M ammonium sulfate, and 20 mM CoCl<sub>2</sub>. For data collection, crystals were frozen in liquid nitrogen in a cryoprotectant solution containing 0.1 M MES (pH 6.0), 1.5 M ammonium sulfate, 15 mM CoCl<sub>2</sub>, and 25% xylitol. All data collections were performed at 100 K. A native data set was collected on ESRF beamline ID29 at a wavelength of 0.9762 Å. The derivative data sets were collected on ESRF beamline ID14-2 at a wavelength of 0.933 Å.

**Structure Determination and Refinement**—Structure determination was carried out using heavy atom derivatives. Native crystals were soaked overnight in the presence of either 1 mM KAu(CN)<sub>2</sub> or 1 mM (CH<sub>3</sub>)<sub>3</sub>PbAc and then frozen in the same cryoprotectant solution used for native crystals. All data were processed with HKL2000 (39). Phase determination by MIRAS was carried out using Auto-SHARP (40) including solvent modification with SOLOMON (41). Model building was performed using TURBO-FRODO and COOT (42). Refinement was done using REFMAC (43). There is one molecule per asymmetric unit. The final refined model has an  $R_{\text{factor}}$  and an  $R_{\text{free}}$  of 22.5% and 26.6%, with good deviations from ideal geometry (supplemental Table 1). Ramachandran



analysis shows 97.7% of residues in favored regions and 2.3% of residues in allowed regions. The refined model and the structure factor amplitudes have been deposited in the Protein Data Bank under the PDB code 2XP1.

**GST Pulldown Experiments**—Yeast extracts were prepared as described previously (44). The GST-fused proteins were lysed in a lysis buffer containing 10 mM Tris (pH 8.0) and 50 mM NaCl. The soluble cellular extract was incubated for 1 h with 50  $\mu$ l of glutathione-Sepharose 4B resin (Amersham Biosciences). The beads were then washed several times in lysis buffer. The amounts of each construct were assessed by SDS-PAGE and Coomassie staining. To perform the pull-down, equivalent amounts of protein bound to the resin were incubated for 2 h in 30  $\mu$ l of yeast extract. After incubation, beads were washed twice with a buffer containing 50 mM potassium chloride, 20 mM K-HEPES (pH 7.9), 0.2 mM EDTA, and 0.5 mM dithiothreitol and resuspended in 50  $\mu$ l of Laemmli buffer. Retention of the largest subunit of the RNAPII was checked by Western blotting using the 7C2 antibody (45) directed against its CTD. Phospho-specificity was assessed using the H5 (Ser<sup>2</sup>-phospho) and H14 (Ser<sup>5</sup>-phospho) antibodies (Covance).

**Yeast Strain Construction**—All TAP-tagged alleles were created by transforming a TAP-NatMx cassette (amplified from pFA6a-C-TAP4-natMX6 (46)) into diploid strain FY2808 (supplemental Table 2). Nat<sup>R</sup> transformants were isolated, and the TAP-tagged constructs were confirmed by PCR and sequencing. The *spt6- $\Delta$ Tandem::TAP* is deleted for the last 201 codons of *SPT6*, encoding the two SH2 domains. The *spt6- $\Delta$ SH2<sub>2</sub>::TAP* is deleted for the last 101 codons of *SPT6*, encoding the SH2<sub>2</sub> domain. Sporulation and tetrad dissection were performed by standard procedures to analyze haploid strains that contained the deletions (47). For *SPT6::TAP* (full-length *SPT6* fused to TAP), cells behaved identically to wild-type cells (a wild-type growth rate and no *spt6* mutant phenotypes detected). Spot tests were performed on the following media: YPD, YPD at 37 °C, YPD at 16 °C, SC-His, SC-Lys, YPD + 150 mM hydroxyurea, YPD + 13  $\mu$ g/ml phleomycin, and YPD + 15 mM caffeine. YPD, SC-His, and SC-Lys were made as described previously (47).

**Western Blot Analysis**—Protein extracts were prepared from strains FY653, FY2796, FY2797, and FY2798 as described previously (48). The TAP tag was detected using the peroxidase anti-peroxidase antibody (1:5000 dilution; Sigma). Spt6 antiserum was a generous gift from Laura McCollough and Tim Formosa (1:2000 dilution in TBST). Pgk1 was used as a loading control and visualized with anti-Pgk1 antiserum (Invitrogen 459250).

## RESULTS

**Spt6 C-terminal Region Contains Tandem SH2 Domains**—The structural and biochemical characterization of the conserved Spt6 C-terminal region has been performed using Spt6 proteins from the fungi-related intracellular parasites *E. cuniculi* and *A. locustae*. The proteins of these organisms are smaller than their orthologs from other organisms, facilitating biochemical and structural analyses, although they retain the essential organization of their orthologs (49, 50). The *A. locus-*

*tae* construct could be expressed in *E. coli*, purified to homogeneity and yielding well diffracting crystals. Structure determination was carried out by soaking native crystals with heavy atom derivatives, and the phase problem was solved by MIRAS. The initial model built in the experimental map was refined against a native dataset at 2.2 Å resolution to an  $R_{\text{factor}}$  of 22.5% and an  $R_{\text{free}}$  of 26.6%, with good deviations from ideal geometry (supplemental Table 1).

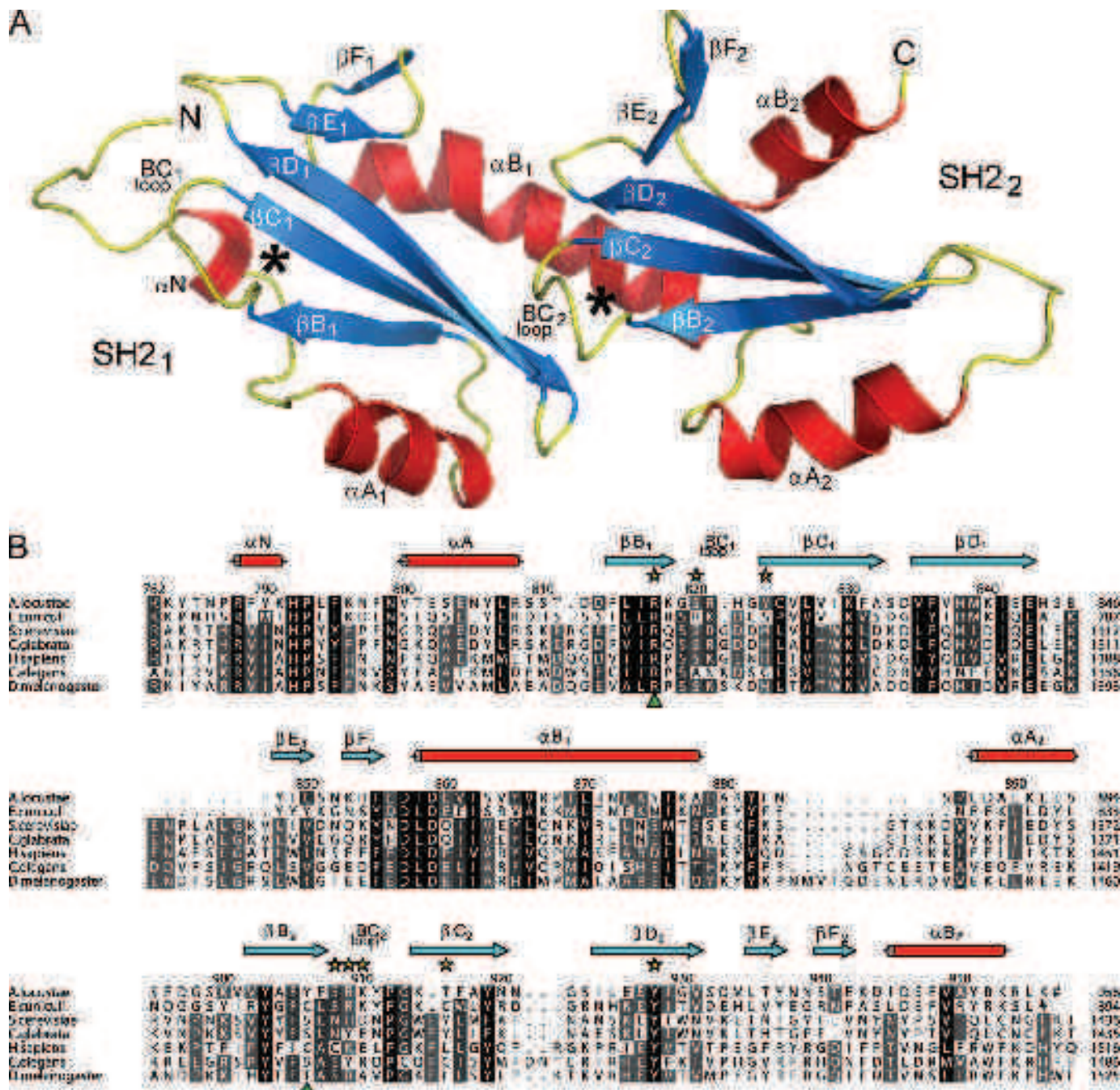
The structure reveals that the C-terminal region of Spt6 does not contain a single SH2 domain but is actually formed by two SH2 domains arranged in tandem, forming a single structural domain (Fig. 1). The presence of an SH2 motif in the second half of the Spt6 C-terminal region was unambiguously assessed by a structural homology search with DALI (51) and is clearly indicated by its secondary structure organization. The two SH2 domains (SH2<sub>1</sub> and SH2<sub>2</sub>) are composed of a central three-stranded antiparallel  $\beta$ -sheet (strands  $\beta$ B– $\beta$ D) sandwiched between N- and C-terminal  $\alpha$ -helices (helices  $\alpha$ A and  $\alpha$ B). The central  $\beta$ -sheets are complemented by a smaller  $\beta$ -hairpin (strands  $\beta$ E– $\beta$ F), reminiscent of the SRC family of SH2 motifs (30).

A specific feature of this tandem SH2 domain is the  $\alpha$ B<sub>1</sub> helix of SH2<sub>1</sub> which is much longer than in standard SH2 domains. Although the N terminus of this helix forms canonical interactions with the SH2<sub>1</sub>, its C-terminal extremity also makes extensive, mostly hydrophobic, interactions with SH2<sub>2</sub> (Fig. 1A). A kink in the second helical turn of this helix, due to an additional residue (Tyr<sup>865</sup>), further brings the two SH2 domains in close vicinity, enabling the interaction between the strands  $\beta$ C<sub>1</sub> and  $\beta$ D<sub>1</sub> of the SH2<sub>1</sub> and the loop BC<sub>2</sub> of the SH2<sub>2</sub>, thus expanding the hydrophobic core made between  $\alpha$ B<sub>1</sub> and the SH2<sub>2</sub>. Therefore, the long  $\alpha$ B<sub>1</sub> helix is essential in providing the scaffold of the tandem SH2. Consequently, both SH2 domains are arranged almost head-to-tail with the two putative phosphoresidue-binding pockets exposed on the same face of the molecule (Fig. 1A).

**SH2<sub>1</sub> Displays Mostly Canonical Features**—Structural analysis of the SH2<sub>1</sub> domain shows that the canonical nature of this domain extends beyond its overall fold. First, an arginine is present on strand  $\beta$ B<sub>1</sub> (Arg<sup>817</sup>) at the bottom of the phosphoresidue-binding pocket (Fig. 2A). An arginine at this position is normally involved in phosphate recognition and represents the only invariant signature of SH2 motifs (30). Second, Arg<sup>817</sup> forms a hydrogen bond network with His<sup>838</sup> on strand  $\beta$ D<sub>1</sub> and Glu<sup>804</sup> on helix  $\alpha$ A<sub>1</sub>, a network often observed in SH2 domains (Fig. 2A). Third, the BC<sub>1</sub> loop contains residues that are compatible with phosphate recognition (e.g. Ser<sup>820</sup>). The same applies for Lys<sup>840</sup>, on the  $\beta$ D<sub>1</sub> strand, which would be perfectly positioned for forming amino-aromatic interactions but which has not been conserved during evolution. On the other hand, as suggested upon the resolution of the *C. glabrata* (cg) SH2<sub>1</sub> structure, the absence of a residue on the helix  $\alpha$ A<sub>1</sub> that may form amino-aromatic interactions with a phosphotyrosine, as well as a shallower phosphoresidue-binding pocket, could explain a favored binding of phosphoserine over phosphotyrosine (37).

Many SH2 domains recognize specifically incoming phosphoproteins by a two-pronged mechanism, where another

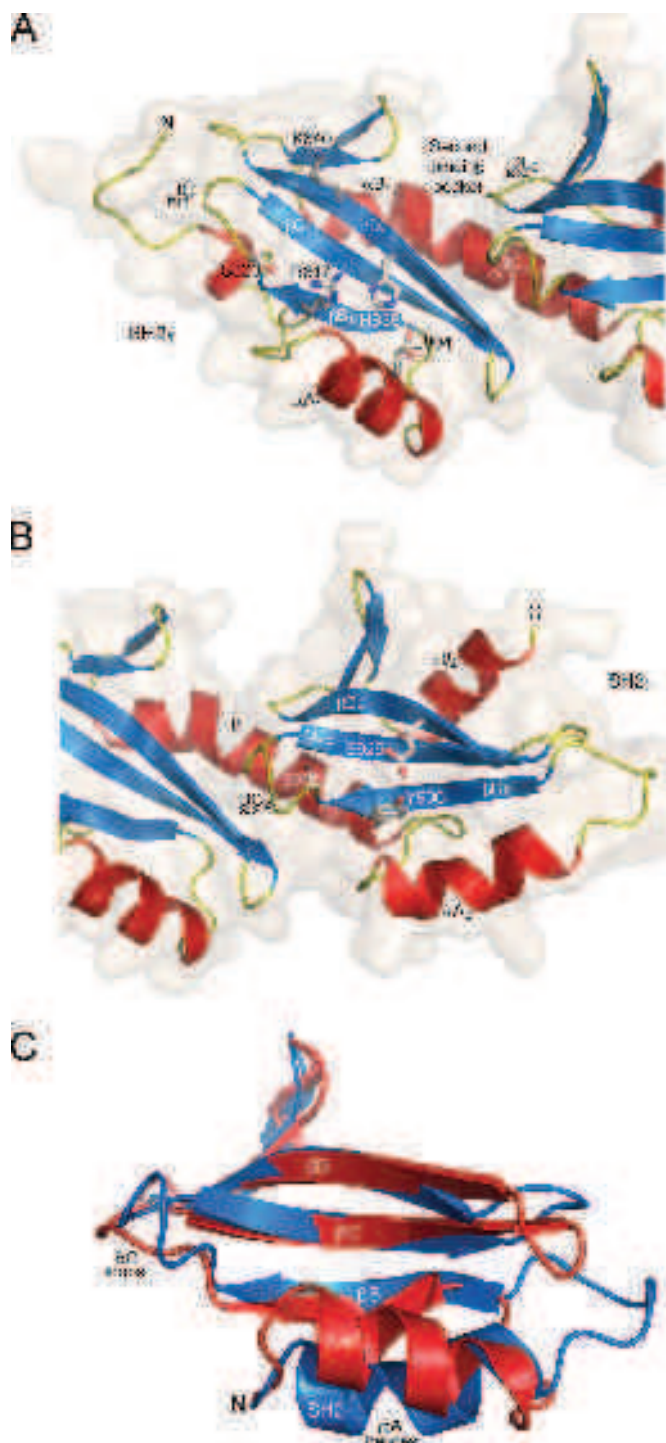
## Noncanonical Tandem SH2 in Spt6 Interacts with RNAPII



**FIGURE 1. Structure and sequence alignment of the conserved C-terminal region of Spt6.** *A*, ribbon representation of the *A. locustae* Spt6 tandem SH2 structure.  $\alpha$ -Helices are colored in red,  $\beta$ -strands in blue, and loops in yellow. The putative phosphoresidue-binding pockets of the two SH2 domains are marked with an asterisk (\*). Secondary structure elements throughout the figures are labeled according to the common SH2 nomenclature (30). All structural figures were created with PyMOL (Delano Scientific). *B*, sequence alignment of the conserved C-terminal region of Spt6. Conservation patterns have been indicated by darker shading (55). Secondary structure elements observed in the *A. locustae* structure are displayed above the sequences as red cylinders ( $\alpha$ -helices) and blue arrows ( $\beta$ -strands). Numbering above the sequences corresponds to *A. locustae*, whereas the numbering at the end of each row relates to the different organisms. Residues shown to coordinate sulfate ions in the structure are labeled with yellow stars. Residues found at the position of the invariant arginine on strand  $\beta$ B of canonical SH2 domains are marked with a green triangle underneath the sequences. Ends of sequences are marked with a number sign (#).

residue, often located in position +3 to the phosphotyrosine, is bound in a second pocket at the surface of the SH2 domain (30). The structure of the cgSH2<sub>1</sub> suggests the presence of such a second pocket which is partially contributed to by residues from the  $\alpha$ B<sub>1</sub> helix (37). However, in the context of the tandem SH2, the residues of the  $\alpha$ B<sub>1</sub> helix that participate in the pocket are different. This is notably due to the kinked

conformation of this helix which is not observed in the cgSH2<sub>1</sub> structure, most likely because of the absence of constraints imposed by the SH2<sub>2</sub> (supplemental Fig. 1). Importantly, our structure reveals that this second binding pocket is complemented by residues of the BC<sub>2</sub> and DE<sub>2</sub> loops from the second SH2 domain. Consequently, this pocket appears more like a groove at the interface between



**FIGURE 2. Structural details of the SH2<sub>1</sub> and SH2<sub>2</sub> domains.** *A*, ribbon representation of the SH2<sub>1</sub> domain. The arginine (R817) that is invariant in canonical SH2 domains is shown as sticks. The two residues (H838 and E804) forming a canonical hydrogen bond network with Arg<sup>817</sup> are also shown as well as other residues (S820 and K840) that might be involved in recognition of an incoming phosphoresidue. The surface of the SH2<sub>1</sub> domain is shown, highlighting the groove formed between the two SH2 domains that define the second binding pocket generally observed in canonical SH2 domains. *B*, ribbon representation of the SH2<sub>2</sub> domain. The tyrosine (Y906) that replaces the invariant arginine of canonical SH2 domains is shown as sticks as well as the glutamate (E926) forming a hydrogen bond with the hydroxyl of this tyrosine. The serine residue (S908) from the BC<sub>2</sub> loop that might be involved in recognition of an incoming phosphoresidue is also displayed. *C*, superposition of the SH2<sub>1</sub> (red) and SH2<sub>2</sub> (blue) domains, highlighting the difference of conformation between the two  $\alpha$ A helices. For clarity, part of the N termini and the  $\alpha$ B helices of the two domains has been removed.

the two SH2 domains where a peptide could make extensive interactions (Fig. 2A).

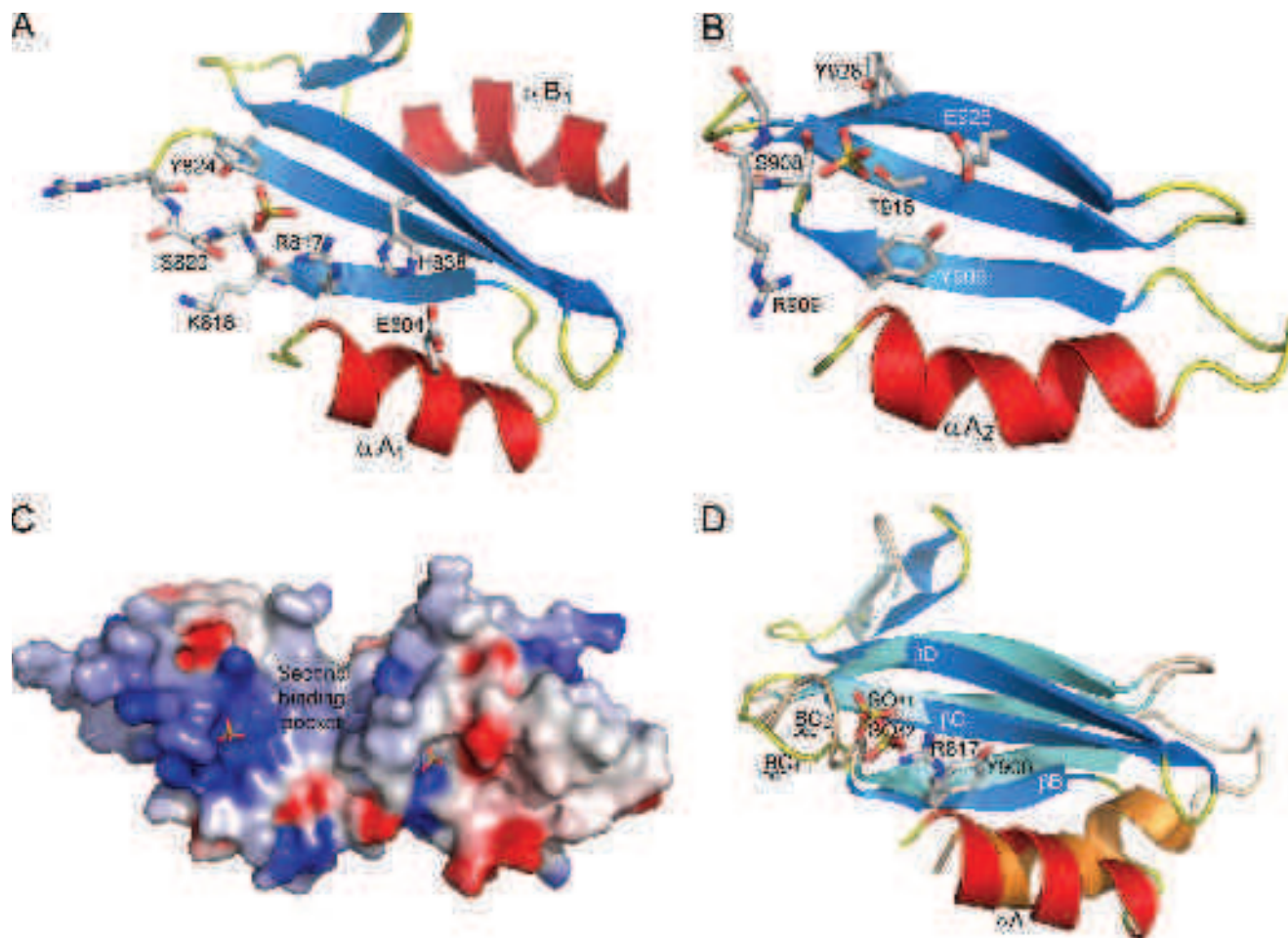
**SH2<sub>2</sub> Is Highly Noncanonical**—In strong contrast to SH2<sub>1</sub>, the structural analysis of the second SH2 domain reveals atypical features despite conserved secondary structure elements. First, the phosphate-binding arginine found on strand  $\beta$ B in other SH2 domains is replaced here by a tyrosine (Tyr<sup>906</sup>) (Fig. 2B). In our structure, the side chain of this tyrosine is not oriented toward the putative phosphate-binding pocket, but toward the carboxylate of Glu<sup>926</sup> of strand  $\beta$ D<sub>2</sub>, which forms a weak hydrogen bond with the hydroxyl of Tyr<sup>906</sup>. These two residues have not been conserved during evolution. In particular, Tyr<sup>906</sup> is generally replaced by smaller residues such as cysteine, threonine, or serine (Fig. 1B). This certainly explains why this second SH2 domain was not identified by sequence analyses. Second, residues on helix  $\alpha$ A<sub>2</sub> and strand  $\beta$ D<sub>2</sub> that could form amino-aromatic interactions with a phosphotyrosine are also absent. Third, analysis of the SH2<sub>2</sub> surface does not give clear evidence for a second binding pocket, which could provide specificity in binding according to a two-pronged mechanism.

Another major difference observed in SH2<sub>2</sub> concerns the orientation of the helix  $\alpha$ A<sub>2</sub> which runs parallel to the strand  $\beta$ B<sub>2</sub> rather than crossing over this strand as seen in canonical SH2 domains, including SH2<sub>1</sub> (Fig. 2C). In fact, a DALI search failed to find other SH2 domains with structural homology to SH2<sub>2</sub> with an  $\alpha$ A helix adopting this conformation. A major effect of this conformational change is the opening of the putative phosphoresidue-binding pocket of the SH2<sub>2</sub> on the side where the  $\alpha$ A<sub>2</sub> helix lies. This is reinforced by the absence of a residue on this helix that could form amino-aromatic interactions with a bound phosphotyrosine.

**Phosphate Recognition by the Tandem SH2 Domains**—The noncanonical nature of the SH2<sub>2</sub> domain clearly raised the question of the ability of this domain to recognize phosphopeptides. However, analysis of the electron density within the putative phosphate-binding pockets of both SH2 domains revealed the presence of additional density which was too large to accommodate single water molecules. Because the crystallization conditions contained sulfate ions that are known to bind readily into phosphate recognition pockets, we modeled two sulfate ions that fitted and refined perfectly into the density (Fig. 3, A and B, and supplemental Fig. 2). The specificity of this binding was confirmed by the fact that no other density for sulfate ion was observed in the rest of the structure and that the electrostatic potential at the surface of the tandem SH2 domains showed that both phosphate-binding pockets are positively charged (Fig. 3C).

Detailed analysis of the recognition of the sulfate ion by the SH2<sub>1</sub> shows that Arg<sup>817</sup> makes a bidentate interaction with the sulfate (Fig. 3A). Further recognition of the sulfate involves interactions with the main chain amide and the hydroxyl of Ser<sup>820</sup> on the BC<sub>1</sub> loop as well as the hydroxyl of Tyr<sup>824</sup> from strand  $\beta$ C<sub>1</sub>. All of these interactions are commonly used by canonical SH2 domains for phosphate recognition, with the exception of the one formed by Tyr<sup>824</sup>, a residue that has not been conserved during evolution. However, superposition of the aSH2<sub>1</sub> and cgSH2<sub>1</sub> domains shows that

## Noncanonical Tandem SH2 in Spt6 Interacts with RNAPII



**FIGURE 3. Close-up views on sulfate recognition by the SH<sub>21</sub> and SH<sub>22</sub> domains.** Views of the phosphate-binding pockets of the SH<sub>21</sub> (A) and the SH<sub>22</sub> (B) domains. The domains are represented as *ribbons* with the exception of the central part of the BC loops, which are represented as *sticks*. The bound sulfate ions and the residues coordinating these ions are shown as *sticks* and labeled. C, electrostatic potential at the surface of the Spt6 tandem SH2 from *A. locustae*. The electrostatic potentials  $-8$  and  $+8 k_B T$  ( $k_B$ , Boltzmann constant;  $T$ , temperature) are colored red and blue, respectively. The bound sulfate ions are shown as *sticks*. A direct path connecting the two phosphoresidue-binding pockets of the SH<sub>21</sub> and SH<sub>22</sub> domains can be observed. The position of the putative second binding pocket of the SH<sub>22</sub> is labeled. D, superposition of the phosphate-binding pockets of the SH<sub>21</sub> and SH<sub>22</sub> domains, respectively as well as the side chains of Arg<sup>817</sup> and Tyr<sup>906</sup> are shown as *sticks*. Deeper positioning of the SO<sub>42</sub> sulfate ion within the SH<sub>22</sub> pocket is clearly seen. This is partly due to the absence of an arginine on strand  $\beta$ <sub>2</sub> that would otherwise cause steric hindrance. For clarity, in B and C, the side chain of Lys<sup>910</sup>, which extends over the phosphate-binding pocket of the SH<sub>22</sub>, has been removed.

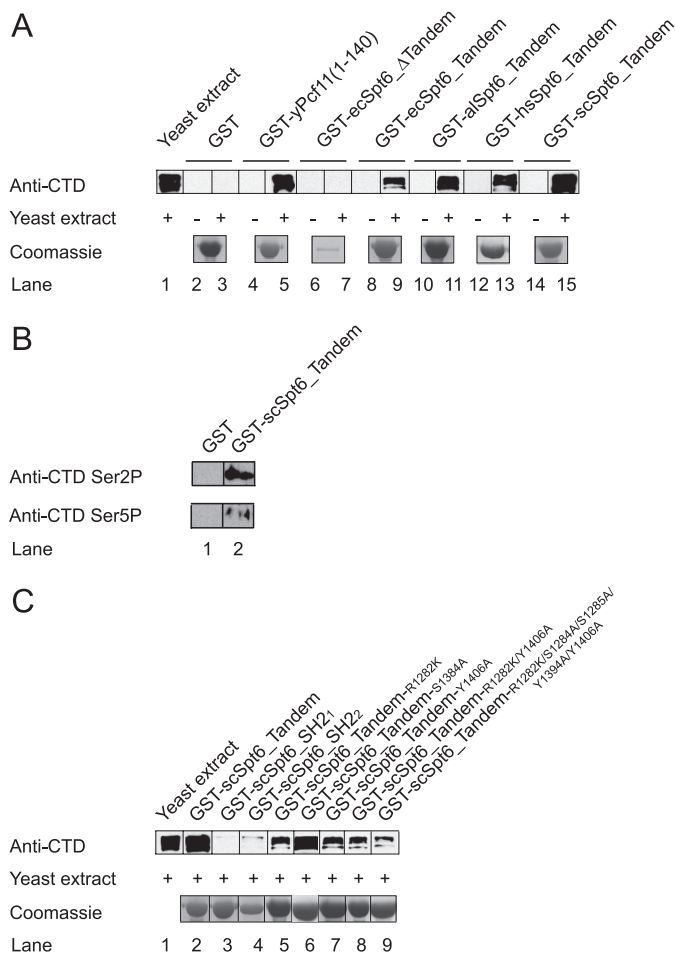
the hydroxyl of Ser<sup>1283</sup> in the BC<sub>1</sub> loop of the cgSH<sub>21</sub> can perfectly play the role of the alTyr<sup>824</sup> hydroxyl (supplemental Fig. 3). These results confirm experimentally that the major determinants for canonical phosphate recognition are fully conserved in the Spt6 SH<sub>21</sub> domain.

In contrast, sulfate recognition by the SH<sub>22</sub> uses both canonical and noncanonical features. First, sulfate recognition involves interactions with the hydroxyl of Ser<sup>908</sup> and the main chain amides of Arg<sup>909</sup> and Lys<sup>910</sup> from the BC<sub>2</sub> loop (Fig. 3B). However, in addition to these canonical features, the hydroxyl of Tyr<sup>906</sup>, which replaces the otherwise invariant arginine on strand  $\beta$ <sub>2</sub>, does not interact with the sulfate ion. Compensating for this loss of interaction, two atypical interactions are formed with the hydroxyls of Thr<sup>916</sup> on the  $\beta$ C<sub>2</sub> strand and Tyr<sup>928</sup> on the  $\beta$ D<sub>2</sub> strand. Therefore, despite its noncanonical features, the SH<sub>22</sub> appears to have kept determinants for phosphate recognition, these determinants reinforcing, however, the atypical character of this domain.

Interestingly, superposition of the SH<sub>21</sub> and SH<sub>22</sub> domains shows that the sulfate ion bound to the SH<sub>22</sub> is located deeper

within the phosphate-binding pocket, about 2 Å away from the position of the sulfate in the SH<sub>21</sub> (Fig. 3D). This shift is due to a different conformation of the BC<sub>2</sub> loop and to the replacement of the invariant arginine on strand  $\beta$ <sub>2</sub> by a tyrosine whose smaller side chain and different orientation enable deeper binding within the pocket. As a consequence of this deeper binding and of the different conformation of the  $\alpha$ A<sub>2</sub> helix, the sulfate ion is more accessible on the side than on the top of the phosphoresidue-binding pocket, contrasting strongly with the SH<sub>21</sub> domain. Strikingly, this side opening creates a direct path between the two phosphoresidue-binding pockets of the SH<sub>21</sub> and SH<sub>22</sub> domains (Fig. 3C).

**Both SH<sub>21</sub> and SH<sub>22</sub> Participate in the interaction of Spt6 with RNAPII**—To examine the role of both SH2 domains in the interaction of Spt6 with RNAPII, we performed GST pull-down experiments on yeast extract using GST-fused Spt6 constructs. Retention of RNAPII by the different constructs was monitored with antibodies directed against the polymerase CTD. Nonspecific binding was excluded because GST alone bound to glutathione-Sepharose was unable to retain



**FIGURE 4. GST pulldowns on yeast extract.** GST pulldowns were carried out using various GST-fused Spt6 constructs in the presence or absence of yeast extract. Binding of RNAPII was monitored by anti-CTD antibodies. The yeast extract control loaded (lane 1) corresponds to 0.1  $\mu$ l of extract, whereas the quantity loaded after pulldown and washes corresponds to the equivalent of 2.5  $\mu$ l of extract. Coomassie-stained samples of the different proteins used are shown. **A**, GST pulldowns made with GST alone (lanes 2 and 3) or fused to the CTD-interacting domain of the yeast termination factor Pcf11 (lanes 4 and 5) or to various constructs of Spt6 from *E. cucuruli* (lanes 6–9), *A. locustae* (lanes 10 and 11), *H. sapiens* (lanes 12 and 13), and *S. cerevisiae* (lanes 14 and 15). Both Pcf11 and the tandem SH2 are able to specifically retain the polymerase. **B**, analysis of phosphospecificity using anti-Ser<sup>2</sup>(P) and anti-Ser<sup>5</sup>(P) antibodies. A slight increase in binding to the Ser<sup>2</sup>-phosphorylated form was observed. **C**, GST pulldowns performed on the yeast Spt6 (i) wild-type tandem SH2 (lane 2), (ii) SH2<sub>1</sub> domain (lane 3), (iii) SH2<sub>2</sub> domain (lane 4), and mutants of phosphate-binding residues from both domains in the context of the entire tandem SH2 (lanes 5–9). The results show that the entire tandem SH2 is required for RNAPII binding. Mutation of phosphate-binding determinants shows a decrease in binding, suggesting that interaction with RNAPII is more extensive. For the *E. cucuruli* (*ec*) constructs, the N-terminal and the tandem SH2 domains span residues 1–725 and 724–894, respectively. The *A. locustae* (*al*) tandem SH2 construct spans residues 783–957. The human (*hs*) construct spans residues 1316–1727. For the *S. cerevisiae* (*sc*) constructs, the tandem SH2, SH2<sub>1</sub>, and SH2<sub>2</sub> domains span residues 1231–1451, 1231–1350, and 1339–1451, respectively.

the RNAPII present in yeast extract (Fig. 4A, lanes 2 and 3). We then used the CTD-interacting domain of yeast Pcf11 as a positive control and showed that this protein is able to retain the RNAPII (Fig. 4A, lanes 4 and 5). We next investigated the capacity of Spt6 to retain RNAPII. A construct of *E. cucuruli* Spt6 lacking only the tandem SH2 showed no retention (Fig. 4A, lanes 6 and 7). In contrast, the tandem SH2 domains from *E. cucuruli*, *A. locustae*, *Homo sapiens*, and *Saccharomyces*

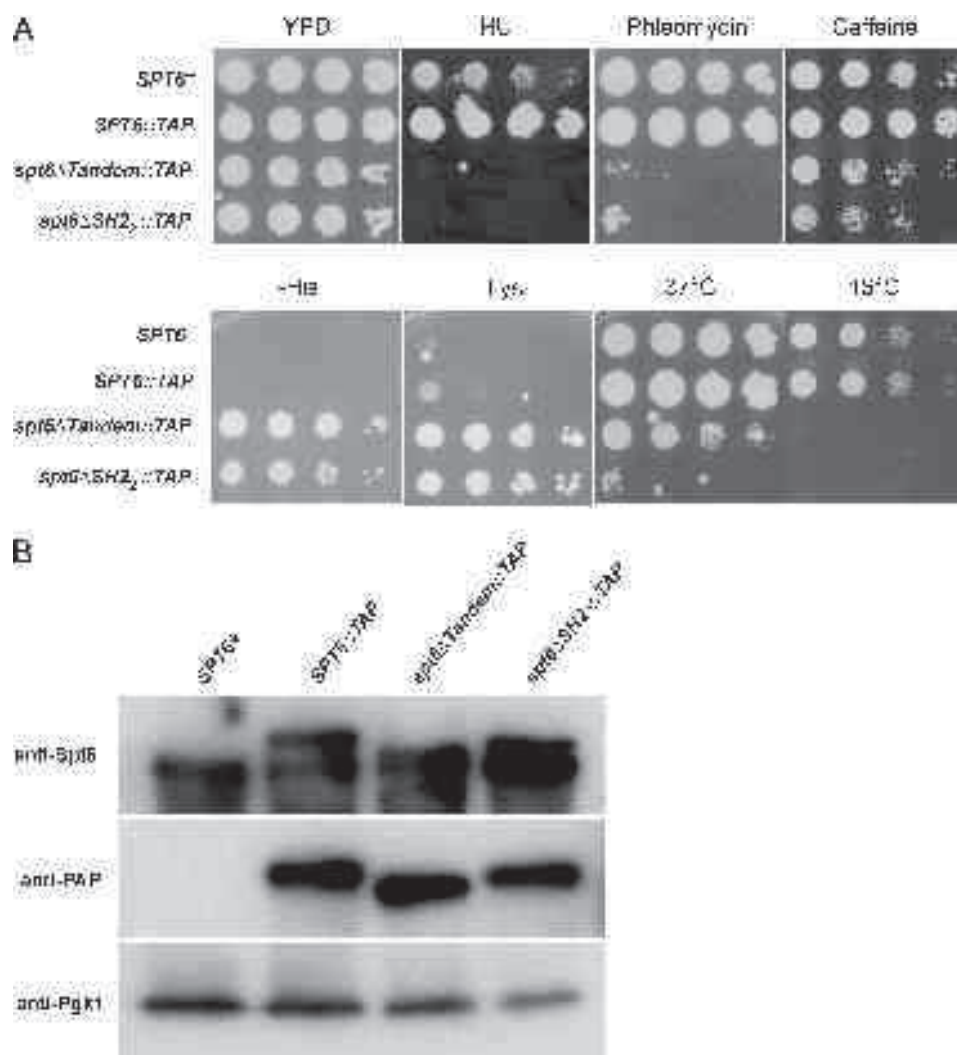
*cerevisiae* were all able to retain the RNAPII (Fig. 4A, lanes 8–15). Antibodies specific for CTD phosphorylation on either Ser<sup>2</sup> or Ser<sup>5</sup> were used for analyzing phosphospecificity, showing an apparent increase in binding for the form phosphorylated on Ser<sup>2</sup> (Fig. 4B).

We next investigated the importance of both SH2 domains for binding to RNAPII. A GST pulldown assay using a yeast construct spanning only the SH2<sub>1</sub> domain was unable to retain significant amounts of RNAPII (Fig. 4C, lane 3). Similarly, yeast SH2<sub>2</sub> alone showed almost no ability to retain RNAPII (Fig. 4C, lane 4). We next asked whether the two phosphate-binding pockets of the tandem SH2 domains are important for recognition. We first changed the invariant arginine scArg<sup>1282</sup> in  $\beta$ B<sub>1</sub> of the yeast SH2<sub>1</sub> domain to lysine. This mutant, although it only partially removes the phosphate-binding determinants of SH2<sub>1</sub>, was impaired about 5-fold in retaining RNAPII (Fig. 4C, lane 5). In contrast, in the case of the yeast SH2<sub>2</sub>, a change of the invariant arginine of the  $\beta$ B<sub>2</sub> strand to alanine (scSer<sup>1384</sup>, equivalent to alTyr<sup>906</sup>), had no effect on RNAPII binding, in agreement with our structural results showing that this residue is most likely not involved in phosphate recognition (Fig. 4C, lane 6).

We therefore altered another putative phospho-binding residue, scTyr<sup>1406</sup>, from the  $\beta$ D<sub>2</sub> strand of the SH2<sub>2</sub> domain to alanine. In this case, the mutant showed the same loss of interaction with RNAPII of that observed with the scR1282K mutant (Fig. 4C, lane 7). Interestingly, a double mutant scR1282K/Y1406A showed no additive effect on RNAPII binding (Fig. 4C, lane 8), suggesting that either the remaining phospho-binding residues are sufficient to enable weak binding or that contacts occurring outside the phospho-binding pockets may stabilize this weak interaction. To verify this hypothesis, we constructed an additional mutant that removes all putative phospho-binding side chains in the tandem SH2 domains (scR1282K/S1284A/S1285A/Y1394A/Y1406A). This mutant also showed a reduced level of RNAPII binding (Fig. 4C, lanes 9). Taken together, these results suggest that the interaction between Spt6 tandem SH2 domains and RNAPII is more extensive than just phosphoresidue recognition.

**Tandem SH2 Domains and SH2<sub>2</sub> Domain Alone Are Critical for Spt6 Function In Vivo**—Our structural and biochemical results show the importance of the tandem SH2 domains for interaction with RNAPII. The SH2<sub>2</sub> domain appears as essential as the SH2<sub>1</sub> for this activity. To investigate the requirement for these domains *in vivo*, we have created two yeast mutants lacking either both of the tandem SH2 domains or only the SH2<sub>2</sub> domain and analyzed them for mutant phenotypes. Our results show that both deletion mutants are viable, although they grow much more slowly than wild type and have several mutant phenotypes, including impaired growth at high (37 °C) and low (16 °C) temperatures, suppression of the insertion mutation *his4-912 $\delta$*  (Spt phenotype (9)), cryptic initiation at a *FLO8-HIS3* reporter (48), and sensitivity to hydroxyurea, phleomycin, and caffeine (Fig. 5A). Our results with the mutant lacking both SH2 domains agree in general with previous studies of this deletion (37). Our finding that a mutant lacking only the SH2<sub>2</sub> domain has severe mutant phe-

## Noncanonical Tandem SH2 in Spt6 Interacts with RNAPII



**FIGURE 5. *In vivo* requirement of the tandem SH2 in yeast Spt6.** *A*, strains were grown to saturation in YPD, serially diluted 10-fold, spotted on the indicated media, and incubated for 6 days. The strains used were FY2809, FY28010, FY2811, and FY2812 (supplemental Table 2). Growth on -His indicates cryptic transcription initiation of the *FLO8-HIS3* reporter (48), and growth on -Lys indicates suppression of the *lys2-128δ* insertion mutation (Spt<sup>-</sup> phenotype) (9). *B*, Western blot analysis of Spt6 levels. Protein extracts were prepared from strains FY653, FY2796, FY2797, and FY2798 (supplemental Table 2) as described previously (48). The *top panel* shows a blot probed with  $\alpha$ -Spt6 antibody (courtesy of Laura McCollough and Tim Formosa). In this *panel*, in *lanes 2, 3, and 4*, two bands are visible because the strains are diploid. The *upper band* is the TAP-tagged Spt6, and the *lower band* is untagged Spt6. The *middle panel* is probed with  $\alpha$ -PAP antibody (Sigma P1291), and the *bottom panel* shows the Pgk1 loading control.

notypes indicates that this noncanonical SH2 domain plays a critical role in Spt6 function.

To determine whether any of the mutant phenotypes might be due to altered levels of Spt6 protein, we performed Western analysis on the diploid strains containing each TAP-tagged allele. Our results (Fig. 5B) show that the deletions do not have a significant effect on Spt6 protein levels. Thus, loss of either both SH2 domains or just the SH2<sub>2</sub> domain severely impairs Spt6 function, demonstrating the importance of these domains *in vivo*.

**Two Consecutive Phosphorylated CTD Repeats Are Not Sufficient for Binding to the Tandem SH2**—The C-terminal region of mammalian Spt6 has been shown to bind directly to the RNAPII CTD upon hyperphosphorylation of the CTD repeats on Ser<sup>2</sup> (17, 23). We reasoned that the recognition of a Ser<sup>2</sup> hyperphosphorylation state most likely implies that the tandem SH2 domains bind consecutive phosphorylated repeats because recognition of remote phosphorylated CTD

repeats could select for low Ser<sup>2</sup> phosphorylation. Modeling shows that two phosphorylated repeats could indeed bind to the tandem SH2 domains (supplemental Fig. 4). In this model, binding would require an extended conformation of the peptide linking both phosphoserines and would also be dependent on the noncanonical features of the SH2<sub>2</sub>, notably the opening between the two phospho-binding pockets and the deeper positioning of the phosphate within the SH2<sub>2</sub> binding pocket, as observed for the sulfate ion in our structure.

We have experimentally investigated this binding mode of two consecutive phosphorylated CTD repeats to the tandem SH2 domains by using microcalorimetry (ITC) and Thermofluor technologies. Peptides bearing no phosphoserines or two consecutive phosphorylations on Ser<sup>2</sup>, Ser<sup>5</sup>, or Ser<sup>7</sup> were used, with a few flanking residues on both sides of the phosphoresidues to enable interaction with the putative second binding pockets of the SH2 domains as in a two-pronged mechanism. Despite the fact that both techniques used rely

on different technologies, none of the peptides was shown to bind to the tandem SH2 domains. These results, which show that two CTD repeats cannot recapitulate RNAPII binding, are in agreement with our mutational data which indicate that the recognition of the phosphorylated RNAPII by the Spt6 C-terminal region requires extensive interaction of these macromolecules.

## DISCUSSION

SH2 domains are found almost exclusively in metazoans where they serve as phosphotyrosine-binding motifs and regulate the activity of protein tyrosine kinases that participate in signaling pathways (52, 53). The discovery that Spt6 contains the only SH2 domain found in yeast (36) and the fact that Spt6 most likely evolved from the bacterial transcription factor Tex (25, 54), suggested that this SH2 domain represents an ancestor of the SH2 domains of multicellular organisms. This view of an ancestral SH2 motif has been reinforced by the finding that the mammalian Spt6 SH2 domain interacts directly with RNAPII upon phosphorylation of serine rather than tyrosine residues in its CTD repeats (17, 23).

We show here that this view is oversimplified and that the Spt6 C-terminal region is actually composed of two SH2 domains arranged in tandem. Superposition of these domains suggests that the tandem arrangement arose by duplication (supplemental Fig. 5). Yet, both domains are highly divergent in the details of their organization. Whereas the SH2<sub>1</sub> domain has characteristics that are often found in SH2 domains from multicellular organisms, the SH2<sub>2</sub> domain harbors features that appear unique in the family of SH2 domains. The conservation of this tandem SH2 throughout evolution, from yeast to man, highlights its functional importance. Our *in vivo* data, which show that the tandem SH2 domains are required for Spt6 function, clearly confirm this view.

The *in vivo* importance of the tandem SH2 domains is in agreement with the functional role of Spt6 during elongation. In particular, the distribution of the Spt6 and Iws1 proteins on the promoters and bodies of genes (24) suggests that these proteins travel together with RNAPII. Previous results on mammalian Spt6 (17, 23) and our current results on yeast Spt6 show that the conserved C-terminal tandem SH2 domain of Spt6 has the ability to bind RNAPII. The known interaction of Iws1 with Spt6 (17–22) provides a means for Iws1 to travel with RNAPII. Importantly, mutation of the mammalian SH2<sub>1</sub> domain causes the same defects as those observed upon mammalian Iws1 depletion (17). Therefore the interaction of the Spt6 tandem SH2 domains with RNAPII is not only critical for Spt6 but most likely also for Iws1 function.

Importantly, our biochemical analyses show that the tandem, but not the individual SH2 domains, interact with RNAPII. As such, the tandem should be seen as a single functional unit. In agreement, it has been shown that binding of doubly phosphorylated peptides to tandem SH2 domains, as in the case of the SH-PTP2 phosphatase and the ZAP-70 protein tyrosine kinase, causes a significant increase in binding affinity (34, 35). Our structural data reveals that both Spt6 SH2 domains have retained determinants for phosphate binding. Mutation of the scSer<sup>1384</sup> residue (equivalent of alTyr<sup>906</sup>),

which is located within the phospho-binding pockets but that is not predicted to bind phosphates, does not affect binding. In contrast, mutation of putative phosphate-binding residues decreases RNAPII binding. Yet, these mutations do not abolish binding as observed with the individual SH2 domains. This suggests that additional interactions are responsible for the observed weaker binding. It cannot be excluded that interactions within the phospho-binding pockets between the Spt6 main chain atoms and the phosphates, such as those observed in our structure, could contribute to this weaker interaction. However, it appears likely that interactions occurring outside the pockets also contribute to interaction and to the weaker binding observed.

The absence of interaction between short phosphorylated tandem CTD repeats and the Spt6 tandem SH2 domains also supports this view. Specifically, these latter experiments were performed with peptides long enough to enable their binding to the second binding pocket of the SH2 domains, assuming a canonical way of binding. Actually, there is very little evidence that the specific recognition of phosphoproteins by SH2 domains is only based on the binding of a phosphoresidue and its flanking residues. So far, the vast majority of binding studies have been carried out on small phosphopeptides rather than full phosphoproteins. However, this view has recently changed with the determination of the structure of the FGFR1/phospholipase C $\gamma$  complex, where a second interaction between a SH2 domain and its target phosphoprotein has been observed and shown to be essential for high affinity binding (33). Our results, together with previous observations made with the mouse Spt6 protein (23), argue in favor of such a hypothesis.

Altogether, the results presented here reveal that the recognition of the phosphorylated CTD of the elongating RNAPII by the C-terminal region of Spt6 does not make use of a simple mechanism based on a single SH2 domain. Rather, this recognition requires a highly intricate mechanism involving a noncanonical tandem SH2 domain. The conservation of this tandem SH2 domain throughout eukaryotic evolution and our *in vivo* analyses demonstrate that this domain is essential for the function of the histone chaperone and elongation factor Spt6 which performs its role in the wake of the RNAPII.

*Acknowledgments*—We thank members of the ESRF-EMBL joint structural biology groups for the use of the ESRF beamline facilities and help during data collection. We also thank Eric Ennifar for help during ITC measurements, Marc Vigneron for the kind gift of the CTD peptides, Laura McCollough and Tim Formosa for Spt6 antisera, Céline Faux and Bertrand Séraphin for help during the preparation of the yeast extract, and Irwin Davidson for fruitful discussions and critical reading of the manuscript.

## REFERENCES

- Li, B., Carey, M., and Workman, J. L. (2007) *Cell* **128**, 707–719
- Perales, R., and Bentley, D. (2009) *Mol. Cell* **36**, 178–191
- Saunders, A., Core, L. J., and Lis, J. T. (2006) *Nat. Rev. Mol. Cell Biol.* **7**, 557–567
- Chapman, R. D., Heidemann, M., Hintermair, C., and Eick, D. (2008) *Trends Genet.* **24**, 289–296

## Noncanonical Tandem SH2 in Spt6 Interacts with RNAPII

5. Buratowski, S. (2009) *Mol. Cell* **36**, 541–546
6. Eglhoff, S., and Murphy, S. (2008) *Trends Genet.* **24**, 280–288
7. Phatnani, H. P., and Greenleaf, A. L. (2006) *Genes Dev.* **20**, 2922–2936
8. Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. (2005) *Genes Dev.* **19**, 1401–1415
9. Winston, F., Chaleff, D. T., Valent, B., and Fink, G. R. (1984) *Genetics* **107**, 179–197
10. Clark-Adams, C. D., and Winston, F. (1987) *Mol. Cell. Biol.* **7**, 679–686
11. Neigeborn, L., Celenza, J. L., and Carlson, M. (1987) *Mol. Cell. Biol.* **7**, 672–678
12. Bortvin, A., and Winston, F. (1996) *Science* **272**, 1473–1476
13. Adkins, M. W., and Tyler, J. K. (2006) *Mol. Cell* **21**, 405–416
14. Kaplan, C. D., Laprade, L., and Winston, F. (2003) *Science* **301**, 1096–1099
15. Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., and Lis, J. T. (2009) *EMBO J.* **28**, 1067–1077
16. Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D. K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., Sato, H., Yamaguchi, Y., Mandal, S. S., Reinberg, D., Wada, T., and Handa, H. (2004) *Mol. Cell. Biol.* **24**, 3324–3336
17. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., and Jones, K. A. (2007) *Genes Dev.* **21**, 160–174
18. Fischbeck, J. A., Kraemer, S. M., and Stargell, L. A. (2002) *Genetics* **162**, 1605–1616
19. Liu, Z., Zhou, Z., Chen, G., and Bao, S. (2007) *Biochem. Biophys. Res. Commun.* **353**, 47–53
20. Gavin, A. C., Börsche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edlmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) *Nature* **415**, 141–147
21. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emil, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002) *Mol. Cell. Biol.* **22**, 6979–6992
22. Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R., 3rd, and Hartzog, G. A. (2003) *Mol. Cell. Biol.* **23**, 1368–1378
23. Yoh, S. M., Lucas, J. S., and Jones, K. A. (2008) *Genes Dev.* **22**, 3422–3434
24. Kim, M., Ahn, S. H., Krogan, N. J., Greenblatt, J. F., and Buratowski, S. (2004) *EMBO J.* **23**, 354–364
25. Kaplan, C. D., Morris, J. R., Wu, C., and Winston, F. (2000) *Genes Dev.* **14**, 2623–2634
26. Andrulis, E. D., Guzmán, E., Döring, P., Werner, J., and Lis, J. T. (2000) *Genes Dev.* **14**, 2635–2649
27. Adelman, K., Wei, W., Ardehali, M. B., Werner, J., Zhu, B., Reinberg, D., and Lis, J. T. (2006) *Mol. Cell. Biol.* **26**, 250–260
28. Saunders, A., Werner, J., Andrulis, E. D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J. T. (2003) *Science* **301**, 1094–1096
29. Pawson, T., and Nash, P. (2003) *Science* **300**, 445–452
30. Kuriyan, J., and Cowburn, D. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288
31. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., et al. (1993) *Cell* **72**, 767–778
32. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., and Kuriyan, J. (1993) *Cell* **72**, 779–790
33. Bae, J. H., Lew, E. D., Yuzawa, S., Tomé, F., Lax, I., and Schlessinger, J. (2009) *Cell* **138**, 514–524
34. Eck, M. J., Pluskey, S., Trüb, T., Harrison, S. C., and Shoelson, S. E. (1996) *Nature* **379**, 277–280
35. Hatada, M. H., Lu, X., Laird, E. R., Green, J., Morgenstern, J. P., Lou, M., Marr, C. S., Phillips, T. B., Ram, M. K., Theriault, K., et al. (1995) *Nature* **377**, 32–38
36. Maclennan, A. J., and Shaw, G. (1993) *Trends Biochem. Sci.* **18**, 464–465
37. Dengl, S., Mayer, A., Sun, M., and Cramer, P. (2009) *J. Mol. Biol.* **389**, 211–225
38. Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P. H., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., Lebbink, J. H., Notenboom, V., Poterszman, A., Rochel, N., Cohen, S. X., Unger, T., Sussman, J. L., Moras, D., Sixma, T. K., and Perrakis, A. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1232–1242
39. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology* (Carter, Jr., C. W., and Sweet, R. M. eds) pp. 307–326, Academic Press, New York
40. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol.* **364**, 215–230
41. Abrahams, J. P., and Leslie, A. G. (1996) *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42
42. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
43. CCP4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
44. Seraphin, B., and Rosbash, M. (1989) *Cell* **59**, 349–358
45. Besse, S., Vigneron, M., Pichard, E., and Puvion-Dutilleul, F. (1995) *Gene Expr.* **4**, 143–161
46. Van Driessche, B., Tafforeau, L., Hentges, P., Carr, A. M., and Vandenhoute, J. (2005) *Yeast* **22**, 1061–1068
47. Rose, M. D., Winston, F., and Hieter, P. (1990) *Laboratory Course Manual for Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
48. Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R., and Winston, F. (2008) *PLoS Biol.* **6**, e277
49. Katinka, M. D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivarès, C. P. (2001) *Nature* **414**, 450–453
50. Romier, C., James, N., Birck, C., Cavarelli, J., Vivarès, C., Collart, M. A., and Moras, D. (2007) *J. Mol. Biol.* **368**, 1292–1306
51. Holm, L., Kääriäinen, S., Rosenström, P., and Schenkel, A. (2008) *Bioinformatics* **24**, 2780–2781
52. Filippakopoulos, P., Müller, S., and Knapp, S. (2009) *Curr. Opin Struct. Biol.* **19**, 643–649
53. Liu, B. A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., and Nash, P. D. (2006) *Mol. Cell* **22**, 851–868
54. Johnson, S. J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S. L., and Hill, C. P. (2008) *J. Mol. Biol.* **377**, 1460–1473
55. Bond, C. S., and Schüttelkopf, A. W. (2009) *Acta Crystallogr. D Biol. Crystallogr.* **65**, 510–512



Les résultats obtenus, présentés dans la publication numéro 1, montrent donc que le domaine carboxy-terminal de Spt6 contient deux domaines SH2 qui sont impliqués dans l'interaction avec l'ARN polymérase II. Cependant, l'absence de détection d'une interaction directe entre ce domaine de la protéine et des peptides du CTD dont les sérines sont phosphorylés nous a poussé à tester l'interaction avec d'autres motifs et notamment la phosphorylation de la tyrosine. Les résultats de ces tests sont présentés ci-dessous.

### 3.4. Etude des interactions du tandem SH2 de la protéine Spt6

#### 3.4.1. Tests de l'interaction entre le tandem SH2 et le CTD de l'ARN polymérase II par co-expression

Parallèlement aux tests de microcalorimétrie et de Thermofluor décrits dans la publication numéro 1, la technique de co-expression a également été utilisée afin de tenter d'observer une interaction entre le CTD de la plus grande sous unité de l'ARN polymérase II et le domaine carboxy-terminal de Spt6. Ainsi, des constructions du CTD de l'ARN polymérase II d'*A. locustae* contenant une à trois répétitions du motif Y-S-P-T-S-P-A (dans cet organisme, les sérines 7 ont été remplacées par des alanines) ont été clonées dans un vecteur pGexNB (par hybridation d'amorces nucléotidiques). Sur ces différentes constructions, les différents états de phosphorylation des sérines en position 2 et 5 ont été mimés par le remplacement de ces sérines par un glutamate, une technique utilisée avec succès à de nombreuses reprises (Maciejewski et al., 1995). Les différentes constructions ainsi produites sont listées dans le Tableau 9.

Des tests de co-expression ont été réalisés entre ces constructions et la construction du domaine carboxy-terminal de la protéine Spt6 d'*A. locustae* (783-957). Afin de valider cette approche par co-expression, une construction du domaine de liaison à l'ARN polymérase II (CID pour CTD Interacting Domain) de la protéine de levure Pcf11 a également été utilisée (construction 1-140). Cette protéine est un facteur de poly-adénylation des ARN messagers, la structure de la protéine homologue humaine (SCAF8) ayant été résolue en complexe avec différentes formes du CTD (Becker et al., 2008). Il a été montré que cette protéine interagissait aussi bien avec la forme non phosphorylée de l'ARN polymérase II qu'avec les

formes phosphorylées, l'interaction avec la forme phosphorylée sur la sérine 2 ayant une plus grande affinité pour la protéine.

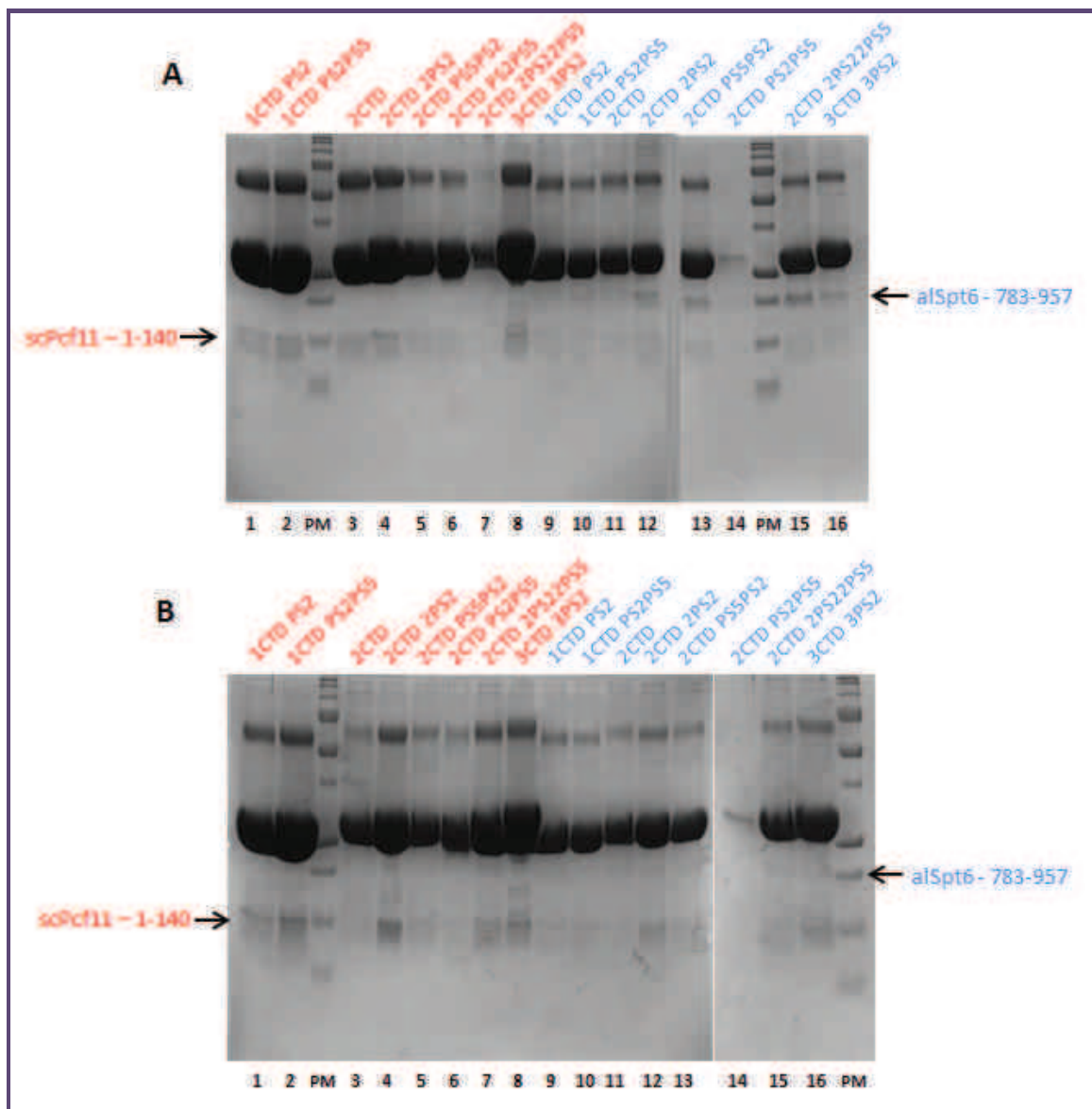
Constructions	Construction
1CTD PS2	Y-E-P-T-S-P-A
1 CTD PS2 PS5	Y-E-P-T-E-P-A
2 CTD	Y-S-P-T-S-P-A-Y-S-P-T-S-P-A
2 CTD 2PS2	Y-E-P-T-S-P-A-Y-E-P-T-S-P-A
2 CTD PS5 PS2	Y-S-P-T-E-P-A-Y-E-P-T-A-P-A
2 CTD PS2 PS5	Y-E-P-T-S-P-A-Y-S-P-T-E-P-A
2 CTD 2PS2 2PS5	Y-E-P-T-E-P-A-Y-E-P-T-E-P-A
3 CTD 3PS2	Y-E-P-T-S-P-A-Y-E-P-T-S-P-A-Y-E-P-T-S-P-A

**Tableau 9 : Différentes constructions de l'extrémité carboxy-terminale de l'ARN polymérase II d'*A. locustae* générées.**

Ces tests de co-expression ont été réalisés en utilisant deux tampons de lyses différents, l'un contenant 50mM NaCl et l'autre 400mM NaCl qui ont permis de tester l'influence de la salinité sur la formation du complexe (Figure 43). Ces tests de co-expression montrent une préférence de liaison de Pcf11 pour les fragments de CTD phosphorylés principalement sur les sérines 2 (lignes 4 et 8), l'interaction étant moins forte quand les sérines en position 5 sont également phosphorylées (lignes 2, 5, 6 et 7). D'autre part, l'interaction formée semble plus forte et plus spécifique lorsque le tampon de lyse contient 400mM NaCl. En effet, certaines interactions peuvent être observées à 50mM NaCl mais les complexes ne sont plus formés à 400mM NaCl (cas du fragment 2CTD, ligne 3) alors que dans d'autre cas l'interaction est identique dans les deux conditions (2CTD 2PS2, ligne 4). Les données obtenues sont en accord avec les informations connues sur le complexe, bien que les complexes observés soient faibles et que l'interprétation de ces résultats ne soit pas évidente. Ainsi, la liaison préférentielle de Pcf11 pour le CTD phosphorylé sur la sérine 2 est bien détectée, un grand nombre de motifs étant reconnus par la protéine.

Le domaine carboxy-terminal de Spt6 semble quant à lui lier préférentiellement les constructions du CTD contenant au moins deux motifs YSPTSPA. De même, le fragment 2CTD non phosphorylé ne retient pas le domaine carboxy-terminal de Spt6 alors que la phosphorylation de la sérine en position 5 ne semble pas gêner la formation du complexe, contrairement à Pcf11. En effet, la construction 2CTD 2PS2-2PS5 (ligne 15) semble avoir une

affinité au moins équivalente à la construction 2CTD 2PS2 (ligne 12). D'autre part, les interactions ne sont détectées qu'à faible concentration saline, ce qui pourrait indiquer que l'interaction mise en place entre Spt6 et le CTD est faible.



**Figure 43 : Co-expression des vecteurs GST-CTD avec les protéines Pcf11 et Spt6**

Différentes constructions des vecteurs GST-CTD ont été co-exprimées avec la construction 1-140 de Pcf11 (annotations rouges) et 783-957 de la protéine Spt6 d'*A. locustae* (annotations bleus). Ces tests de co-expression ont été réalisés dans deux conditions de lyse distinctes : la première solution est composée de 50 mM NaCl, 10mM Tris pH8 (A), la seconde est composée de 400mM NaCl, 10mM Tris pH8 (B). Il faut noter que dans le cas de la co-expression entre les domaines SH2 de Spt6 et le fragment 2CTD PS2PS5 (ligne 14) aucune bande n'est visible sur le gel du fait d'un problème de culture.

Ces tests tendent à confirmer que la liaison du domaine carboxy-terminal de Spt6 au CTD de l'ARN polymérase II requiert au moins deux répétitions du motif YSPTSPA et la phosphorylation de ces motifs. Cependant, aucune information claire concernant le motif de phosphorylation reconnu spécifiquement par Spt6 n'a pu être obtenue. En effet, les complexes observés sont faibles et seules de faibles différences d'intensité sont observées entre les bandes correspondant à la protéine Spt6 retenue par les différentes constructions du CTD.

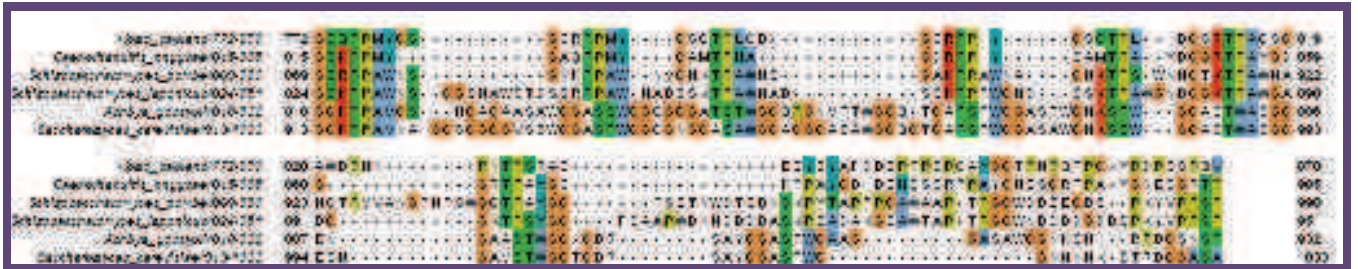
### 3.4.2. Test de l'interaction du tandem SH2 avec des peptides portant des phosphotyrosines

Aucune interaction directe entre le tandem SH2 et des peptides du CTD dont les sérines sont phosphorylées n'a pu être détectée par microcalorimétrie et Thermofluor. Pour cette raison, bien que les données publiées montrent que le domaine carboxy-terminal de Spt6 n'interagit pas avec le CTD de l'ARN polymérase II phosphorylé sur la tyrosine (Yoh et al., 2007), des tests d'interaction avec des peptides phosphorylés sur ce résidu ont été réalisés. L'interaction entre le domaine carboxy-terminal de la protéine Spt6 et un peptide contenant deux répétitions du motif YSPTSPS phosphorylé sur sa tyrosine centrale a donc été testée par microcalorimétrie et par thermofluor (séquence YSPTSPSpYSPTSPS, peptide synthétisé). La protéine a été purifiée dans un tampon contenant 400mM NaCl, 10mM Tris pH8 et a été diluée après concentration afin de tester deux concentrations saline (50 et 400mM NaCl) lors de la réalisation des tests d'interaction. De plus, deux concentrations protéiques ont été testées (50 et 100µM), le peptide utilisé pour les mesures étant en large excès (500µM). Cependant aucune interaction entre la protéine et le peptide n'a été observée avec aucune de ces deux techniques.

### 3.4.3. Interaction avec la protéine Spt5

La protéine Spt5 fait partie du complexe DSIF (pour DRB Sensitivity Inducing Factor) (Wada et al., 1998). Ce complexe intervient dans la régulation de l'élongation de la transcription et la maturation des ARN messagers. Il est recruté à la fin de l'étape d'initiation de la transcription et la phosphorylation du domaine carboxy-terminal de Spt5 par PTEF-b permet le passage en phase d'élongation productive (Sims et al., 2004). En effet, l'extrémité carboxy-terminale de Spt5 contient des éléments répétés proches du CTD de l'ARN polymérase II, mais de séquences différentes. Ces séquences sont très divergentes mais peuvent être caractérisées par un motif thréonine-proline chez les eucaryotes et une séquence

consensus (SAWGGQ) chez les levures (Figure 44). Chez la levure, il a été démontré que la kinase Bur1, homologue à p-TEF-b, phosphorylait les sérines présentes dans ces séquences répétées (Liu et al., 2009).



**Figure 44 : Alignement de séquences de l'extrémité carboxy-terminal de Spt5 (adapté de Pfam)**

*L'extrémité carboxy-terminale de Spt5 contient une région contenant des motifs répétés. La conservation de séquence entre ces motifs est faible. Chez les eucaryotes supérieurs un motif thréonine – proline est retrouvé alors que chez la levure une séquence consensus (SAWGGQ) est retrouvée. Le CTD de la protéine Spt5 de Schizosaccharomyces pombe est proche de celui des eucaryotes supérieurs.*

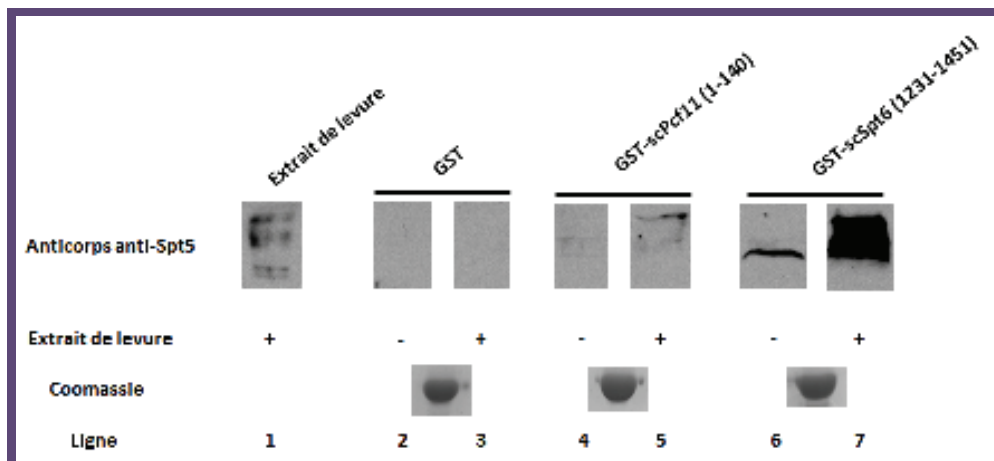
Une interaction entre Spt5 et Spt6 a été suggérée par des expériences d'immunoprécipitation (Lindstrom et al., 2003) et a été confirmée par la suite mais n'a pas été caractérisée biochimiquement. Ainsi, (i) sachant que le domaine carboxy-terminal de Spt6 semble lier le CTD phosphorylé de l'ARN polymérase II dont la structure est proche du CTD de Spt5, (ii) que des sérines du CTD de Spt5 sont phosphorylées au cours de la transcription et (iii) connaissant l'interaction entre ces deux protéines, l'implication du tandem SH2 de Spt6 dans cette interaction a été testée par pulldown et par co-expression.

#### 3.4.3.1. Test de l'interaction Spt6/Spt5 par GST pulldown

Des expériences de pulldown ont été réalisées avec le domaine carboxy-terminal de la protéine Spt6 de levure. Cette construction de la protéine a été produite en fusion avec une étiquette de purification GST et la résine a été incubée avec un extrait de levure. Après plusieurs lavages, les échantillons ont été déposés sur gel d'acrylamide et la présence de Spt5 dans la fraction retenue a été détectée par western blot avec des anticorps dirigés contre le domaine carboxy-terminal de Spt5.

L'extrait de levure utilisé pour les expériences de pulldown contient bien la protéine Spt5, le signal obtenu étant très faible comme le montre la ligne 1 (Figure 45). De plus, la reconnaissance aspécifique de l'anticorps dirigé contre Spt5 avec l'étiquette de purification GST a été exclue, aucune bande ne pouvant être détectée par western blot (lignes 2 et 3) lorsque la protéine GST couplée à la résine est incubée avec l'extrait de levure. De plus, la

construction 1-140 de la protéine Pcf11 qui n'interagit pas avec Spt5 ne retient ni l'anticorps ni la protéine Spt5, confirmant l'absence de liaison aspécifique. Ces expériences de pulldown montrent par contre clairement un effet d'enrichissement de la protéine Spt5 lorsque l'extrait de levure est incubé avec le domaine carboxy-terminal de Spt6. Ces expériences suggèrent donc qu'un complexe pourrait être formé entre le domaine carboxy-terminal de Spt6 et la protéine Spt5.

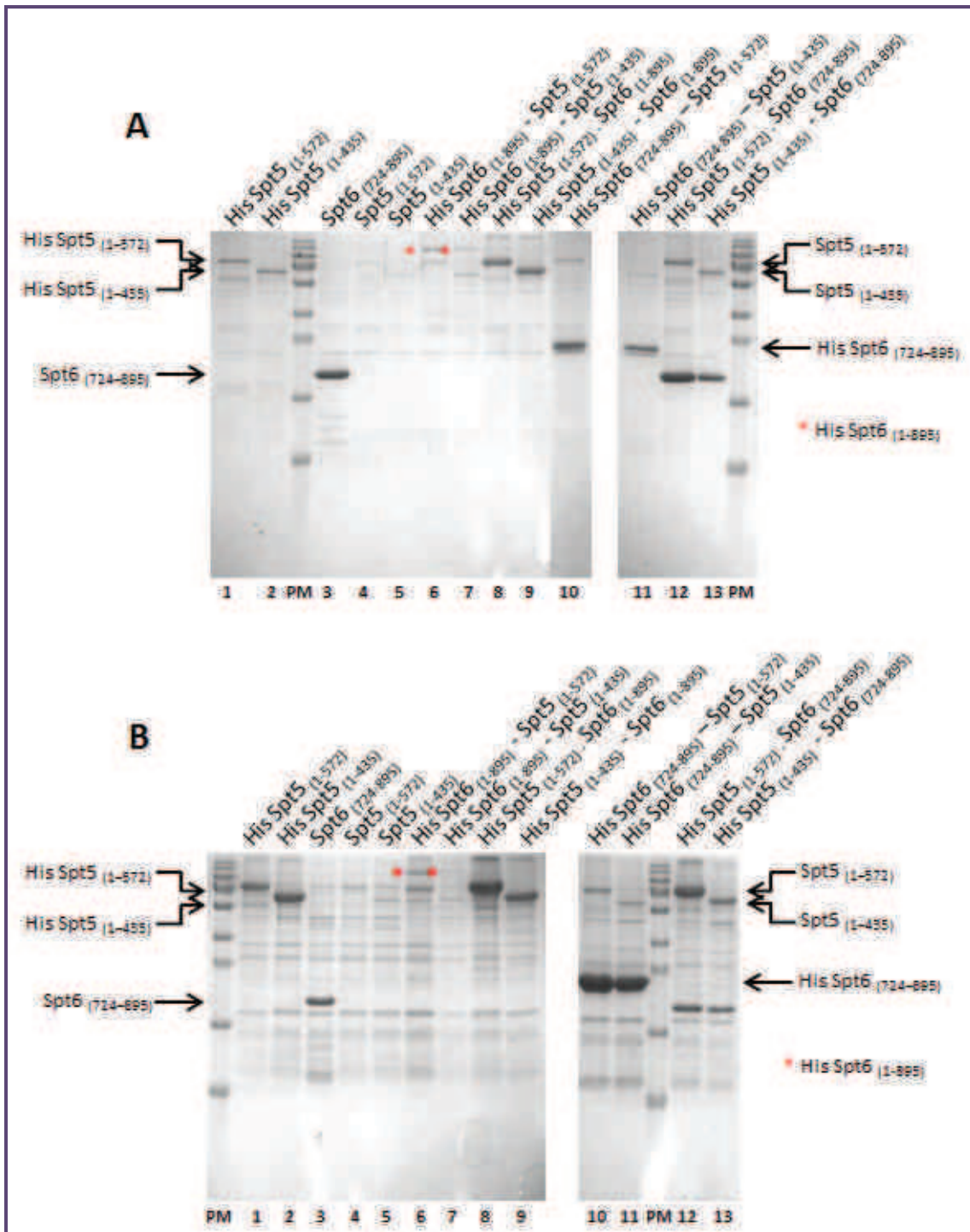


**Figure 45 : Interaction Spt6/Spt5 - expériences de pulldown**

*Des quantités équivalentes des différentes protéines liées à la résine de purification sont incubées avec l'extrait de levure. La présence de la protéine Spt5 dans la fraction retenue est détectée par Western blot avec un anticorps dirigé contre l'extrémité carboxy-terminale de la protéine. L'absence de reconnaissance aspécifique par l'anticorps est vérifiée par dépôt de protéine purifiée non incubée avec l'extrait, la quantité de protéine incubée ayant été normalisée par coloration d'un gel d'acrylamide au bleu de coomassie.*

### 3.4.3.2. Test de l'interaction Spt6/Spt5 par co-expression

Dans la perspective de produire ce complexe et de définir les bornes du fragment de Spt5 impliqué dans cette interaction, des tests de co-expression ont été réalisés dans la bactérie. Pour ces tests, les protéines d'*Encephalitozoon cuniculi* ont été utilisées. Deux constructions de Spt5 ont été clonées dans les vecteurs pNEA-tH et pNCS, l'une contenant le gène de la protéine entière, la seconde ne contenant que la partie amino-terminale de la protéine (résidus 1 à 435, les répétitions du motif S-T/A-W-G-G-A/Q ont été supprimées). Ces tests ont été réalisés à partir de cultures réalisées en utilisant un milieu de culture Graffinity (2,5 mL de culture). Ces cultures ont ensuite été dupliquées afin de tester l'influence de la concentration saline du tampon de lyse sur la formation du complexe. Ainsi, deux tampons de lyses ont été testés, l'un contenant 50mM NaCl et 10mM Tris pH8, l'autre 400mM NaCl et 10mM Tris.



**Figure 46 : Interaction Spt6/Spt5 – tests de co-expression**

Différentes constructions des protéines Spt6 et Spt5 ont été co-transformées dans des bactéries BL21[DE3] pRARE2 pour la réalisation de tests de co-expression. Deux conditions de lyse ont été testées afin d'observer l'effet de la concentration saline sur la formation du complexe ((A) tampon contenant 50mM NaCl, 10mM Tris pH8 et (B) tampon contenant 400mM NaCl, 10mM Tris)

Ces tests montrent que les deux constructions de Spt5 sont exprimées, leur solubilité étant meilleure à 400mM NaCl (lignes 1 et 2). Par contre une légère interaction aspécifique à

la résine de ces deux constructions est observée dans les deux tampons de lyses utilisés (ligne 4 et 5). Le fragment du domaine carboxy-terminal de la protéine Spt6 utilisé (construction 724-895) montre quant à lui une forte liaison aspécifique à la résine (ligne 3), cette information étant importante pour l'interprétation des résultats de ces tests d'expression.

La co-expression de la protéine Spt6 entière permet de retenir faiblement les deux constructions de Spt5 (lignes 6 à 9). Cette interaction est principalement observable à 50mM NaCl lorsque la protéine Spt6 porte l'étiquette de purification (lignes 6 et 7). Dans ce cas les bandes correspondant à Spt5 ne proviennent pas de l'interaction aspécifique de la protéine avec la résine puisque leur intensité est supérieure à celle observée dans les lignes 4 et 5 (expression de la protéine ne portant pas d'étiquette de purification avec la résine de purification). Dans le cas où la protéine Spt5 porte l'étiquette de purification (lignes 8 et 9), l'intensité plus forte des bandes correspondant à Spt5 peut être expliquée par la solubilisation de la protéine par sa co-expression avec Spt6, ce qui suggérerait qu'un complexe est formé. Dans ce cas, les bandes correspondant à Spt6 (1-895) ne sont pas observées, ce qui est probablement dû à la co-migration de la protéine avec les deux constructions de Spt5.

Une interaction faible peut également être observée dans le cas où le domaine carboxy-terminal de Spt6 est co-exprimé avec les deux constructions de Spt5 (lignes 10 à 13). Alors que la protéine Spt6 portant une étiquette de purification est fortement retenue sur la résine, seules de faibles bandes correspondant aux constructions de Spt5 sont observées (lignes 10 et 11). Inversement, l'expression des constructions de Spt5 portant une étiquette de purification permet effectivement de retenir la construction de Spt6 mais à des niveaux équivalents à la liaison aspécifique de Spt6 à la résine (comparer les lignes 12 et 13 à la ligne 4).

Ainsi, le complexe observé lors des expériences de pulldown n'a pu être reproduit de manière claire dans les bactéries. En effet, les complexes observés sont faibles, la liaison aspécifique des protéines à la résine rendant l'interprétation des résultats difficile. La modification de l'étiquette de purification utilisée permettrait d'obtenir une vision plus claire de la formation du complexe. De plus, l'expression en bactéries n'est peut-être pas adaptée à la production de ce complexe. En effet, l'absence de modifications post-traductionnelles chez la bactérie, et notamment de phosphorylation de la protéine Spt5 pourrait expliquer l'absence d'interaction entre les deux protéines puisque le tandem de domaines SH2 de Spt6 est censé



reconnaître spécifiquement les motifs phosphorylés. Le complexe pourrait ainsi être formé dans un système d'expression eucaryote tel que le système baculovirus ou la levure.

### 3.5. Discussion

La résolution de la structure du domaine carboxy-terminal de la protéine Spt6 d'*A. locustae* a montré que ce domaine contient deux domaines SH2 organisés en tandem. Le premier domaine présente toutes les caractéristiques des domaines SH2, les résidus de la poche de fixation du groupement phosphate étant conservés ainsi que la poche de reconnaissance spécifique. Seuls les résidus impliqués dans la formation des interactions amino-aromatiques sont en partie mutés. Cette poche de fixation du groupement phosphate étant moins profonde, elle pourrait donc lier des résidus phosphosérine. Le second domaine SH2 est quant à lui hautement non canonique, quasiment aucun résidu de la poche de fixation du phosphate n'étant conservé et la seconde poche étant absente. Nos données suggèrent cependant que ces deux domaines sont nécessaires à la fonction de la protéine et à la reconnaissance de l'ARN polymérase II, bien qu'aucune interaction directe forte entre des fragments de CTD phosphorylés et le tandem SH2 n'ait pu être détectée *in vitro*.

#### 3.5.1. Comparaison des structures du domaine carboxy-terminal de Spt6 d'*A. locustae*, de *C. glabrata* et de *S. cerevisiae*

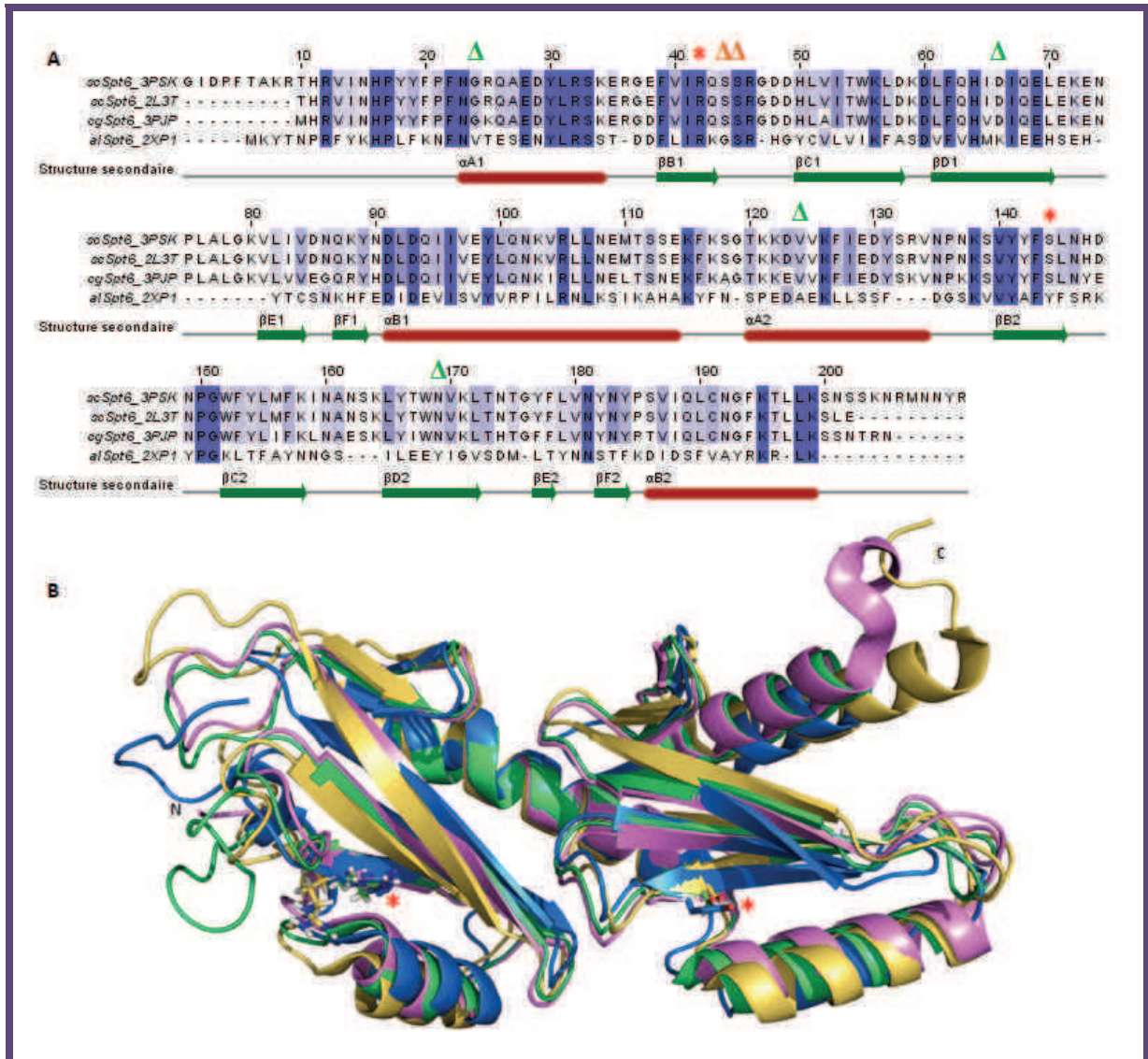
Nos données sont en accord avec les structures des domaines SH2 des protéines Spt6 de *Candida glabrata* (Dengl et al., 2009; Sun et al., 2010) et de *Saccharomyces cerevisiae* (Close et al., 2011; Liu et al., 2011) résolues récemment. En effet, les repliements adoptés par les trois structures sont identiques au repliement de la structure d'*A. locustae* (Figure 47B), les valeurs de déviation standard entre ces quatre structures, présentées dans le Tableau 10, étant modérées.

Calcul par rapport à <i>Antonospora locustae</i> (2XP1)	<i>Candida glabrata</i> (3PJP)	<i>Saccharomyces cerevisiae</i> (3PSK)	<i>Saccharomyces cerevisiae</i> (2L3T) – RMN
<b>RMSD calculés sur le tandem SH2 (Å)</b>	2,006	1,692	2,355
<b>RMSD calculés sur SH2<sub>1</sub> (Å)</b>	1,505	1,376	1,784
<b>RMSD calculés sur SH2<sub>2</sub> (Å)</b>	1,804	1,707	1,907

**Tableau 10 : Valeurs de déviations standard entre les structures des domaines SH2.**

*Les valeurs de RMSD ont été déterminées grâce au logiciel MATRAS à partir de l'alignement deux à deux des structures des chaînes principales (C-alpha).*

Les caractéristiques canoniques du premier domaine SH2 et non canonique du second domaine sont confirmées par les trois structures (Figure 47).



**Figure 47 : Alignement des séquences et superposition des structures de domaines SH2 de la protéine Spt6.**

Les séquences des quatre domaines carboxy-terminaux des protéines dont les structures ont été résolues ont été alignées (A), (\*) position de l'Arg  $\beta\text{B}$  conservée, ( $\Delta$ ) position des résidus de la boucle BC fixant le phosphate, ( $\Delta$ ) position des résidus réalisant l'interaction de type amino-aromatique. (B) superposition des quatre structures de tandems SH2 de Spt6 résolues (vert : scSpt6, 3PSK, bleu : alSpt6, 2XP1, magenta : cgSpt6, 3PJP et ocre : scSpt6 RMN, 2L3T), (\*) indique la position des résidus  $\beta\text{B}$  de fixation du groupement phosphate.

En effet, dans le cas du domaine SH2<sub>1</sub>, toutes possèdent une arginine sur le brin  $\beta\text{B}$  (Figure 47) et des résidus sérines dans la boucle BC qui sont compatibles avec la reconnaissance d'un groupement phosphate. Par contre, les structures de *C. glabrata* et *S.*

*cerevisiae* montrent que les résidus de l'hélice  $\alpha A$  et du brin  $\beta D$  normalement impliqués dans l'interaction de type amino-aromatique ne sont pas compatibles avec la mise en place de ce type d'interaction, une glycine remplaçant l'arginine de l'hélice  $\alpha A$  et un aspartate remplaçant la lysine du brin  $\beta D$  (des résidus valines et lysines respectivement étant retrouvés dans la structure d'*A. locustae*). Les poches de fixation sont également moins profondes que dans les domaines SH2 reconnaissant des phosphotyrosines, en accord avec la reconnaissance des phosphosérines. Dans les deux structures déterminées par diffraction des rayons X, des ions sulfates (3PSK) et acétates (3PJP) sont retrouvés dans cette poche de fixation du domaine SH2<sub>1</sub>, confirmant que ce premier domaine SH2 des protéines Spt6 pourraient lier des résidus phosphorylés. La poche de spécificité de ces domaines SH2<sub>1</sub> est également présente.

Le caractère non canonique du second domaine SH2 a également été confirmé par les structures de *C. glabrata* et *S. cerevisiae*. Tout d'abord, les résidus permettant la fixation du groupement phosphate retrouvés dans les domaines SH2 canoniques sont également absents dans ces structures. En effet, (i) l'arginine du brin  $\beta B$  est remplacée par des sérines, (ii) les résidus sérine/thréonine de la boucle BC remplacés par des résidus plus encombrés qui ne peuvent pas interagir avec un groupement phosphate et (iii) les résidus formant l'interaction amino-aromatique sont remplacés par des valines ( $\alpha A$ ) et des asparagines ( $\beta D$ ). De plus, les structures de *Saccharomyces cerevisiae* et *Candida glabrata* présentent une tyrosine sur le brin  $\beta C$  qui comble cette poche de fixation de la phosphotyrosine, suggérant que le domaine SH2<sub>2</sub> ne participe pas à la formation d'un complexe par cette poche. Ce dernier résultat est intéressant car il va à l'encontre de nos observations et conclusions. D'autre part, la poche de spécificité n'est pas détectable sur ce second domaine SH2. Ainsi, la comparaison des structures des domaines SH2<sub>2</sub> des protéines Spt6 prouve que le caractère non canonique observé sur ce domaine de la protéine d'*A. locustae* est commun aux domaines SH2<sub>2</sub> des protéines Spt6 et pose clairement la question de son rôle biologique.

### 3.5.2. Interaction des tandems SH2 avec des peptides phosphorylés

Les différentes équipes ayant résolu ces structures ont également testé la capacité de ces tandems SH2 à reconnaître des peptides du CTD de l'ARN polymérase II. Des constantes de dissociation ont ainsi pu être mesurées entre les tandems SH2 et des peptides par anisotropie de fluorescence (Sun et al., 2010; Close et al., 2011) et résonance magnétique

nucléaire (Liu et al., 2011) (Tableau 11), alors que nos tests d'ITC n'ont pas permis de montrer une interaction entre les protéines et les peptides. En effet, les techniques d'anisotropie de fluorescence et de RMN permettent de déterminer des constantes d'affinité plus faibles que la technique de microcalorimétrie (de l'ordre du  $\mu\text{M}$ ).

Peptide	cgSH2 (Sun et al., 2010) Anisotropie de fluorescence	scSH2 (Close et al., 2011) Anisotropie de fluorescence	scSH2 RMN (Liu et al., 2011) RMN
<b>Tampon de la protéine</b>	10mM NaCl, 20mM HEPES pH8, 5mM DTT	100mM NaCl, 20mM Tris- HCl pH 8, 5% glycérol	100mM NaCl, 10mM Bis- Tris propane pH7, 1mM DTT, 1mM EDTA
<b>Non phospho- rylé</b>	Pas d'interaction	$K_d > 1000$ , interaction négligeable	Pas d'interaction
<b>pTyr</b>	nd	110,2 +/- 9,1 PS(pY)SPTSPS(pY)SPTSPS	1100 +/- 100 TSPS(pY)SPTSPS
<b>1pSer2</b>	34,7 +/- 2,2 SY(pS)PTSPSYSPTSPS 46,0 +/- 2,3 SYSPTSPSY(pS)PTSPS	nd	3200 +/- 100 SPSY(pS)PTSPS
<b>2pSer2</b>	10,4 +/- 20 SY(pS)PTSPSY(pS)PTSPS	197,1 +/- 39,69 PSY(pS)YSPTSPSY(pS)PTSP S	1500 +/- 200 SY(pS)PTSPSY(pS)PTSPS 1200 +/- 100 TSPSY(pS)PTSPSYSPTSPSY( pS)PTSPS
<b>1pSer5</b>	nd	nd	5300 +/- 100 SYSPT(pS)PSYSPT
<b>2pSer5</b>	nd	245,4 +/- 29,4 PSYSPT(pS)PSYSPT(pS)PS	nd
<b>1pSer7</b>	nd	nd	8000 +/- 500 PTSP(pS)YSPTS
<b>1pSer2/ 1pSer5</b>	nd	nd	1400 +/- 100 SPSY(pS)PT(pS)PSYSPT
<b>2pSer2/ 2pSer5</b>	nd	23,0 +/- 2,2 T(pS)PSY(pS)PT(pS)PSY(pS) PTSPS	nd

**Tableau 11 : Mesure des constantes de dissociation entre les tandems SH2 et de peptides phosphorylés mimant le CTD de l'ARN polymérase II ( $\mu\text{M}$ ).**

(*nd* : non déterminé), les séquences des peptides utilisés pour la réalisation de ces mesures sont indiquées sous la valeur de la constante dissociation.

Ces mesures de constantes de dissociation permettent de tirer deux conclusions concernant l'interaction de Spt6 avec les peptides du CTD de l'ARN polymérase II. Tout d'abord, elles montrent que le tandem de domaine SH2 de Spt6 lie préférentiellement les peptides contenant deux répétitions du motif YSPTSPS dont les sérines en position 2 sont phosphorylées, la phosphorylation additionnelle de la sérine 5 permettant de renforcer

l'interaction. Elles montrent également que le domaine de la protéine peut fixer des peptides dont les tyrosines sont phosphorylées avec une affinité équivalente aux peptides phosphorylés en sérine 2. Cependant, les interactions mises en place sont plus faibles que celles réalisées par les domaines SH2. Ces résultats sont en accord avec les observations que nous avons réalisées lors de nos tests de co-expression (partie 3.4.1) au cours desquels le domaine carboxy-terminal de Spt6 était co-exprimé avec des fragments du CTD dont les phosphorylations étaient mimées par des résidus glutamate.

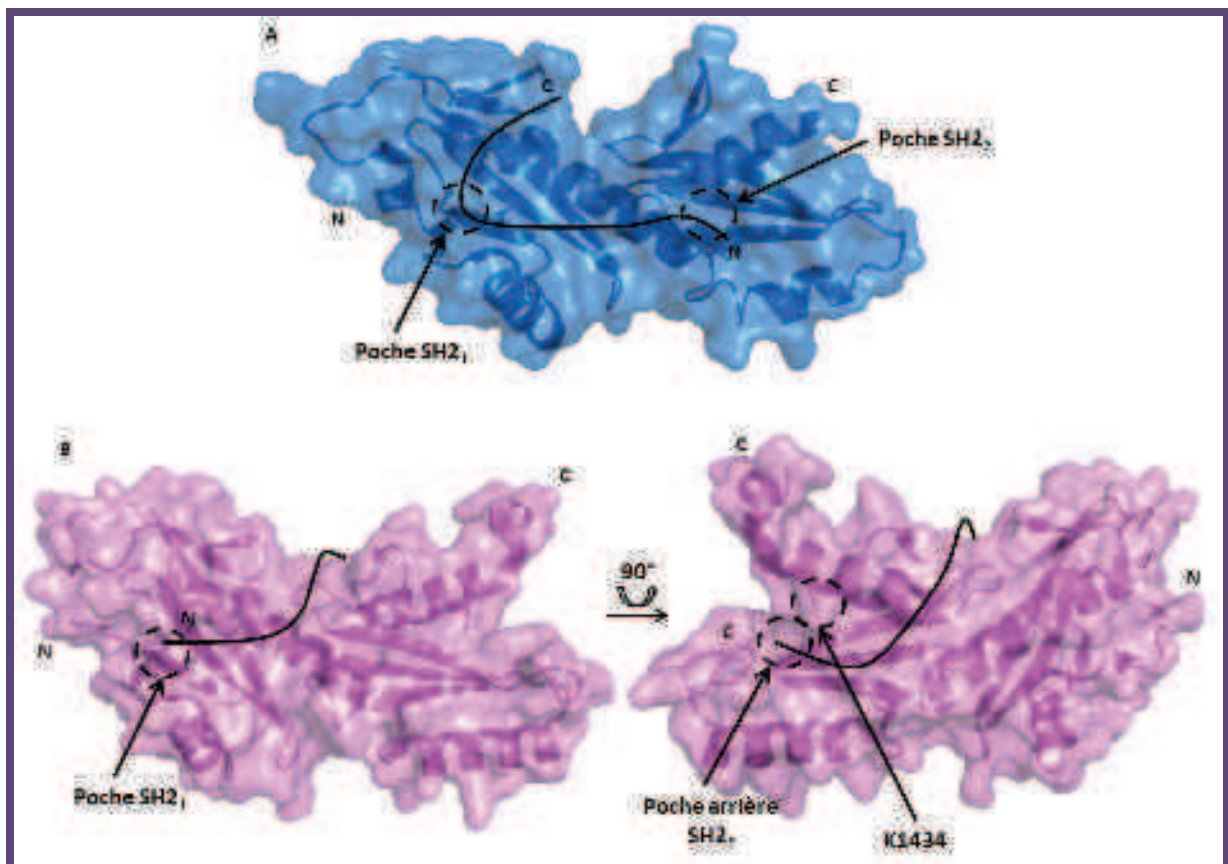
### 3.5.3. Modèle de l'interaction entre le tandem SH2 de Spt6 et le CTD de l'ARN polymérase II

La comparaison des quatre structures disponibles soulève également la question de la capacité du second domaine SH2 à reconnaître des groupements phosphate. Le modèle que nous avons proposé s'appuie sur la présence d'un groupement phosphate dans la poche de fixation de la phosphotyrosine du domaine SH2<sub>2</sub>. Notre modèle suggère donc qu'un premier résidu phosphorylé interagit avec la seconde poche de fixation de la phosphotyrosine (poche SH2<sub>2</sub>), le peptide rejoignant la première poche qui interagit avec le second résidu phosphorylé (poche SH2<sub>1</sub>) (Figure 48A). Nous avons confirmé ce modèle par la réalisation d'expériences de GST-pulldown utilisant une construction du tandem SH2 de *A. locustae* dans laquelle les résidus impliqués dans l'interaction ont été mutés. Ces expériences montrent un affaiblissement du complexe.

Contrairement à la structure de *A. locustae*, les structures de *S. cerevisiae* et *C. glabrata* révèlent la présence d'un résidu tyrosine dans la poche de fixation de la phosphotyrosine du domaine SH2<sub>2</sub> qui n'est par conséquent pas accessible pour la mise en place d'une interaction. Au contraire, Sun et ses collaborateurs (Sun et al., 2010) décrivent une poche conservée et chargée positivement sur l'autre face du domaine SH2<sub>2</sub>, (Figure 48B, structure de droite) cette poche étant notamment formée par une lysine (1434) conservée qui pourrait participer à l'interaction d'un résidu phosphorylé (poche arrière). De plus, la mutation de cette lysine augmente la constante de dissociation mesurée avec un peptide contenant deux répétitions du motif Y-S-P-T-S-P-S dont les sérines 2 sont phosphorylées.

D'autre part, la réalisation d'expériences de titration par RMN entre le tandem SH2 de levure et des peptides phosphorylés mimant le CTD de l'ARN polymérase (Liu et al., 2011) confirme que les résidus de la poche de fixation du groupement phosphate du domaine SH2<sub>2</sub>

n'interviennent pas dans la fixation des peptides. En effet, l'enregistrement de spectres  $^{15}\text{N}$ -HSQC en présence et en l'absence de peptides permet d'observer des changements dans les valeurs de déplacements chimiques des résidus impliqués dans l'interaction. Ainsi, les déplacements chimiques sont modifiés sur l'autre face du domaine  $\text{SH2}_2$ , au niveau de l'hélice  $\alpha\text{B2}$ . Les observations de Sun et ses collaborateurs, ainsi que les résultats de RMN présentés par Liu et ses collaborateurs permettent de proposer un autre modèle dans lequel le peptide pourrait lier le domaine de liaison du groupement phosphate du domaine  $\text{SH2}_1$ , puis rejoindrait l'autre face du second domaine  $\text{SH2}$  pour lier la poche arrière de domaine (Figure 48A).



**Figure 48 : Différents mode de liaison entre le tandem SH2 de Spt6 et le CTD phosphorylé de l'ARN polymérase II.**

(A) Mode de liaison du CTD de l'ARN polymérase II au tandem SH2 de Spt6 dans la publication 1 (B) mode de liaison proposé par (Sun et al., 2010), (l'orientation de la structure de la vue de gauche est identique à celle de la Figure 47B),

Ainsi, le caractère non canonique du domaine  $\text{SH2}_2$  de Spt6 a conduit à l'élaboration de différentes hypothèses concernant le mode de liaison entre le tandem SH2 et le CTD de l'ARN polymérase II (Figure 48). Notre modèle est compatible avec celui suggéré par Sun et

Lieu concernant la fixation dans la poche arrière du second SH2<sub>2</sub>, puisque le peptide modélisé pourrait très bien rejoindre cette poche arrière, sa lysine invariante étant également retrouvée dans la protéine de *A. locustae*. A l'heure actuelle aucune information ne permet de favoriser l'un de ces modèles de liaison. Cependant l'observation que l'hyperphosphorylation du CTD de l'ARN polymérase II (Yoh et al., 2007, 2008), et donc la phosphorylation de deux motifs consécutifs des répétitions YSPTSPS, est requise pour l'interaction avec Spt6 pourrait favoriser notre modèle. En effet, il est le seul qui permette la mise en place de ce type d'interaction entre deux résidus phosphorylés qui sont peu éloignés.

#### 3.5.4. Structure des domaines SH2 et interaction avec le CTD de l'ARN polymérase - Conclusion

Notre étude et celles menées par les autres équipes sur les domaines SH2 des protéines Spt6 de différents organismes ont permis de montrer que le repliement du domaine carboxy-terminal de la protéine a été conservé. Le rôle biologique important de cette région de la protéine a également été confirmé, sa délétion conduisant à l'apparition de phénotypes thermosensibles et de défauts dans la phase d'élongation de la transcription (sensibilité à la 6-azauracyl) (Sun et al., 2010). Ils n'ont par contre pas permis de conclure clairement quant au degré de phosphorylation du CTD de l'ARN polymérase II requis pour l'interaction avec Spt6, ni sur la zone d'interaction du second domaine SH2.

Seule la résolution de la structure d'un complexe comprenant le tandem SH2 de Spt6 et des peptides phosphorylés permettrait finalement de connaître précisément le mode de liaison entre la protéine et le CTD de l'ARN polymérase II. Ces résultats confirment également que le mode de recrutement de Spt6 au cours de la transcription nécessite des déterminants qui restent à définir. En effet, le CTD de l'ARN polymérase II subit un grand nombre de modifications au cours de la transcription qui permettent la régulation de cette étape de la synthèse des protéines : phosphorylation des sérines 2,5 et 7, de la thréonine et de la tyrosine et isomérisation de la proline. Ainsi, l'obtention d'une vision claire des motifs reconnus est rendue difficile par le nombre extrêmement important de combinaisons de modifications possibles.

#### 3.5.5. Interaction du tandem SH2 de Spt6 avec Spt5

Nos résultats ont également permis de suggérer l'implication du tandem SH2 de Spt6 dans son interaction avec Spt5 (tests de pulldown paragraphe 3.4.3.1), ce complexe ayant été

montré précédemment (Lindstrom et al., 2003). Cependant, pour l'instant, aucune preuve d'une interaction directe entre ces deux protéines n'a pu être fournie. En effet, les tests de co-expression réalisés ne permettent pas d'affirmer clairement qu'un complexe entre ces deux protéines est formé, notamment du fait de l'interaction aspécifique des protéines à la résine de purification. La modification de l'étiquette de purification (pour une étiquette plus spécifique, comme la GST) et le changement du système d'expression pourraient permettre d'observer l'interaction entre les deux protéines. Alternativement, la phosphorylation *in vitro* de Spt5 pourrait être réalisée afin de tester l'interaction par des expériences de type pulldown. En effet, des tests réalisés dans l'équipe d'Anne-Catherine Dock-Bregeon ont montré qu'un facteur PTEF-b recombinant pouvait phosphoryler le CTD de la protéine Spt5 de *E. cuniculi* qui pourrait alors être utilisée pour la réalisation d'expériences de retards sur gel. Cette approche pourrait également être utilisée afin de tester l'interaction du tandem SH2 de Spt6 avec des constructions du CTD de l'ARN polymérase II phosphorylées *in vitro*, la phosphorylation de la sérine 5 pouvant être réalisée par la sous-unité CAK (composée des sous-unités Cdk7-CycH-Mat1) du facteur de transcription TFIIH et celle de la sérine 2 par PTEF-b.

Le domaine carboxy-terminal de Spt6 participe au recrutement de la protéine par l'ARN polymérase II hyperphosphorylée. Cependant, alors que les études réalisées *in vivo* par pulldown suggèrent que c'est la phosphorylation de la sérine 2 du CTD qui est reconnue par Spt6 (Yoh et al., 2007), nos études structurales ainsi que celles des autres équipes montrent que l'interaction implique des déterminants qui ne sont pas encore connus (Sun et al., 2010; Liu et al., 2010; Close et al., 2011). En effet, la mesure des constantes de dissociation entre des tandems SH2 et des peptides mimant le CTD de l'ARN polymérase II phosphorylé montre que plusieurs motifs de phosphorylation peuvent être reconnus par Spt6. De plus, la résolution de la structure du domaine carboxy-terminal de Spt6 montre que, alors que le premier domaine SH2 peut réaliser des interactions canoniques, le caractère hautement non canonique du second domaine ne permet pas de prédire le mode de liaison mis en place. Ainsi, deux modèles d'interaction ont pu être proposés. Cependant, à l'heure actuelle, aucun de ces modèles ne peut être validé et seule la résolution de la structure du domaine carboxy-terminal de Spt6 en complexe avec un peptide phosphorylé pourrait permettre de définir le mode d'interaction entre Spt6 et l'ARN polymérase II. De plus, la présence d'une autre



protéine pourrait permettre de renforcer l'interaction entre Spt6 et l'ARN polymérase II cette protéine restant encore à identifier bien que de nombreux candidats puissent être proposés.

Nos travaux ont également permis de mettre en évidence l'implication de tandem SH2 de Spt6 dans l'interaction avec Spt5, bien qu'aucune preuve d'une interaction directe n'ait pour l'instant pu être apportée. Cette interaction est cohérente avec le mode de recrutement de Spt6 au cours de la transcription, cependant son rôle reste à définir. En effet, si le domaine carboxy-terminal de Spt6 est impliqué dans une interaction avec le CTD de l'ARN polymérase II, une autre région du domaine carboxy-terminal de Spt6 doit être impliquée dans la formation du complexe, Spt5 pourrait alors permettre d'augmenter l'affinité de Spt6 pour l'ARN polymérase II.

Récemment, le domaine carboxy-terminal de Spt6 a également été impliqué dans l'interaction avec d'autres protéines : la cytidine déaminase AID dans les cellules humaines (Okazaki et al., 2011) et la sous-unité Rpa43 de l'ARN polymérase I chez la levure (Beckouët et al., 2011). Dans le cas de l'interaction avec AID, qui est une protéine spécifique des eucaryotes supérieurs, la région de Spt6 localisée en position carboxy-terminale du tandem SH2 pourrait permettre cette interaction. En effet, ce domaine de Spt6 est spécifique des mammifères, comme AID. Cependant, des tests d'expression qui ont été réalisés en collaboration avec l'équipe de Bernardo Reina n'ont pas permis de mettre en évidence une interaction entre ces deux protéines, notamment du fait de la forte insolubilité de AID. L'interaction avec Rpa43 a été définie bien plus récemment par des expériences de double hybride et n'a pas été caractérisée pour l'instant. Cette interaction suggère une implication de Spt6 dans la transcription réalisée par l'ARN polymérase I, ce qui n'avait pas encore été proposé.

En conclusion, le domaine carboxy-terminal de Spt6 est impliqué dans un nombre divers de mécanismes et d'interactions, la plupart d'entre elles n'étant pour l'instant pas caractérisées, notamment en termes moléculaires et fonctionnels. D'autre part, l'interaction de Spt6 avec l'ARN polymérase II, bien que reconnue fonctionnellement, reste peu comprise en termes moléculaires.

**Chapitre 4**  
**Etude du complexe Spt6/lws1 impliqué**  
**dans l'export des ARN messagers**

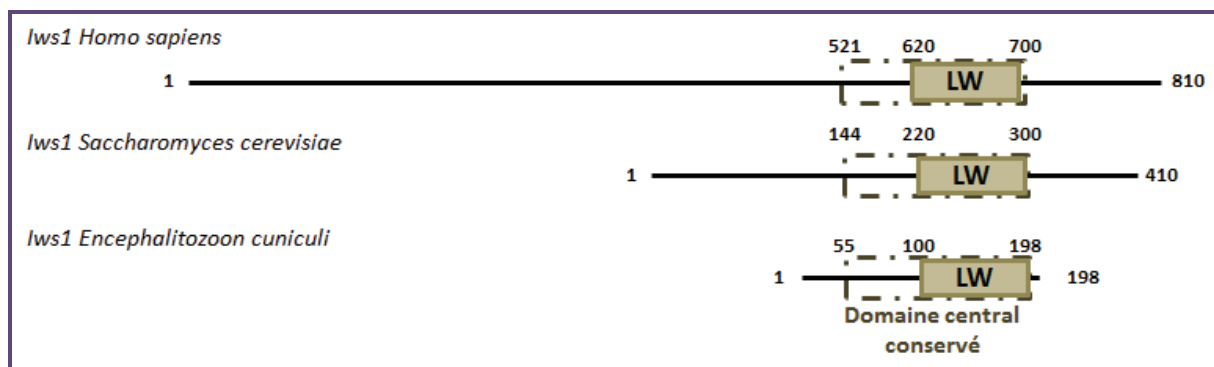


## 4.1. Introduction

Le complexe formé par Spt6 et Iws1 a été identifié lors de la réalisation d'une étude ayant pour but de déterminer les différentes protéines impliquées dans l'élongation de la transcription (Krogan et al., 2002). Au début du projet, l'interaction entre ces deux protéines avait partiellement été localisée au cours d'expériences de pulldown réalisées sur des extraits de cellules HeLa avec des constructions des protéines humaines recombinantes (Yoh et al., 2007). Ces expériences ont permis de montrer que l'extrémité amino-terminale de Spt6 (résidus 1 à 485) interagit avec le domaine central conservé de Iws1 (résidus 523 à 819). Ce chapitre décrit brièvement les caractérisations biochimiques, structurales et fonctionnelles de ce complexe, ces résultats ayant été publiés dans « The EMBO Journal » (Diebold et al., 2010a). D'autre part une interaction directe entre Iws1 et l'ARN polymérase II a été testée, dont les résultats sont également présentés dans ce chapitre. Finalement, nos données seront comparées aux données publiées par d'autres équipes ayant caractérisé biochimiquement, structuralement et fonctionnellement ce complexe (Pujari et al., 2010; McDonald et al., 2010).

## 4.2. Brève description des résultats

Dans un premier temps, notre but a été de localiser précisément les régions de ces deux protéines impliquées dans la formation du complexe. Nous nous sommes appuyés sur les alignements multiples des séquences de Spt6 et Iws1 afin d'identifier les régions homologues des régions décrites précédemment. Ces alignements nous ont permis de définir des constructions homologues des protéines du parasite *Encephalitozoon cuniculi* (notées ecSpt6 et ecIws1). La réalisation de tests de co-expression a permis dans un premier temps de confirmer que les constructions définies permettaient bien de former le complexe Spt6/Iws1 en utilisant les protéines du parasite. Dans un second temps, ces tests nous ont permis de définir plus précisément les zones impliquées dans la formation du complexe. Nous avons ainsi montré que le domaine conservé de ecIws1 (résidus 55 à 198) interagissait avec un très court fragment de ecSpt6 (résidus 53 à 71) (Figure 1 de la publication numéro 2). Le fragment d'interaction de ecIws1 correspond bien au domaine central retrouvé dans les protéines de levure et humaines, l'extrémité carboxy-terminale de la protéine étant plus courte dans la protéine de *E. cuniculi* (Figure 49). Afin de comprendre comment le complexe Spt6/Iws1 était formé, nous avons entrepris de résoudre sa structure en utilisant les constructions des protéines définies précédemment.



**Figure 49 : Organisation de la protéine Iws1**

Les recherches de domaines par Interproscan permettent de définir la présence d'un domaine LW.

Trois formes cristallines ont pu être obtenues (Figure 50). Dans ce cas, le problème de phase a été résolu par remplacement moléculaire en utilisant la structure de la construction (55 – 198) de la protéine Iws1 d'*Encephalitozoon cuniculi* qui avait été résolue par le Dr. Michael Koch au sein de l'équipe (Koch et al., 2010, publication présentée en annexe 8.8, page 289 et structure présentée dans la publication numéro 2).



**Figure 50 : Formes cristallines obtenues lors de la résolution de la structure du complexe Spt6/Iws1 par diffraction des rayons X.**

La résolution de la structure du domaine central de Iws1 montre que ce domaine est formé d'un motif HEAT, constitué de deux hélices  $\alpha$  ( $\alpha A$  et  $\alpha B$ ), suivi de deux motifs de type armadillo (notés ARM1 et ARM2), chacun de ces motifs étant formé par trois hélices ( $\alpha 1$  à  $\alpha 3$ ). La résolution de la structure du complexe Spt6/Iws1 a permis de montrer que Spt6 interagissait avec la surface des motifs ARM1 et ARM2 de Iws1 de manière bipartite (Figure 4 de la publication numéro 2). La première zone d'interaction est majoritairement hydrophobe et fait intervenir à la fois des résidus de l'hélice alpha de l'extrémité amino-terminale ( $\alpha N$ ) du fragment de Spt6 ainsi qu'un motif hydrophobe Ile-Phe (IF) de Spt6 qui interagit avec une poche hydrophobe à la surface de Iws1. A la suite du motif IF de Spt6, un motif Gly-x-Gly

conservé sert de « linker » avec la seconde zone d'interaction de Spt6. Dans cette seconde zone d'interaction, un mélange de liaisons hydrogènes et de contacts hydrophobes permettent de renforcer la liaison de Spt6 à Iws1.

Afin de valider la structure cristallographique du complexe Spt6/Iws1, des mutations ponctuelles ont été introduites dans chacune des deux protéines, en remplaçant les résidus impliqués dans l'interaction par des alanines. Ainsi, la mutation de résidus dans les deux zones d'interaction de Spt6 abolit la formation du complexe (Figure 5A de la publication numéro 2). D'autre part, des mutations ponctuelles introduites dans la protéine Iws1 ne font qu'affaiblir le complexe.

Suite à ces résultats, afin de comprendre le rôle du complexe Spt6/Iws1 *in vivo*, nous avons poursuivi nos travaux par des études de génétique de levure en collaboration avec le laboratoire du Dr. Fred Winston à Harvard. Dans un premier temps, afin de valider la zone d'interaction définie sur la protéine d'*E. cuniculi*, la région homologue de la protéine Spt6 de levure a été clonée dans des vecteurs d'expression. Des tests de co-expression ont alors été réalisés entre cette construction et la protéine Iws1 de levure entière ou le domaine central conservé de Iws1 de levure. Ces travaux ont permis de confirmer la formation du complexe (Figure 5B de la publication numéro 2). De plus, la co-expression d'un fragment de Spt6 de levure où le motif IF est muté abolit également la formation du complexe. Les mutations introduites dans les protéines d'*E. cuniculi* semblent avoir le même effet *in vitro* sur les protéines de levure et peuvent donc être utilisées pour tenter d'empêcher la formation du complexe *in vivo*. Ainsi, les mutations mises en évidence sur la protéine Spt6 de *E. cuniculi* ont été introduites par l'équipe de Fred Winston dans un vecteur codant le gène de Spt6 de levure. Ces vecteurs ont été utilisés pour transformer des levures dans lesquelles l'expression du gène de Spt6 endogène a été supprimée. Dans la majorité des cas, les mutations conduisent à l'apparition de phénotypes sévères, tels que la thermosensibilité, des défauts de transcription reflétés par l'apparition d'un phénotype de type Spt- et la sensibilité à la caféine et la phléomycine. Ainsi, la zone d'interaction de Spt6 avec la protéine Iws1 est bien nécessaire à la fonction de la protéine *in vivo*.

### 4.3. Publication numéro 2

Les données supplémentaires relatives à cette publication sont présentées en Annexe 8.4, page 272

# The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIIS, Elongin A and Med26

Marie-Laure Diebold<sup>1,4</sup>, Michael Koch<sup>2,4</sup>,  
Erin Loeliger<sup>3</sup>, Vincent Cura<sup>1</sup>,  
Fred Winston<sup>3</sup>, Jean Cavarelli<sup>1</sup>  
and Christophe Romier<sup>1,\*</sup>

<sup>1</sup>Département de Biologie et Génomique Structurales, IGBMC (Institut de Génétique et Biologie Moléculaire et Cellulaire), UDS, CNRS, INSERM, Illkirch Cedex, France, <sup>2</sup>Institut für Biochemie, Universität zu Köln, Köln, Germany, <sup>3</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA

**Binding of elongation factor Spt6 to Iws1 provides an effective means for coupling eukaryotic mRNA synthesis, chromatin remodelling and mRNA export. We show that an N-terminal region of Spt6 (Spt6N) is responsible for interaction with Iws1. The crystallographic structures of *Encephalitozoon cuniculi* Iws1 and the Iws1/Spt6N complex reveal two conserved binding subdomains in Iws1. The first subdomain (one HEAT repeat; HEAT subdomain) is a putative phosphoprotein-binding site most likely involved in an Spt6-independent function of Iws1. The second subdomain (two ARM repeats; ARM subdomain) specifically recognizes a bipartite N-terminal region of Spt6. Mutations that alter this region of Spt6 cause severe phenotypes *in vivo*. Importantly, the ARM subdomain of Iws1 is conserved in several transcription factors, including TFIIIS, Elongin A and Med26. We show that the homologous region in yeast TFIIIS enables this factor to interact with SAGA and the Mediator subunits Spt8 and Med13, suggesting the molecular basis for TFIIIS recruitment at promoters. Taken together, our results provide new structural information about the Iws1/Spt6 complex and reveal a novel interaction domain used for the formation of transcription networks.**

*The EMBO Journal* (2010) 29, 3979–3991. doi:10.1038/emboj.2010.272; Published online 5 November 2010  
*Subject Categories:* chromatin & transcription; structural biology

*Keywords:* crystallography; mRNA export; RNA polymerase II; transcription elongation; yeast genetics

## Introduction

Eukaryotic RNA polymerase II (RNAPII) requires the timely and concerted recruitment of many factors to enhance

\*Corresponding author. Département de Biologie et Génomique Structurales, IGBMC (Institut de Génétique et Biologie Moléculaire et Cellulaire), UDS, CNRS, INSERM, 1 rue Laurent Fries, B.P. 10142, Illkirch Cedex 67404, France. Tel.: +33 38 854 5798;

Fax: +33 38 865 3276; E-mail: romier@igbmc.fr

<sup>4</sup>These authors contributed equally to this work

Received: 10 September 2010; accepted: 18 October 2010; published online: 5 November 2010

pre-initiation complex (PIC) formation, promoter clearance, transcription elongation and to overcome the transcriptional barriers imposed by chromatin (Saunders *et al.*, 2006; Li *et al.*, 2007). Furthermore, recruitment by elongating RNAPII of factors involved in mRNA processing, mRNA surveillance and mRNA export enables a direct and tight coupling between transcription, mRNA maturation and mRNA export (Perales and Bentley, 2009).

The Iws1/Spt6 complex participates in this coupling by acting in transcription elongation, chromatin remodelling and mRNA export. Spt6 is a putative histone chaperone that interacts directly with histone H3, and appears to promote nucleosome reassembly at core promoters and within the body of genes in the wake of RNAPII (Bortvin and Winston, 1996; Winkler *et al.*, 2000; Kaplan *et al.*, 2003; Adkins and Tyler, 2006). Spt6 is also an elongation factor that enhances elongation rates of RNAPII both *in vitro* and *in vivo* (Endoh *et al.*, 2004; Yoh *et al.*, 2007; Ardehali *et al.*, 2009). Yet, the role of Spt6 in elongation cannot be fully recapitulated by its histone chaperone activity as elongation enhancement also occurs on naked DNA templates (Endoh *et al.*, 2004; Yoh *et al.*, 2007). A recent study has shown that both Spt6 and Iws1 are associated genome-wide across transcribed regions (Mayer *et al.*, 2010).

Interaction between Spt6 and the Iws1/Spn1 protein has been reported both in yeast and in mammals (Krogan *et al.*, 2002; Lindstrom *et al.*, 2003; Yoh *et al.*, 2007). The involvement of Spt6 in mRNA export requires its association with Iws1, as Iws1 interacts directly with the mRNA export factor REF1/Aly (Yoh *et al.*, 2007). Furthermore, either depletion of Iws1 or a mutation within the SH2 domain of Spt6, which impairs the ability of the Iws1/Spt6 complex to interact with elongating RNAPII, results in splicing defects and nuclear retention of bulk poly(A)<sup>+</sup> mRNAs in mammalian cells (Yoh *et al.*, 2007). Mammalian Iws1 is also required for the recruitment of the lysine methyltransferase, HYPB/Setd2, which trimethylates H3K36 across transcribed regions (Yoh *et al.*, 2008).

Iws1 also functions independently of Spt6 by binding constitutively to the postassembly regulated *CYC1* gene, thus preventing the recruitment of the Swi/Snf complex and repressing transcription of this gene (Fischbeck *et al.*, 2002; Zhang *et al.*, 2008). A single amino-acid change in yeast Iws1 (K192N) prevents its recruitment to the *CYC1* gene, leading to the constitutive recruitment of Swi/Snf and increased transcription, even in non-inducing conditions. The direct interaction of Iws1 with the Ser5-phosphorylated RNAPII has been suggested for the constitutive recruitment of Iws1 (Zhang *et al.*, 2008).

Despite the wealth of data on Spt6 and Iws1, their functional roles remain largely unknown at the molecular level. So far, only the structures of the Spt6 tandem SH2 domains (Diebold *et al.*, 2010; Sun *et al.*, 2010) and of the putative

bacterial ancestor of Spt6, Tex (Johnson *et al*, 2008), have shed some light on the structural organization of Spt6. As for Iws1, part of the conserved region of Iws1 is homologous in sequence to the N-terminal domains of the elongation factors TFIIS and Elongin A, and of the co-activator Med26 (Wery *et al*, 2004; Ling *et al*, 2006). In TFIIS, this domain enables TFIIS recruitment to gene promoters by the co-activators SAGA and Mediator, in which it favours the formation of active PICs (Pan *et al*, 1997; Wery *et al*, 2004; Prather *et al*, 2005; Guglielmi *et al*, 2007; Kim *et al*, 2007). Yet, the NMR structures of the yeast and mouse TFIIS N-terminal domains adopt totally different folds (PDB code 1EO0; Booth *et al*, 2000; PDB code 1WJT, RIKEN Structural Genomics/Proteomics Initiative), precluding our understanding of the role of this domain.

Here, we present the biochemical, crystallographic and functional characterization of Iws1 and the Iws1/Spt6 complex from *Encephalitozoon cuniculi* and *Saccharomyces cerevisiae*. We show that Iws1 possesses two binding subdomains (HEAT and ARM subdomains) defined by the N- and C-termini of its conserved region. Another study has also recently identified these domains from the crystallographic analysis of *S. cerevisiae* Iws1 (Pujari *et al*, 2010). The HEAT subdomain has determinants for recognizing large negatively charged ions or small molecules such as phosphates, suggesting that it may serve to recognize phosphoproteins. This subdomain is specific to Iws1 and is most likely responsible for the Spt6-independent function of Iws1. The ARM subdomain of Iws1 specifically recognizes an N-terminal region of Spt6 (Spt6N). In yeast, we show that mutations that alter this region of Spt6 cause severe phenotypes, suggesting that the interaction between Spt6 and Iws1 is critical *in vivo*.

Importantly, the ARM subdomain of Iws1 encompasses the region of sequence homology with TFIIS, Elongin A and Med26. We show that the homologous subdomain from yeast TFIIS is able to form complexes with the Spt8 and Med13 subunits of SAGA and the Mediator, suggesting a molecular basis for TFIIS recruitment at promoters. These complexes combine interactions that are observed within the Iws1/Spt6N complex, along with interactions that are specific to the TFIIS/Spt8 and TFIIS/Med13 complexes. Taken together, our results provide a structural characterization of the Iws1/Spt6 complex and highlight a specific interaction domain shared by other transcriptional effectors.

## Results

### **A small N-terminal region of Spt6 is sufficient to interact with Iws1**

Yeast (*S. cerevisiae*; *sc*) and human (*Homo sapiens*; *hs*) Spt6 proteins are very large polypeptides that are not easily amenable to biochemical and structural studies. To overcome this problem, we use the fungi-related intracellular parasite, *E. cuniculi* (*ec*), as a model organism as its proteins are generally shorter than their eukaryotic orthologues (Katinka *et al*, 2001; Romier *et al*, 2007; Diebold *et al*, 2010). Spt6 and Iws1 are present in *E. cuniculi* and show strong sequence conservation with their yeast and human orthologues despite their shorter length (Figure 1A; Supplementary Figure 1).

Our experiments show that full-length *E. cuniculi* Spt6 and Iws1 form a complex when co-expressed in *Escherichia coli*. Singly expressed *ec*Spt6, N-terminally tagged with a

poly-histidine sequence, is retained on a cobalt affinity resin (Figure 1B, lane 1). In contrast, singly expressed, but untagged *ec*Iws1 does not bind to the same resin, showing that no non-specific binding occurs (Figure 1B, lane 2). However, upon co-expression of his-tagged *ec*Spt6 with untagged *ec*Iws1, both proteins are retained on the resin, revealing the formation of a complex between these two proteins (Figure 1B, lane 3).

Additional co-expression experiments defined regions in each protein sufficient for their interaction. First, the conserved region of Iws1 (*ec*Iws1 residues 55–198), which is sufficient for the essential functions of Iws1 in yeast (Fischbeck *et al*, 2002), is also sufficient for binding to *ec*Spt6 (Figure 1B, lane 5). In the case of *ec*Spt6, as observed for its mammalian homologue (Yoh *et al*, 2007), the N-terminal region (*ec*Spt6 residues 1–71; *ec*Spt6N<sub>1–71</sub>), but not the remaining C-terminal region (*ec*Spt6 residues 71–894), is involved in Iws1 binding (Figure 1B, compare lanes 6 and 8).

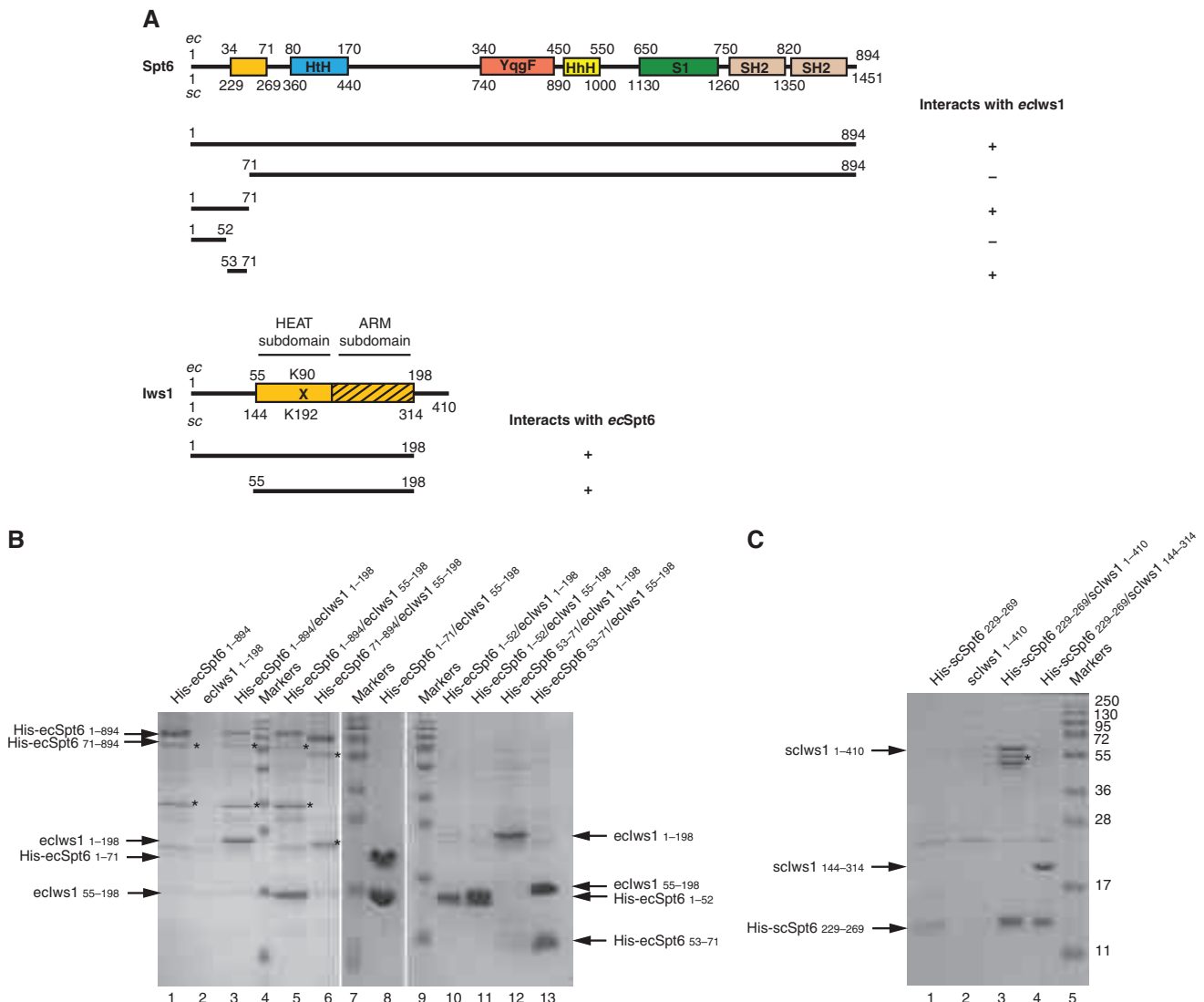
Multiple alignment of Spt6 sequences shows that its N-terminal region contains a small conserved region (*ec*Spt6 residues 53–71; *ec*Spt6N<sub>53–71</sub>; Supplementary Figure 1A). Co-expression experiments showed that this small conserved region of Spt6 is sufficient for binding *ec*Iws1 (Figure 1B, lanes 10–13). These results were confirmed by looking for a similar interaction between the yeast proteins: an *sc*Spt6N<sub>229–269</sub> construct, similar to *ec*Spt6N<sub>53–71</sub>, was able to form a complex with either full-length *sc*Iws1 or its conserved region (residues 144–314) (Figure 1C).

### **The conserved region of Iws1 is formed by HEAT and ARM repeats**

To characterize precisely the interaction between Spt6 and Iws1, we solved the X-ray crystallographic structures of these proteins and their complex from *E. cuniculi*. Owing to rapid proteolytic N-terminal degradations occurring during purification, only the stable *ec*Iws1<sub>55–198</sub> construct was considered and could be readily crystallized on its own in different space groups. For Spt6, due to yield and C-terminal proteolytic degradation problems with the full-length protein, only N-terminal constructs (encoding *ec*Spt6N residues 1–71, 34–71 and 53–71) were used in complex with *ec*Iws1<sub>55–198</sub>. Three different crystal forms were obtained: one form with the *ec*Spt6N<sub>34–71</sub> construct and two forms with the *ec*Spt6N<sub>53–71</sub> construct. All crystal forms were unrelated, providing an unbiased view of the interaction between Spt6N and Iws1.

The structure of *ec*Iws1<sub>55–198</sub> was solved by multiple anomalous dispersion using selenomethionine derivatives. The model was refined against a 2.25 Å resolution native data set to an R-factor of 20.4% and an R-free of 26.5% (Supplementary Table I). The three structures of the complexes were then solved by molecular replacement using the structure of free Iws1. The Spt6N regions were built into the additional density and the structures of the *ec*Spt6N<sub>53–71</sub>/*ec*Iws1<sub>55–198</sub> (forms 1 and 2) and *ec*Spt6N<sub>34–71</sub>/*ec*Iws1<sub>55–198</sub> complexes were refined to 1.95, 2.10 and 1.75 Å resolution to R-factors of 21.0, 20.0 and 19.4% and R-frees of 23.6, 24.7 and 23.6%, respectively (Supplementary Table II). For Spt6N and Iws1, almost all residues could be seen in density, apart from a few N- and/or C-terminal residues of both proteins. A single exception was observed for the longer *ec*Spt6N<sub>34–71</sub>





**Figure 1** A small N-terminal region of Spt6 is sufficient to retain the full length or the conserved domain of Iws1. **(A)** Schematic view of the putative domain architecture of Spt6 and Iws1. The domains of Spt6 and Iws1 characterized in this study are shown as light orange boxes. This domain in Iws1 is composed of two subdomains (HEAT subdomain and ARM subdomain) and corresponds to the conserved region of Iws1 sufficient for Iws1 function in yeast (Fischbeck *et al*, 2002). The Iws1 invariant lysine (*scK192/ecK90*) involved in the Spt6-independent function of Iws1 is marked by an 'X'. The Iws1 region homologous to TFIIS, Elongin A and Med26 N-terminal domains is hatched. The domains (HhH, YqgF, HhH and S1) that have been putatively assigned to the Spt6 core domain based on the structure of the bacterial Tex protein (Johnson *et al*, 2008) are shown. The two SH2 domains from the tandem SH2 domains at the C-terminus of Spt6 (Diebold *et al*, 2010; Sun *et al*, 2010) are indicated. The results of co-expression experiments shown in **(B)** are summarized below the proteins. *E. cucurbit* (*ec*) and *S. cerevisiae* (*sc*) numbering are shown. **(B)** Deciphering of *E. cucurbit* Iws1/Spt6 complex formation upon (co-) expression in *E. coli* of various constructs of both proteins and purification by affinity chromatography. All samples are analysed on SDS-PAGE. The fainter band for His-ecSpt6N<sub>53-71</sub> in lane 12 compared with lane 13 is due to the lower amount of soluble complex obtained. Construct boundaries are indicated. Spt6 degradation products are marked with an '\*'. **(C)** Characterization of the *S. cerevisiae* Iws1/Spt6N interaction based on the data obtained with the *E. cucurbit* proteins. The faint band for His-scSpt6N<sub>229-269</sub> in lane 1 is most likely due to the poor solubility of this construct when expressed alone. Iws1 degradation products are marked with an '\*'. Molecular weights are shown and are the same throughout the figures.

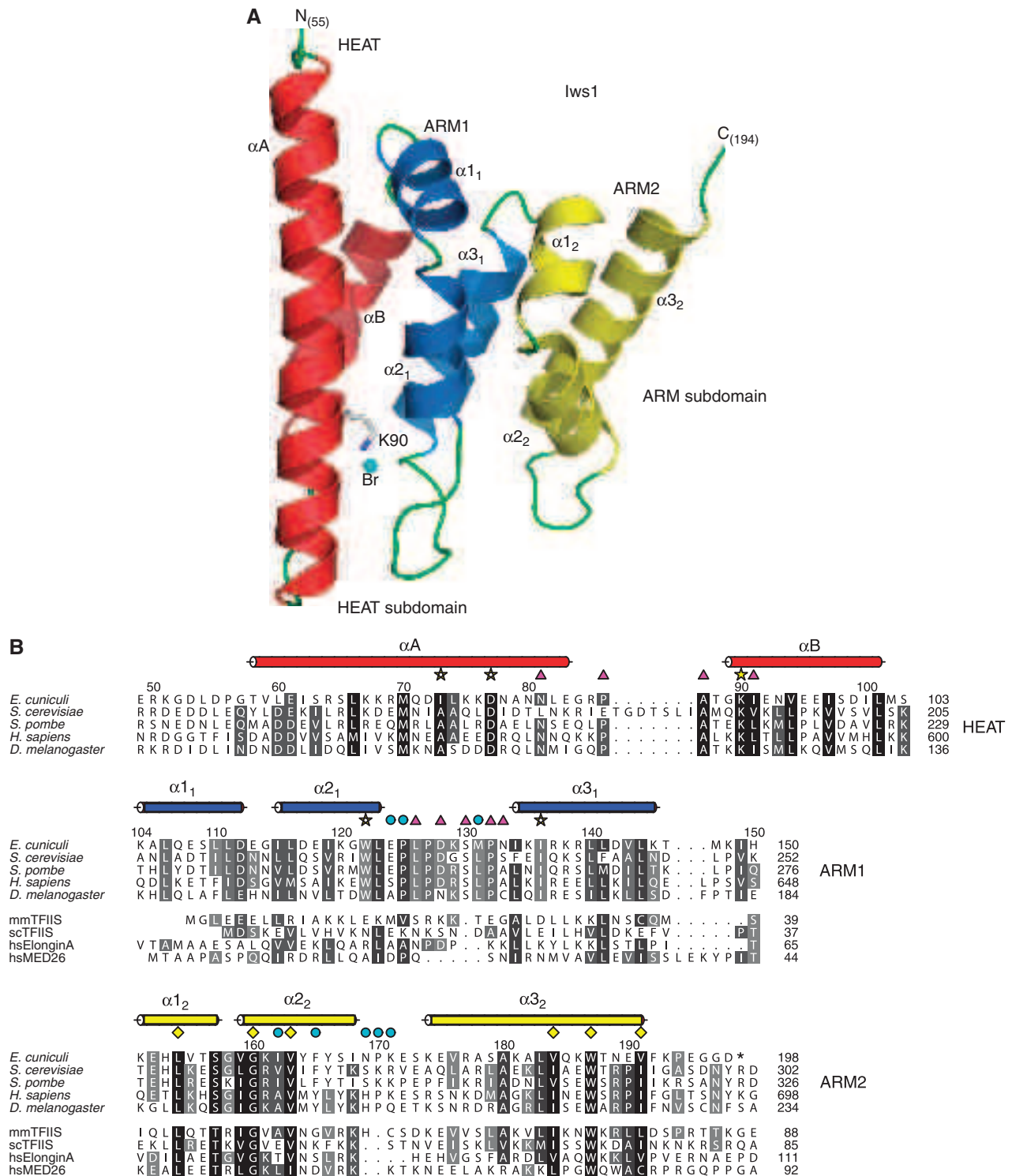
construct in which the first 11 N-terminal residues could not be seen.

Structural analysis of *ecIws1*<sub>55-198</sub> in the different crystal forms reveals that the Iws1 conserved region is formed by a single HEAT repeat followed by two ARM repeats (ARM1 and ARM2) (Figure 2). HEAT and ARM motifs are related protein/protein interaction modules that are often found repeated several times in many proteins (Andrade *et al*, 2001). The HEAT motif is composed of an  $\alpha$ -helical hairpin formed by two  $\alpha$ -helices ( $\alpha$ A and  $\alpha$ B), whereas the ARM motif is composed of a short  $\alpha$ -helix ( $\alpha$ 1), which is almost perpendi-

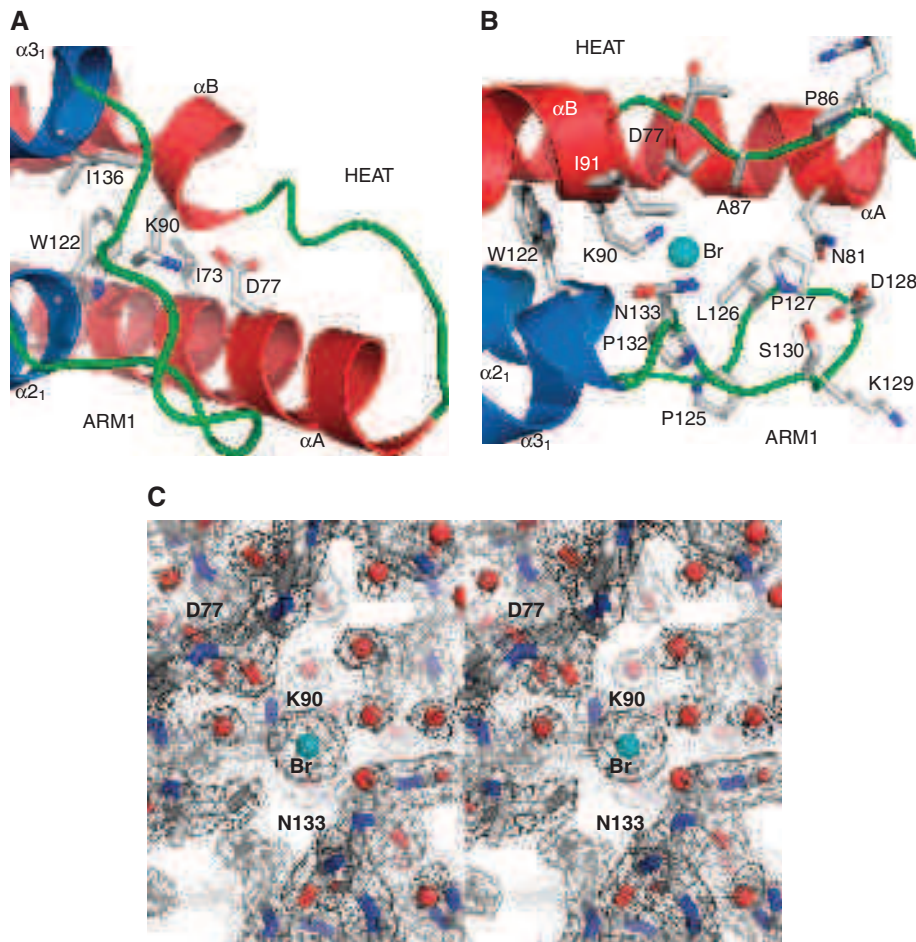
cular to a helical hairpin formed by two  $\alpha$ -helices ( $\alpha$ 2 and  $\alpha$ 3). Only minor structural differences are observed for Iws1 between all crystal forms, implying that binding of Spt6N does not induce large conformational changes.

**A conserved binding subdomain (HEAT subdomain) formed by the Iws1 HEAT repeat and part of the ARM1 repeat contains the invariant *ecK90/scK192***

In yeast Iws1, mutation of the invariant lysine K192 to asparagine leads to the rescue of TBP mutants that possess postrecruitment defects, and prevents constitutive



**Figure 2** Structure of *E. cucuruli* lws1. (A) Ribbon representation of *ec*lws1<sub>55–198</sub> crystal structure.  $\alpha$ -helices of the HEAT, ARM1 and ARM2 motifs are coloured red, blue and yellow, respectively. Loops are coloured green. The two subdomains of lws1 are indicated. The invariant K90 is displayed as sticks and coloured according to atom type. The anion bound to K90  $Na^+$  is shown as a cyan sphere. The first and last residues observed in the density are labelled. This colour scheme is used throughout the figures unless otherwise stated. The ribbon figures have been made with PYMOL (version 0.99; DeLano Scientific). (B) Multiple sequence alignment of the conserved region of lws1 (top five rows) with the TFIIS, Elongin A and Med26 N-terminal domains (bottom three rows; not shown for the HEAT repeat, which is lws1 specific). mm, *Mus musculus*; sc, *S. cerevisiae*; hs, *H. sapiens*. Sequence similarities are indicated by shading. Observed  $\alpha$ -helices in the *ec*lws1<sub>55–198</sub> structure are shown above the sequences as cylinders coloured as in (A). Numbering above the sequences correspond to *E. cucuruli*, whereas the numbering at the end of each row relates to the different organisms. K90 and the residues involved in its packing at the interface of the HEAT and ARM1 repeats are labelled with yellow stars. Residues forming the highly conserved putative phosphate-binding domain of lws1, together with K90, are labelled by magenta triangles. Residues involved in Spt6 IR1 and IR2 binding are labelled with yellow diamonds and cyan circles, respectively. Alignment features are identical in all figures unless otherwise stated. Alignments were created with ALINE (Bond and Schuttelkopf, 2009).



**Figure 3** The HEAT-subdomain Iws1. (A) Close-up view of K90 interactions at the interface of Iws1 HEAT and ARM1 repeats. The side chains involved in hydrophobic interactions and hydrogen bonding are shown. (B) Close-up view of the HEAT subdomain formed by Iws1 HEAT and ARM1 repeats. Residues forming the surface are displayed. The bromide anion (Br) is shown as a cyan sphere. (C) Stereo view of the structure at the bromide-binding site. The bromide (Br) and water molecules are shown as cyan and red spheres, respectively. The 2Fo-Fc electron density map of the refined structure is shown and contoured at 1.5  $\sigma$ .

recruitment of Iws1 to the *CYCl* gene promoter in an Spt6-independent manner (Fischbeck *et al*, 2002; Zhang *et al*, 2008). The equivalent lysine of *ecIws1*, K90, is found at the start of helix  $\alpha B$  in the HEAT repeat (Figures 2A and 3A). The K90 side chain is found at the interface of the tips of both HEAT and ARM1 repeats. The aliphatic part of the K90 side chain is at the centre of a hydrophobic core formed by the side chains of HEAT  $\alpha A$  I73, ARM1  $\alpha 2_1$  W122 and ARM1  $\alpha 3_1$  I136. Furthermore, K90 N $\epsilon$  is fixed through a hydrogen bond with the carboxylate of invariant HEAT  $\alpha A$  D77 that also hydrogen bonds with the K90 main chain amide (Figure 3A).

Importantly, the hydrogen bond formed between the carboxylate of D77 and the K90 N $\epsilon$  positions the latter atom precisely at the bottom of a pocket formed by both tips of the HEAT and ARM1 repeats, making it solvent accessible. Strikingly, most of the constituent residues of this pocket are highly conserved (Figures 2B and 3B). Inspection of this pocket reveals the direct binding of an ion to K90 N $\epsilon$ , which is further coordinated by the N133 main chain amide as well as one or two water molecules. In most structures, the bound ion is most likely a chloride ion provided by the protein buffer. In one structure, however, a much larger electron

density was observed reflecting the replacement of the chloride ion by a negatively charged bromide ion provided by the crystallization conditions (Figure 3B and C). Interestingly, this ion-binding pocket would be large enough to accommodate a phosphate at the position of the bromide. Furthermore, the atoms coordinating the bromide ion (K90 N $\epsilon$ , N133 main chain amide and two water molecules) would be perfectly placed to coordinate an incoming phosphate. As one of the coordinating water molecules is located at the entrance of the pocket, it could also be replaced by the oxygen of a phosphorylated residue. Taken together, these results suggest that this conserved binding domain of Iws1 could be specific for phosphoproteins.

To further investigate the role of K90, four mutants were created: K90R, K90A, K90D and K90N. All mutants could be expressed and purified. Despite a two- to four-fold decrease in yield for most of the mutants, none showed an increased tendency to aggregate as assessed by dynamic light scattering analyses. However, temperature-dependent unfolding of these proteins using the ThermoFluor technology showed that the K90D and K90N mutants have greater instability (Supplementary Figure 2A). We next tested the effect of these mutants on the formation of the complex with Spt6N. Upon

co-expression of the mutants with *ecSpt6N*<sub>53–71</sub>, no complex was observed with the K90D mutant and the level of complex formation appeared to be reduced with the K90N mutant (Supplementary Figure 2B, lanes 1–5). Surprisingly, none of the mutants impaired the binding of the longer *ecSpt6N*<sub>1–71</sub> construct (Supplementary Figure 2B, lanes 7–11). Taken together, our results suggest that the effects of the K90 mutants are local and suggest a more extensive interface between Iws1 and Spt6N.

### **Bipartite binding of Spt6N to Iws1 ARM repeats (ARM subdomain)**

Analysis of the structures of the Iws1/Spt6N complex confirms the hypothesis of extended Iws1–Spt6N interactions, showing that the region of *ecSpt6N* interacting with Iws1 encompasses residues 45–67, forming contacts with both ARM repeats, but not with the HEAT repeat (Figure 4A and B). This region is composed of an N-terminal  $\alpha$ -helix ( $\alpha$ N) that interacts with Iws1 ARM2, pointing towards the ARM2  $\alpha$ <sub>2</sub> helix.  $\alpha$ N stops abruptly at the invariant glycine 58, which enables the peptide to go round the Iws1  $\alpha$ <sub>2</sub> helix. The C-terminal part of the Spt6N peptide adopts a rather extended conformation followed by a short-helical turn ( $\alpha$ C) and interacts with both ARM repeats (Figure 4B).

As such, the interaction region (IR) of Spt6N with Iws1 can be divided into two sub-regions that we termed IR1 (Spt6N  $\alpha$ N up to the invariant F57/G58 motif) and IR2 (C-terminal to G58) (Figure 4A). The interaction between the Spt6N IR1 region and Iws1 is dictated mostly by hydrophobic interactions. The Spt6N  $\alpha$ N helix packs against the C-terminus of Iws1 ARM2  $\alpha$ <sub>3</sub><sub>2</sub> helix (Figure 4C). This interaction is essentially hydrophobic and involves Iws1 V191 as well as several residues at the N-terminus of Spt6N  $\alpha$ N. These latter residues are solvent exposed and their contribution to the hydrophobic core is mediated by their main chain C $\beta$  and/or C $\gamma$  atoms. This most likely explains why the sequence of this helix is not evolutionarily conserved as only its  $\alpha$ -helical propensity appears important. In agreement, secondary structure predictions of *S. cerevisiae* and *H. sapiens* Spt6N indicate an  $\alpha$ -helix precisely at the position of *ecSpt6N*  $\alpha$ N, ending at the invariant FG motif.

At the C-terminal end of Spt6N  $\alpha$ N, I56 and F57 interact with a hydrophobic cavity of Iws1 formed by residues L154, G160, V163, V184, W187 and V191 of the ARM2 repeat (Figure 4D). The perfect conservation of the hydrophobic character of the I56/F57 motif, as well as the ability of the *ecSpt6N*<sub>53–71</sub> construct to retain its interaction with Iws1 despite the large truncation of the  $\alpha$ N helix, demonstrates the importance of the binding of the I56/F57 motif to the hydrophobic cavity of Iws1.

In contrast, the interaction between the Spt6N IR2 region and Iws1 is composed of both hydrophobic and hydrogen bond contacts. Following the I56/F57 motif, Spt6N G58 has the function of helix breaker, necessary to prevent steric clashes between Iws1 and an elongated  $\alpha$ N helix of Spt6N. The four residues following G58 do not make extensive contacts with Iws1, but serve as a linker for the peptide to go around the  $\alpha$ <sub>2</sub> helix of ARM2. The next interaction between both proteins is mediated through the hydrogen bond formed between the hydroxyl of Spt6 Y63 and the carboxylate of Iws1 E124 (Figure 4E). This interaction is reinforced by hydrophobic contacts made between the Y63

side chain and Iws1 I162 and F165. Other interactions involve Spt6N tyrosine Y65 whose hydroxyl forms a water-mediated interaction with Iws1 L126 carboxyl, and Spt6 V66 and L67, which form hydrophobic contacts with residues of Iws1 ARM1 and ARM2 repeats.

Interestingly, both the ion-binding HEAT subdomain and the Spt6N-binding ARM subdomain of Iws1 are connected by a positively charged channel (Figure 4F). Specifically, the K90-binding pocket and the Spt6 IR2-binding surface of Iws1 are relatively close, suggesting that binding of a protein to the former pocket could prevent binding of Spt6 to the latter surface, or vice versa.

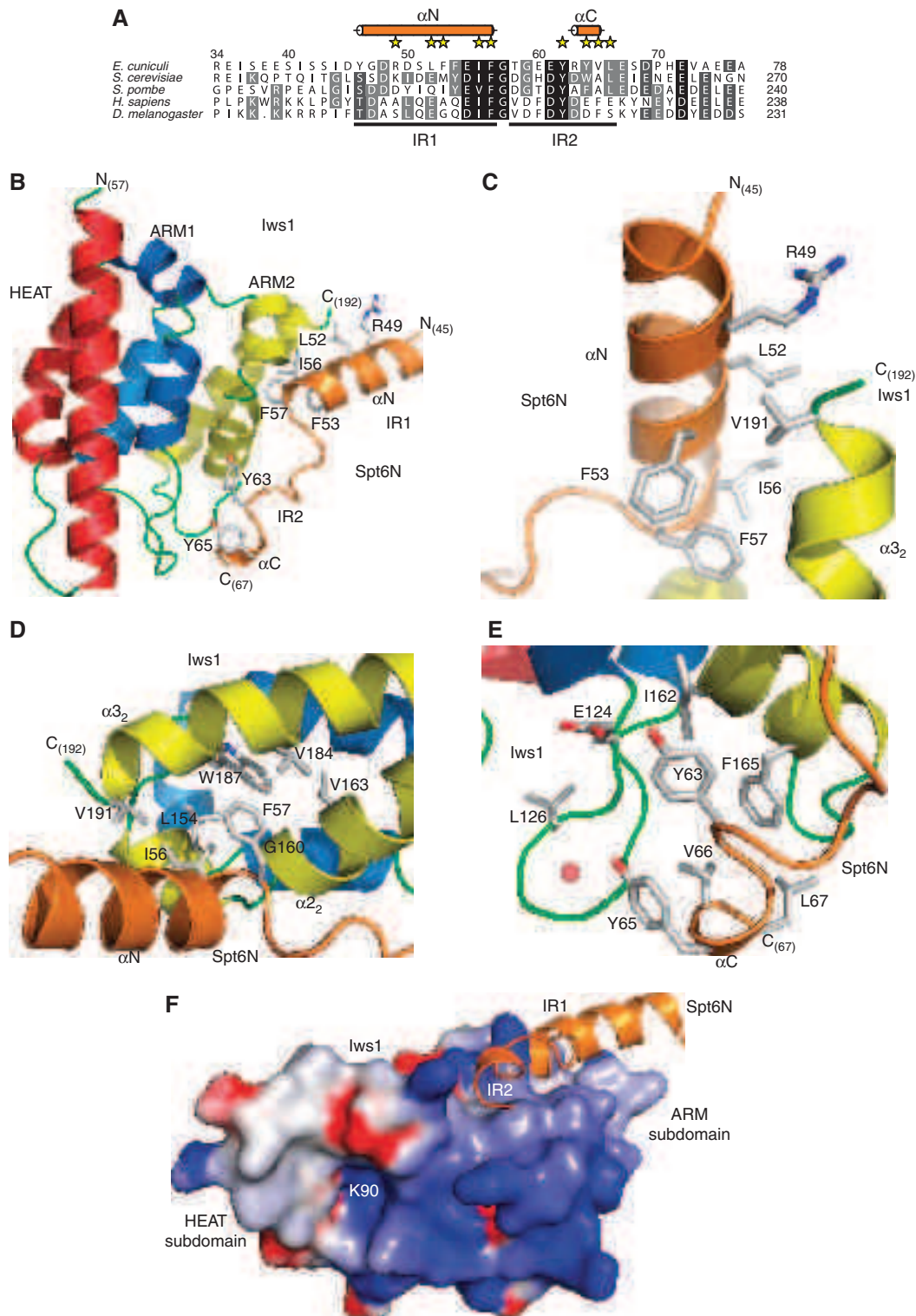
### **Point mutations within Spt6 Iws1-binding region cause Spt<sup>-</sup> and Ts<sup>-</sup> phenotypes *in vivo***

We next investigated the interaction between Spt6 and Iws1 by mutational analysis, using the *ecSpt6N*<sub>1–71</sub> construct, which encompasses the full region of interaction of *ecSpt6N* with Iws1. Based on our structural data, single and double mutants of *ecSpt6N* and *ecIws1* were made. At the Spt6N IR1/Iws1 interface, Iws1 residues G160 and V191 were changed to bulkier residues (single mutants G160Y, V191Y and V191W), whereas Spt6N residues I56 and F57 were changed to smaller alanine residues (double mutant I56A/F57A). At the IR1/IR2 transition, both Spt6N G58 and G60 were changed to alanines (double mutant G58A/G60A). At the Spt6N IR2/Iws1 interface, Iws1 E124 was changed to either serine or alanine (E124S and E124A), whereas Spt6N Y63 and Y65 were changed to alanines (single mutant Y63A and double mutant Y63A/Y65A).

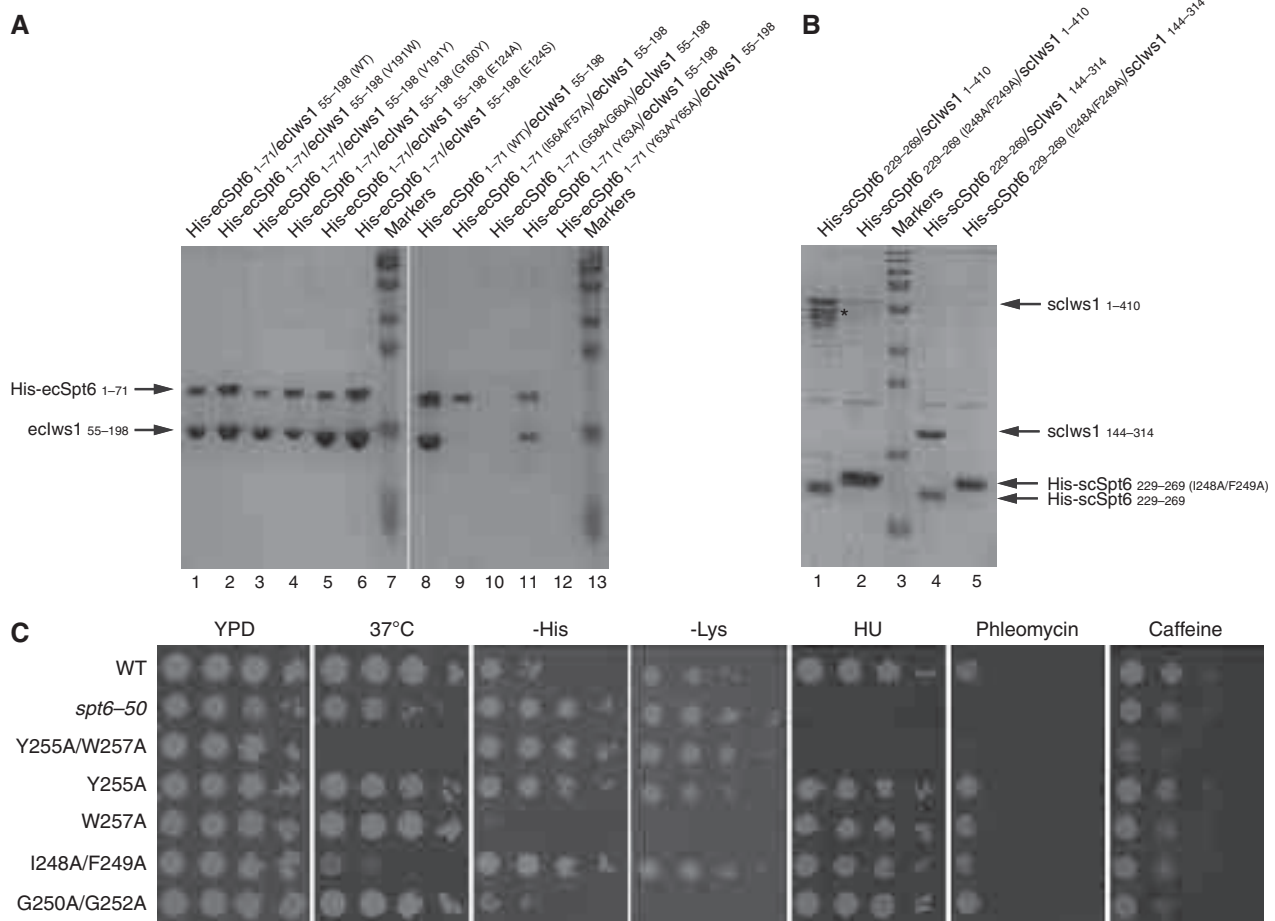
Upon co-expression, none of the Iws1 mutants affected complex formation, showing that single mutations in Iws1 are not sufficient to destroy the complex with Spt6N (Figure 5A, lanes 1–6). In contrast, all Spt6N mutants, with the exception of the Y63A mutant, led to the loss of the complex (Figure 5A, lanes 8–12). Some of these Spt6N mutants appeared insoluble (Figure 5A, lanes 10 and 12), suggesting that the loss of the complex may be due to insolubility. However, the I56A/F57A mutant was still soluble, showing that, in this case, the loss of interaction with Iws1 was due to the mutations. In agreement, the same mutation introduced in the yeast Spt6N<sub>229–269</sub> construct (I248A/F249A) leads to a drastic loss of interaction with *scIws1* (Figure 5B).

As the Spt6N mutants appeared to have the strongest effect on the formation of the Iws1/Spt6N complex, we analysed the effect of these mutations in *S. cerevisiae in vivo*, comparing them with wild type and to a previously studied mutant, *spt6–50*. Initially, three mutants were constructed for the full-length yeast Spt6 protein: I248A/F249A, G250A/G252A and Y255A/W257A (equivalent to the I56A/F57A, G58A/G60A and Y63A/Y65A mutants of *ecSpt6*, respectively). When these mutant forms of full-length Spt6 were expressed *in vivo* as the only form of the protein, all of the yeast strains were viable, showing that all of the mutants still possess some Spt6 function (Figure 5C). The G250A/G252A mutant displayed a wild-type phenotype (Figure 5C), suggesting that either this form of Spt6 is functional *in vivo* or that it is redundant with some other function.

In contrast to the G250A/G252A mutant, the strains harbouring the I248A/F249A and Y255A/W257A mutants displayed several mutant phenotypes. Both mutants showed a



**Figure 4** Crystal structure of *E. cuniculi* *lws1*/Spt6N complex. **(A)** Multiple sequence alignment of the *lws1*-binding region of Spt6N. Both Spt6N IR1 and IR2 sub-regions are indicated below the sequences. Residues whose side chains are involved in *lws1* binding are labelled with yellow stars. Spt6N  $\alpha$ -helices observed in the *lws1*/Spt6N structures are shown as orange cylinders. **(B)** Ribbon representation of the Spt6N<sub>34–71</sub>/*lws1*<sub>55–198</sub> structure. Spt6N is coloured orange. Most Spt6N side chains interacting with *lws1* are shown. **(C)** Close-up view of the interaction between Spt6N  $\alpha$ N and *lws1*  $\alpha$ 3<sub>2</sub> helices. **(D)** Close-up view of Spt6N IF motif binding to *lws1* hydrophobic cavity. **(E)** Close-up view of Spt6N IR2 binding to *lws1*. The red sphere represents a water molecule. **(F)** GRASP (Nicholls *et al*, 1991) representation of the electrostatic potential at the surface of *E. cuniculi* *lws1*. The electrostatic potentials  $-8$  and  $+8$   $k_B T$  ( $k_B$ , Boltzmann constant;  $T$ , temperature) are coloured red and blue, respectively. The Spt6N region binding to *lws1* is shown as orange ribbon. K90 N $\epsilon$  is located within a cavity and is labelled (K90).



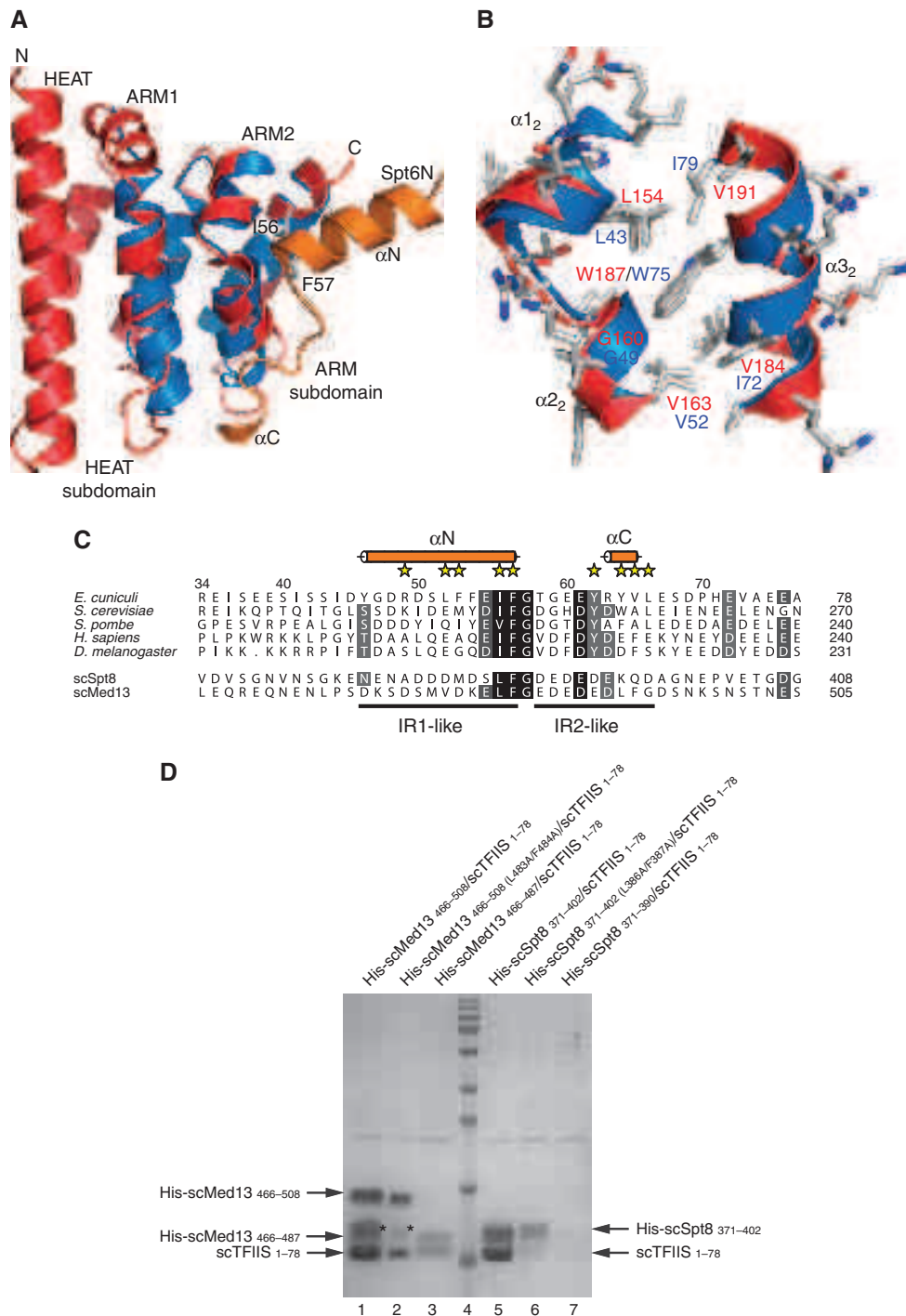
**Figure 5** Mutational and *in vivo* analysis of the Iws1/Spt6 complex. **(A)** Analysis of the *E. cucurbiti* Iws1/Spt6N complex by mutation of Iws1- and Spt6N-specific residues involved in complex formation. Stronger effects are observed by mutating the Spt6N protein. **(B)** Analysis of the full-length *S. cerevisiae* Iws1/Spt6 complex by mutation of the IF motif, confirming the result observed for the *E. cucurbiti* Iws1/Spt6N complex. The mutant migrates more slowly than the wild-type construct. Degradation products are marked by an \*. **(C)** Analysis of Spt6 mutants within the Iws1-binding region of full-length *S. cerevisiae* Spt6. Strains were grown to saturation in YPD, serially diluted 10-fold and spotted on the indicated media. Included for comparison was *spt6-50*, a previously characterized mutant.

strong temperature-sensitive phenotype at 37°C (Figure 5C). Furthermore, these two mutants caused a marked Spt<sup>-</sup> phenotype (suppression of auxotrophies caused by the insertion mutations *his4-912δ* and *lys2-128δ*), indicative of transcription defects. The Y255A/W257A mutant had the broadest range of phenotypes, as it was sensitive to both hydroxyurea and phleomycin, and modestly sensitive to caffeine (Figure 5C). To determine whether the phenotypes of Y255A/W257A required both amino-acid changes, the single mutants were constructed and tested for the same set of phenotypes. Compared with the double mutant, the effects of the Y255A and W257A single mutants were less drastic. Neither of them showed sensitivity to high temperature, hydroxyurea, phleomycin or caffeine. Although the Y255A mutant shows a partial Spt<sup>-</sup> phenotype (His<sup>+</sup> phenotype), the W257A mutant displays a tighter His<sup>-</sup> phenotype than even the wild-type strain (Figure 5C). These results are in agreement with our co-expression tests in which we showed that the *ecSpt6N* Y63A mutant had less effect on the interaction with Iws1 than the Y63A/Y65A mutant. Taken together, these results confirm the importance of the Iws1-interacting region of Spt6 for the function of this protein *in vivo*.

### The IR1-binding region of Iws1 is structurally conserved in TFIIS

Part of the conserved region of Iws1 has been shown to share sequence similarity with the N-terminal domains of TFIIS, Elongin A and Med26 (Wery *et al*, 2004; Ling *et al*, 2006; Figure 2B). Our structural data reveals that these domains correspond to the ARM subdomain of Iws1, with the HEAT subdomain of Iws1 being specific to this protein. We have tentatively superimposed our Iws1 structure on the mouse and yeast TFIIS Domain I structures (PDB code 1WJT, RIKEN Structural Genomics/Proteomics Initiative; and PDB code 1E00; Booth *et al*, 2000). We observed a very good fit between *ecIws1* and the mouse N-terminal domain of TFIIS (Figure 6A). Specifically, the upper parts of the ARM repeats superimpose extremely well, whereas the lower parts, notably the loop regions between helices  $\alpha 2$  and  $\alpha 3$  of the repeats, show much more structural divergence.

Surprisingly, no satisfying fit was found with the yeast protein (PDB code 1E00). Careful inspection of the yeast TFIIS Domain I structure revealed the presence of several charged side chains in hydrogen bonding distance to hydrophobic side chains. To further investigate this discrepancy, we



**Figure 6** Specific interaction of TFIIIS N-terminal domain with Spt8 and Med13. **(A)** Superposition of *E. cuculici* Iws1 (red) and mouse TFIIIS Domain I (blue) structures. For comparison, the position of Spt6N (orange) bound to Iws1 is shown. **(B)** Superposition of the hydrophobic (IF-binding) pockets of *E. cuculici* Iws1 (red) and mouse TFIIIS domain I (blue). All side chains are shown. **(C)** Multiple sequence alignment of the Iws1-binding region of Spt6N (top five rows) and of the putative TFIIIS-binding region of *S. cerevisiae* (sc) Spt8 and Med13 (bottom two rows). **(D)** Reconstitution by co-expression of the complexes formed between *S. cerevisiae* TFIIIS (1–78) and Med13 (466–508), and *S. cerevisiae* TFIIIS (1–78) and Spt8 (371–402). Importance of the IF motifs and acidic regions on complex formation is addressed by using specific mutants. Degradation products of Med13 are labelled with an ‘\*’.

have established a model of the yeast TFIIIS Domain I using the mouse structure and the alignment provided in Figure 2B (model available upon request). Inspection of the model reveals that the yeast TFIIIS Domain I can perfectly adopt the fold observed for the mouse protein. Specifically, no steric clashes or any unconventional structural features are

observed, suggesting that the current yeast TFIIIS Domain I structure is most likely incorrect.

The most striking conservation between Iws1 and mouse TFIIIS Domain I concerns the hydrophobic pocket binding the I56/F57 motif of Spt6 (Figure 6B). At the sequence level, this is reflected by the almost perfect conservation between the

residues forming the pocket, with only a few conservative mutations. Therefore, both at the sequence and structural levels, the Spt6 IR1-binding region of Iws1 is almost completely conserved in TFIIS. In contrast, the IR2-binding region of Iws1 is different between these two proteins. Sequence analysis shows that the same pattern of conservation/divergence occurs in Elongin A and Med26 (Figure 2B).

TFIIS has been shown to be recruited to promoters in a SAGA- and Mediator-dependent manner (Prather *et al*, 2005; Kim *et al*, 2007) and two-hybrid analyses have shown that the yeast TFIIS N-terminal domain interacts with SAGA subunit Spt8 and Mediator subunit Med13 (Wery *et al*, 2004). In this latter study, the TFIIS IRs of Spt8 and Med13 have been delimited to 100 and 200 residues, respectively. Alignment of these two regions revealed a single patch of conservation composed of a Leucine-Phenylalanine-Glycine sequence (LFG motif) followed by acidic residues. Furthermore, despite poor sequence conservation, both sequences upstream of the LFG motif are predicted to be  $\alpha$ -helical. Therefore, these two proteins contain a region highly reminiscent of the Spt6 IR1 region (Figure 6C). In contrast, the IR2 region is not conserved in these proteins and is replaced in both proteins by a stretch of aspartate and glutamate residues.

We investigated the possibility of an interaction between Spt8 and Med13 with TFIIS by co-expressing the scSpt8<sub>371–402</sub> and scMed13<sub>466–508</sub> (IR1/IR2)-like motifs with the yeast TFIIS N-terminal domain (residues 1–78). Formation of a complex was observed in both cases (Figure 6D, lanes 1 and 5; Supplementary Figure 3). Interestingly, upon mutation of the LF motifs, the interaction with scTFIIS<sub>1–78</sub> was apparently weakened with scMed13<sub>466–508</sub> and abolished with scSpt8<sub>371–402</sub> (Figure 6D, lanes 2 and 6), suggesting that residues of scMed13<sub>466–508</sub>, which are not conserved in scSpt8<sub>371–402</sub>, strengthen the interaction. Furthermore, removal of the IR2-like regions of these two proteins by using shorter constructs encompassing only two acidic residues after the LFG motif (scSpt8<sub>371–390</sub> and scMed13<sub>466–487</sub>) abolishes interaction in the case of Spt8, but only weakens the interaction in the case of Med13 (Figure 6D, lanes 3 and 7), demonstrating the importance of the IR2-like region in Spt8 and Med13 and suggesting that N-terminal residues in Med13 interact more extensively with scTFIIS<sub>1–78</sub>. Taken together, these results indicate that Spt8 and Med13 form specific interactions with TFIIS N-terminal domain. These complexes rely on features observed not only in the formation of the Iws1/Spt6 complex (recognition of the IR1 region), but also on TFIIS-specific features.

## Discussion

Despite the progress made in the functional characterization of Spt6 and Iws1, the roles of these proteins in molecular terms are not well understood. We have started addressing this issue by studying the Spt6 and Iws1 proteins from the fungi-related parasite *E. cuniculi*. We show here that the interaction between Spt6 and Iws1 only requires a small N-terminal region of Spt6 (*E. cuniculi* residues 45–67), which agrees with and refines previous results obtained with the mammalian versions of these proteins (Yoh *et al*, 2007). Furthermore, the crystallographic structures of Iws1 and Iws1 bound to this N-terminal region of Spt6 reveal the presence of two distinct subdomains in Iws1. An independent

determination of the Iws1 crystallographic structure (Pujari *et al*, 2010) fits well with the structure that we determined.

The first subdomain (HEAT subdomain) is specific to Iws1 and is formed by its N-terminal HEAT repeat and part of the following ARM repeat. At the heart of this domain is an invariant lysine (*E. cuniculi* K90, *S. cerevisiae* K192), which is instrumental in the Spt6-independent constitutive recruitment of Iws1 to promoters in yeast (Fischbeck *et al*, 2002; Zhang *et al*, 2008). The N $\epsilon$  atom of this lysine is located at the bottom of a highly conserved pocket that has the ability to preferentially bind negatively charged ions as large as phosphates. In agreement, it has been proposed that Ser5-phosphorylated RNAPII CTD repeats directly recruit yeast Iws1 in a K192-dependent manner (Zhang *et al*, 2008). Yet, our preliminary investigations *in vitro* have failed to characterize a strong interaction between Iws1 and the phosphorylated CTD of RNAPII. This suggests that such an interaction requires another partner that could either stabilize a direct interaction between Iws1 and the RNAPII CTD or, alternatively, bridge both molecules.

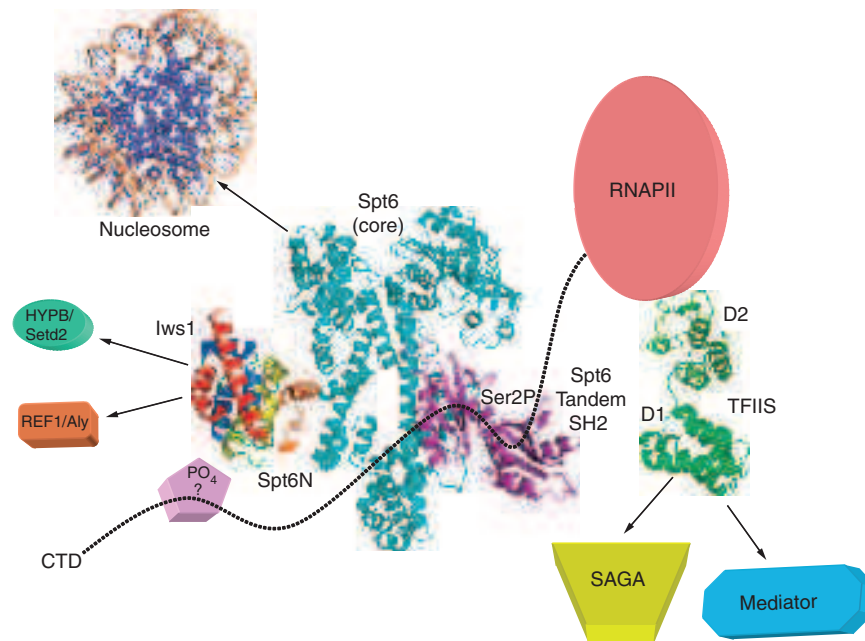
The second subdomain (ARM subdomain) of Iws1 is defined by its two ARM repeats and forms a specific complex with a small N-terminal region of Spt6 (Spt6N). Biochemical and structural analysis of Spt6 recognition by Iws1 revealed two anchoring regions in Spt6N, IR1 and IR2, which are both important for proper complex formation. Furthermore, introduction of amino-acid changes in the context of the full-length yeast Spt6 protein *in vivo* showed that two mutants of the IR1 and IR2 regions, I248A/F249A and Y255A/W257A, respectively, display strong Spt<sup>-</sup> and Ts<sup>-</sup> phenotypes, thus highlighting the importance of the full Iws1-binding region of Spt6 *in vivo*.

Interestingly, the HEAT subdomain of Iws1, which most likely modulates the Spt6-independent role of Iws1, and the IR2-binding region of the Iws1 ARM subdomain are relatively close to each other (Figure 4F). Furthermore, almost all residues of the highly conserved loop, which connects helices  $\alpha_2$  and  $\alpha_3$  of the Iws1 ARM1 repeat, participate either in the HEAT subdomain or in the IR2-binding region of Iws1 ARM subdomain (Figure 2B). Thus, the two Iws1 subdomains appear strongly linked, which suggests that the switch between the Spt6-independent and Spt6-dependent functions of Iws1 occurs through steric hindrance as well as by possible structural rearrangements. It cannot be excluded that a longer C-terminal construct of Spt6 might interact directly with the HEAT subdomain of Iws1. Our co-expression experiments suggest, however, that this interaction is highly unlikely.

Another striking result of our studies is the almost perfect structural conservation between the IR1-binding region of the Iws1 ARM subdomain and the mouse TFIIS N-terminal domain (Domain I). In yeast, the TFIIS Domain I has recently been associated with the newly discovered function of TFIIS in PIC formation and stabilization through its recruitment by the SAGA and Mediator complexes (Prather *et al*, 2005; Guglielmi *et al*, 2007; Kim *et al*, 2007). In agreement, a direct interaction has been observed between yeast TFIIS Domain I and SAGA and Mediator respective subunits, Spt8 and Med13 (Wery *et al*, 2004).

The strong structural homology observed between the IR1-binding region of Iws1 ARM subdomain and TFIIS Domain I suggests that both domains bind their targets in an analogous manner. We show here that Spt8 and Med13 contain a single





**Figure 7** Model of the transcription networks made by Iws1/Spt6 and TFIIS. Based on the results presented in this article as well as previously published data, the figure provides a current model of the Iws1/Spt6 complex and recapitulates our current knowledge of the transcription networks affected by Iws1/Spt6 and TFIIS, including homologous domains (ARM subdomain for Iws1 and Domain I for TFIIS). The model for Spt6 has been assembled from (i) the structure of the bacterial Tex protein that shares sequence homology with the core domain of Spt6 (cyan; Johnson *et al*, 2008), (ii) the recently discovered C-terminal tandem SH2 domains of Spt6 (purple; Diebold *et al*, 2010; Sun *et al*, 2010) and (iii) the Iws1/Spt6N complex (coloured as in previous figures; this manuscript). The N-terminal acidic domain of Spt6 is not included. Interactions of Iws1/Spt6 and TFIIS with different transcriptional effectors are shown with arrows. The C-terminal domain of the RNA polymerase II (CTD) is shown as dotted line. Binding of the Spt6 tandem SH2 repeats to Ser2-phosphorylated CTD repeats has been labelled with 'Ser2P'. The putative CTD binding to the HEAT subdomain of Iws1, possibly through a phosphorylated residue and the requirement of an unknown factor, has been labelled with '?'. The PDB codes for the structures used to make the figure are 1AOI, 1WJT, 2XP1, 3BZC, 3GTM and 3GXW.

patch of conserved sequence that is composed of a highly conserved IR1 motif followed by a stretch of acidic residues. Co-expression of these conserved regions from yeast Spt8 and Med13 with yeast TFIIS Domain I results in complex formation, and requires their IR1-like regions as well as their specific acidic regions. Altogether, these results suggest a specific interaction between TFIIS and Spt8, and TFIIS and Med13, thereby suggesting a molecular basis for TFIIS recruitment at promoters, in order to participate in active PIC formation, possibly by recruitment of RNAPII through TFIIS Domain II.

Importantly, in yeast, only TFIIS Domain I shares sequence homology with the Iws1 IR1-binding region. In contrast, in humans, this homology extends not only to the different isoforms of TFIIS, but also to the isoforms of Elongin A, to Med26 and to two other nuclear proteins, PNUTS and PIBP, whose functions are less well characterized (Wery *et al*, 2004; Ling *et al*, 2006). Surprisingly, in humans, no homologue of yeast Spt8 has been found to date and the TFIIS-binding domain of yeast Med13 does not seem to be conserved in its human homologue. However, in agreement with what has been observed in yeast, both human TFIIS and Elongin A IR1-binding regions are sufficient to interact with the RNAPII holoenzyme containing various PIC components as well as the Mediator subunit Cdk8 (Pan *et al*, 1997). This suggests that one or several subunits of the human holoenzyme have evolved specific sequence for the binding of TFIIS and Elongin A N-terminal domains, possibly recapitulating the role of Spt8 and Med13 in yeast.

The results presented here provide a detailed view of the Iws1/Spt6 complex. Specifically, we show the presence of two specific conserved subdomains within Iws1, one that may recognize a phosphoprotein and a second subdomain specific for Spt6 recognition. Our structural, mutational and *in vivo* functional data demonstrate a bipartite binding of Spt6 to this second subdomain of Iws1. Importantly, part of the recognition determinants within Iws1 for Spt6 are conserved in yeast TFIIS and suggest the molecular basis for recruitment of TFIIS by SAGA Spt8 and Mediator Med13 to promoters. Therefore, our results deepen our understanding of the transcription networks made by the Iws1/Spt6 complex and TFIIS (Figure 7). Furthermore, the broad distribution of the Iws1 (Spt6 binding) ARM subdomain in proteins of higher eukaryotes reveals the larger functional implication of this domain in these organisms. Our study paves the way for deciphering these additional interaction networks, which will require further structural, biochemical and *in vivo* investigations in order to better understand the role of these networks in transcriptional regulation.

## Materials and methods

### Expression and purification

The constructs and mutants used were created by standard PCR procedures and inserted in the pET-MCN series of multi-expression vectors (Romier *et al*, 2006), using *NdeI* and *BamHI* restriction sites. All proteins were expressed or co-expressed in *E. coli* BL21[DE3] cells containing the pRARE vector (Novagen). For single expression, proteins were fused with an N-terminal (His)<sub>6</sub> tag.

For co-expression, one protein was his-tagged and the other untagged and complex formation was monitored by retention of the untagged protein on affinity resin (Talon Metal affinity resin, Clontech) through its interaction with the tagged protein. Following expression, cells were resuspended in lysis buffer (10 mM Tris pH 8.0 with either 50 mM NaCl for the different complexes or 400 mM NaCl for Iws1 alone) and then lysed by sonication.

For small-scale (2 ml) interaction analysis, soluble fractions were incubated for 1 h with 25  $\mu$ l Talon Metal affinity resin (Clontech) and washed with lysis buffer. Laemmli buffer (30  $\mu$ l) was then added to the resin for analysis on SDS-PAGE. For large-scale purification, to produce the protein and complexes for crystallization trials, the soluble fraction was recovered by high-speed centrifugation and incubated with Talon Metal affinity resin for 1 h. The resin was then extensively washed with lysis buffer and resuspended in 2 ml of this buffer, and bovine thrombin was added overnight for cleaving off the histidine tag. The supernatant was recovered and applied onto a gel filtration column Hiload 16/60 Superdex 75 (Amersham Pharmacia) equilibrated with lysis buffer supplemented with 2 mM dithiothreitol. The purified proteins/complexes were concentrated on an Amicon 10K system (Millipore). For production of selenomethionylated Iws1, protein expression was carried out in minimal medium containing selenomethionine.

### Biophysical characterization

Dynamic Light Scattering experiments were performed using a Dynapro-MS (Protein Solutions). Thermal Shift Assay experiments (ThermoFluor) were performed using the Mini Opticon Real Time PCR Detection System from Bio-Rad. The fluorescent reporter used was Sypro Orange (Invitrogen; final concentration 25  $\times$ ). The temperature was raised from 20 to 90°C in increments of 1°C every minute. Wild type and mutant Iws1 proteins were used at a concentration of 1 mg/ml.

### Crystallization

Crystallization trials were performed using the sitting and hanging drop vapour diffusion method. The different crystallization conditions for the *ecIws1<sub>55-198</sub>* construct are reported elsewhere (Koch *et al*, 2010). The *ecSpt6<sub>53-71</sub>/ecIws1<sub>55-198</sub>* complex (10 mg/ml) gave two crystal forms. The first form was obtained at 24°C in the presence of 22% PEG 3350 (Fluka) and 0.2 M NaBr. The second crystal form was grown at 4°C in the presence of 0.1 M sodium acetate pH 5, 18% PEG 1500 (Fluka) and 0.05 M MgCl<sub>2</sub>. The *ecSpt6<sub>34-71</sub>/ecIws1<sub>55-198</sub>* complex (10 mg/ml) was crystallized at 17°C in the presence of 33% PEG 1500 (Fluka) and 0.1 M of the succinic acid/sodium dihydrogen phosphate/glycine system, pH 7 (Hampton). For data collection, the crystals were frozen into liquid nitrogen, either directly (*ecSpt6<sub>34-71</sub>/ecIws1<sub>55-198</sub>*) or after cryoprotection within the crystallization solution supplemented with 20% PEG 200 (*ecIws1<sub>55-198</sub>* and *ecSpt6<sub>53-71</sub>/ecIws1<sub>55-198</sub>* form 2) or 20% glycerol (*ecSpt6<sub>53-71</sub>/ecIws1<sub>55-198</sub>* form 1).

### Data collection, structure determination, model building and refinement

All data obtained in this project were collected at 100 K on ESRF beamlines ID14-1, ID14-2, ID14-4, ID23-1, ID29 and SOLEIL beamline PROXIMA1. All data were processed and scaled using HKL2000 (Otwinowski and Minor, 1997). The *ecIws1<sub>55-198</sub>* structure was solved by performing three-wavelength MAD experiments on selenomethionylated protein-containing crystals (Supplementary Table I). Phasing was performed with SHARP (La Fortelle and Bricogne, 1997). Model building was carried out using

TURBO-FRODO and COOT (Emsley and Cowtan, 2004), and refinement with REFMAC (CCP4, 1994). The structures of the complexes were solved by molecular replacement with PHASER (McCoy *et al*, 2007) using the *ecIws1<sub>55-198</sub>* structure as a search model. The Spt6 protein constructs were built into the additional electron density and the structures refined using identical procedures as for *ecIws1<sub>55-198</sub>*. All structures showed good deviations from ideal geometry (Supplementary Tables I and II), with no Ramachandran outliers. The structures described in this project have been deposited under the PDB codes 2xpl, 2xpn, 2xpo and 2xpp.

### Yeast genetics

Mutations encoding the amino-acid changes I248A/F249A, G250A/G252A, Y255A/W257A, Y255A and W257A in *S. cerevisiae* SPT6 were each individually introduced into a copy of SPT6 on a pRS414-based CEN plasmid. Mutagenesis was performed using the Quikchange Lightning Site-directed Mutagenesis kit (Agilent). The mutated plasmids were used to transform *S. cerevisiae* strain FY857 (*MAT $\alpha$  his4-912 $\delta$  lys2-128A leu2A1 ura3-52 trp1A63 spt6A::LEU2* containing pCC11 (SPT6 in a CEN URA3-marked plasmid)) using standard protocols. A 5-FOA plasmid shuffle (Boeke *et al*, 1984) was performed to obtain a strain that contained only the plasmid with the *spt6* mutant allele. Spot tests were performed on the following media, all incubated at 30°C unless otherwise indicated: YPD, YPD at 37°C, SC-his, SC-lys, YPD + 150 mM hydroxyurea, YPD + 13  $\mu$ g/ml phleomycin and YPD + 15 mM caffeine. YPD, SC-his and SC-lys were made as previously described (Rose *et al*, 1990). All strains are listed in Supplementary Table III. The plasmids they contain are indicated in parentheses.

### Note added in proof

An independent study (McDonald SM, Close D, Xin H, Formosa T and Hill CP. *Mol Cell*, 10.1016/j.molcel.2010.11.014) has also determined a very similar structure of an Iws1/Spt6 complex; the results of the two studies are highly complementary.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank Michel Werner for providing the yeast TFIIIS cDNA, Jean-Marie Wurtz and Jean-Paul Renaud for technical discussions, Irwin Davidson for critical reading of the manuscript and members of the Structural Genomics Platform of IGBMC for setting up automated procedures. We thank members of the ESRF-EMBL joint structural biology groups and members of SOLEIL for the use of their beamline facilities and for help during data collection. MLD is supported by a PhD grant from the French Research Ministry. This work was supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Université de Strasbourg (UDS), the European Commission SPINE2-Complexes project (contract n° LSHG-CT-2006-031220) and NIH grant GM32967.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

Adkins MW, Tyler JK (2006) Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol Cell* **21**: 405–416  
Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P (2001) Comparison of ARM and HEAT protein repeats. *J Mol Biol* **309**: 1–18  
Ardehali MB, Yao J, Adelman K, Fuda NJ, Petesch SJ, Webb WW, Lis JT (2009) Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. *EMBO J* **28**: 1067–1077  
Boeke JD, LaCroute F, Fink GR (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity

in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**: 345–346  
Bond CS, Schuttelkopf AW (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* **65**(Part 5): 510–512  
Booth V, Koth CM, Edwards AM, Arrowsmith CH (2000) Structure of a conserved domain common to the transcription factors TFIIIS, elongin A, and CRSP70. *J Biol Chem* **275**: 31266–31268  
Bortvin A, Winston F (1996) Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272**: 1473–1476

- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**(Part 5): 760–763
- Diebold ML, Loeliger E, Koch M, Winston F, Cavarelli J, Romier C (2010) A non-canonical tandem SH2 enables interaction of elongation factor SPT6 with RNA polymerase II. *J Biol Chem* (doi:10.1074/jbc.M110.146696)
- Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Part 12 Part 1): 2126–2132
- Endoh M, Zhu W, Hasegawa J, Watanabe H, Kim DK, Aida M, Inukai N, Narita T, Yamada T, Furuya A, Sato H, Yamaguchi Y, Mandal SS, Reinberg D, Wada T, Handa H (2004) Human Spt6 stimulates transcription elongation by RNA polymerase II *in vitro*. *Mol Cell Biol* **24**: 3324–3336
- Fischbeck JA, Kraemer SM, Stargell LA (2002) SPN1, a conserved gene identified by suppression of a postrecruitment-defective yeast TATA-binding protein mutant. *Genetics* **162**: 1605–1616
- Guglielmi B, Soutourina J, Esnault C, Werner M (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation *in vivo*. *Proc Natl Acad Sci USA* **104**: 16062–16067
- Johnson SJ, Close D, Robinson H, Vallet-Gely I, Dove SL, Hill CP (2008) Crystal structure and RNA binding of the Tex protein from *Pseudomonas aeruginosa*. *J Mol Biol* **377**: 1460–1473
- Kaplan CD, Laprade L, Winston F (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301**: 1096–1099
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivares CP (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**: 450–453
- Kim B, Nesvizhskii AI, Rani PG, Hahn S, Aebersold R, Ranish JA (2007) The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci USA* **104**: 16068–16073
- Koch M, Diebold ML, Cavarelli J, Romier C (2010) Crystallization and preliminary crystallographic analysis of eukaryotic transcription and mRNA export factor Iws1 from *Encephalitozoon cuniculi*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**(Part 2): 207–210
- Krogan NJ, Kim M, Ahn SH, Zhong G, Kobor MS, Cagney G, Emili A, Shilatifard A, Buratowski S, Greenblatt JF (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol* **22**: 6979–6992
- La Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In: *Methods in Enzymology*, Carter JCW, Sweet RM (eds), Vol. 276, pp 472–494. New-York: Academic Press
- Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* **128**: 707–719
- Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA, Yates III JR, Hartzog GA (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol Cell Biol* **23**: 1368–1378
- Ling Y, Smith AJ, Morgan GT (2006) A sequence motif conserved in diverse nuclear proteins identifies a protein interaction domain utilised for nuclear targeting by human TFIIS. *Nucleic Acids Res* **34**: 2219–2229
- Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17**: 1272–1278
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**(Part 4): 658–674
- Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296
- Otwinowski Z, Minor W (1997) Processing X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology*, Carter JCW, Sweet RM (eds), Vol. 276, pp 307–326. New-York: Academic Press
- Pan G, Aso T, Greenblatt J (1997) Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J Biol Chem* **272**: 24563–24571
- Perales R, Bentley D (2009) ‘Cotranscriptionality’: the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**: 178–191
- Prather DM, Larschan E, Winston F (2005) Evidence that the elongation factor TFIIS plays a role in transcription initiation at GAL1 in *Saccharomyces cerevisiae*. *Mol Cell Biol* **25**: 2650–2659
- Pujari V, Radebaugh CA, Chodaparambil JV, Muthurajan UM, Almeida AR, Fischbeck JA, Luger K, Stargell LA (2010) The transcription factor Spn1 regulates gene expression via a highly conserved novel structural motif. *J Mol Biol* (doi:10.1016/j.jmb.2010.09.040)
- Romier C, Ben Jelloul M, Albeck S, Buchwald G, Busso D, Celie PH, Christodoulou E, De Marco V, van Gerwen S, Knipscheer P, Lebbink JH, Notenboom V, Poterszman A, Rochel N, Cohen SX, Unger T, Sussman JL, Moras D, Sixma TK, Perrakis A (2006) Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr D Biol Crystallogr* **62**(Part 10): 1232–1242
- Romier C, James N, Birck C, Cavarelli J, Vivares C, Collart MA, Moras D (2007) Crystal structure, biochemical and genetic characterization of yeast and *E. cuniculi* TAF(II)5 N-terminal domain: implications for TFIID assembly. *J Mol Biol* **368**: 1292–1306
- Rose MD, Winston F, Hieter P (1990) *Laboratory Course Manual for Methods in Yeast Genetics*. New York: Cold Spring Harbor
- Saunders A, Core LJ, Lis JT (2006) Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7**: 557–567
- Sun M, Lariviere L, Dengl S, Mayer A, Cramer P (2010) A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II CTD. *J Biol Chem* (doi:10.1074/jbc.M110.144568)
- Wery M, Shematorova E, Van Driessche B, Vandenhoute J, Thuriaux P, Van Mullem V (2004) Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J* **23**: 4232–4242
- Winkler M, aus Dem Siepen T, Stamminger T (2000) Functional interaction between pleiotropic transactivator pUL69 of human cytomegalovirus and the human homolog of yeast chromatin regulatory protein SPT6. *J Virol* **74**: 8053–8064
- Yoh SM, Cho H, Pickle L, Evans RM, Jones KA (2007) The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21**: 160–174
- Yoh SM, Lucas JS, Jones KA (2008) The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22**: 3422–3434
- Zhang L, Fletcher AG, Cheung V, Winston F, Stargell LA (2008) Spn1 regulates the recruitment of Spt6 and the Swi/Snf complex during transcriptional activation by RNA polymerase II. *Mol Cell Biol* **28**: 1393–1403

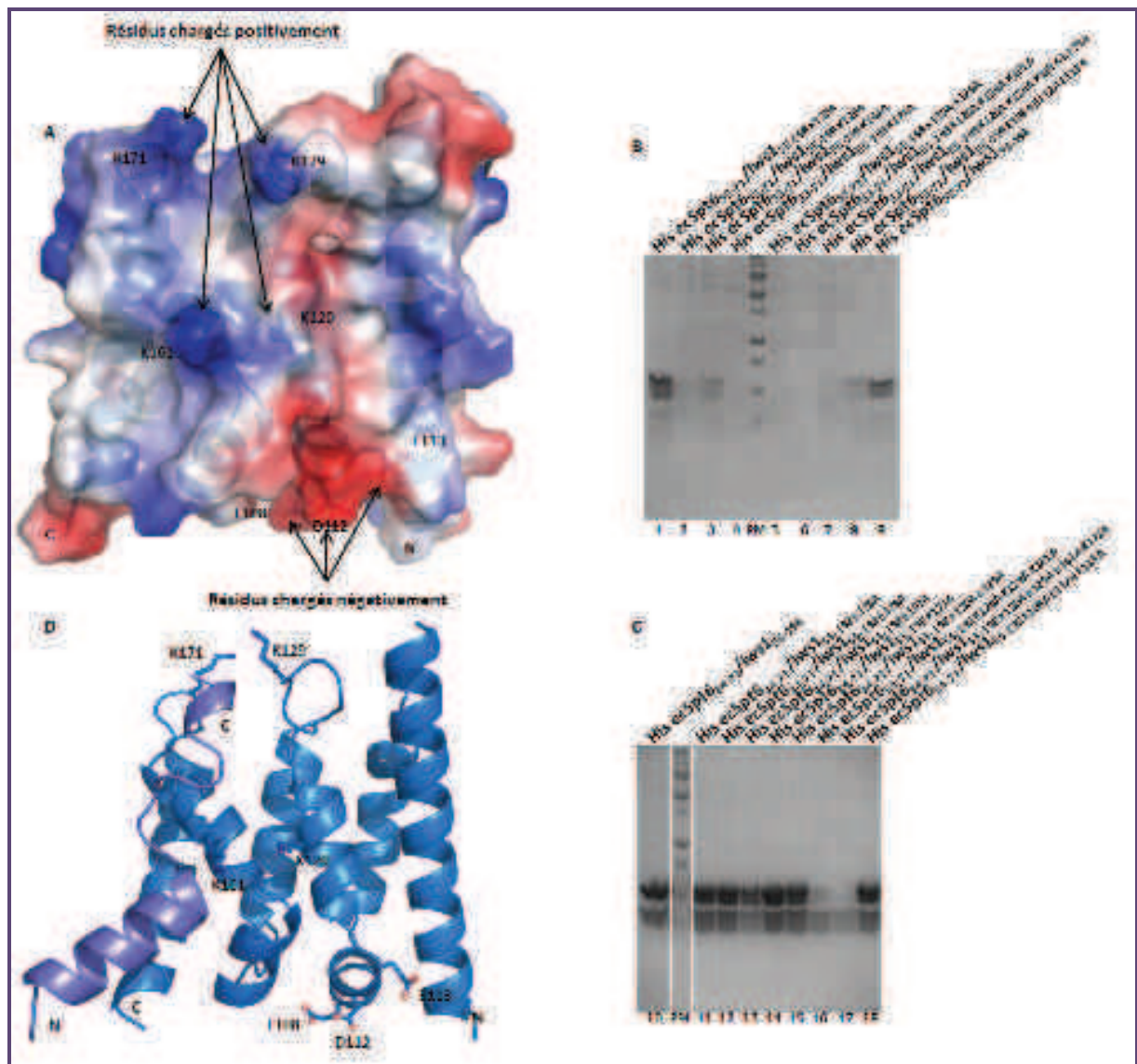
Les résultats obtenus, présentés dans la publication numéro 2 montrent donc qu'un court fragment de l'extrémité amino-terminale de Spt6 interagit fortement avec le domaine central conservé de la protéine Iws1. De plus, la mutation de résidus contenus dans ce fragment abolit la formation du complexe *in vitro* et causes de sévères phénotypes lorsque ces mutations sont introduites dans des levures. En nous appuyant sur la connaissance de la structure du complexe Spt6/Iws1 et sur les données publiées dans la littérature, différents autres complexes ont été testés, les résultats de ces tests étant présentés ci-dessous.

#### 4.4. Etude des complexes formés par Spt6 et Iws1

L'analyse des structures des protéines Iws1 et Spt6 nous a permis d'identifier la présence d'éléments caractéristiques, dont nous avons testé le rôle biochimiquement. Nous avons ainsi testé la liaison de protéines Iws1 mutées avec des fragments de Spt6 mais également l'interaction de Iws1 avec l'ARN polymérase II, l'histone méthyltransférase Set2 et le facteur d'export Yra1p. En effet, les études de génétique de levure ont montré que Iws1 était recrutée au niveau des promoteurs des gènes en l'absence de Spt6, suggérant une fonction indépendante de celle du complexe Spt6/Iws1, en interaction avec l'ARN polymérase II (Fischbeck et al., 2002). De plus, il a été démontré dans des cellules de souris que Iws1, lorsqu'elle interagit avec Spt6, permettait le recrutement de Setd2, homologue de Set2 chez la levure, et REF1/Aly, homologue de Yra1p chez la levure, au cours de la transcription (Yoh et al., 2007, 2008).

##### 4.4.1. Complexe Spt6/Iws1 – mutants de surface de Iws1

Alors que la structure du complexe Spt6/Iws1 n'avait pas été résolue, l'implication des résidus de surface conservés de Iws1 dans la formation du complexe a été testée par co-expression. L'analyse des alignements de séquence de la protéine Iws1 d'*E. cuniculi* a permis de mettre en évidence la présence, à la surface de la protéine de régions conservées fortement chargées, aussi bien positivement que négativement (Figure 51A). Afin de comprendre le rôle de ces régions, nous avons muté les résidus qui les composaient puis testé la capacité de ces protéines mutées à interagir avec deux constructions de Spt6. Ces tests ont été réalisés par co-expression entre les constructions 34-71 et 53-71 de ecSpt6 et le domaine central de ecIws1, dont les différents résidus mutés sont : (i) les lysines 120, 129, 161 et 171 pour les résidus chargés positivement et (ii) les glutamates 108 et 113 et l'aspartate 112 pour les résidus chargés négativement (Figure 51A).



**Figure 51 : Co-expression de Spt6 avec des mutants de surface de Iws1**

(A) Régions chargées et conservées de la surface de la protéine Iws1 d'*E. cuniculi*. (B-C) Co-expression du domaine central de la protéine Iws1 d'*E. cuniculi* portant différentes mutations des résidus de surface chargés et conservés avec le fragment 53-71 (B) et 34-71 (C) de Spt6. (D) Localisation des résidus mutés de la surface de Iws1 (bleu) par rapport au fragment de Spt6 (mauve).

Alors que la mutation des résidus de surface n'a que peu d'effet sur la co-expression avec un fragment long de Spt6 (construction 34-71), la co-expression des mutants de Iws1 avec un fragment court de Spt6 (construction 53-71) a un effet important sur la formation du complexe (Figure 51 B et C). En effet, lors de la co-expression avec le fragment 53-71 de Spt6, quasiment toutes les mutations affaiblissent ou abolissent la formation du complexe. Seule la mutation de la lysine 120 ne semble pas avoir d'effet sur le complexe (ligne 1), la mutation de la lysine 161 (ligne 3) et des résidus chargés négativement (E108A, D112A,

E113A, ligne 8) affaiblissant fortement le complexe. Dans le cas de la co-expression avec la construction 34-71 de Spt6 seule la mutation de trois (ligne 16) ou quatre (ligne 17) des lysines de surface de Iws1 affaiblit le complexe, cet affaiblissement du complexe pouvant refléter une diminution de la solubilité de la protéine Iws1 mutée. La mutation des résidus de surface conservés de Iws1 affaiblit donc fortement le complexe Spt6/Iws1 dans le cadre d'un complexe contenant un fragment court de Spt6, montrant que les résidus de l'extrémité amino-terminale de ce fragment sont impliqués dans la formation du complexe.

Ces différentes observations peuvent être expliquées par l'analyse de la structure du complexe Spt6/Iws1, résolue par la suite (Figure 51D). En effet, dans le cas de la co-expression du fragment 34-71 de Spt6 avec Iws1, des contacts hydrophobes sont mis en place entre l'extrémité amino-terminale de l'hélice  $\alpha$ N de Spt6 et le motif ARM2 de Iws1. Ces contacts sont absents lors de la co-expression du fragment 53-71, la mutation des résidus de Iws1 à proximité du site d'interaction va alors déstabiliser le complexe plus facilement (cas des lysines 129, 161 et 171). La mutation de la lysine 120 de Iws1 n'a que peu d'effet sur la formation d'un complexe contenant le fragment court de Spt6, ce résidu n'étant pas localisé au niveau de la zone d'interaction avec Spt6, contrairement aux autres lysines. Les résidus de la zone chargée négativement (E108, D112 et E113) ne sont pas non plus localisés au niveau de la zone d'interaction, l'affaiblissement de l'interaction entre Iws1 et le fragment court de Spt6 résultant probablement d'une diminution de la solubilité de la protéine Iws1.

Ces résultats montrent que, bien que la zone d'interaction entre Spt6 et Iws1 soit très localement définie, des résidus autour de ces motifs participent également à la formation du complexe. Ainsi, la zone d'interaction entre ces deux protéines est relativement étendue.

#### 4.4.2. Interaction de Iws1 avec l'ARN polymérase II

Une interaction génétique entre l'ARN polymérase II et Iws1 a été plusieurs fois démontrée (Fischbeck et al., 2002; Zhang et al., 2008), impliquant notamment la lysine 192 de Iws1, dont la mutation chez la levure perturbe la régulation de l'initiation de la transcription.

La résolution de la structure de Iws1 et des complexes Spt6/Iws1, a permis d'observer que le groupement amine de la lysine correspondante chez *E. cuniculi* (K90) se trouve au fond d'une poche qui fixe un ion chargé négativement (brome ou chlore). Les résidus présents dans cette poche pourraient accommoder un groupement phosphate (Figure 3, publication 2). De plus, la mutation de la lysine 90 de ecIws1 déstabilise la protéine, ce qui est reflété par une

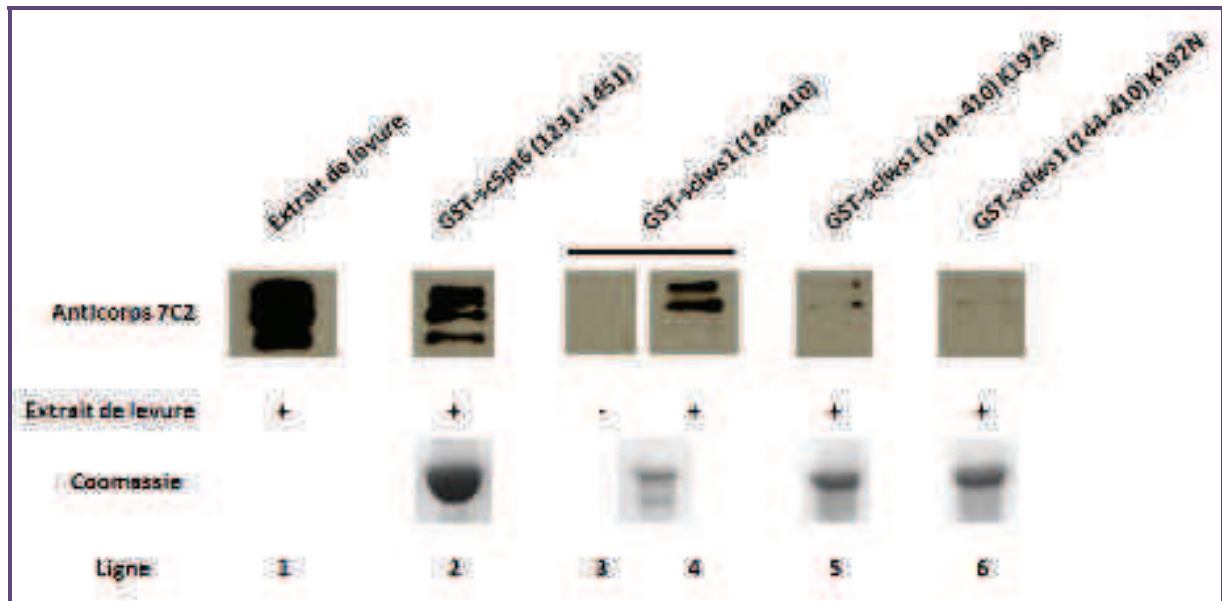
diminution de la température de fusion des mutants mesurée par Thermofluor et affaiblit le complexe formé avec Spt6 (publication 2). En effet, la co-expression de très courts fragments de Spt6 (construction 53-71) avec des mutants de K90D et K90N de Iws1 abolit la formation du complexe.

Ces données structurales et biochimiques montrent que la lysine 192 de scIws1 pourrait jouer un rôle important dans la fonction de la protéine, en interagissant notamment avec des protéines phosphorylées. Afin de voir l'implication de cette lysine dans l'interaction avec l'ARN polymérase II, des expériences de pulldown ont été réalisées avec des protéines Iws1 mutées. D'autre part, l'observation au niveau structural de la capacité de Iws1 à lier des protéines phosphorylées nous a poussé à également tester son interaction avec des peptides phosphorylés mimant le CTD de l'ARN polymérase II.

#### 4.4.2.1. Test de l'interaction avec l'ARN polymérase II par pulldown

Des expériences de GST pulldown ont été réalisées afin de confirmer l'implication de la lysine 192 de Iws1 dans l'interaction avec l'ARN polymérase II. Une construction de la protéine Iws1 de levure contenant le domaine central et le domaine carboxy-terminal (construction 144-410) a été définie et clonée dans un vecteur pGexNB permettant d'exprimer la protéine en fusion avec une étiquette GST. Des mutations ponctuelles remplaçant la lysine 192 par une alanine (K192A) ou une asparagine (K192N) ont été introduites dans le gène. Ces protéines ont été produites, purifiées et incubées avec un extrait de levure. Après lavage, la fraction retenue est déposée sur gel d'acrylamide et la présence de l'ARN polymérase II est révélée par western blot en utilisant un anticorps dirigé contre le CTD de l'ARN polymérase II (anticorps 7C2).

Malgré la faible quantité de protéine Iws1 disponible pour la réalisation de ces expériences, une interaction entre l'ARN polymérase II et Iws1 est bien mise en évidence (ligne 4, Figure 52). La force de cette interaction semble être presque équivalente à l'interaction entre le domaine carboxy-terminal de la protéine Spt6, utilisée ici comme contrôle (ligne 2, Figure 52). La mutation de la lysine affaiblit fortement cette interaction comme le montre la diminution du signal détecté par western blot (ligne 5 et 6, Figure 52).



**Figure 52 : Tests de l'interaction Iws1/ARN polymérase II – expériences de pull-down.**

La construction 144-410 de la protéine *scIws1* fusionnée à la GST a été incubée avec un extrait de levure. La présence de l'ARN polymérase II dans la fraction retenue a été révélée par western blot avec des anticorps dirigés contre le CTD de l'ARN polymérase II.

Ces expériences montrent bien la capacité de Iws1 à interagir avec l'ARN polymérase II et l'implication de la lysine 192 dans la formation de cette interaction. Cependant, le mode de détection utilisé pour mettre ces interactions en évidence ne permet pas de savoir si elle est directe ou indirecte. En effet, l'interaction Iws1/ARN polymérase II pourrait être réalisée à travers l'interaction de chacune des protéines avec un partenaire commun.

#### 4.4.2.2. Interaction de Iws1 avec des peptides phosphorylés mimant le CTD de l'ARN polymérase II

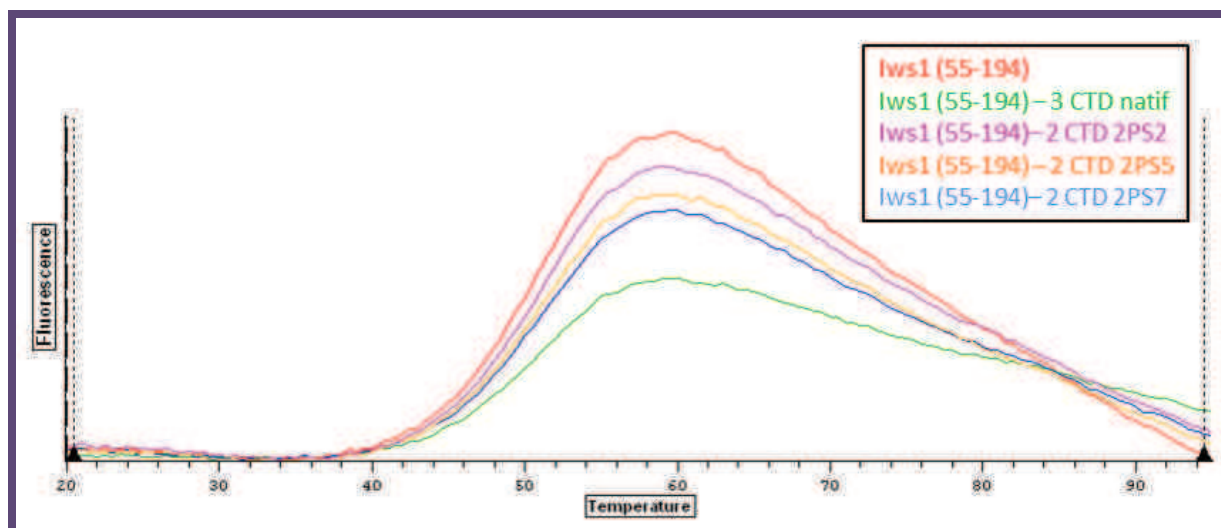
Afin de tester si une interaction directe pouvait être mise en place entre Iws1 et le CTD de l'ARN polymérase II, des expériences de microcalorimétrie de Thermofluor ont été réalisées entre des peptides phosphorylés du CTD de l'ARN polymérase II et la construction du domaine central de la protéine Iws1 d'*E. cuniculi* (construction 55-194). Cette interaction a été testée avec un peptide non phosphorylé mais également avec des peptides phosphorylés sur les sérines en position 2, en position 5 et en position 7. La séquence de ces quatre peptides est présentée dans le Tableau 12.



Peptide	Séquence
<b>3CTD natif</b>	YSPTSPSYSPSTSPSYSPSTSPS
<b>2CTD PS2</b>	PTSPSY <sub>p</sub> SPTSPSY <sub>p</sub> SPTSPSY
<b>2CTD PS5</b>	PSYSPT <sub>p</sub> SPSPSYSPST <sub>p</sub> SPSPSYSPST
<b>2CTD PS7</b>	YSPTSP <sub>p</sub> SYSPSTSP <sub>p</sub> SYSPST

**Tableau 12 : Séquence des peptides utilisés pour les expériences de microcalorimétrie et Thermofluor entre Iws1 et le CTD de l'ARN polymérase II.**

La protéine utilisée pour ces expériences a été concentrée à 50 $\mu$ M, les peptides étant deux fois plus concentrés (100 $\mu$ M).



**Figure 53 : Interaction entre Iws1 et des peptides phosphorylés du CTD de l'ARN polymérase II – expériences de Thermofluor.**

50 $\mu$ M de protéine Iws1 d'*E. cuniculi* ont été mélangés à 100 $\mu$ M des différents peptides phosphorylés. La température de fusion mesurée de la protéine Iws1 seule et en présence des peptides phosphorylés est de 50,5°.

Ni les expériences de Thermofluor (Figure 53) ni les expériences d'ITC (courbes non présentées) n'ont permis de révéler d'interaction entre Iws1 et les peptides phosphorylés du CTD de l'ARN polymérase. En effet, les expériences de Thermofluor n'ont pas révélé d'effet de stabilisation de la protéine en présence des peptides, la température de fusion mesurée étant de 50,5°C aussi bien lorsque la protéine est seule que lorsqu'elle est mélangée aux différents peptides.

Ainsi, bien que nos tests d'interaction par pulldown aient permis de montrer une interaction, directe ou indirecte, d'une construction de Iws1 avec l'ARN polymérase II, les

tests réalisés *in vitro* n'ont pas permis de mettre en évidence une interaction directe. En effet, les expériences de microcalorimétrie et Thermofluor n'ont pas révélé d'interaction directe entre le domaine central de Iws1 et des peptides courts du CTD de l'ARN polymérase II. Ces résultats permettent de formuler deux hypothèses : soit cette interaction est indirecte, elle pourrait notamment être médiée par une autre protéine de la machinerie transcriptionnelle, soit elle n'implique pas le CTD de l'ARN polymérase II, mais une autre région ou protéine de ce très grand complexe protéique.

#### 4.4.3. Interaction de Iws1 avec l'histone méthyltransférase Set2 et le facteur d'export Yra1

Une interaction entre Iws1 et l'histone méthyltransférase Setd2 (Yoh et al., 2008) et le facteur d'export REF1/Aly (Yoh et al., 2007) a été montrée dans les cellules de souris par des expériences de pulldown. Ces expériences, réalisées avec diverses constructions de la protéine Iws1, suggèrent que REF1/Aly interagit avec l'extrémité carboxy-terminale de Iws1, en aval du motif LW, alors que Setd2 interagit avec une région en amont du domaine central conservé.

Afin de localiser précisément ces zones d'interactions et caractériser biochimiquement ces complexes des tests de co-expression dans la bactérie ont été entrepris en utilisant les protéines de levure. En effet, les gènes homologues de ces protéines n'ont pas été détectés chez *E. cuniculi* par recherche de similarité de séquence dans une banque contenant les séquences codantes du parasite. Cela explique d'ailleurs que les zones d'interaction de Iws1 d'*E. cuniculi* avec Setd2 et REF1/Aly n'ont pas été conservée dans cette protéine. Ainsi les gènes des protéines homologues de levure ont été clonés : les protéines Set2, homologue de Setd2 et Yra1p, homologue de REF1/Aly. La protéine Yra1p étant constituée d'un seul domaine de type RRM (RNA Recognition Motif) le gène de la protéine entière a été cloné dans les vecteurs d'expressions pnEA-tH et pnCS. La protéine Set2 est par contre constituée de plusieurs domaines, trois constructions de la protéine ont donc été réalisées. Une première construction contient les motifs AWS, SET et POST-SET qui portent l'activité méthyltransférase de la protéine (43-271), une seconde contient le domaine central de la protéine dans lequel aucun domaine n'a été identifié (262-463) et une dernière construction contient les motifs WW et SRI, qui permettent la liaison de la protéine à la sous unité Rpb1 de l'ARN polymérase II (468-733). Ces différentes constructions ainsi que le gène codant pour la

protéine entière ont également été clonées dans les vecteurs pnEA-tH et pnCS. Ces vecteurs codant pour Set2 et Yra1p ont été co-transformés avec trois constructions de la protéine Iws1 : (i) la protéine entière (1-410), (ii) une construction qui ne contient pas le domaine carboxy-terminal de la protéine (1-314) et (iii) une construction ne contenant que le domaine central conservé de la protéine (144-314).

Ces tests de co-expression ont été réalisés dans un milieu Graffinity, en testant plusieurs tampons de lyse (50 et 400 mM NaCl, 10 mM Tris, pH8). Cependant, malgré le test de plusieurs constructions et concentrations salines aucun complexe n'a pu être mis en évidence (gels non présentés).

Différentes hypothèses peuvent être formulées afin d'expliquer cette absence d'interaction. Tout d'abord, des modifications post-traductionnelles pourraient être nécessaires à la formation de ces complexes. La réalisation de ces tests d'expression dans un système d'expression eucaryote (système baculovirus ou cellules mammifères) pourrait ainsi permettre l'observation d'une interaction. Il se pourrait également que ces complexes ne soient formés que dans les cellules de mammifère, puisqu'ils ont été mis en évidence dans des cellules de souris. Des tests d'expression utilisant les protéines humaines ont donc été entrepris qui n'ont pas encore donnés de résultats, la grande taille des gènes rendant les étapes de clonage difficiles.

## 4.5. Discussion

Notre étude a permis de caractériser biochimiquement, structuralement et fonctionnellement le complexe Spt6/Iws1. Nous avons également démontré l'implication de la lysine 192 de Iws1 dans l'interaction avec l'ARN polymérase, bien que l'interaction n'ait pas pu être définie précisément.

### 4.5.1. Comparaison des structures de Iws1 et du complexe Spt6/Iws1

Deux études proches ont mis en évidence des résultats similaires concernant la structure de Iws1 (Pujari et al., 2010) et du complexe Spt6/Iws1 (McDonald et al., 2010) de levure. Ainsi, la résolution de la structure du domaine central de Iws1 de levure montre un repliement identique à celui du domaine de la protéine de *E. cuniculi* (Figure 54), les valeurs de déviation standard étant faibles (1,985Å par rapport à la structure 3NFQ, 1,922Å par

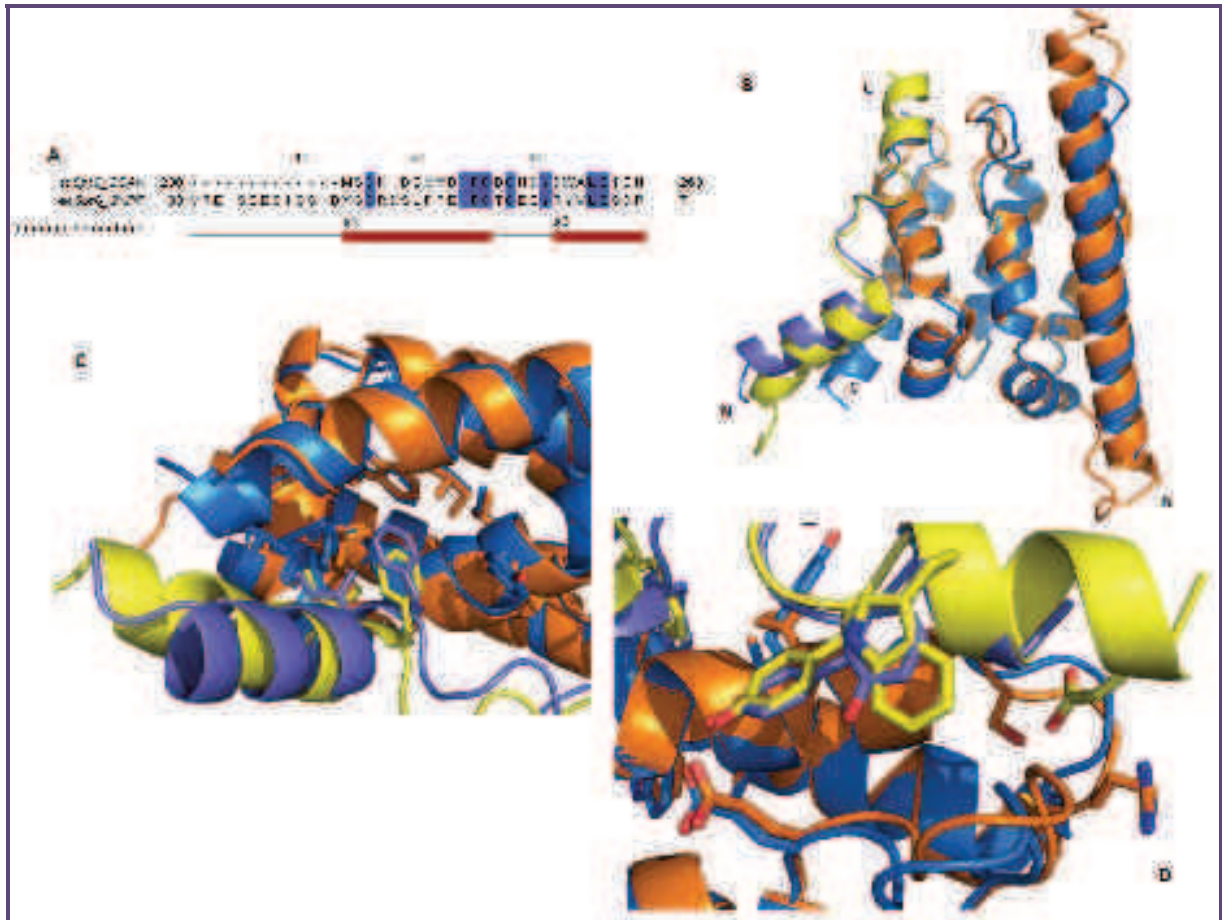
rapport à la structure 3O8Z). La structure de *E. cuniculi* présente principalement une hélice  $\alpha A$  plus courte que celle de *S. cerevisiae* (Figure 54B) mais les résidus formant le cœur hydrophobe de la protéine sont conservés.



**Figure 54 : Alignement des séquences et des structures des protéines Iws1 dont les structures ont été résolues.**

(A) Les séquences des constructions des protéines Iws1 de *E. cuniculi* et *S. cerevisiae* qui ont été cristallisées ont été alignées avec le programme MAFFT.  $\Delta$  indique la position de la K192 (B) Superposition des structures du domaine central de Iws1, les résidus conservés formant le cœur hydrophobe étant représentés en bâtons (vert : 3NFQ, orange : 3O8Z et bleu : 2XPL).

La construction de la protéine Spt6 de levure cristallisée en complexe avec Iws1 est fortement similaire à la construction de *E. cuniculi* que nous avons utilisée (Figure 55A), montrant que les déterminants nécessaires à la mise en place du complexe sont conservés au cours de l'évolution. La résolution de la structure du complexe Spt6/Iws1 de levure montre également un mode de liaison identique entre les deux protéines (Figure 55B). Le mode de reconnaissance hydrophobe du motif IF de Spt6 par Iws1 est identique chez *E. cuniculi* et *S. cerevisiae*, les résidus formant la poche de Iws1 étant fortement conservés (Figure 55C).



**Figure 55 : Alignement des séquences et des structures des complexes Spt6/Iws1 dont les structures ont été résolues.**

Les séquences des fragments de la protéine Spt6 ont été alignées en se basant sur la superposition des structures des complexes (A). Dans le cas des structures de *E. cuniculi*, seule la séquence du fragment le plus long est présentée. (B) superposition des structures des complexes Spt6/Iws1 de *E. cuniculi* (teintes bleues, 2XPP) et *S. cerevisiae* (teintes oranges, 3OAK), (C) et (D) représentant des agrandissements des zones de contact I et II respectivement.

D'autre part, la seconde zone d'interaction fait également intervenir des contacts hydrophobes additionnés de contacts électrostatiques (Figure 55D). Seule une tyrosine de Spt6 de *E. cuniculi* est remplacée chez la levure par un résidu tryptophane. Ce résidu adopte une orientation similaire à la tyrosine et réalise les mêmes contacts hydrophobes. Par contre l'interaction électrostatique médiée à travers une molécule de l'hydroxyle de cette tyrosine avec le groupement amine de la chaîne principale de la leucine 126 de eIws1 est perdue. La structure du complexe de levure fait par contre apparaître la présence d'un résidu glutamate à l'extrémité de l'hélice  $\alpha$ C qui réalise des interactions électrostatiques avec des résidus du motif ARM2 (Figure 55D).

Ainsi, les différentes structures de Iws1 et du complexe Spt6/Iws1 résolues montrent des repliements et des modes de liaisons dans le cas du complexe similaires, confirmant les données que nous avons obtenues avec les protéine d'*E. cuniculi*.

#### 4.5.2. Caractérisation fonctionnelle – expression de mutants des protéines Iws1 et Spt6

Différentes mutations des protéines Spt6 et Iws1 ont été réalisées dans chacune des deux publications présentant les structures de Iws1 (Pujari et al., 2010) et du complexe Spt6/Iws1 (McDonald et al., 2010). Les phénotypes associés à ces mutations sont présentés dans le Tableau 13, les mutations réalisées dans la publication 2 étant également présentées. Dans le cas du complexe Spt6/Iws1, l'effet de ces mutations sur la formation du complexe *in vitro* a été testé par la réalisation de mesures d'affinités par microcalorimétrie. Ces mesures ont notamment permis de définir la constante de dissociation du complexe Spt6/Iws1 natif qui est de 170 nM.

L'introduction de ces mutations dans les gènes de Spt6 et Iws1 montre clairement une implication de ces deux protéines dans les mécanismes de transcription. La mutation des résidus de Spt6 conduit dans la plupart des cas à l'apparition de phénotypes Spt-, typiques de défauts de la transcription, et de thermosensibilité. Seule la mutation du résidu W257 ne produit pas de phénotype, le complexe étant probablement encore formé. Parmi les différentes mutations testées, c'est la double mutation des résidus Y255 et W257 qui conduit à l'apparition des phénotypes les plus importants suggérant que ces deux résidus sont fortement impliqués dans la formation du complexe Spt6/Iws1 et sa fonction. La mesure de la constante de dissociation du complexe natif Spt6/Iws1 montre que l'interaction entre les deux protéines est forte (McDonald et al., 2010), ce qui peut être expliqué par les nombreux contacts hydrophobes et électrostatiques mis en place au niveau de la zone d'interaction.

Mutation - référence	Phénotypes	Formation du complexe Spt6/Iws1
<b><i>Spt6 Y255A W257A</i>, publication 2</b>	Thermosensibilié, Spt-, sensibilité à l'hydroxyurée, phléomycine et caféine	non
<b><i>Spt6 Y255A</i>, publication 2</b>	Spt-	non
<b><i>Spt6 W257A</i>, publication 2</b>	Comme le WT	nd
<b><i>Spt6 I248A F249A</i>, publication 2</b>	Thermosensibilité, Spt-	non
<b><i>Spt6 G250A G252A</i>, publication 2</b>	Spt-	non
<b><i>Spt6 F249A</i></b> , (McDonald et al., 2010)	Thermosensibilité, Spt-	Affaiblissement du complexe, Kd=10µM
<b><i>Spn1 R263D</i></b> (McDonald et al., 2010)	Spt-	Affaiblissement du complexe, Kd= 29,7µM
<b><i>Spn1 F267E</i></b> (McDonald et al., 2010)	Thermosensibilité, Spt-	Pas de complexe détecté
<b><i>Spn1 D172G</i></b> (Pujari et al., 2010)	Thermosensibilité, perte de l'interaction avec l'ARNPII et Spt6, augmentation du taux de transcription du gène CYC1	-
<b><i>Spn1 L218P</i></b> (Pujari et al., 2010)	Thermosensibilité, perte de l'interaction avec l'ARNPII et Spt6, augmentation du taux de transcription du gène CYC1	-
<b><i>Spn1 I180T/L218P</i></b> (Pujari et al., 2010)	Thermosensibilité (interaction avec l'ARNPII et Spt6 non testée)	-
<b><i>Spn1 I180T</i></b> (Pujari et al., 2010)	Comme WT	-

**Tableau 13 : Phénotypes associés aux mutations de Spt6 et Iws1.**

*Dans la publication 2, la capacité des mutants de Spt6 à lier Iws1 a été testée par co-expression, alors qu'elle a été testée par microcalorimétrie par McDonald et al., 2010. (nd, non déterminée, WT souche sauvage)*

Les mutations introduites dans Iws1 montrent aussi des effets importants. Les mutations R263D et F267E introduites dans la publication (McDonald et al., 2010) sont localisées au niveau du deuxième site d'interaction de Spt6 avec Iws1. *In vitro*, ces mutations affaiblissent fortement le complexe voire l'abolissent (cas de la mutation F267E), *in vivo* elles conduisent à l'apparition de phénotypes sévères, montrant notamment un défaut dans le mécanisme de la transcription (phénotype Spt-). Les mutations D172G, I280T/L218P et L218P conduisent toutes à l'apparition de phénotypes thermosensible chez les levures (Pujari et al., 2010), les phénotypes Spt- n'ayant pas été testés. De plus, les mutations D172G et

L218P induisent une perte de l'interaction de Iws1 avec la protéine Spt6 (testé par immunoprécipitation), le taux d'expression de ces protéines mutantes étant diminué par rapport à la protéine native (à température non permissive). La perte de l'interaction lors de la mutation de l'aspartate 172, qui est localisé loin de la zone d'interaction définie suggère soit une interaction plus étendue entre Spt6 et Iws1 ou alors une déstabilisation de la protéine mutée conduisant à une perte de l'interaction.

Ces différents mutants montrent clairement que la protéine Iws1 et le complexe Spt6/Iws1 ont des fonctions essentielles chez les levures, ces fonctions étant encore à préciser. Les études réalisées dans des cellules de souris ont suggéré un rôle dans la régulation épigénétique de la transcription et l'export des ARN messagers via une interaction avec l'histone méthyltransférase Setd2 et le facteur d'export REF1/Aly. Cependant ces liens n'ont pour l'instant pas été confirmés chez la levure, nos tests d'expression n'ayant pas permis de mettre en évidence ces complexes. D'autres tests vont donc être nécessaires avant de pouvoir conclure quant à la formation de ces complexes.

#### 4.5.3. Interaction de Iws1 avec l'ARN polymérase II

L'interaction entre Iws1 et l'ARN polymérase II a également été suggérée dans les deux études réalisées sur la protéine Iws1 et le complexe Spt6/Iws1. Aucune interaction directe n'a pu être mise en évidence, aussi bien dans notre étude que dans l'étude réalisée par Mc Donald et ses collaborateurs. En effet, des tests de microcalorimétrie, de Thermofluor (notre étude) et d'anisotropie de fluorescence (McDonald et al., 2010) n'ont pas permis de mettre en évidence une interaction directe entre le domaine central de la protéine et des peptides phosphorylés. Cependant, *in vivo*, une interaction a pu être montrée par immunoprécipitation (Pujari et al., 2010) et GST pulldown (notre étude), l'introduction de mutations dans Iws1 et notamment au niveau de la lysine 192 abolissant cette interaction. Ainsi, Iws1 interagit avec l'ARN polymérase II, cette interaction pouvant être indirecte, et éventuellement médiée par Spt6. Il sera donc intéressant et important de comprendre si cette interaction est renforcée ou mutuellement exclusive avec la fixation de Spt6, les données de levure et humaines étant apparemment contradictoire sur ce point (Fischbeck et al., 2002; Zhang et al., 2008; Yoh et al., 2007, 2008).



Nous avons caractérisé biochimiquement, structuralement et fonctionnellement le complexe Spt6/Iws1 en utilisant les protéines du parasite *E. cuniculi*. La résolution de la structure de ce même complexe composé des protéines de levure par une autre équipe montre une forte conservation de l'architecture des deux protéines et de leur mode de liaison. Les différentes études de génétique de levure qui ont été réalisées montrent clairement l'implication de Spt6 et Iws1 dans les mécanismes transcriptionnels. Ces différents résultats obtenus à partir de levures ou de parasites doivent cependant encore être confirmés chez les eucaryotes supérieurs. En effet, dans les cellules de souris, Iws1 interagit avec Spt6 uniquement dans le cas où Spt6 est préalablement liée à l'ARN polymérase II, suggérant que le mode de liaison des protéines dans ces organismes diffère. Une confirmation pourrait être amenée par la réalisation de tests d'expression utilisant les protéines humaines.

L'analyse des différentes structures qui contiennent Iws1, et notamment du fait de la présence d'une large poche chargée positivement à la surface de la protéine, nous a conduit à tester l'interaction de Iws1 avec l'ARN polymérase II phosphorylée. Bien qu'une interaction ait pu être détectée *in vivo*, aucune interaction directe entre la protéine et des peptides phosphorylés n'a pu être observée. Il semble donc que Iws1 interagissent bien avec l'ARN polymérase II, comme cela a déjà été proposé chez la levure, mais cette interaction pourrait être médiée par une autre protéine. Il paraît peu probable que cette protéine soit Spt6, puisque Iws1 peut être recrutée indépendamment de Spt6 chez la levure. Ainsi, Iws1 est très fortement liée aux mécanismes transcriptionnels, avec ou indépendamment de Spt6, son mode de recrutement restant encore à définir lorsque Spt6 est absente.

Nos travaux nous ont également amenés à tester une interaction entre Iws1 et plusieurs protéines impliquées dans les mécanismes épigénétiques et l'export : Set2 et Yra1p. Cependant, nos tests de co-expression n'ont pas permis de mettre en évidence la formation d'un complexe entre Iws1 et ces deux protéines, ces premiers tests ayant été réalisés dans les bactéries. La répétition de ces tests dans un système d'expression eucaryote est donc nécessaire afin de conclure quant à la formation de ces complexes. Cependant, les interactions Iws1/Setd2 et Iws1/REF1/Aly ont été mises en évidence dans des cellules de souris et il apparaît que le mode de recrutement des protéines impliquées dans la transcription pourrait différer entre les eucaryotes supérieurs et la levure, comme c'est éventuellement le cas pour le complexe Spt6/Iws1. Ainsi, la réalisation de tests d'expression utilisant les protéines humaines pourrait s'avérer nécessaire.

## **Chapitre 5**

# **Caractérisation biochimique, fonctionnelle et structurale de l'extrémité amino- terminale de Spt6 et des complexes Spt6/TFIIS, Spt8/TFIIS et Med13/TFIIS**



## 5.1. Introduction – brève description des résultats

La résolution de la structure du complexe Spt6/Iws1, a permis de mettre en évidence la présence, dans le domaine central conservé de Iws1, de deux domaines de type « ARM repeat ». Ces domaines constituent le domaine LW qui a été identifié dans Iws1 et qui est également retrouvé dans le domaine amino-terminal de TFIIS, Med26 et la protéine Elongin A (Ling et al., 2006). La superposition du domaine central de Iws1 au domaine amino-terminal de TFIIS (également appelé domaine I) montre une forte homologie structurale entre ces deux domaines. Cette homologie est notamment très importante au niveau de la poche hydrophobe à la surface de Iws1 qui interagit avec le motif IF de Spt6, les résidus formant cette poche étant très conservés, la seconde zone d'interaction du complexe Spt6/Iws1 n'étant pas conservée sur TFIIS (Figure 6 A et B, publication 2). De plus, un motif IF est présent dans la séquence de deux interactants de TFIIS : Med13, une sous-unité du complexe Mediateur, et Spt8, une sous-unité de SAGA, ces deux complexes étant impliqués dans la régulation de la transcription (Figure 6C, publication 2). Des constructions de Spt8, Med13 et du premier domaine de TFIIS ont été clonées et insérées dans les vecteurs d'expressions. Les tests réalisés ont montré que des fragments courts des protéines Spt8 et Med13 qui contiennent les motifs IF sont capables de retenir la protéine TFIIS (Figure 6D, publication 2). Alors que la mutation du motif de Spt8 abolit son interaction avec TFIIS, la mutation de ce motif dans le fragment de la protéine Med13 ne fait qu'affaiblir l'interaction avec TFIIS, probablement du fait de la présence d'un second motif IF dans le fragment de Med13.

Ces résultats suggèrent aussi que d'autres complexes pourraient être formés, notamment Spt6/TFIIS, Med13/Iws1 et Spt8/Iws1. Alors qu'aucun lien n'a jamais été décrit entre SAGA, le Mediateur et Iws1, il existe une interaction génétique entre Spt6 et TFIIS qui pourrait être expliquée par l'interaction de ces deux protéines (Hartzog et al., 1998). Nous avons donc tenté de comprendre le rôle de l'extrémité amino-terminale de Spt6 en combinant des approches de co-expression, des études structurales et des études fonctionnelles en testant notamment l'interaction de Spt6 avec TFIIS.

Cependant, aucun des complexes Spt8/Iws1 et Med13/Iws1, testés par co-expression, n'ont pu être mis en évidence (données non présentées), suggérant que le mode de liaison entre Med13/Spt8 et TFIIS, et Spt6 et Iws1 est différent. Cette différence de liaison pourrait notamment être expliquée par la présence d'un segment poly-ED à la suite des motifs IF des

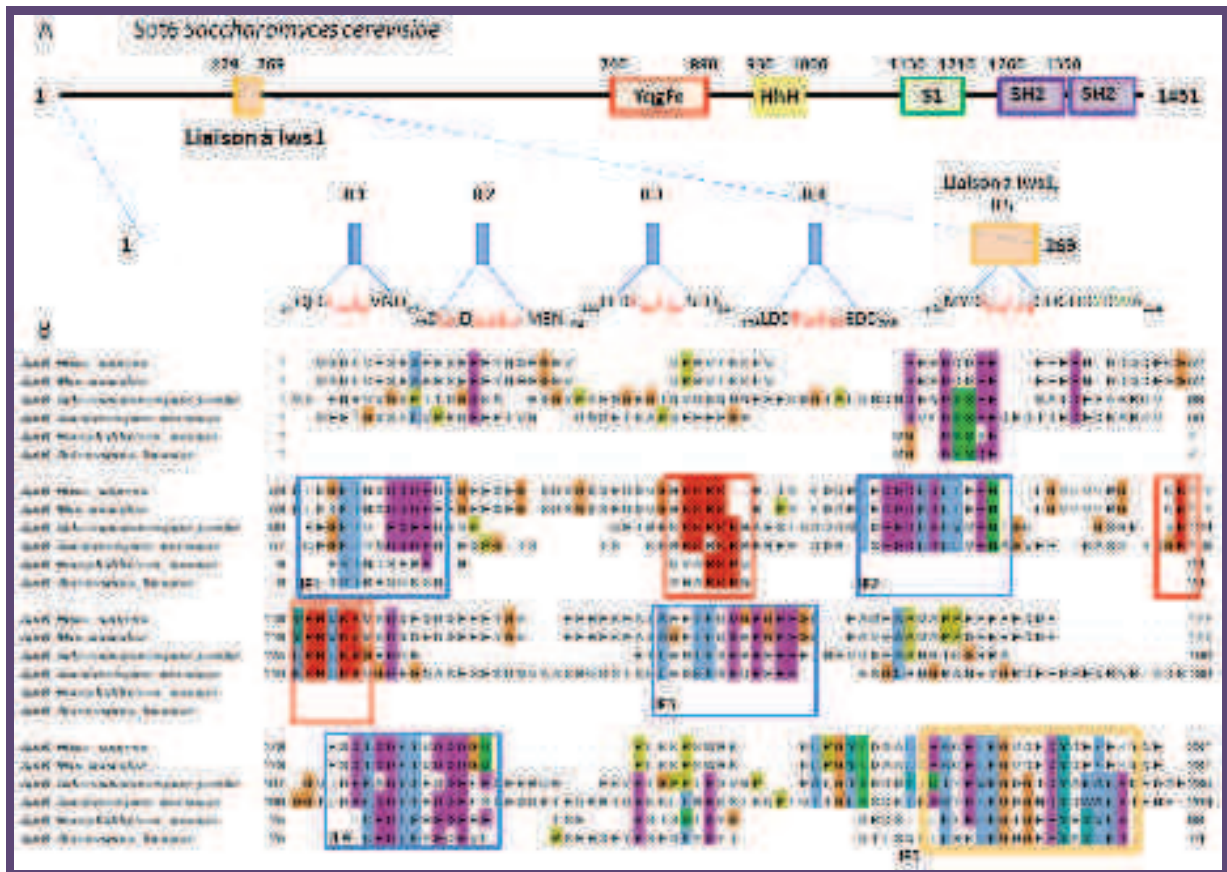
protéines Spt8 et Med13, dont la suppression conduit à la perte des complexes Spt8/TFIIS et Med13/TFIIS (publication 2). Ainsi, l'analyse biochimique et structurale des complexes formés par TFIIS va permettre de comprendre comment ils sont formés *in vivo*.

D'autre part, l'extrémité amino-terminale de Spt6 qui contient le site d'interaction à Iws1 est très peu caractérisée. L'étude de sa séquence montre qu'une grande part de cette région est composée de résidus chargés négativement, entrecoupés de motifs hydrophobes dont la séquence est proche du site de liaison de Iws1. D'autre part les prédictions de régions désordonnées montrent que cette région n'est probablement pas structurée. Il semblerait donc que le repliement de cette région de Spt6 soit induit par la liaison de ses partenaires, comme c'est visiblement le cas pour la zone d'interaction à Iws1. Nous avons donc tenté de comprendre quel était le rôle de cette région de Spt6 et plus particulièrement des motifs hydrophobes qu'elle contient.

## 5.2. L'extrémité amino-terminale de Spt6 contient plusieurs motifs hydrophobes conservés

Nos observations suggèrent que Spt6 pourrait recruter non seulement Iws1 mais aussi TFIIS, voire d'autres protéines à motif LW. Nous avons également constaté que le domaine amino-terminal de Spt6 contient d'autres motifs hydrophobes qui pourraient recruter Iws1 et TFIIS.

Pour cela, les séquences homologues à la protéine Spt6 de levure ont été identifiées par une recherche d'homologie de séquence avec l'algorithme Psi-blast. Les séquences homologues détectées ont ensuite été alignées avec le programme Mafft. L'analyse de l'alignement multiple de séquence montre que certaines régions de la protéine sont fortement conservées dans tous les organismes, les séquences reliant ces différents motifs conservés étant plus variables. Ainsi, les séquences des protéines d'*E. cuniculi* et *A. locustae* sont plus courtes que les séquences des eucaryotes supérieurs, l'extrémité amino-terminale de ces protéines étant plus courtes, de même que les séquences reliant les régions conservées, ce qui est probablement le reflet de l'évolution réductive subit par ces organismes (Peyretailade et al., 2011). Cependant, des motifs conservés peuvent être mis en évidence (Figure 56).



**Figure 56 : Motifs hydrophobes de l'extrémité amino-terminale de Spt6**

(A) Organisation des domaines constituant la protéine Spt6 de *Saccharomyces cerevisiae*. Les domaines YqgFc, S1 et le premier domaine SH2 ont été identifiés à l'aide d'Interproscan, les domaines de liaison à Iws1 et le deuxième domaine SH2 ayant été complétés à partir des résultats obtenus précédemment (publications numéro 1 et 2). (B) Alignement de séquence multiple de l'extrémité amino-terminal de la protéine Spt6, généré avec le programme Mafft. Cet alignement permet de mettre en évidence les cinq motifs hydrophobes notés IF1 à IF5 (ce dernier correspond au site de reconnaissance de Iws1).

Ainsi, cinq motifs hydrophobes, notés IF1 à IF5, peuvent être identifiés dans les séquences des protéines Spt6 de tous les organismes, certains de ces motifs étant ponctuellement absents des séquences de certains organismes. C'est notamment le cas des motifs IF2 et IF3 qui sont absents des protéines d'*E. cuniculi* et *A. locustae* (Figure 56). Ces motifs hydrophobes sont invariablement entourés de régions riches en résidus chargés négativement (aspartate et glutamate), les régions séparant ces motifs étant peu conservées au niveau de la séquence. Le motif noté IF5 correspond à la zone d'interaction avec la protéine Iws1 qui a été définie précédemment (publication numéro 2). En plus des cinq motifs hydrophobes, deux motifs conservés chargés positivement sont également observés qui

pourraient éventuellement constituer un signal de localisation nucléaire (NLS) ou être impliqués dans des interactions protéine-protéine.

Trois constructions de l'extrémité amino-terminale de la protéine Spt6 de levure ont été définies afin de tester le rôle de ces différents motifs (Figure 57A). En effet, du fait de l'absence de certains motifs hydrophobes dans les séquences de la protéine Spt6 d'*E. cuniculi* et *A. locustae*, la protéine de levure a été utilisée pour la réalisation de ces tests. Les constructions définies ne contiennent pas le site de liaison d'Iws1 (motif IF5) et permettent de supprimer successivement chaque motif (Figure 57A). Ces constructions ont par la suite été utilisées afin de tester par co-expression les interactants de ces différents motifs.

### 5.3. Etude des interactions réalisées par l'extrémité amino-terminale de Spt6

Deux interactants principaux ont été testés dans un premier temps : les protéines Iws1 et TFIIS. Le but de ces tests est de définir (i) si la protéine Iws1 reconnaît d'autres motifs hydrophobes que le motif IF5 mais également (ii) si la protéine TFIIS interagit avec l'extrémité amino-terminale de Spt6.

#### 5.3.1. Iws1 n'interagit qu'avec le motif IF5 de Spt6

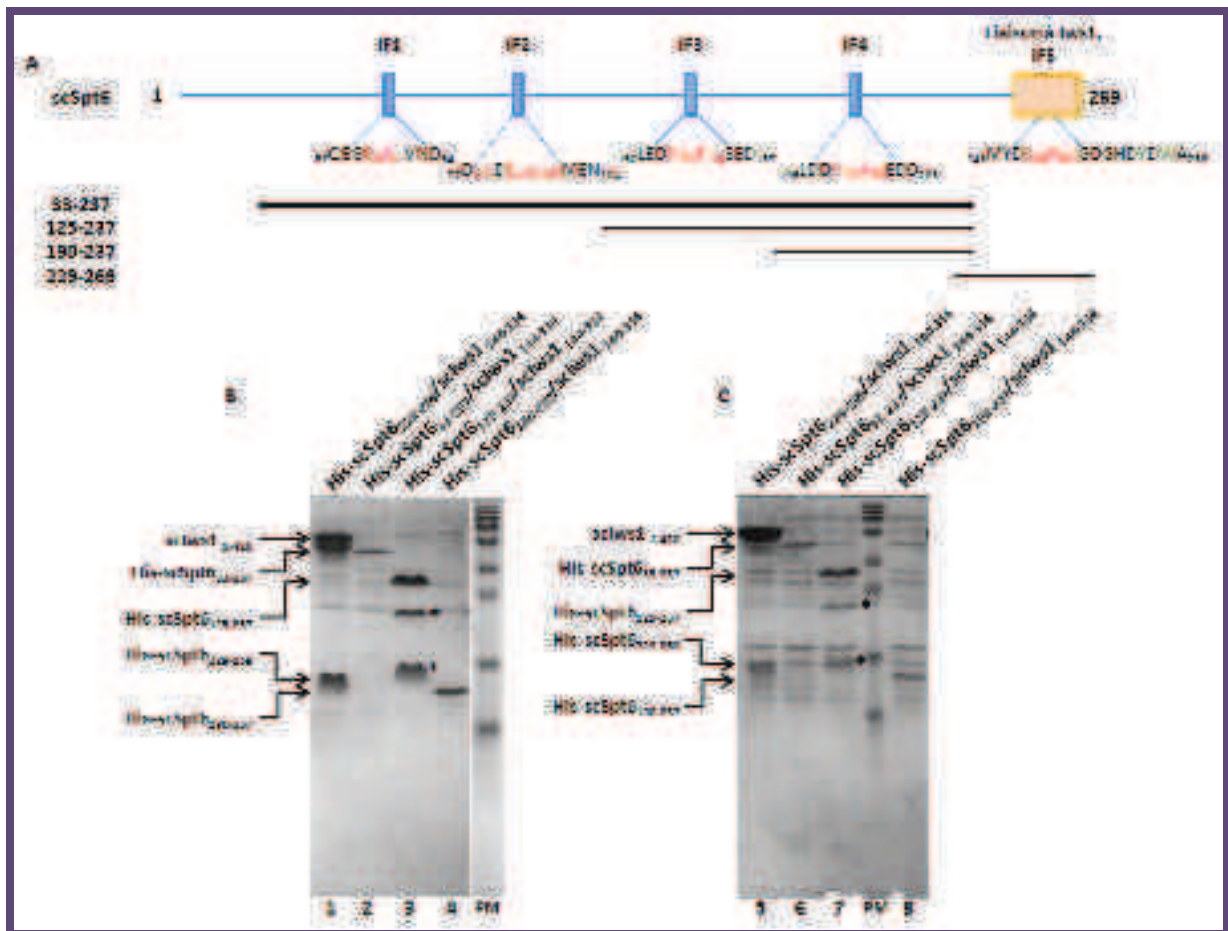
Dans un premier temps, l'interaction de Iws1 avec les différents motifs hydrophobes a été testée. En effet, l'homologie de séquence entre ces différents motifs hydrophobes pourrait suggérer que Iws1 pourrait interagir avec plusieurs motifs.

Des tests de co-expression ont donc été réalisés avec une construction de la protéine Iws1 entière de levure (construction 1-410) et les différentes constructions de l'extrémité amino-terminale de la protéine Spt6 (Figure 57A). Deux conditions de lyses ont été testées contenant 50mM et 400mM de sel, le complexe Spt6/Iws1 défini précédemment étant formé à forte concentration saline.

Ces tests confirment que le complexe Spt6/Iws1 décrit précédemment (publication numéro 2) est formé aussi bien à haut sel (ligne 5) qu'à bas sel (ligne 1), le complexe étant plus soluble lorsque la concentration saline est basse. Par contre, la co-expression de la protéine scIws1 entière avec les constructions (33-237) (ligne 2 et 6), (125-237) (ligne 3 et 7) et (190-237) (ligne 4 et 8) de Spt6 ne permet pas de former un complexe stable entre les deux protéines, confirmant que seul le motif IF5 définit précédemment (publication numéro 2)

permet la formation d'un complexe stable. Ces tests montrent également que les fragments contenant les motifs IF1 à IF4 (construction 33-237) et IF4 (construction 190-237) sont faiblement solubles, la construction contenant les motifs IF3 et IF4 (construction 125-237) étant bien exprimée mais dégradée en deux fragments.

Ainsi, la protéine Iws1 n'interagit qu'avec le motif IF5 de la protéine Spt6. Cette spécificité pourrait s'expliquer par la présence des résidus Tyrosine 255 et Tryptophane 257 de la seconde zone d'interaction de Spt6 (Figure 57A) qui sont spécifiques au motif IF5 et qui permettent la formation de liaisons hydrogènes et de contacts hydrophobes à la surface de la protéine Iws1 (cf. structure du complexe Spt6/Iws1, publication 2).



**Figure 57 : Test de l'interaction de Iws1 avec les motifs hydrophobes de l'extrémité amino-terminale de Spt6**

Les différents fragments de l'extrémité amino-terminale portant une étiquette de purification (A) ont été co-exprimés avec la protéine Iws1 entière, exprimée sous forme native. Deux tampons de lyse ont été testés, contenant 50mM NaCl (B) et 400mM NaCl (C). (\*) Dégradation du fragment 125-237 de Spt6 de levure.



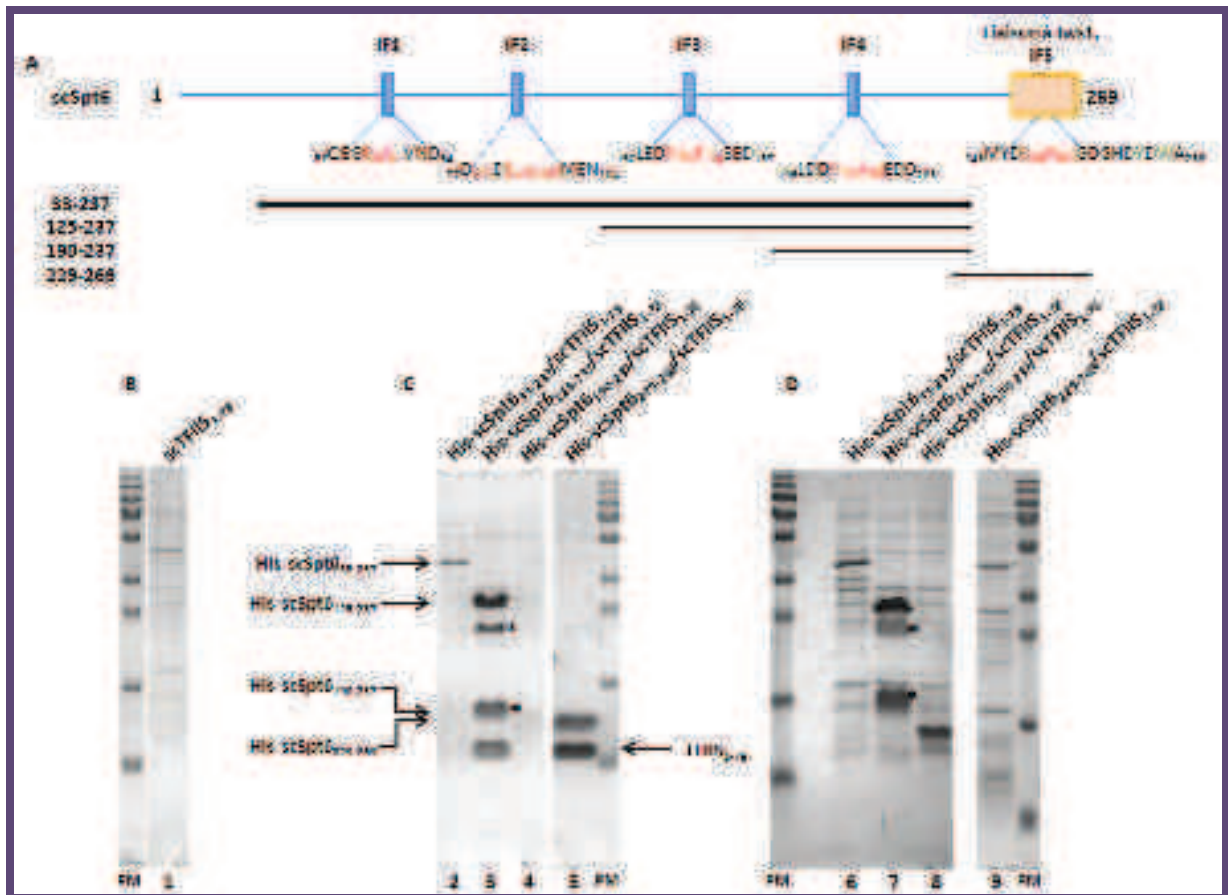
### 5.3.2. Interaction avec TFIIIS

L'interaction entre Spt6 et TFIIIS a été suggérée par une interaction génétique entre ces deux protéines et les différentes homologues structurales décrites plus haut. Afin de vérifier si un complexe entre Spt6 et TFIIIS était formé, des tests de co-expression ont été réalisés entre TFIIIS et les différentes constructions de l'extrémité amino-terminale de Spt6.

#### 5.3.2.1. TFIIIS interagit avec deux motifs hydrophobes conservés de Spt6

Dans un premier temps, afin de définir si un complexe stable était formé entre Spt6 et TFIIIS, des tests de co-expression ont été réalisés entre le domaine amino-terminal de TFIIIS (domaine I résidus 1-78) et (i) les trois constructions de l'extrémité amino-terminale de Spt6 ainsi qu'avec (ii) la construction du motif IF5 de Spt6 (résidus 229-269) qui interagit avec Iws1 (Figure 58). En effet, bien que la formation d'un complexe entre le motif IF5 de Spt6 et Iws1 ait déjà été caractérisé, on ne peut exclure que ce motif de Spt6 interagisse également avec TFIIIS.

Ces tests montrent que deux fragments de Spt6 interagissent avec TFIIIS, mais uniquement lorsque la concentration saline est faible. L'un de ces fragments contient le motif IF5 (ligne 5), qui est également reconnu par Iws1. En effet, la présence de deux bandes stœchiométriques sur le gel montre que l'interaction observée ne peut résulter d'une interaction de TFIIIS avec la résine, ce qui a également été exclu expérimentalement (Figure 58B, ligne 1). Le second fragment qui interagit avec TFIIIS contient les motifs IF3 et IF4 (ligne 3), l'absence d'interaction avec le motif ne contenant que le domaine IF4 (ligne 4) montre qu'il s'agit du motif IF3 qui interagit avec TFIIIS. Ces deux complexes ne sont pas formés lorsque la concentration saline est élevée (lignes 7 et 9), suggérant que l'interaction mise en place est faible et probablement majoritairement ionique.



**Figure 58 : Interaction de TFIIIS avec les motifs hydrophobes de l'extrémité amino-terminale de Spt6.**

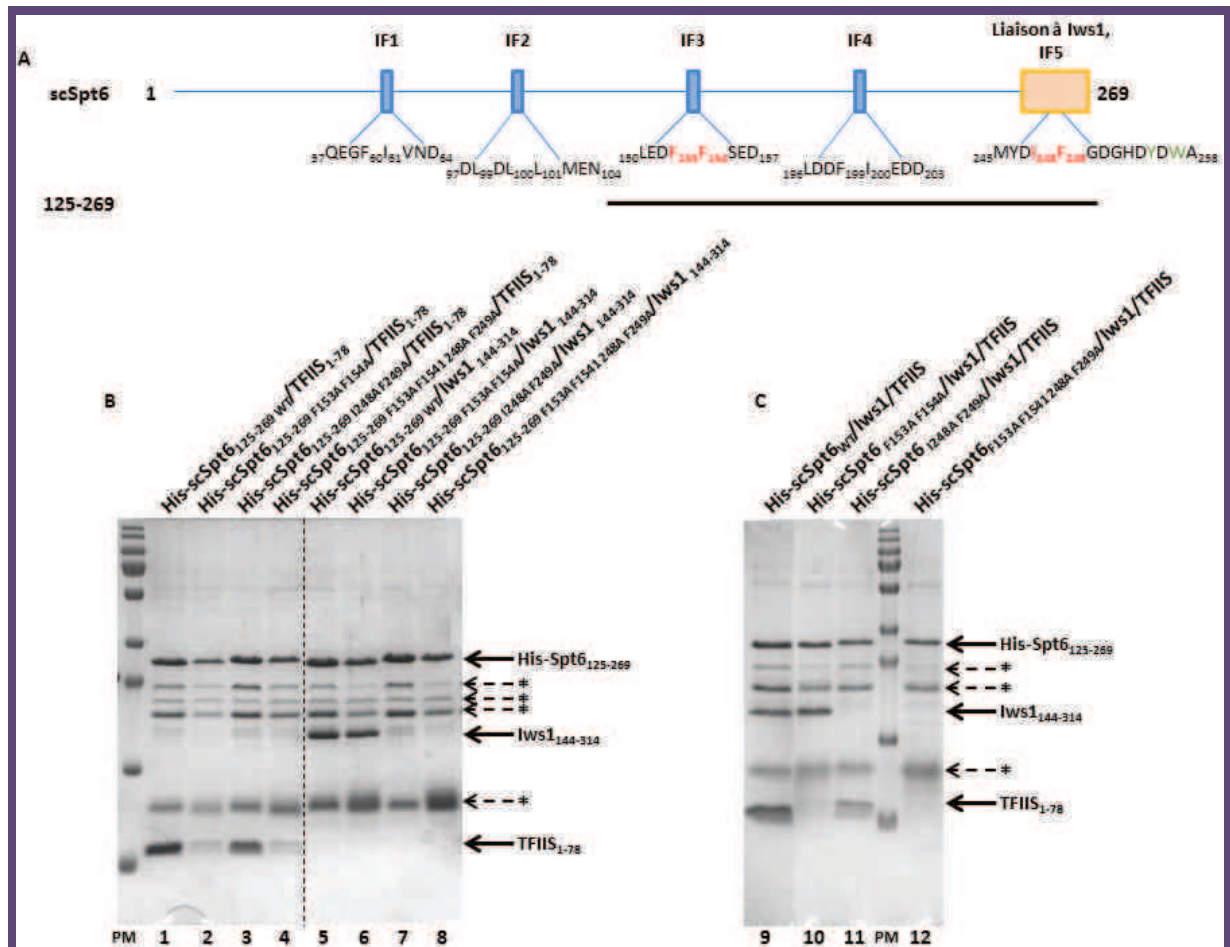
(A) Bornes des différentes constructions du domaine amino-terminal de la protéine Spt6 utilisées pour la réalisation des tests de co-expression. (B) Test de l'interaction aspécifique de TFIIIS à la résine. Co-expression des différentes constructions de Spt6 avec la construction 1-78 de TFIIIS (C-D), le tampon de lyse utilisé contenant 50mM de sel (C) et 400mM de sel (D). (\*) Dégradation du fragment 125-237 de Spt6.

Ces premiers tests de co-expression montrent clairement que la protéine TFIIIS interagit avec deux motifs de l'extrémité amino-terminale de Spt6. Cependant, la capacité du fragment IF5 de Spt6 à interagir aussi bien avec Iws1 qu'avec TFIIIS soulève la question de la spécificité des interactions mises en place et du mode de liaison des protéines.

### 5.3.2.2. Spécificité de l'interaction Spt6/TFIIIS

Afin de tester la spécificité des deux interactions définies, des expériences de co-expression entre un fragment de Spt6 contenant les motifs IF3, IF4 et IF5 (construction 125-269) et la protéine TFIIIS ont été réalisées. Pour cela, le fragment sauvage de Spt6 a été utilisé ainsi que différentes constructions où chaque motif IF a été muté l'un après l'autre (Figure 59). Les différents fragments ont ensuite été co-exprimés avec le domaine amino-terminal de

TFIIS (construction 1-78) mais également avec le domaine central de scIws1 (construction 144-314). Une seule concentration saline a été testée, le complexe Spt6/TFIIS n'étant formé qu'à faible concentration, le complexe Spt6/Iws1 étant également formé à cette concentration.



**Figure 59 : Spécificité des complexes Spt6/TFIIS et Spt6/Iws1 - Interaction de TFIIS et Iws1 avec un fragment muté de Spt6.**

(A) Le fragment de Spt6 125-269 qui contient les motifs IF3, IF4 et IF5 mutés ont été co-exprimés (B) avec les protéines TFIIS et Iws1 individuellement (C) mais également avec les deux protéines en même temps. (\*) Dégradation du fragment 125-269 de Spt6.

La co-expression avec les fragments mutés de Spt6 montre que TFIIS interagit préférentiellement avec le motif IF3. En effet, lorsque les résidus F153 et F154 du motif IF3 sont mutés, TFIIS est retenu plus faiblement par Spt6 (ligne 2), tandis que la mutation des résidus I248 et F249 du motif IF5 ne semble pas affecter la formation de ce complexe (ligne 3). De plus la mutation des deux motifs hydrophobes (ligne 4) ne permet pas d'abolir complètement la formation du complexe Spt6/TFIIS, une faible quantité de TFIIS étant encore retenue de manière identique à ce qui est observé lors de la mutation des résidus du motif IF3 (ligne 2). Cette interaction résiduelle suggère que la zone d'interaction ne se limite

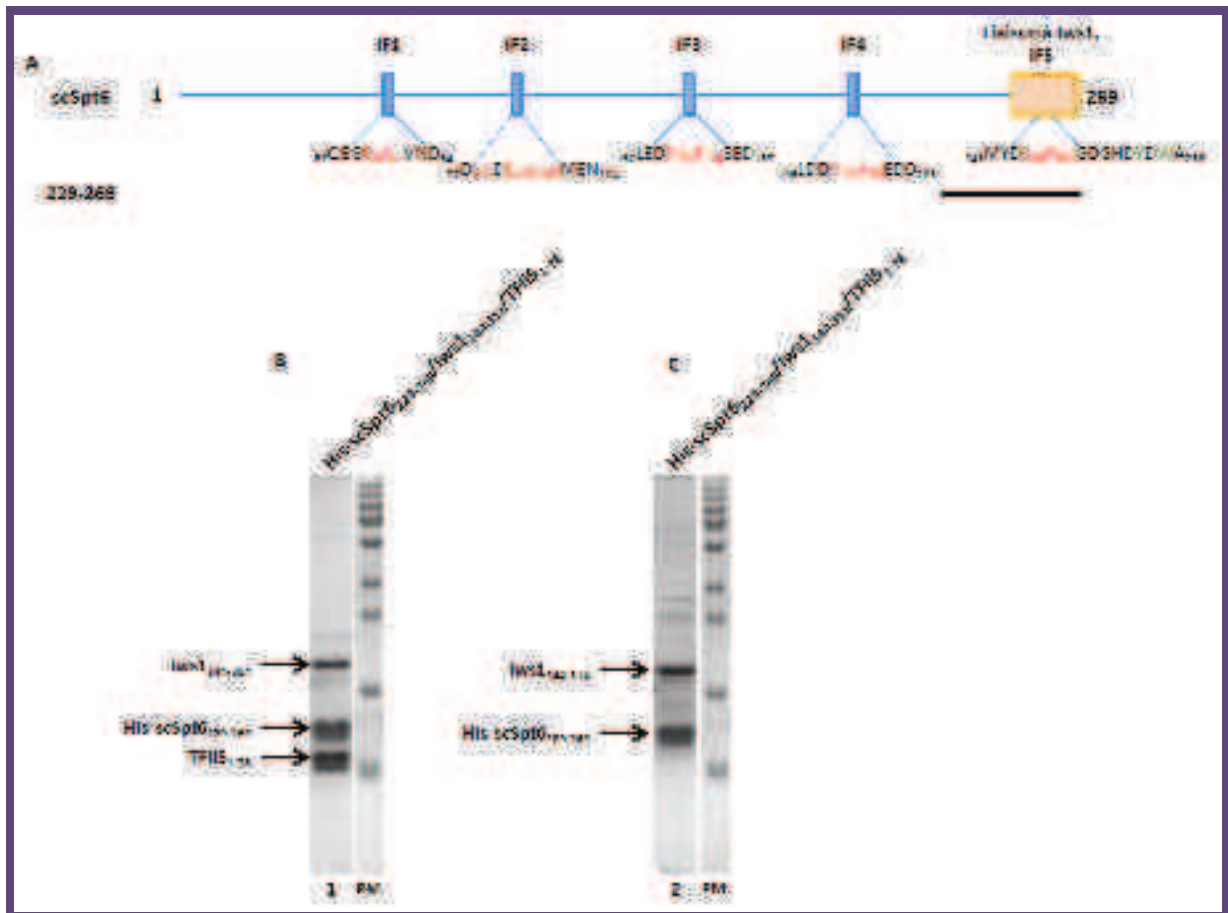
pas aux résidus mutés mais est plus étendue et implique d'autres résidus des deux protéines (l'absence d'interaction aspécifique de TFIIIS ayant été montrée précédemment, Figure 58B)

Les différents fragments mutés de Spt6 ont également été co-exprimés avec le domaine central de la protéine Iws1. Ces tests de co-expression montrent que les fragments de Spt6 dont les motifs IF5 sont mutés n'interagissent pas avec Iws1 (ligne 7 et 8), la mutation du motif IF3 n'affectant pas la formation du complexe (ligne 6). Dans ce cas Iws1 apparaît comme reconnaissant spécifiquement le motif IF5, confirmant nos données structurales et biochimiques antérieures.

Ces tests suggèrent donc que Spt6 interagit avec TFIIIS via son motif IF3 alors que la protéine interagit avec Iws1 via son motif IF5. L'implication de deux motifs distincts dans la formation de ces deux complexes suggère qu'un complexe entre ces trois protéines pourrait être formé. Les gènes codant pour la protéine TFIIIS et le domaine central de Iws1 ont donc été concaténés afin de pouvoir co-exprimer ces deux protéines sous leur forme native. Ce vecteur a ensuite été utilisé afin de réaliser des tests de co-expression avec les mêmes fragments mutés de la protéine Spt6 utilisés précédemment (Figure 59C), ces tests s'apparentant à des expériences de compétition.

La co-expression de Iws1/TFIIIS avec la forme native du fragment 125-269 de Spt6 montre clairement qu'un complexe formé des trois protéines peut être formé (Figure 59C, ligne 9), la mutation des motifs IF3 et IF5 abolissant la formation de ce complexe ternaire (ligne 12). Individuellement, la mutation du motif IF5 empêche la fixation de la protéine Iws1, la protéine TFIIIS étant toujours retenue par le fragment de Spt6 (ligne 11). D'autre part, la mutation du motif IF3 n'empêche pas la fixation d'Iws1 mais n'abolit pas complètement l'interaction avec TFIIIS, l'interaction observée étant très faible (ligne 10). Ainsi, l'interaction entre TFIIIS et le motif IF3 ne peut pas être complètement supprimée par la mutation de deux résidus dans ce motif, comme observé précédemment.

Une question restante était de comprendre pourquoi le fragment de TFIIIS est capable de fixer fortement le motif IF5 de Spt6, comme nous l'avons observé. Pour cela, des tests de co-expression ont été réalisés entre la construction native du motif IF5 de Spt6 (construction 229-269), connue pour interagir avec chacune des protéines, le domaine amino-terminal de TFIIIS (construction 1-78) et le domaine central de Iws1 (construction 144-314). Deux concentrations salines ont été testées puisque le complexe Spt6/TFIIIS est formé à bas sel alors que le complexe Spt6/Iws1 est formé aussi bien à haut sel qu'à bas sel.



**Figure 60 : Co-expression du complexe Spt6 IF5/Iws1/TFIIS**

*Le fragment 229-269 de Spt6, qui contient le motif IF5 seul (A) a été co-exprimé avec les protéines Iws1 et TFIIS. Deux concentrations salines ont été testées au cours de la lyse des cellules et des lavages : (B) 50mM NaCl, 10mM tris pH8 et (C) 400mM NaCl, 10mM Tris pH8.*

Alors qu'à forte concentration saline seul le complexe Spt6 IF5/Iws1 est formé (Figure 60C), à basse concentration saline (50mM NaCl) un complexe entre les trois protéines peut être formé (Figure 60B, ligne 1). En effet, trois bandes de même intensité sont observées sur le gel ce qui exclut la liaison aspécifique à la résine de l'une des protéines ne portant pas d'étiquette de purification. Il semble donc que la co-expression des trois protéines dans le cadre d'un fragment court de Spt6 contenant uniquement le motif IF5 fait apparaître une interaction qui n'est pas observée lors de la co-expression avec une construction longue de Spt6 (Figure 59C). En effet lors de la co-expression de Iws1 et TFIIS avec la construction 125-269 de Spt6 dont le motif IF3 est muté, un complexe stœchiométrique entre Spt6 et Iws1 était observé, TFIIS étant retenu plus faiblement. Une hypothèse est que les séquences qui entourent le motif IF5, qui sont présentes dans la construction longue de Spt6 et pas dans la

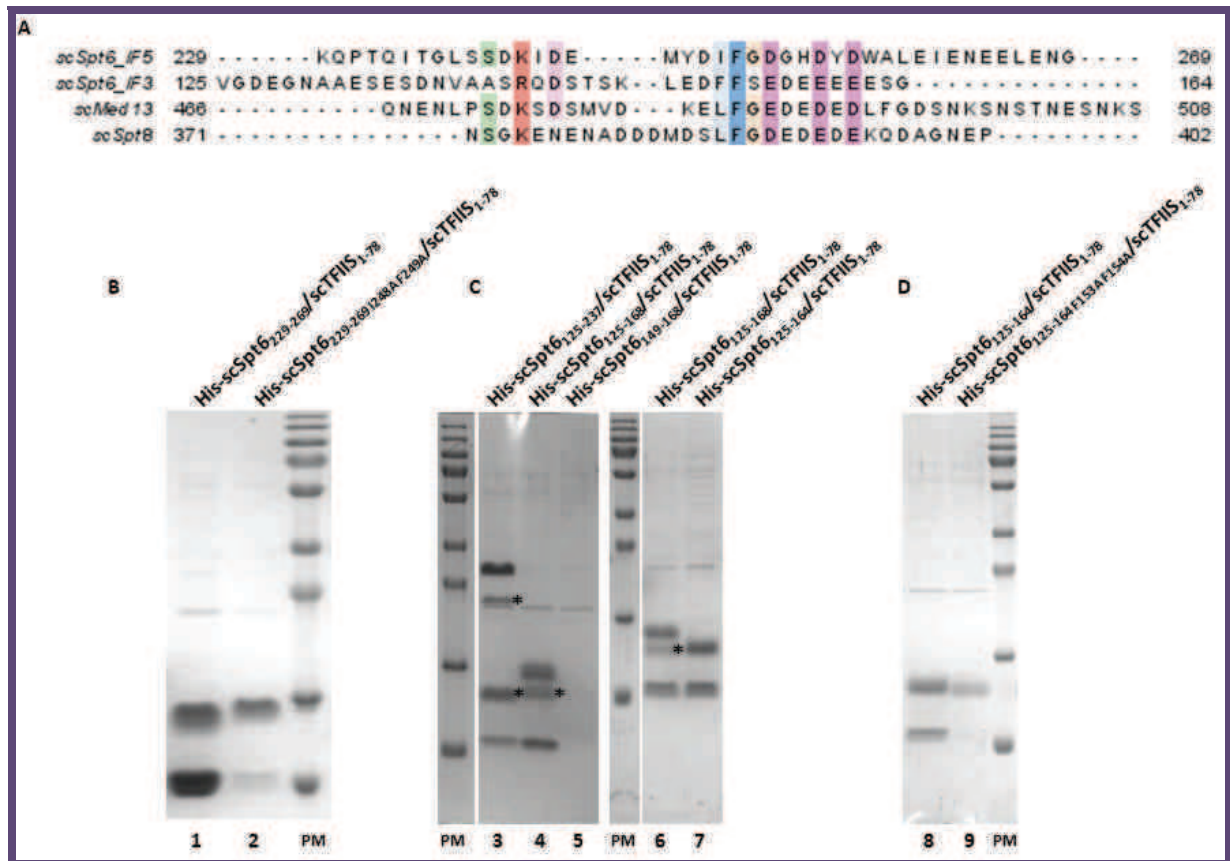
plus courte, pourraient destabiliser le complexe Spt6 IF5/TFIIS et ainsi favoriser la liaison de TFIIS sur le motif IF3

En résumé, l'ensemble de ces résultats montre que (i) Iws1 interagit avec Spt6 uniquement via le motif IF5 de la protéine, (ii) que TFIIS interagit avec deux motifs de Spt6 mais semble lier préférentiellement le motif IF3, et (iii) qu'un complexe peut être formé entre TFIIS, Iws1 et Spt6 contenant les deux motifs IF3 et IF5. Un certain nombre de questions restent en suspens, montrant l'intérêt d'une caractérisation structurale des complexes Spt6/TFIIS qui devrait permettre de comprendre comment ces complexes sont formés.

### 5.3.2.3. Caractérisation biochimique des complexes Spt6/TFIIS

Un problème posé par cette étude est que la construction longue de Spt6 contenant les motifs IF3, IF4 et IF5 est immédiatement dégradée au cours de la purification, empêchant la caractérisation structurale du complexe ternaire. Nous avons donc décidé d'entreprendre la caractérisation structurale des complexes Spt6 IF3/TFIIS et Spt6 IF5/TFIIS, ainsi que du complexe Spt8/TFIIS que nous avons caractérisé dans la publication numéro 2. Dans un premier temps, les régions minimales d'interaction des complexes Spt6/TFIIS ont été définies afin de pouvoir produire des complexes stables. Dans le cas du complexe Spt6 IF5/TFIIS, la construction 229-269 de la protéine Spt6 décrite précédemment a été utilisée. En effet, cette construction de la protéine ne subit pas de dégradation et la mutation des résidus I247 et F248 affaiblit fortement le complexe (Figure 61B).

Dans le cas du complexe Spt6 IF3/TFIIS, différentes constructions contenant le motif IF3 de scSpt6 ont été clonées afin de définir quelle région de Spt6 permet de produire un complexe stable. Ces différentes constructions ont été définies à partir d'un alignement multiple entre les motifs IF3 et IF5 de Spt6 et les motifs IF des protéines Med13 et Spt8 interagissant avec TFIIS (Figure 61A). Des tests de co-expression ont été réalisés entre ces constructions et le domaine I de TFIIS (construction 1-78) en testant uniquement une condition saline basse (Figure 61C).



**Figure 61 : Caractérisation biochimique des complexes Spt6 IF3/TFIIS et Spt6 IF5/TFIIS.**

(A) Alignement des séquences des fragments des protéines Spt6 (IF3 et IF5), Spt8 et Med13 interagissant avec TFIIS. (B) Co-expression du fragment Spt6<sub>229-269</sub> (IF5) dont les résidus IF ont été muté avec TFIIS<sub>1-78</sub>. (C) Différentes constructions contenant le motif IF3 de Spt6 ont été co-exprimées avec le domaine I de TFIIS<sub>1-78</sub> afin de définir la zone d'interaction minimale de Spt6 (D). Co-expression du fragment minimal d'interaction muté (F153A F154A) avec la protéine TFIIS<sub>1-78</sub>.

La co-expression des différentes constructions du motif IF3 avec TFIIS montre qu'une extension amino-terminale relativement longue est nécessaire à la formation du complexe. En effet, une construction ne possédant que huit résidus avant le motif hydrophobe ne permet pas de retenir la protéine TFIIS (Figure 61C, ligne 5), suggérant que cette région est nécessaire au repliement du fragment de Spt6 ou que des contacts sont formés entre la région en amont du motif IF3 et TFIIS. Ces tests ont également permis de définir les bornes d'un fragment court de Spt6 qui interagit avec TFIIS et n'est pas dégradé (ligne 7). Le fragment ainsi défini possède une longue extension en amont du motif IF et s'arrête juste après l'extension poly-ED en aval du motif (la séquence de ce fragment est présentée sur la Figure 61A), ce fragment migrant au même niveau que le plus petit produit de dégradation de la construction 125-269 de Spt6. L'introduction des mutations F153A et F154A dans ce fragment minimal abolit la formation du complexe (ligne 9).

Un fragment court de Spt6 a ainsi pu être défini permettant la formation d'un complexe stable Spt6 IF3/TFIIS. La définition de ce complexe minimal permet la production d'un échantillon utilisable pour la réalisation d'études structurales qui aideront à comprendre (i) quel motif de Spt6 est reconnu par TFIIS *in vivo* et (ii) le mode de liaison entre ces deux protéines, nos données biochimiques semblant indiquer qu'il diffère du mode de liaison mis en place par le complexe Spt6/Iws1.

## 5.4. Structure du complexe Spt8/TFIIS et implications pour les complexes Spt6/TFIIS

Des études structurales ont donc été entreprises sur trois complexes : (i) le complexe Spt8<sub>371-402</sub>/TFIIS dont la caractérisation biochimique est décrite dans la publication 2, (ii) le complexe Spt6<sub>125-164</sub>/TFIIS décrit plus haut et (iii) le complexe Spt6<sub>229-269</sub>/TFIIS.

### 5.4.1. Purification et essais de cristallisation des complexes Spt6 IF3/TFIIS, Spt6 IF5/TFIIS et Spt8/TFIIS

Les trois complexes ont été produits dans un milieu 2LB, la production des protéines étant induite par l'ajout de 0,75mM d'IPTG. De la même manière qu'au cours des tests d'expression, la protéine TFIIS est exprimée sous forme native, les fragments de Spt8, Spt6 IF3 et Spt6 IF5 portant une étiquette de purification (His)<sub>6</sub> suivie d'un site de coupure par la protéase thrombine à leur extrémité amino-terminale. Après centrifugation des cultures, les cellules sont resuspendues dans un tampon de lyse contenant 50mM NaCl, 10mM Tris pH 8,0, les tests de co-expression réalisés précédemment ayant montré que le complexe n'était formé qu'à faible concentration saline. Le protocole de purification réalisé comprend (i) une étape de chromatographie d'affinité, (ii) une coupure à la thrombine et finalement (iii) une étape de chromatographie d'exclusion afin d'obtenir un échantillon homogène. Après concentration de l'échantillon, des tests de cristallisation ont été réalisés. Un grand nombre de conditions de cristallisation, de températures de cristallisation et de concentration protéique ont ainsi été testées mais aucun cristal n'a pu être obtenu. La structure de la protéine TFIIS résolue par RMN montre la présence d'un grand nombre de résidus lysine à la surface de la protéine, la flexibilité des chaînes latérales de ces résidus ainsi que leur charge pouvant expliquer l'impossibilité de cristalliser cette protéine. Afin de tenter de remédier à ce problème, les trois complexes ont été méthylés sur ces lysines selon un protocole décrit en

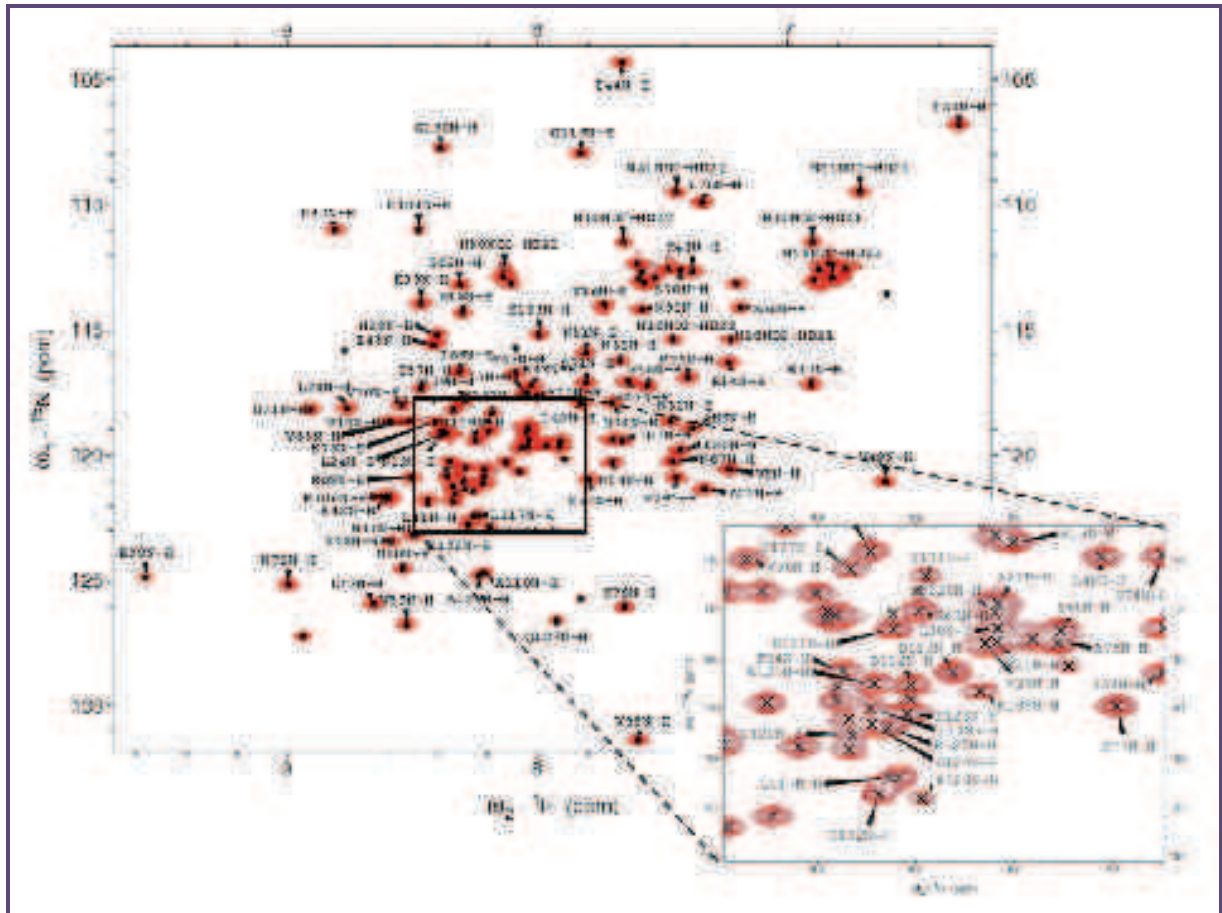


annexe (8.4, page 272) et issu de la publication (Walter et al., 2006), l'introduction de groupement méthyle permettant la formation de contacts hydrophobes. Les tests de cristallisation réalisés à partir de ces protéines méthylées n'ont pas non plus permis d'obtenir des cristaux. La résonance magnétique nucléaire a donc été utilisée pour tenter de résoudre la structure de ces complexes.

#### 5.4.2. Résolution de la structure du complexe Spt8/TFIIS par RMN

Le premier complexe dont la structure a été étudiée est le complexe Spt8/TFIIS. Ce complexe a été produit dans un milieu minimum permettant l'incorporation d'atomes  $^{15}\text{N}$  et  $^{13}\text{C}$ , puis purifié selon un protocole standard. Les spectres permettant l'assignement séquentiel, la détermination des effets Overhauser nucléaires et des phénomènes de relaxation  $T_1/T_2$  ont été enregistrés.

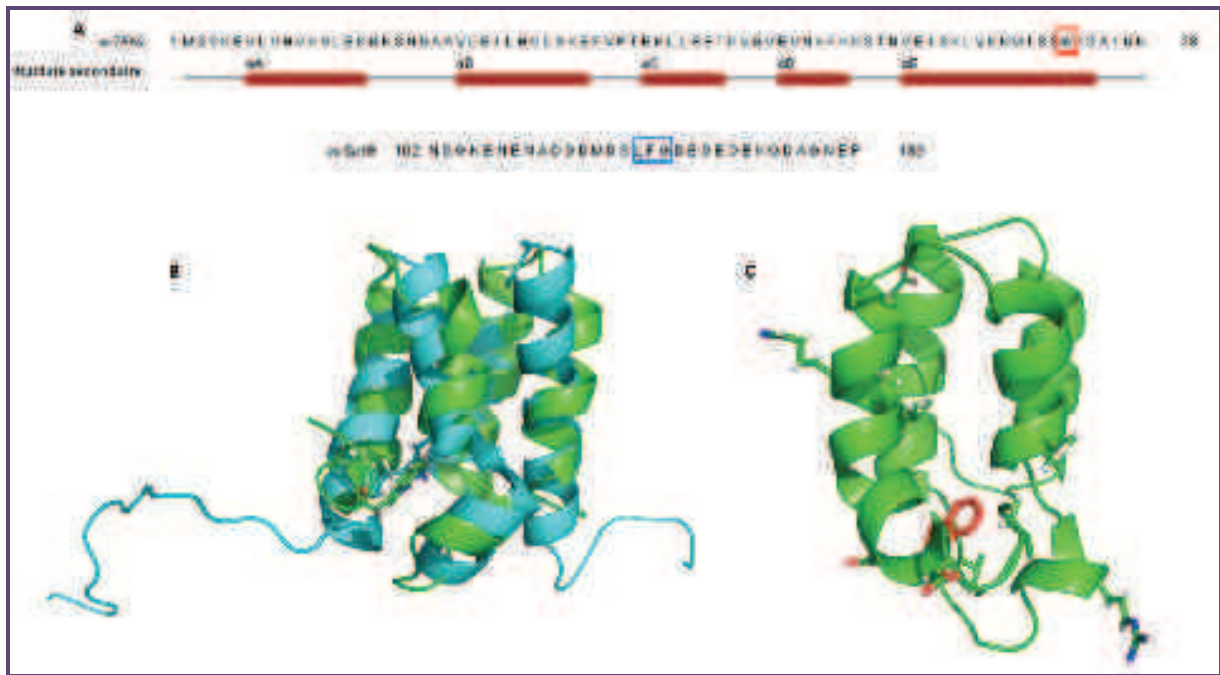
Le spectre  $^{15}\text{N}$ -HSQC enregistré montre la présence de pics de corrélation séparés, qui peuvent donc être attribués. Seule une région centrale du spectre présente un nombre importants de pics de corrélation dont les déplacements chimiques sont très proches mais qui peuvent quand même être distingués (Figure 62). Les différents spectres tridimensionnels ont été utilisés afin de déterminer les déplacements chimiques des atomes qui constituent ce complexe. Ainsi, des déplacements chimiques ont été attribués pour environ 80% des atomes de la structure. La connaissance de ces déplacements chimiques a ainsi permis d'établir un premier modèle provisoire de la structure du complexe grâce au logiciel UNIO.



**Figure 62 : Annotation du spectre 15N-HSQC du complexe Spt8/TFIIS.**

*Les spectres di- et tridimensionnels du complexe Spt8/TFIIS ont été enregistrés sur un spectromètre 700MHz. Les déplacements chimiques de 80% des atomes qui constituent le complexe ont été attribués.*

Ce premier modèle montre une très forte homologie structurale entre la sous-unité TFIIS de notre complexe et la structure du domaine amino-terminal de TFIIS qui a été résolue par RMN (code pdb : 1WJT, résolue par le consortium RIKEN Structural Genomics, non publiée, Figure 63). Dans chacune de ces deux structures, un motif structural de type « ARM repeat » peut être identifié (hélices C, D et E), les hélices  $\alpha A$  et B adoptant un repliement proche des deux dernières hélices du motif ARM1 de Iws1. Dans la publication numéro 2, en nous basant sur l'homologie structurale entre Iws1 et TFIIS, nous avons suggéré que la structure publiée du domaine amino-terminal de TFIIS de levure (Booth et al., 2000) était probablement fautive. Notre premier modèle de la structure de TFIIS de levure tend à confirmer cette hypothèse (Figure 63).



**Figure 63 : Modèle de la structure du domaine amino-terminal de TFIIS résolue par RMN**

(A) séquence des fragments des protéines TFIIS et Spt8 utilisés pour former le complexe Spt8/TFIIS.

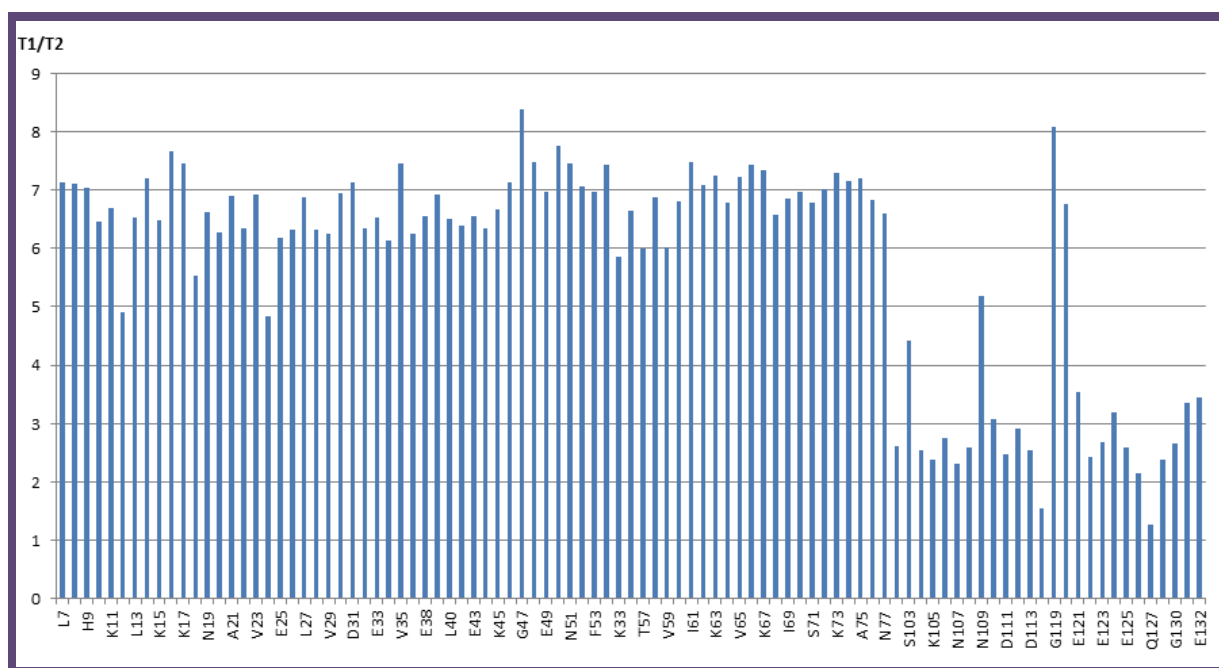
Les éléments de structure secondaire identifiés sur la structure de TFIIS qui a été résolue sont indiqués, le tryptophane qui constitue le fond de la poche hydrophobe est indiqué en rouge, le motif IF (séquence LFG) de Spt8 étant indiqué en bleu (B) Superposition de la structure du domaine amino-terminal de TFIIS de souris (cyan) sur la structure de TFIIS de levure issue de notre complexe (vert). (C) Résidus dont les déplacements chimiques sont fortement modifiés sur les spectres 15N-HSQC que nous avons et du domaine amino-terminal de TFIIS de levure (Booth et al., 2000).

Ce premier modèle ne permet pas encore de positionner le fragment de Spt8 à la surface de TFIIS. En effet, une inspection manuelle des contacts NOE intermoléculaire va être nécessaire afin de définir précisément quelle région de TFIIS est impliquée dans la formation du complexe.

Cependant, la comparaison des spectres 15N-HSQC que nous avons obtenus et celui obtenu par Booth et ses collaborateurs lors de l'étude du domaine amino-terminal de TFIIS de levure permet de formuler des hypothèses quant à la zone d'interaction. En effet, l'interaction d'un peptide à la surface de la protéine va modifier l'environnement des résidus impliqués dans l'interaction, des modifications des déplacements chimiques étant alors attendus qui vont conduire au déplacement des pics de corrélation sur le spectre 15N-HSQC. L'observation des déplacements de ces pics va permettre de suggérer une zone d'interaction (Figure 63B). Cette analyse ne montre pas de modification des déplacements chimiques du tryptophane conservé de TFIIS qui forme le fond de la poche hydrophobe dans laquelle le motif IF de Spt6 interagit dans le complexe Spt6/Iws1. Par contre les résidus qui forment les parois de cette poche sont

fortements déplacés, suggérant qu'ils participent à l'interaction. D'autres résidus à la surface de TFIIS présentent également de fortes modifications de leur déplacement chimique, leur implication dans la formation du complexe devant encore être confirmée. Il semble cependant que les résidus à la surface du motif ARM2 ne sont pas impliqués dans la formation du complexe, comme c'est le cas dans le complexe Spt6/Iws1.

Le mode de liaison différent du complexe Spt8/TFIIS par rapport au complexe Spt6/Iws1 est confirmé par l'identification de feuillets beta dans le peptide de Spt8. En effet, l'analyse des déplacements chimiques du peptide montre que sa structure n'est pas hélicoïdale, les déplacements chimiques étant plutôt caractéristiques de feuillets beta. L'implication des résidus du motif IF de Spt8 (résidus L117/F118, Figure 63A) dans la formation du complexe est confirmée par l'analyse des résultats de dynamique de relaxation T1/T2 (Figure 64).



**Figure 64 : Relaxation T1/T2 du complexe Spt8/TFIIS**

*L'intensité des pics de corrélation du spectre 15N-HSQC a été mesurée après différents temps de relaxation T1 et T2, permettant de définir la dynamique du complexe. Résidus 7-77 : TFIIS, résidus 103 – 132 : Spt8*

En effet, ces études montrent que la sous-unité TFIIS de notre complexe est très peu flexible, ce qui est reflété par un rapport T1/T2 élevé alors que, globalement, le fragment de Spt8 est plus flexible. Cependant, dans ce fragment, une zone contenant les résidus du motif IF ainsi que quelques résidus en amont (notamment l'aspartate 120) semble également fixés,

les valeurs du rapport T1/T2 pour ces résidus étant plus élevés, en accord avec une possible interaction de ces résidus à la surface de TFIIIS. Ainsi, les résidus qui constituent la séquence poly-ED semblent être impliqués dans l'interaction.

Ces différentes données préliminaires suggèrent donc déjà que le mode de liaison mis en place par TFIIIS avec les protéines contenant un motif hydrophobe est différent du mode de liaison du complexe Spt6/Iws1, bien que le même motif IF soit impliqué dans l'interaction.

Pour comprendre si le mode de liaison particulier observé dans le cas du complexe Spt8/TFIIIS est également mis en place dans le cas des deux complexes Spt6/TFIIIS, nous avons également entrepris des études structurales par RMN sur ces deux complexes. Pour cela, les deux complexes impliquant des fragments de Spt6 ont également été produits et purifiés et des spectres 2D <sup>15</sup>N HSQC ont été enregistrés. La superposition de ces spectres au spectre obtenu dans le cas du complexe Spt8/TFIIIS montre que la position de la majorité des pics de corrélation qui appartiennent à TFIIIS est fortement conservée suggérant que la structure globale de TFIIIS est proche de celle décrite précédemment. Il reste cependant encore à enregistrer les spectres permettant l'attribution des déplacements chimiques des différents atomes qui constituent ces deux complexes.

Ces premières données structurales confirment donc que le repliement du domaine amino-terminal de TFIIIS est constitué d'un motif de type « ARM repeat », comme le montrait la structure de la protéine de souris. De plus, ces premières données indiquent que le mode de liaison de Spt8 à la surface de TFIIIS est différent de celui observé sur le complexe Spt6/Iws1, ce qui était relativement inattendu. En effet, Spt8 semble adopter un repliement en feuillet β, les résidus de TFIIIS impliqués dans l'interaction n'étant visiblement pas localisés au niveau du motif ARM de TFIIIS. Par contre il semblerait que les résidus en aval du motif IF de Spt8 participent à la formation du complexe.

## 5.5. Etude fonctionnelle des motifs hydrophobes conservés de l'extrémité amino-terminale de Spt6

Afin de comprendre le rôle des différents motifs hydrophobes de l'extrémité amino-terminale de Spt6 *in vivo*, ces différents motifs ont été mutés et l'effet de leur mutation a été observé chez la levure. Ces différentes mutations ponctuelles (Tableau 14) ont été introduites dans un vecteur de levure contenant le gène entier de Spt6 et ont ensuite été transformés par

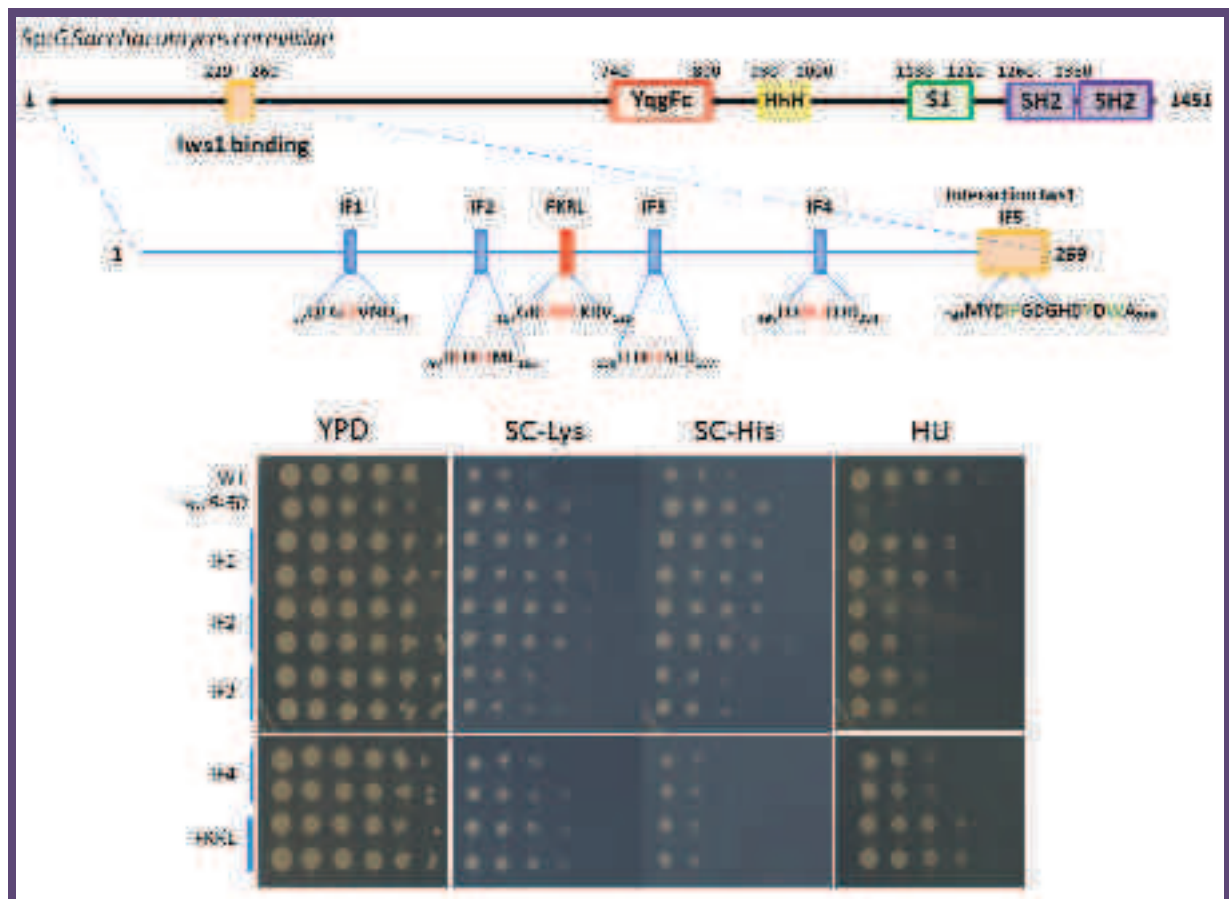
Erin Loeliger de l'équipe de Fred Winston, dans une souche dont le gène de Spt6 endogène a été supprimé de manière similaire au travail effectué dans les publications 1 et 2.

Motif muté	Mutations ponctuelles introduites
<b>IF1</b>	F60A, I61A
<b>IF2</b>	L98A, L100A, L101A
<b>IF3</b>	F153A F154A
<b>IF4</b>	D198A, F199A, I200A
<b>FKRL</b>	F119A, K120A, R121A, L122A

**Tableau 14 : Caractérisations fonctionnelles de l'extrémité amino-terminale de Spt6 – mutations ponctuelles introduites.**

Ces différentes souches transformées ont alors été étalées sur des milieux contenant des agents chimiques induisant des stress (Figure 65). Seuls les tests conduisant à l'apparition d'un phénotype sont présentés ici, la souche *spt6-50* servant de contrôle négatif. Cette souche contient une version tronquée de Spt6 qui provient d'une mutation ponctuelle qui introduit un codon stop dans le domaine carboxy-terminal de Spt6 (K1274 Stop), les phénotypes de cette souche ayant déjà été décrits de nombreuses fois (Hartzog et al., 1998).

Ces tests montrent que les différentes mutations ponctuelles introduites conduisent à l'apparition de divers défauts selon le motif qui a été muté. Aucune de ces mutations ne conduit à l'apparition de phénotypes thermosensibles comme c'était le cas lors de la mutation de la zone d'interaction avec Iws1 (publication 2). La mutation des motifs IF1 et IF2 conduit à l'apparition de phénotypes Spt- (croissance sur la plaque SC-Lys) et à l'initiation de la transcription à partir de sites d'initiation cryptique (croissance sur la plaque SC-His). La mutation des motifs IF3 et IF4 conduit également à l'apparition de phénotypes Spt- qui sont plus faibles que ceux observés avec la souche *spt6-50* (Figure 65, Tableau 15). Ces deux souches présentent par contre une sensibilité à l'hydroxyurée, une drogue qui conduit à l'apparition de défauts au cours de la réplication de l'ADN. Le mutant FKRL présente des effets beaucoup plus faibles et principalement un phénotype Spt- très faible.



**Figure 65 : Phénotype des différents mutants de Spt6**

Les souches contenant les gènes de Spt6 mutés ont été étalées sur différents milieux qui vont induire des stress chez la levure et vont permettre de déduire leur phénotype. Ces différents phénotypes sont récapitulés dans le Tableau 15. SC-Lys : milieu complet sans lysine, SC-His : milieu complet sans histidine et HU milieu contenant de l'hydroxyurée. Ces tests ont été réalisés par Erin Loeliger de l'équipe du Dr. Fred Winston, Harvard Medical School.

Mutant	Phénotype Spt-	Initiation cryptique	Sensibilité à l'hydroxyurée
IF1	Oui	Oui	Non
IF2	Oui	Oui	Faible
IF3	Faible	Non	Oui
IF4	Faible	Non	Oui
FKRL	Faible	Non	Non
spt6-50	Oui	Oui	Oui

**Tableau 15 : Principaux phénotypes causés par la mutation des motifs hydrophobes de l'extrémité amino-terminale de Spt6**

Ainsi, la mutation de ces différents motifs conduit à l'apparition de phénotypes divers mais dans l'ensemble plutôt modérés. Ces effets réduits pourraient être expliqués par

l'absence d'une abolition complète des complexes formés *in vivo*, la mutation de régions plus larges pouvant s'avérer nécessaire pour l'observation de phénotypes forts. De plus, il n'est pas exclu non plus que plusieurs de ces motifs puissent être nécessaires à la mise en place d'une même fonction biologique. Si cela devait être le cas, seule la mutation de tous les motifs impliqués dans une même fonction conduirait à l'apparition de phénotypes forts.

## 5.6. Discussion

L'analyse de la séquence de l'extrémité amino-terminale de Spt6 nous a amené à identifier plusieurs motifs hydrophobes dans cette partie de la protéine qui est globalement fortement chargée négativement. Nos résultats nous ont également amené à proposer une interaction entre deux motifs hydrophobes du domaine amino-terminal de Spt6 et le domaine I de TFIIS, l'un de ces motifs interagissant également avec Iws1. Des tests de compétition ont permis de suggérer que TFIIS interagissait préférentiellement avec le motif IF3 dans un grand fragment de Spt6 long mais interagit également avec le motif IF5 dans le cadre d'un fragment court de Spt6, même en présence de Iws1. Ces différentes constatations biochimiques ont donc soulevé la question du mode de liaison de Spt6 à TFIIS. Des études structurales par RMN ont donc été entreprises, la cristallisation de ces complexes s'avérant impossible. A l'heure actuelle, seul un premier modèle partiel de la structure du complexe Spt8/TFIIS a pu être calculé, la position de Spt8 n'étant pour l'instant pas définie précisément. Ce premier modèle fait apparaître une forte homologie structurale entre notre structure de TFIIS et la structure de la protéine de souris qui ne se superpose pas avec la structure de la protéine de levure qui a été publiée précédemment (Booth et al., 2000). Le repliement du fragment de Spt8 a également pu être déduit de nos données RMN et il semble qu'il adopte une structure en feuillet beta, contrairement au fragment de Spt6 qui adopte une structure hélicoïdale lorsqu'il interagit avec Iws1. Ces données préliminaires semblent donc indiquer que le mode de liaison de Spt8 à TFIIS est différent du mode de liaison de Spt6 à Iws1, ce qui pourrait notamment expliquer nos observations biochimiques concernant la formation de plusieurs complexes Spt6/TFIIS. En effet, l'extension poly-ED en aval du motif IF sur Spt8 semble participer à l'interaction Spt8/TFIIS en formant notamment un brin  $\beta$ . Ainsi, la résolution de la structure des complexes Spt6/TFIIS permettra d'observer le mode d'interaction mis en place, et d'aider à identifier potentiellement le motif de Spt6 qui pourrait interagir avec TFIIS *in vivo*, et, enfin, de comprendre comment un complexe Spt6 IF5/TFIIS/Iws1 peut être formé *in vitro*.



Chez la levure, il a été démontré que le domaine amino-terminal de Spt6 était nécessaire à la croissance de ces organismes, la délétion des résidus 2 à 205 conduisant à l'apparition d'un phénotype non viable (données non publiées du Dr. Fred Winston). Afin d'observer l'implication de cette région de la protéine *in vivo* et de tester le rôle des différents motifs hydrophobes qu'elle contient, des mutations ponctuelles dans ces motifs ont été introduites dans la levure. Ces mutations conduisent à l'apparition de phénotypes faibles, montrant que ces motifs ont un rôle biologique important, mais qu'ils ne sont pas seuls impliqués dans la mise en place de ces fonctions. L'extrémité amino-terminale de Spt6 pourrait ainsi jouer un rôle de recrutement de différents facteurs au cours de la transcription, chacun des motifs pouvant reconnaître spécifiquement un de ces facteurs. Notre étude a déjà permis d'identifier deux interactants potentiels de cette région de Spt6 : Iws1 et TFIIS. De plus, une interaction génétique entre Spt6 et le complexe PAF a déjà été mise en évidence chez la levure, des souches portant des mutations des protéines de ce complexe et de Spt6 étant non viables (Kaplan et al., 2005). La compréhension du rôle du domaine amino-terminal de Spt6 va donc nécessiter l'identification de ses différents interactants, ainsi que la définition précise de leurs sites de liaison.

## **Chapitre 6**

# **Interaction de Spt6 avec les histones – co-expression de l'octamère d'histone**



## 6.1. Introduction

Le rôle de chaperonne d'histone de la protéine Spt6 a très tôt été décrit. Ce rôle a été suggéré par (i) l'observation d'une interaction directe entre Spt6 et la paire d'histone H3/H4 et (ii) la capacité de la protéine à reconstituer partiellement des nucléosomes sur un plasmide en présence des protéines histones (Bortvin et Winston, 1996). L'implication de Spt6 dans la reconstitution des nucléosomes au cours de la transcription a été confirmée par l'apparition de transcrits courts chez la levure lorsqu'une version mutée de Spt6 été exprimée, ces transcrits résultants d'une initiation de la transcription réalisée à partir de sites cryptiques (Kaplan et al., 2003). Bien que ce rôle de chaperonne d'histone de la protéine Spt6 soit bien défini, aucune information concernant l'interaction de la protéine avec les histones ni son mode d'action dans ce mécanisme n'est disponible. Afin de répondre à ces questions, des tests de co-expression entre Spt6 et les paires ou l'octamère d'histone ont été entrepris. Dans un premier temps, nous avons mis à profit la technique de co-expression afin d'exprimer les paires et l'octamère d'histone dans le but de tester facilement l'interaction de ces protéines avec leurs protéines chaperonnes, cette approche ayant déjà été exploitée partiellement dans le cas de la co-expression des paires d'histones dans les bactéries (Anderson et al., 2010). Cet outil a par la suite été utilisé afin de tester leur interaction avec Spt6.

## 6.2. Brève description des résultats

Les gènes codant pour les protéines histones H2A, H2B, H3 et H4 de *Xenopus laevis* ont été clonés dans les vecteurs d'expression pnEA-tH et pnCS. Différentes constructions de ces protéines ont ainsi été clonées : (i) le gène codant pour la protéine entière mais aussi (ii) des constructions ne contenant pas les queues amino-terminales de ces protéines qui ne sont pas nécessaires à la formation de l'octamère d'histone. Des tests de co-expression ont alors été réalisés afin de reconstituer les paires d'histones H3/H4 et H2A/H2B. Deux concentrations salines ont été testées, avec soit 1M NaCl ou 2M NaCl, qui sont les concentrations auxquelles les histones sont solubles du fait de leur forte charge positive. Ces premiers tests ont permis de définir sur quelle protéine histone placer l'étiquette de purification afin de former et purifier ces différents complexes et les conditions salines nécessaire à leur production sous forme soluble. Ces premiers tests ont permis de définir : (i) que la paire H2A/H2B est plus soluble lorsque le tampon de lyse contient 1M NaCl et n'est purifiée en grande quantité que quand la protéine H2A porte l'étiquette de purification, et que (ii) la paire H3/H4 est plus

soluble lorsque le tampon de lyse contient 2M NaCl et n'est purifiée que lorsque la protéine H4 porte l'étiquette de purification. Ces premières observations ont permis de définir la stratégie de clonage à adopter lors de la concaténation des gènes des protéines qui a pour but de co-exprimer les quatre protéines en même temps. La réalisation de nouveaux tests de co-expression utilisant ces vecteurs codant pour les paires d'histone montre que les quatre protéines sont co-purifiées lorsque le tampon de lyse contient 2M NaCl, quatre bandes individuelles et stœchiométriques pouvant être visualisées lorsque l'histone H2A porte l'étiquette de purification. Les deux segments de vecteur codant pour les constructions H3/H4 et H2B/HisH2A ont alors été concaténés.

Ce vecteur codant pour les quatre protéines a été utilisé afin de produire l'octamère d'histone dans les bactéries. Après purification par affinité, la filtration sur gel révèle la présence de deux espèces dans le mélange protéique : (i) une espèce correspond à l'octamère d'histone, quatre bandes distinctes pouvant être visualisées sur gel d'acrylamide dénaturant et (ii) une espèce qui correspond à la paire d'histone H2A/H2B qui a été produite en excès par rapport à la paire H3/H4. Cette purification a permis d'obtenir un peu plus de 1mg de complexe après cette étape de filtration sur gel.

La production de l'octamère d'histone par co-expression offre une alternative à la production des histones à partir des corps d'inclusions qui est couramment utilisée. En effet cette dernière technique est relativement longue à mettre en œuvre, bien qu'elle permette d'obtenir de grandes quantités de complexe. A l'inverse, la co-expression permet d'obtenir rapidement de petites quantités d'octamère, notamment dans le but de réaliser des études fonctionnelles, peu gourmandes en matériel biologique (autour du  $\mu\text{M}$ ). La publication numéro 3 ci-après récapitule ces résultats ainsi que de nombreux autres résultats obtenus grâce à la technique de multi-expression développée dans le groupe de recherche du Dr. Christophe Romier.

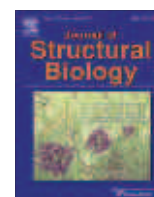
### 6.3. Publication numéro 3

Les données supplémentaires relatives à cette publication sont présentées en Annexe 8.6, page 281



Contents lists available at ScienceDirect

## Journal of Structural Biology

journal homepage: [www.elsevier.com/locate/yjsbi](http://www.elsevier.com/locate/yjsbi)

## Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer

Marie-Laure Diebold<sup>a,1</sup>, Sébastien Fribourg<sup>b,c,1</sup>, Michael Koch<sup>a</sup>, Thibaud Metzger<sup>a</sup>, Christophe Romier<sup>a,\*</sup>

<sup>a</sup> Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Département de Biologie Intégrative, Institut National de Santé et de Recherche Médicale (INSERM) U964/Centre National de Recherche Scientifique (CNRS) UMR 1704/Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France

<sup>b</sup> Institut Européen de Chimie et Biologie, INSERM U869, 2 rue Robert Escarpit, 33607 Pessac, France

<sup>c</sup> Université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux, France

### ARTICLE INFO

#### Article history:

Available online 12 February 2011

#### Keywords:

Co-expression  
*Escherichia coli*  
 Protein complexes  
 SAGA transcriptional activator  
 Histone octamer  
 Nucleosome

### ABSTRACT

Macromolecular complexes are responsible for most of the essential mechanisms in cells, leading to a broad interest in their purification and characterization. Co-expression is now widely recognized as a major technique for assembling multiprotein complexes and many co-expression systems are currently available for performing co-expression experiments in different hosts. However, comparative knowledge on co-expression strategies is still crucially lacking. Using versatile co-expression systems for *Escherichia coli*, the pET-MCN and pET-MCP series, and ternary protein complexes as examples, we demonstrate how to successfully delineate correct co-expression strategies. Specifically, an appropriate, complex-dependent approach alleviates stoichiometry imbalance and yield problems, and even failure in producing complexes. Importantly, some of the parameters influencing co-expression strategies appear independent of the expression host, thus having implications for co-expression in eukaryotic hosts. By further using these strategies, we show that co-expression in *E. coli* enables reconstitution of protein complexes as large as the deubiquitination module of the SAGA transcription factor and the histone octamer.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Macromolecular complexes play key roles in most, if not all, cellular processes as highlighted by high-throughput studies on prokaryotic and eukaryotic organisms (Benschop et al., 2010; Devos and Russell, 2007; Gavin et al., 2006; Kuhner et al., 2009). As a consequence, complexes have become the major focus of numerous studies. Yet, biochemical, structural, and functional characterization of such complexes necessitate dedicated techniques for their production in sufficient quantity and quality.

The co-expression technique is now widely recognized as a major method for reconstituting multiprotein complexes, even for high-throughput studies (Graslund et al., 2008; Busso et al., this issue; Vijayachandran et al., this issue). Landmarks of this technique are co-solubilisation, specificity, characterization of protein/protein interactions, delineation of minimal interaction domains and co-purification (Fribourg et al., 2001). Many systems are now available for co-expressing two or more proteins in *Escherichia coli* and in eukaryotic cells (see for instance Berger et al., 2004; Bieniossek et al., 2009; Held et al., 2003; Kriz et al.,

2010; Romier et al., 2006; Scheich et al., 2007; Tan et al., 2005; Tolia and Joshua-Tor, 2006; Busso et al., this issue).

To date, *E. coli* remains the most prominent expression and co-expression host due to the ease of its manipulation and its rapid growth. Co-expression strategies range from the use of several vectors with a single gene each, to the use of a single vector harboring several genes that are under the control of either a single promoter (polycistron) or several promoters (individual cassettes) (Perrakis and Romier, 2008; Romier, 2008; Romier et al., 2006). Interestingly, the various co-expression systems currently available make use of either polycistron or individual cassettes based on either single or multiple vectors strategies. Comparison of some of these systems on well characterized complexes shows that they are all capable of producing complexes (Busso et al., this issue).

Our own co-expression system, the pET-MCN series (Perrakis and Romier, 2008; Romier, 2008; Romier et al., 2006), which has been developed within the frame of the SPINE program, also shows a good behavior in this comparative work. This is in agreement with the successful usage which has been made of the pET-MCN series since several years for the production of complexes for structural studies in our and in other laboratories (Bono et al., 2004, 2006, 2010; Boskovic et al., 2007; Buchwald et al., 2010; Diebold et al., 2010; Fribourg et al., 2003; Jeyaprakash et al., 2007; Lorentzen et al., 2007; Romier et al., 2003). Nonetheless, for some

\* Corresponding author. Fax: +33 3 88 65 32 76.

E-mail address: [romier@igbmc.fr](mailto:romier@igbmc.fr) (C. Romier).

<sup>1</sup> These authors contributed equally to this work.

complexes we studied, we encountered occasional difficulties in obtaining sufficient amounts of complexes, mostly due to low yield, poor stoichiometry or even failure in complex assembly. One good example is provided by the human NFYA/NFYB/NFYC (NFY) transcription factor that was reconstituted using the pET-MCN series in small scale co-expression tests (Romier et al., 2006), but could not be purified to homogeneity when scaled up (Romier, unpublished data). Intriguingly, we noticed that these problems were likewise encountered when other systems were used such as the commercially available pET-DuET (Novagen) series (Romier et al., 2006 and data not shown).

Currently these specific problems cannot be explained, notably due to a lack of comprehensive studies on parameters governing multi-expression. In particular, the question of paramount importance whether the co-expression systems used or, alternatively, the complexes themselves are causing failure in the experiment, cannot be adequately addressed to date. To address this question, and thus provide essential guidelines for future complex reconstitutions, we report here a comparative study by varying most of the major parameters supposed to influence co-expression: tag position, use of multiple versus single vector strategies, use of single versus multiple promoters on single vectors, variation of the order of the genes or promoters on single vectors. Our study on well-characterized complexes reveals that the design of a complex-dependent co-expression strategy enhances complex production. Based on these results, we further validated this approach by reconstituting by co-expression large complexes involved in transcription such as the deubiquitination module of the SAGA transcription factor and the histone octamer.

## 2. Material and methods

### 2.1. Small-scale expression tests

For small-scale expression, *E. coli* BL21[DE3] cells (Novagen) were transformed with the respective multi-expression vectors and plated on LB/agar supplemented with the corresponding antibiotics. For the deubiquitination module of SAGA and the histone octamer, the pRare2 vector (Novagen) coding for *E. coli* rare tRNAs was also used for co-transformation. A few colonies were then used to inoculate 2 to 10 ml minicultures at 37 °C which were then grown at 37 °C up to an OD<sub>600</sub> of 0.4. The cultures were cooled for 15 min at room temperature. Complex production was then induced by adding IPTG to a final concentration of 1 mM. The cultures were then transferred to a 25 °C shaker for overnight expression.

After expression, cells were resuspended in 1 ml Lysis Buffer (10 mM Tris/HCl pH 8.0 with 50 mM, 400 mM, 1 M or 2 M NaCl, depending on the complex considered) and lysed by sonication. After centrifugation to remove cells debris, the supernatants were loaded on Talon affinity resin (Clontech) pre-equilibrated in Lysis Buffer. Loaded resin was incubated for 1 h at 4 °C and shaken mildly. Supernatants were removed and the affinity resin was washed twice with 1 ml Lysis Buffer. Finally, the affinity resin was resuspended in 30 µl of Laemmli Buffer and the samples boiled prior to analysis by SDS-PAGE. The proteins retained on the beads were revealed by Coomassie staining.

### 2.2. Large-scale purification

Large-scale protein expression was carried out as for small-scale expression tests but using 1 or 3 l cultures. After sonication in Lysis Buffer (same as in small co-expression tests) and centrifugation to clear the lysate, the supernatant was incubated for 1 h at 4 °C with 0.5 ml Talon (Clontech) affinity resin. The resin was then

transferred into a small Bio-Rad column and washed extensively with Lysis Buffer. The resin was resuspended in 1 ml lysis buffer and 50 µl thrombin (1 u/µl; MP Biomedicals) was added overnight for liberating the complex from the resin by cleavage of the histidine tag.

For the YMP and NFY complexes, the supernatant was recovered and applied onto a Superdex 75 10/300 GL column (GE Healthcare) for gel filtration. In case of the histone octamer, a Superdex 200 16/60 column (GE Healthcare) was used. Peak fractions were analyzed by SDS-PAGE and Coomassie staining.

## 3. Results

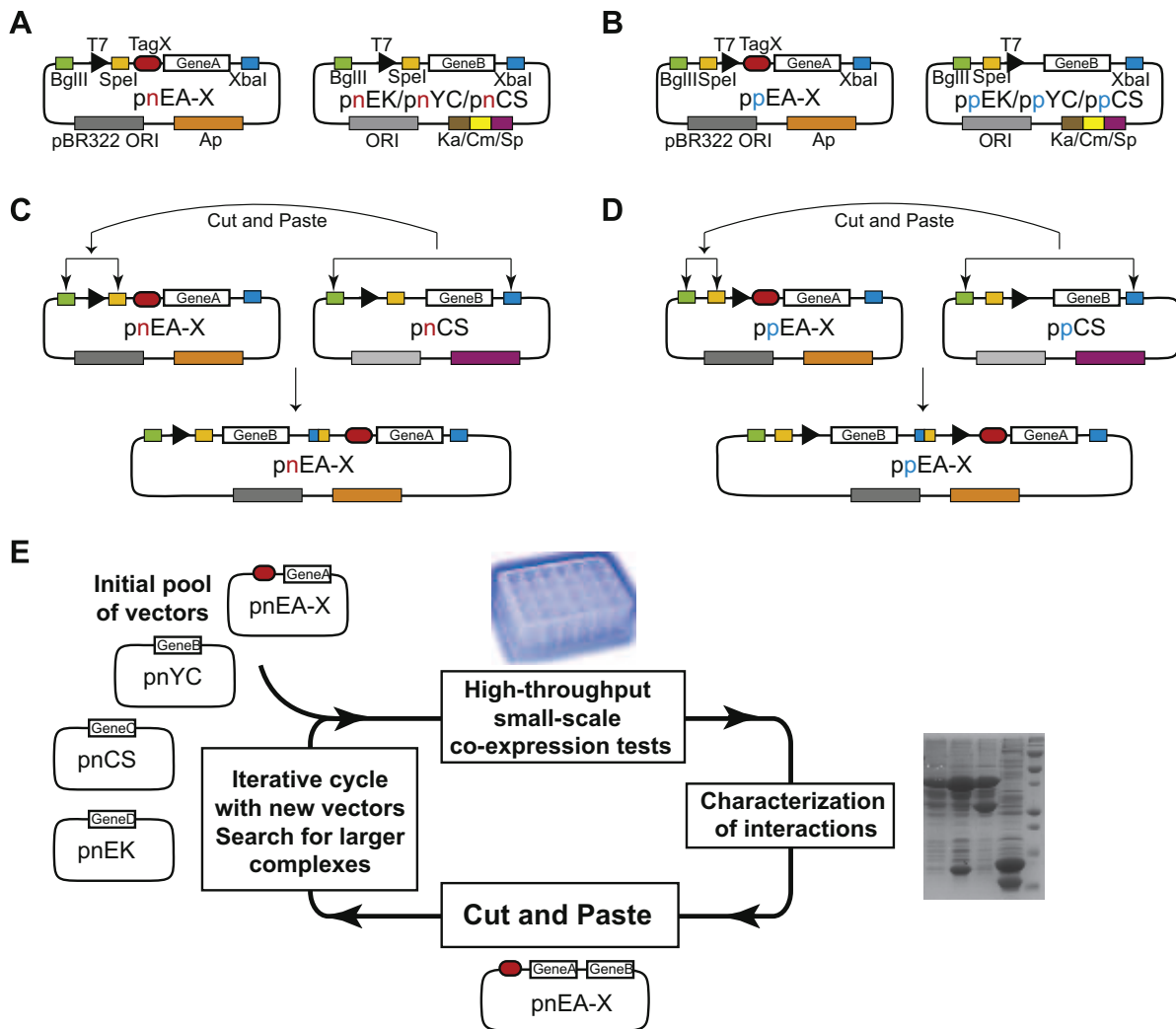
### 3.1. Improvement of the pET-MCN series and development of the pET-MCP series

Our initial work on co-expression has suggested that some major parameters can influence the outcome of co-expression experiments (Fribourg et al., 2001; Romier et al., 2006). These include notably (i) the use of several versus single vectors, (ii) the use of mono- versus multi-promoter plasmids, (iii) the order of genes or promoters on single vectors, and (iv) the influence of affinity purification tag placement on complex formation and stability. Comprehensive testing of these parameters has however been impaired by the lack of a single co-expression system enabling all these parameters to be tested in parallel. To this end, we have extended our pET-MCN co-expression system and developed the pET-MCP series.

The initial pET-MCN (**p**ET **M**ulti-**C**loning and **e**xpression) series was based on two kind of vectors (pnYC and pnEA-X, X means any fusion) that could be used for pairwise co-expression. In addition, single vectors could also be produced by concatenation of the initial vectors, leading to a single vector harboring several genes under the control of a single promoter (Romier et al., 2006). Based on these vectors and on the fact that the concatenation procedure was regenerative, a specific strategy had been designed enabling the reconstitution of protein complexes by successive cycles of co-expression experiments and vector concatenations (Romier, 2008; Romier et al., 2006). This strategy is particularly well-suited for structural studies since the use of the multiple vector approach enables the rapid testing of different constructs for each protein studied.

One limitation of this series came from the fact that, initially, only pairwise co-expression experiments could be performed since only two vectors were available (pnYC and pnEA-X; X any fusion). To address this issue, two other vectors were created that were compatible with the pnEA-X and pnYC vectors: pnCS and pnEK (Fig. 1; Supplementary data Fig. 1; Supplementary Website). These new vectors extended the initial testing of interaction by a multiple vector approach to four different proteins. Furthermore, the initial strategy for complex reconstitution could be extended and simplified to speed up the reconstitution process (Fig. 1; Supplementary data Fig. 2; Supplementary Website). In addition, a semi-automated approach was set up to cope with the combinatorial created by this strategy (Vijayachandran et al., this issue).

However, this extended pET-MCN series could not be used in an approach combining a single vector with multiple promoters. Therefore, by slightly modifying our pET-MCN vectors, we created the pET-MCP vector series (**p**ET **M**ulti-**C**loning **P**romoter). The small changes introduced enabled, upon vector concatenation, every gene to be under the control of its own T7 promoter, leading to the creation of single vectors with multiple T7 promoters (Fig. 1; Supplementary data Fig. 1; Supplementary Website). The pET-MCP series retained otherwise all other features of the pET-MCN series and could also be used for the reconstitution strategy developed



**Fig. 1.** Overview of pET-MCN and pET-MCP vector series. Simplified view of (A) pET-MCN and (B) pET-MCP vector series. pnEA-X and ppEA-X vectors provide expression of N- and/or C-terminally X-tagged proteins (X represents any tag available). pnEK/ppEK, pnYC/ppYC and pnCS/ppCS vectors provide expression of untagged proteins. All vectors have different antibiotic resistances. Positions of major vector elements are shown. Example of concatenation of (C) pET-MCN and (D) pET-MCP vectors using specific restriction sites using a regenerative cut/restriction and paste/ligation mechanism. Other ways of concatenation are available allowing (i) variation of genes or promoters order and (ii) concatenation of pET-MCN and pET-MCP vectors (Supplementary Website). (E) Iterative complex reconstitution strategy using either vector series. Small-scale co-expression tests use a semi-automated high-throughput protocol (Vijayachandran et al., this issue). More detailed information on the pET-MCN and pET-MCP vector series is provided in Supplementary data Figs. 1 and 2, and Supplementary Website.

for the pET-MCN series. In addition, both the pET-MCN and pET-MCP series are compatible, enabling the creation of hybrid vectors harboring multiple promoters, some of them controlling the expression of a single gene and others controlling the expression of several genes (Supplementary Website).

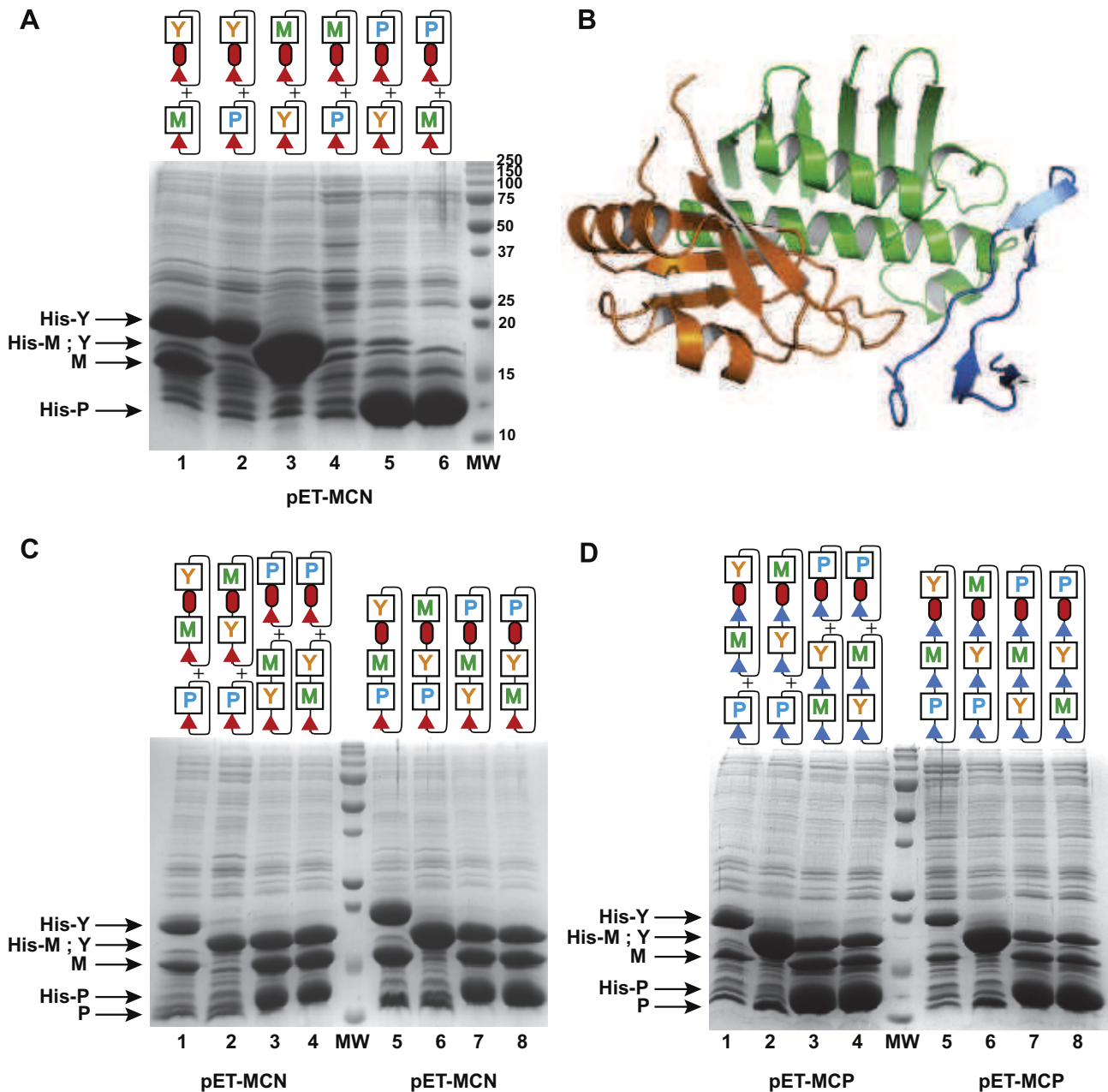
We then used the pET-MCN and pET-MCP series to study the influence on the outcome of co-expression experiments of the various parameters mentioned earlier. Two already characterized ternary complexes were chosen as case studies for this comparative work: (i) the eukaryotic Y14/Mago/Pym (YMP) complex, from the exon-exon junction complex (EJC), that had already been partially reconstituted using the pET-MCN series, leading to its structure determination (Bono et al., 2006; Fribourg et al., 2003) and (ii) the human NFYA/NFYB/NFYC (NFY) transcription factor mentioned above. For this study, all genes from both complexes were first cloned separately into pET-MCN and pET-MCP vectors that provide either expression of a protein with an N-terminal polyhistidine tag or expression of an untagged protein. Complex reconstitution was then monitored using the strategy designed for the pET-MCN and pET-MCP series. To clearly decipher the assembly of both com-

plexes, we initially decided to use a pairwise co-expression approach with two vectors.

### 3.2. Case study: the Y14/Mago/Pym (YMP) complex

Initial pairwise small-scale co-expression tests of the YMP-subcomplexes using two vectors revealed a strong interaction between Y14 and Mago, independent of the tag position. Solubilisation of Mago upon binding of Y14 was clearly observed (Fig. 2A, compare lanes 1–4). Although Pym was soluble, interaction was not observed with either Y14 or Mago, in agreement with the complex structure showing that the Y14/Mago complex, but not the individual subunits, provides an extensive interaction surface for Pym (Fig. 2A, lanes 5–6; Fig. 2B; also described in Bono et al., 2004). Based on the results of the pairwise expression experiments, single vectors harboring both genes coding for Y14 and Mago were produced by concatenation (procedure as described in Fig. 1C and D). In these vectors, both genes were placed under control of either a single promoter (pET-MCN) or, alternatively, in two separate expression cassettes, each with its own promoter





**Fig. 2.** Reconstitution of the Y14/Mago/Pym (YMP) ternary complex by multi-expression. SDS-PAGE analysis of soluble YMP sub- and entire complexes retained on affinity resin. The different combinations of vectors used are depicted above each lane. Y14, Mago and Pym are represented by Y, M and P, respectively. T7 promoters for the pET-MCN and pET-MCP series are depicted as red and blue arrowheads, respectively. The N-terminal polyhistidine tag (denoted His in gel annotations) coding sequence is shown as red ellipse. (A) Analysis of pairwise protein interactions using two vectors. (B) Ribbon representation of the Y14/Mago/Pym complex crystal structure (PDB 1rk8). Pym (blue) interacts with both Y14 (orange) and Mago (green), explaining the results observed in (A). (C) YMP complex reconstitution using two or a single vector (pET-MCN series). (D) YMP complex reconstitution using two or a single vector (pET-MCP series).

(pET-MCP). Furthermore, to assess the influence of the polyhistidine tag on complex formation, different vector versions were made where either Y14, or Mago, or neither of them were tagged.

In the second round of two-vector co-expression experiments, production of the entire ternary YMP-complex was attempted. In case of the pET-MCN series, complexes were produced irrespective of tag placement (Fig. 2C; lanes 1–4). However, differences in stoichiometry – assessed from the large scale purification described below – were observed, resulting notably in low quantities of Pym when Y14 or Mago were tagged. In strong contrast, better stoichiometry was observed when Pym was tagged, with higher yields for all three proteins. A different situation was observed

with the pET-MCP series: stoichiometry imbalances were observed irrespective of tag placement, the tagged protein appearing always in excess as compared to the untagged proteins (Fig. 2D; lanes 1–4). In contrast, varying in some cases the order of the genes or of the promoters on the vectors harboring two genes did not affect yields or stoichiometry.

Finally, single vectors harboring all three genes under the control of either a single (polycistronic; pET-MCN) or several (individual expression cassettes; pET-MCP) promoters were created, again varying tag placement. Co-expression of the YMP-complex using a single vector did not reveal major differences as compared to the experiments when two separate vectors were used (Fig. 2C

and D). Notably, the stoichiometry imbalances remained basically the same. Only in the case of the pET-MCN series, more Pym appeared present in the complex when either Y14 or Mago was tagged. Varying the order of the genes or of the promoters on the vectors had once again no effect.

Thus, placing the tag on Pym appeared to provide the best results notwithstanding the strategy selected. Therefore, larger cultures of the YMP-complex were produced by using a single vector of the pET-MCN series. The complex was purified by affinity and size exclusion chromatography. Analytical gel filtration resulted in two major peaks, corresponding to ternary YMP-complex (large peak), and excess Pym (small peak) (Supplementary data Fig. 3A). Although the analytical gel filtration column was not as resolutive as a preparative column, both species could be separated, showing in this case the interest in tagging the smallest subunit of the complex.

Finally, we asked whether the use of three independent vectors, each coding for a single gene, had an effect on complex formation. As shown in Supplementary data Fig. 4A, and as observed for ternary complex co-expression from one or two vectors, stoichiometry differences were observed depending on the protein bearing the tag. Interestingly, the yield and stoichiometry for the complexes produced with the three vectors appeared similar to the ones observed in the case of expression from a single vector of the pET-MCN series (compare Fig. 2C, lanes 5–8 and Supplementary Fig. 4A). Specifically, the stoichiometry imbalance observed with the pET-MCP series in the case where Pym is tagged is not observed anymore. This shows that the differences of stoichiometry observed between the pET-MCN and pET-MCP series in the single vector case cannot be fully accounted for by the production of a single mRNA (pET-MCN/one vector) versus multiple mRNAs (pET-MCP/one vector or pET-MC(N or P)/three independent vectors).

### 3.3. Case study: the NFYA/NFYB/NFYC (NFY) complex

An identical approach was applied to the NFY complex that had been successfully assembled in small-scale using a polyhistidine tag on the N-terminus of the NFYA subunit (Romier et al., 2006), but that we failed to obtain in large scale complex production (data not shown). Initial pairwise mini-scale co-expression experiments evidenced interaction between the NFYB and NFYC subunits and co-solubilisation of both subunits was observed. In contrast, interaction with NFYA could not be detected (Fig. 3A).

After creating single vectors harboring both NFYB and NFYC genes, varying the number of promoters and the tag position, co-expression of the NFY complex using two vectors was attempted. Similar to our observations for the YMP complex, all combinations produced varying amounts of ternary complex (Fig. 3B). This time, stoichiometry of the purified complexes appeared identical for the two vector series used. However, major stoichiometry imbalances were observed when the affinity tag was placed on NFYA (Fig. 3B, lanes 1–2 and 7–8). When either NFYB or NFYC were tagged, a slight excess of the NFYB/NFYC sub-complex was observed (Fig. 3B, lanes 3–6 and 9–12).

After generating single vectors harboring all three genes, co-expression experiments were performed with the tag placed on either NFYB or NFYC, resulting in almost stoichiometric production and successful ternary complex formation (Fig. 3C). Once again, no major differences were observed between the pET-MCN and pET-MCP series.

However, the picture changed dramatically when we expressed the NFY complex using three independent vectors. In this case, a nice complex was observed when either NFYA or NFYB were tagged, but not when NFYC was tagged (Supplementary Fig. 4B). Yet, large scale production of the complex using three vectors and with NFYA N-terminally his-tagged failed once again to

produce complex. Taken together, this series of experiments suggests that the reconstitution strategy initially used for NFY, i.e. tagging N-terminally NFYA, is not the most suitable. Rather, placing the tag on either NFYB or NFYC results in better overall yield and balanced stoichiometry.

In agreement, expression from a single pET-MCN vector, containing all three genes with the tag placed on NFYB, resulted in large quantities of intact ternary complex that could be purified from large scale cultures by affinity purification followed by gel filtration (Supplementary Fig. 3B). However, as observed on the analytical gel filtration profile, the peak of the complex appeared double. We soon realized that this peak was composed of two species that could not be separated: a NFYA/NFYB/NFYC ternary complex and a dimer of the NFYB/NFYC heterodimer. In the end, the solution found to obtain a homogeneous NFY ternary complex was to His-tag NFYA C-terminally, leading to a trimeric complex that could easily be separated from the excess of the small monomeric NFYA protein by size exclusion chromatography (data not shown). This result clearly highlights the importance of tag position for complex reconstitution.

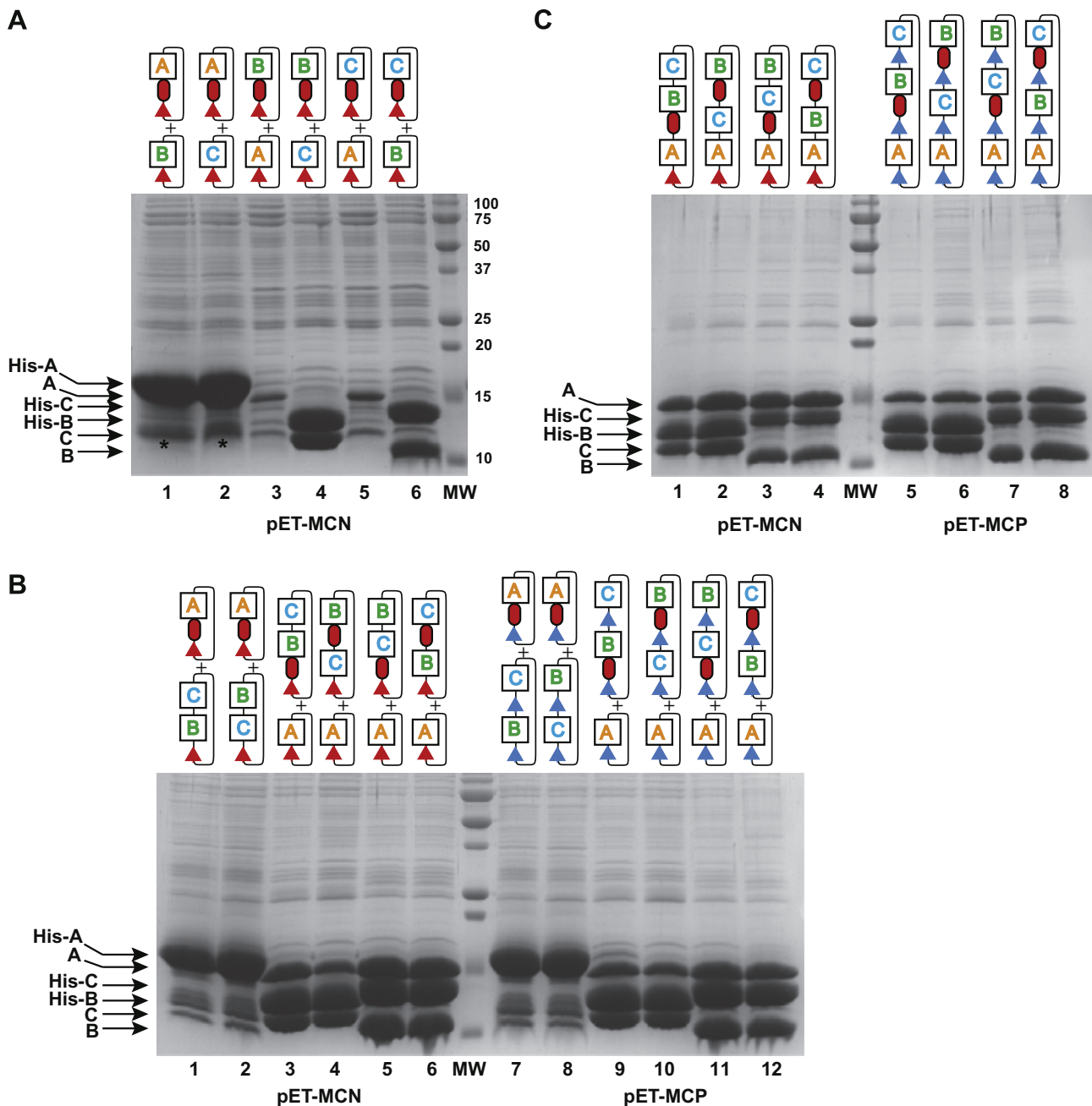
### 3.4. Study of the deubiquitination module (Sgf11/Sus1/Ubp8/Sgf73) of the yeast SAGA complex

The preceding examples have been carried out on small, already characterized ternary complexes, revealing the importance of specific co-expression strategies to be used to obtain sufficient amounts of sample for further studies. We next asked whether this approach could also be used for larger, uncharacterized complexes. We chose to study the reconstitution of the deubiquitination module (DUB) of the transcription factor SAGA with our pET-MCN series. This module deubiquitinates histone H2B lysine 123 (in yeast) or lysine 120 (in humans), leading to transcription activation. In yeast, this module is formed by four subunits: scSgf11, scSus1, scUbp8 and scSgf73, which all have human counterparts (hsAtaxin7-L3, hsEny2, hsUsp22 and hsAtaxin7, respectively), and form a large complex of 150 kDa (Helmlinger et al., 2004; Köhler et al., 2006, 2008; Zhao et al., 2008). Although the Ubp8 subunit bears the catalytic activity of this complex, this subunit on its own is inactive and requires its association with the three other subunits in order to be fully active within the SAGA complex (reviewed in Rodriguez-Navarro, 2009).

Whereas scSgf11 and scSus1 are small proteins, scUbp8 and scSgf73 are larger proteins (Fig. 4A). Specifically, these two latter proteins appear to have two conserved domains: an N-terminal zinc-binding domain and a C-terminal catalytic domain for scUbp8, and an N-terminal and a middle (Sca7) zinc-binding domains for scSgf73. In addition, scSgf73 contains a long poorly conserved C-terminal domain whose function is not fully understood.

Since the role of these proteins and their domains were not known, several constructs were made for each of the four proteins and positioning of the tag was also addressed in our study. For clarity, however, and although the position of the tag was shown once again to influence the outcome of the reconstitution experiments, only a subset of our results is presented here. Notably, only the full-length scSgf11 and scSus1 proteins are considered, whereas two constructs are used for both scUbp8 and scSgf73: one spanning only the first N-terminal domain (scUbp8-Zn and scSgf73-Zn) and a second spanning both conserved domains (scUbp8 and scSgf73-ZnSca7).

Upon pairwise expression of the different constructions, a single complex was observed that was formed between scSgf11 and scSus1, showing a strong solubilisation of the former protein upon complex formation (Fig. 4B, lanes 1–5). In contrast, all other combinations did not lead to soluble complex formation. Based on this initial result, both vectors coding for the native scSgf11 and scSus1



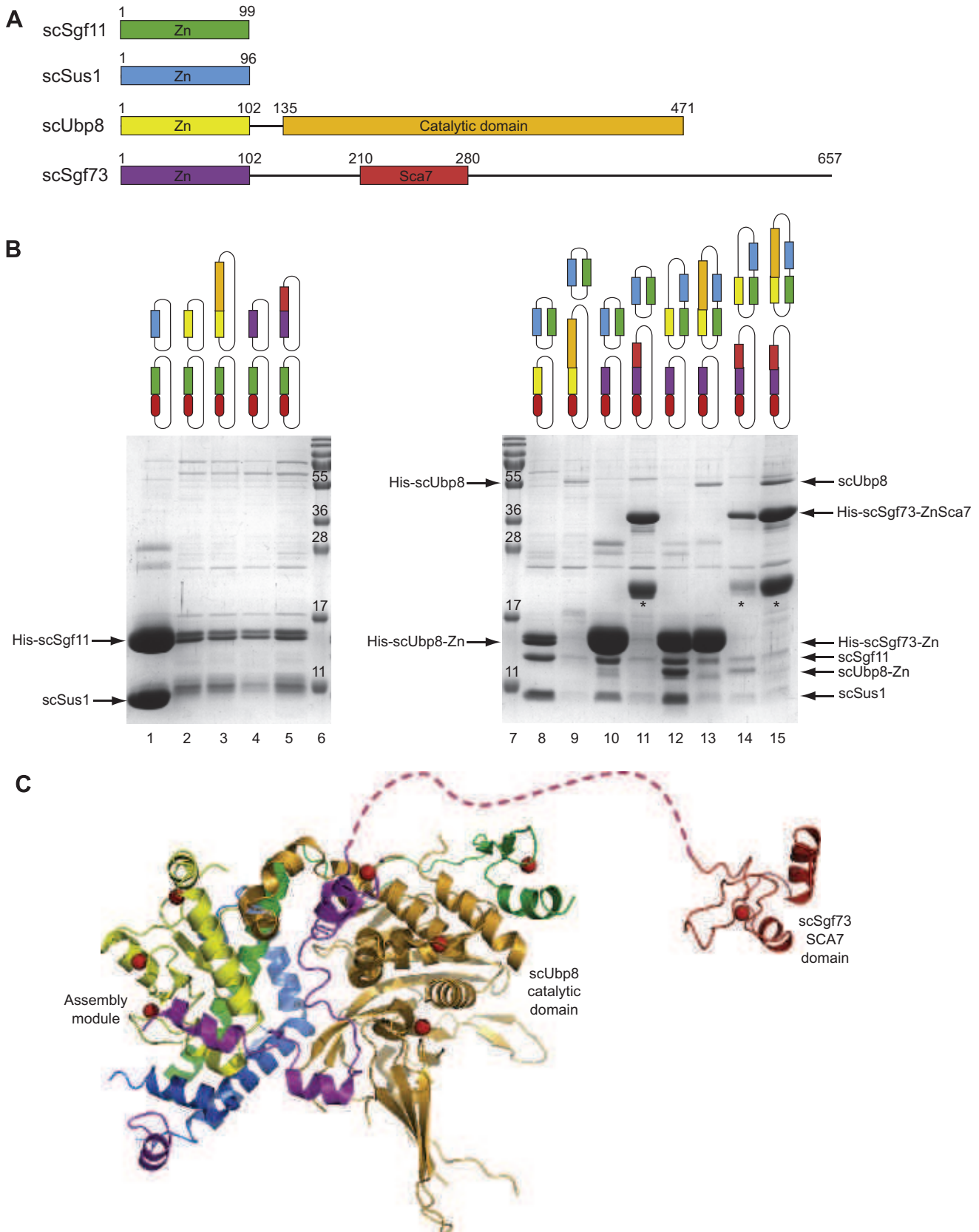
**Fig. 3.** Reconstitution of the NFYA/NFYB/NFYC (NFY) ternary complex by multi-expression. SDS-PAGE analysis of soluble NFY sub- and entire complexes retained on affinity resin. Vector elements are shown as in Fig. 2. NFYA, NFYB and NFYC are represented by A, B and C, respectively. (A) Analysis of pairwise protein interactions using two vectors. In lanes 1 and 2, NFYA degradations are marked with a '\*' and give a false impression of interaction. In lanes 3 and 5, unspecific binding of untagged NFYA protein to the resin is observed. (B) NFY complex reconstitution using two vectors of the pET-MCN (lanes 1–6) and pET-MCP (lanes 7–12) series. (C) NFY complex reconstitution using a single vector of the pET-MCN (lanes 1–4) and pET-MCP (lanes 5–8) series.

proteins were concatenated and the resulting vector used for ternary complex formation.

Interestingly, upon co-expression trials, the scSgf11/scSus1 complex could associate with the different constructs of scUbp8 and scSgf73 to form ternary complexes (Fig. 4B, lanes 8–11). In addition, in both cases only the N-terminal domains of scUbp8 and scSgf73 were sufficient for ternary complex assembly (Fig. 4B, lanes 8 and 10). Importantly, the use of these shorter constructs was important since, in the case of the longer constructs, the yield of the complexes appeared rather low, suggesting some instability of these larger complexes and showing that in case only larger constructs would have been considered for this

study, formation of ternary complexes could potentially have been missed.

Since the complex with scUbp8 appeared more stoichiometric than the one with scSgf73, we decided to concatenate the vectors coding for the native scUbp8 constructs together with the vector encoding both scSgf11 and scSus1 subunits. In a last cycle of co-expression trials, complexes formed by the four proteins could be observed (Fig. 4B, lanes 12–15). Interestingly, once again only the N-terminal domains of scUbp8 and scSgf73 were sufficient for complex assembly. However, although the use of the longer constructs decreased yields, the quantity of complex formed was apparently higher in the case of the four- than the three-protein



**Fig. 4.** Reconstitution by co-expression with the pET-MCN vector series of the SAGA deubiquitination (DUB) module (scSgf11/scSus1/scUbp8/scSgf73). (A) Domain organization of the four proteins composing the DUB module. Residue numbering for the domains is shown above the proteins. The color code used for the different proteins/domains is conserved throughout the figure. (B) SDS-PAGE analysis of soluble DUB sub and entire complexes retained on affinity resin. The different combinations of vectors and constructs (colored squares) are depicted above each lane. The N-terminal polyhistidine tag (denoted His in gel annotations) coding sequence is shown as a red ellipse. C-terminal degradation products of the scSgf73-ZnSca7 construct are marked with '\*'. (C) Crystallographic structure (PDB 3m99) of the scSgf11/scSus1/scUbp8/scSgf73-Zn complex linked (dotted line) to the NMR structure (PDB 2kkk) of the human Ataxine7 Sca7 domain (homologous to scSgf73-Sca7). The structure shows how interwoven all subunits are, explaining the requirement of the use of the co-expression technique for complex assembly.

complex, suggesting that stabilization of the complex occurs when the four subunits are present. Yet, stoichiometry imbalance remained, even when other subunits were tagged (data not shown).

The recent structures of the complex formed by full-length scSgf11, scSus1, scUbp8 together with the N-terminal domain of scSgf73, as well as the structure of the human Sca7 domain have now shed light on the results obtained by co-expression (Bonnet et al., 2010; Köhler et al., 2010; Samara et al., 2010; Fig. 4C). As observed in the co-expression experiments, complex assembly requires the formation of a scSgf11/scSus1 sub-complex with whom the N-terminal domain of scUbp8 and the very N-terminus of scSgf73 can interact, forming the so-called assembly module (Fig. 4C). The catalytic domain of scUbp8 is somewhat remote from the assembly module, but the N-terminal domain of scSgf73 binds at the interface of these two structural entities, thus explaining the stabilizing effect observed upon formation of the full complex.

The structure of the complex does not however include the longer scSgf73 construct which spans the Sca7 domain whose structure has been solved and which has been shown to interact with the H2A/H2B histone pair and could thus participate in the catalytic mechanism of the DUB module (Bonnet et al., 2010). The C-terminal degradation of the long scSgf73 construct observed in our co-expression trials (Fig. 4B, lanes 11, 14 and 15) suggests that the Sca7 domain does not associate tightly with the rest of the complex or may only do so in presence of the ubiquitinated nucleosome substrate.

### 3.5. Reconstitution of the histone octamer

Chromatin modification and chromatin remodeling mechanisms are currently subject to a very strong focus from many laboratories worldwide. Therefore, the requirement for reconstituted nucleosomes or nucleosomal arrays is very high. The reconstitution of homogeneous nucleosomes remains a complicated task since it requires (i) to reconstitute the 110 kDa histone octamer, (ii) to purify the nucleosomal DNA, and (iii) to assemble nucleosomes by mixing together and dialyzing the histone octamer and the DNA (Lüger et al., 1999a,b).

Already at the step of the histone octamer reconstitution, many laboratories express all histone proteins (H2A, H2B, H3 and H4) independently as inclusion bodies and then reconstitute the octamer using time-consuming refolding strategies (Lüger et al., 1999a,b). Interestingly, co-expression has been shown to be useful in reconstituting H2A/H2B and H3/H4 dimers (Anderson et al., 2010; Levenstein and Kadonaga, 2002). Yet, to our knowledge, nobody has tried to reconstitute the full histone octamer by co-expression.

We have used the pET-MCN series to attempt the reconstitution by co-expression of the histone octamer. After cloning the different histone genes in the pNEA-tH (encoding N-terminal His-tag fusion) and pNCS (providing native proteins) vectors, vector concatenation was then immediately used to create vectors encoding the H2A/H2B and H3/H4 histone pairs. Different vectors were made to produce the pairs without any protein tagged, or with either H2A or H4 tagged. The choice of the protein bearing the tag had been made upon initial binary co-expression tests looking for the best positioning of the affinity his-tag (data not shown). Vectors harboring all four histone genes were then created by concatenation, with the his-tag-encoding sequence placed in front of either H2A or H4.

Co-expression tests were then carried out on either the pairs alone or the four histones. For purification, high salt buffers were used (either 1 M or 2 M NaCl) since the histone octamer can only be reconstituted in presence of high concentration of salt (Lüger et al., 1999a,b). Fig. 5A shows that the solubility of the H2A/H2B pair is much higher in 1 M than in 2 M NaCl (compare lanes 1

and 5), whereas the solubility of the H3/H4 pair is less affected by salt concentration (lanes 2 and 6). In case of the four histones, quaternary complexes are observed whatever the histone tagged (H2A or H4) and the salt concentration. However, at 1 M salt an excess of the H2A/H2B pair is observed when H2A is tagged, excess which diminishes at higher salt concentration due to the lower solubility of the pair at 2 M NaCl (Fig. 5A, compare lanes 3 and 7). In the case where H4 is tagged, complexes appear also stoichiometric, but the co-migration of all proteins complicates stoichiometry interpretation (Fig. 5A, lanes 4 and 8).

Since the histone complex obtained at 2 M NaCl when H2A was tagged (Fig. 5A, lane 7) appeared stoichiometric and since all four histones could be easily visualized, we carried out a large-scale purification of this complex by affinity purification followed by gel filtration. As shown in Fig. 5B, two major peaks were observed on the gel filtration profile that corresponded to complex formed by the four histones and an excess of the H2A/H2B pair. This initial purification attempt already yielded a few milligrams of the histone octamer, demonstrating the feasibility in reconstituting this octamer by multi-expression in *E. coli*.

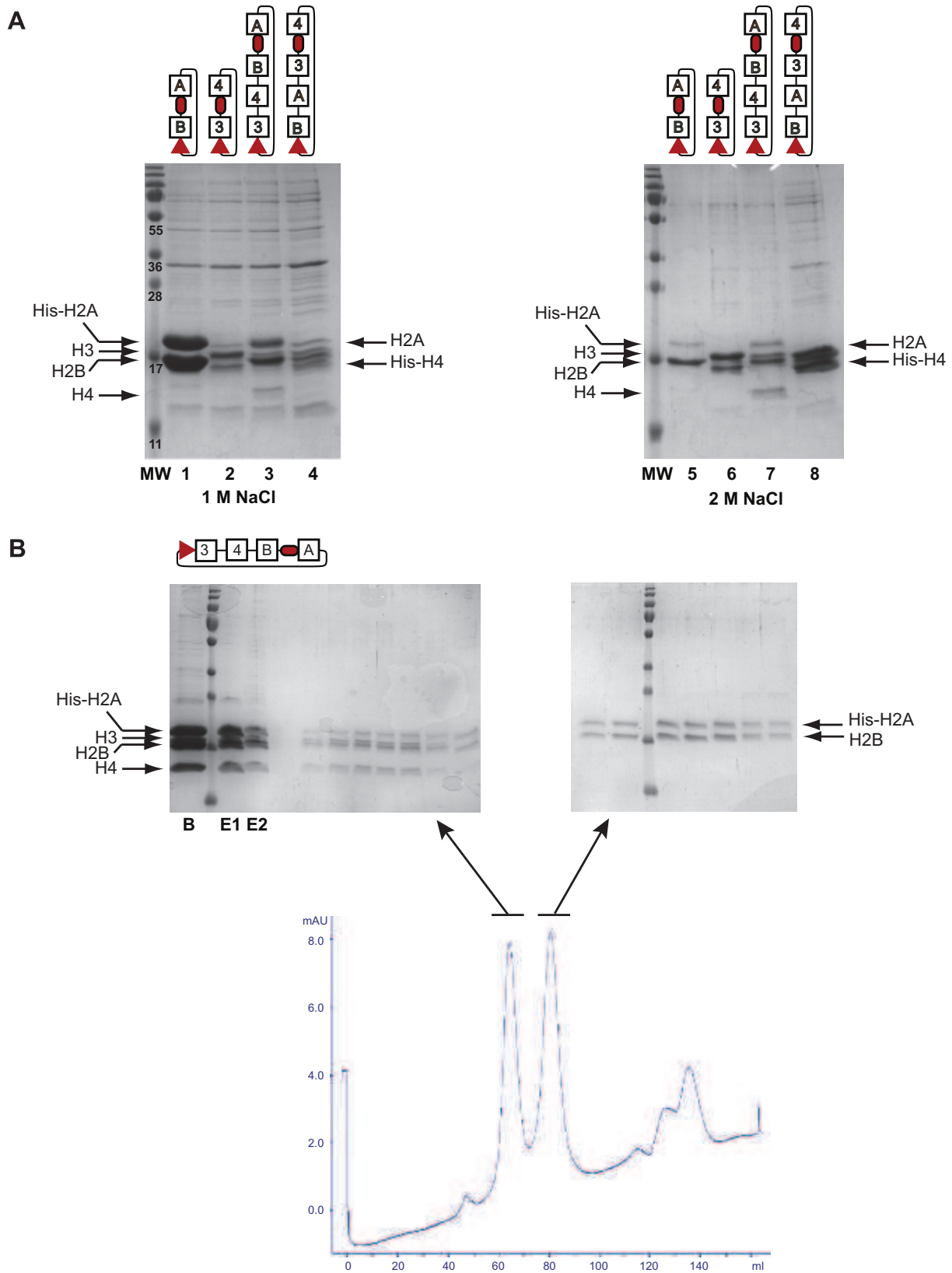
## 4. Discussion

Over the past decade, the co-expression technique has established itself as a method of choice for assembling complexes, notably when large amounts of homogeneous samples are required. Interestingly, however, many complexes reconstituted by co-expression still appear to be binary complexes. When larger complexes are to be studied, it is still relatively frequent that these are reconstituted from co-expressed binary complexes that are then further assembled *in vitro* with other binary complexes or single subunits expressed on their own. This reveals the progresses that can be made in using correct co-expression strategies.

Our results show that one likely reason for the restriction of the co-expression technology to mostly small complexes comes from the fact that the intricacy of complex assembly by co-expression increases with the number of subunits considered. The results obtained on the YMP and NFY complexes, although they address the reconstitution of smaller complexes, both in terms of size and number of subunits, provide clear examples of this increasing complexity. Specifically, these results highlight the importance for the choice of the strategy in reconstituting complexes by co-expression. Choosing the wrong strategy can definitively hamper or even prevent the reconstitution of a complex even in small-scale, or complicate its large scale purification. These results indicate that the optimal co-expression solution is varying from case to case and depends on the complexes themselves.

Notably, in both cases presented, but also for many other complexes studied in our laboratory, appropriate placement of the affinity tag is of paramount importance for complex formation. Clearly, location of the tag, at the N- or C-terminus, also affects the outcome of the co-expression experiment as mentioned in case of the NFY complex. It can be expected that using a variety of tags/fusions (polyhistidine, Strep, GST...) will also influence the experimental outcome, also due to steric considerations. Interestingly, other parameters, such as the use of single vectors harboring single versus multiple promoters, appear to affect complex quantity and stoichiometry alike, in a complex-dependent manner.

On the other hand, parameters such as the order of the genes or of the promoters on single vectors did not impact on the complex expressions presented here. It cannot be excluded, however, that this parameter plays a more significant role in other cases, notably when trying to express larger proteins. Finally, it is also likely that the co-expression system used may influence the reconstitution process and thus the reconstitution strategy (see Busso et al., this issue). Taken together, these results provide an in-depth



**Fig. 5.** Reconstitution of the histone octamer by co-expression with the pET-MCN system. (A) SDS-PAGE analysis of the soluble H2A/H2B and H3/H4 histone pairs and of the entire histone octamer complex retained on affinity resin after co-expression of the pairs or the octamer in *E. coli*. The different combinations of vectors used are depicted above each lane. A, H2A; B, H2B; 3, H3; 4, H4. The vector elements are displayed as in Fig. 2. The salt concentration in the lysis and purification buffer is indicated underneath the gels. The sizes of the molecular weight markers are shown and are the same for all gels in the figure. (B) Gel filtration profile and SDS-PAGE analysis of the gel filtration fractions upon large scale purification of the co-expressed histone octamer. The first peak of the gel filtration contains the histone octamer, the second peak contains the excess of the H2A/H2B pair. Lane B represents the complex retained on cobalt affinity beads. E1 and E2 are the elution fractions obtained upon washing of the affinity beads with an imidazole solution. All purification steps are performed in presence of 2 M NaCl.

understanding of the use of the co-expression technique, showing that a case that appears not suitable for co-expression may in fact be perfectly suitable, thus saving precious time on a reconstitution project.

The reconstitution of the DUB module by combinatorial co-expression experiments provides another proof-of-concept of the importance of the co-expression technique and demonstrates that this technology can be used for obtaining sufficient amounts of large macromolecular complexes. In fact, the structure of the SAGA DUB module clearly shows how closely interwoven the different subunits are, showing that the co-expression technique is essential for the reconstitution of this kind of complex. In agreement, both teams that reported the structure of this module used co-expression of the four subunits for complex reconstitution (Köhler et al., 2010; Samara et al., 2010).

The reconstitution of the histone octamer by co-expression also participates to the demonstration that co-expression in *E. coli* can ease large complex reconstitution. Here again, tagging of the right subunit, but also the choice of the right buffer, is important for obtaining a complex with balanced stoichiometry. Our preliminary purification trial showed that large scale purification can be carried out on this complex too and a few milligrams of purified histone octamer were obtained. It is most likely that improvement of the purification protocol will provide higher yield of histone octamer. However, the quantities already obtained should be sufficient to reconstitute enough nucleosomes for many applications. As such, the reconstitution by co-expression of the histone octamer provides a fast and attractive alternative to the refolding procedure used in many laboratories.

Several conclusions can be drawn from these studies. First, there is a definite requirement for a large set of experiments in a combinatorial approach towards successful co-expression of a particular complex. This experimental throughput increases dramatically with the number of subunits present in a complex, and calls for automated approaches to cope with the number of trials to be performed (Vijayachandran et al., this issue). However, it is highly likely that reconstitution of many complexes will only be feasible by co-expression and will rely on these automated approaches. Second, a flexible co-expression system is necessary, which allows for expedient investigation of many different strategies in parallel. Importantly, our results imply that the inherent characteristics of the protein complex studied influence strongly the parameters that lead to successful complex reconstitution. Therefore, it is most likely that the requirements for automated combinatorial approaches, detailed here for expression in *E. coli*, will be also similar for complex expressions in eukaryotic hosts.

### Conflict of Interest

The authors declare no conflict of interest.

### Author's contributions

C.R. designed and constructed the multi-expression vectors. C.R., M.-L.D., S.F., M.K. and T.M. carried out the testing of the vectors. C.R. oversaw the project and wrote the manuscript.

### Acknowledgments

The authors are grateful to Fulvia Bono and Elena Conti for the kind gift of the genes encoding the Y14, Mago and Pym proteins, to Jacques Bonnet and Didier Devys for the collaborative work and useful discussions on the SAGA DUB complex, to Ali Hamiche for the kind gift of the genes encoding the histones, and to Mailys Unterfinger for her contribution to this work. The authors thank

Imre Berger for the critical reading of the manuscript and useful discussions. The authors also thank the IGBMC, Institut de Génétique et de Biologie Moléculaire et Cellulaire, for assistance. MLD is supported by a PhD grant from the French Research Ministry. This work was supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Université de Strasbourg (UDS), the ANR PIRIBIO ANR-09-PIRI-0031-01 grant, the AVENIR/ARC grant (SF), and the European Commission SPINE2-Complexes Project (contract number LSHG-CT-2006-031220).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2011.02.001. The Supplementary Website describing in detail the use of the pET-MCN and pET-MCP series can be found at [http://igbmc.fr/Romier\\_Multi-Expression](http://igbmc.fr/Romier_Multi-Expression).

### References

- Anderson, M., Huh, J.H., Ngo, T., Lee, A., Hernandez, G., Pang, J., Perkins, J., Dutnall, R.N., 2010. Co-expression as a convenient method for the production and purification of core histones in bacteria. *Protein Expr. Purif.* 72, 194–204.
- Benschop, J.J., Brabers, N., van Leenen, D., Bakker, L.V., van Deutekom, H.W., van Berkum, N.L., Apweiler, E., Lijnzaad, P., Holstege, F.C., Kemmeren, P., 2010. A consensus of core protein complex compositions for *Saccharomyces cerevisiae*. *Mol. Cell* 38, 916–928.
- Berger, I., Fitzgerald, D.J., Richmond, T.J., 2004. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* 22, 1583–1587.
- Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., Romier, C., Berger, P., Richmond, T.J., Steinmetz, M.O., Berger, I., 2009. Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods* 6, 447–450.
- Bonnet, J., Wang, Y.H., Spedale, G., Atkinson, R.A., Romier, C., Hamiche, A., Pijnappel, W.W., Timmers, H.T., Tora, L., Devys, D., Kieffer, B., 2010. The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. *EMBO Rep.* 11, 612–618.
- Bono, F., Ebert, J., Unterholzner, L., Guttler, T., Izaurralde, E., Conti, E., 2004. Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep.* 5, 304–310.
- Bono, F., Ebert, J., Lorentzen, E., Conti, E., 2006. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126, 713–725.
- Bono, F., Cook, A.G., Grunwald, M., Ebert, J., Conti, E., 2010. Nuclear import mechanism of the EJC component Mago-Y14 revealed by structural studies of importin 13. *Mol. Cell* 37, 211–222.
- Boskovic, J., Coloma, J., Aparicio, T., Zhou, M., Robinson, C.V., Mendez, J., Montoya, G., 2007. Molecular architecture of the human GINS complex. *EMBO Rep.* 8, 678–684.
- Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., Conti, E., 2010. Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl. Acad. Sci. USA* 107, 10050–10055.
- Devos, D., Russell, R.B., 2007. A more complete, complexed and structured interactome. *Curr. Opin. Struct. Biol.* 17, 370–377.
- Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., Romier, C., 2010. The structure of an lws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. *EMBO J.* 29, 3979–3991.
- Fribourg, S., Romier, C., Werten, S., Gangloff, Y.G., Poterszman, A., Moras, D., 2001. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli*. *J. Mol. Biol.* 306, 363–373.
- Fribourg, S., Gatfield, D., Izaurralde, E., Conti, E., 2003. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat. Struct. Biol.* 10, 433–439.
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dimpfelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., Superti-Furga, G., 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631–636.
- Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H.W., Savchenko, A., Yee, A., Edwards, A., Vincenzelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo,

- S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang, M., Montelione, G.T., Stuart, D.L., Owens, R.J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., Gunsalus, K.C., 2008. Protein production and purification. *Nat. Methods* 5, 135–146.
- Held, D., Yaeger, K., Novy, R., 2003. New coexpression vectors for expanded compatibilities in *E. coli*. *Innovations* 18, 4–6.
- Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorselaer, A., Wurtz, J.M., Mandel, J.L., Tora, L., Devys, D., 2004. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum. Mol. Genet.* 13, 1257–1265.
- Jeyaprasath, A.A., Klein, U.R., Lindner, D., Ebert, J., Nigg, E.A., Conti, E., 2007. Structure of a Survivin-Borealin-INCENP core complex reveals how chromosomal passengers travel together. *Cell* 131, 271–285.
- Köhler, A., Pascual-Garcia, P., Llopis, A., Zapater, M., Posas, F., Hurt, E., Rodriguez-Navarro, S., 2006. The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitination through its interaction with Ubp8 and Sgf11. *Mol. Biol. Cell* 17, 4228–4236.
- Köhler, A., Schneider, M., Cabal, G.G., Nehrbass, U., Hurt, E., 2008. Yeast Ataxin-7 links histone deubiquitination with gene gating and mRNA export. *Nat. Cell Biol.* 10, 707–715.
- Köhler, A., Zimmerman, E., Schneider, M., Hurt, E., Zheng, N., 2010. Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* 141, 606–617.
- Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., Berger, P., 2010. A plasmid-based multigene expression system for mammalian cells. *Nat. Commun.* 1, 120.
- Kuhner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., Castano-Diez, D., Chen, W.H., Devos, D., Guell, M., Norambuena, T., Racke, I., Rybin, V., Schmidt, A., Yus, E., Aebbersold, R., Herrmann, R., Bottcher, B., Frangakis, A.S., Russell, R.B., Serrano, L., Bork, P., Gavin, A.C., 2009. Proteome organization in a genome-reduced bacterium. *Science* 326, 1235–1240.
- Levenstein, M.E., Kadonaga, J.T., 2002. Biochemical analysis of chromatin containing recombinant *Drosophila* core histones. *J. Biol. Chem.* 277, 8749–8754.
- Lorentzen, E., Dziembowski, A., Lindner, D., Seraphin, B., Conti, E., 2007. RNA channelling by the archaeal exosome. *EMBO Rep.* 8, 470–476.
- Lüger, K., Rechsteiner, T.J., Richmond, T.J., 1999a. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* 119, 1–16.
- Lüger, K., Rechsteiner, T.J., Richmond, T.J., 1999b. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* 304, 3–19.
- Perrakis, A., Romier, C., 2008. Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview. *Methods Mol. Biol.* 426, 247–256.
- Rodriguez-Navarro, S., 2009. Insights into SAGA function during gene expression. *EMBO Rep.* 10, 843–850.
- Romier, C., 2008. Protein complexes assembly by multi-expression in bacterial and eukaryotic hosts. In: Sussman, J.L. (Ed.), *Structural Proteomics*. World Scientific Publishing Company, London, pp. 233–250.
- Romier, C., Cocchiarella, F., Mantovani, R., Moras, D., 2003. The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *J. Biol. Chem.* 278, 1336–1345.
- Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P.H., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., Lebbink, J.H., Notenboom, V., Poterszman, A., Rochel, N., Cohen, S.X., Unger, T., Sussman, J.L., Moras, D., Sixma, T.K., Perrakis, A., 2006. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallogr. D: Biol. Crystallogr.* 62, 1232–1242.
- Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E., Wolberger, C., 2010. Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* 328, 1025–1029.
- Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., Bussow, K., 2007. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* 35, e43.
- Tan, S., Kern, R.C., Selleck, W., 2005. The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expr. Purif.* 40, 385–395.
- Tolia, N.H., Joshua-Tor, L., 2006. Strategies for protein coexpression in *Escherichia coli*. *Nat. Methods* 3, 55–64.
- Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H.G., Krasnov, A., Georgieva, S.G., Schule, R., Takeyama, K., Kato, S., Tora, L., Devys, D., 2008. A TFIIIC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. *Mol. Cell* 29, 92–101.



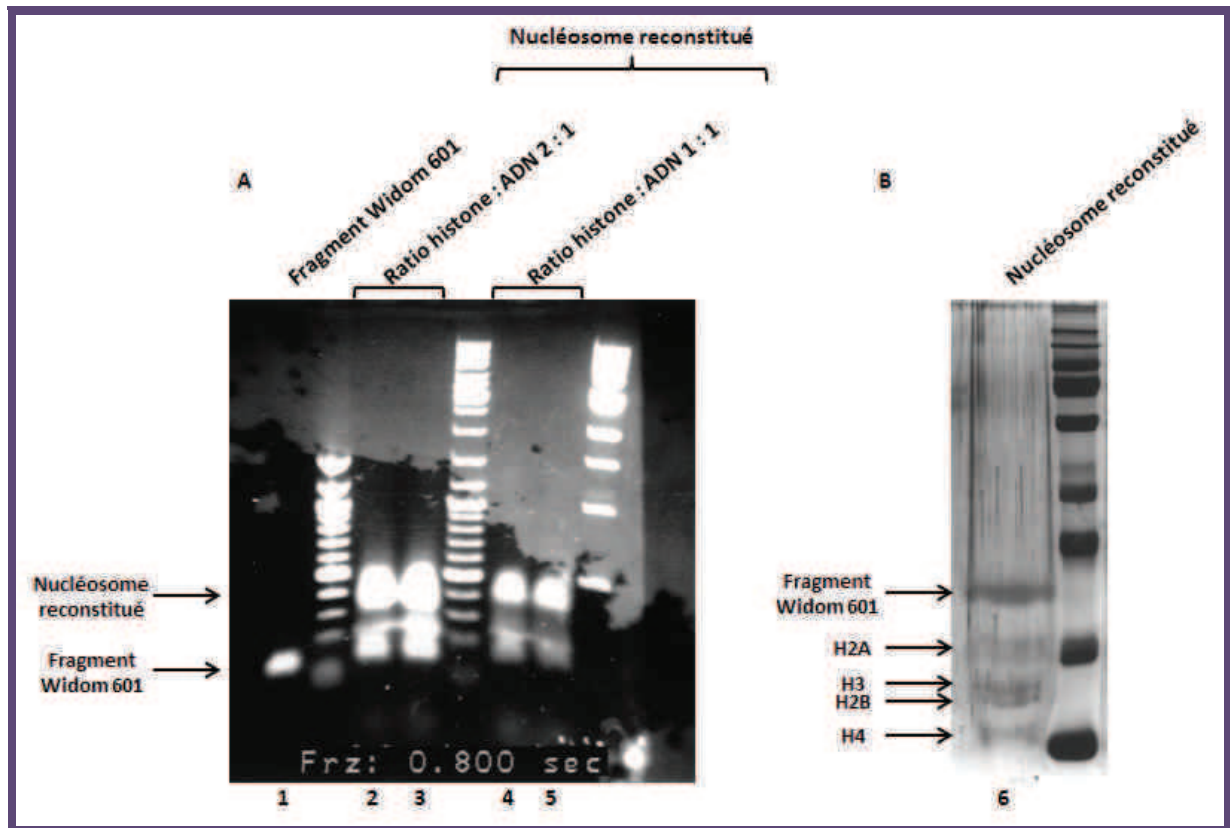
## 6.4. Validation de l'octamère d'histone co-exprimé – reconstitution du nucléosome

Afin de valider notre technique de co-expression de l'octamère d'histone et notamment de vérifier que l'octamère d'histone est bien formé, des tests ont été réalisés qui ont notamment eu pour but d'enrouler un ADN autour de cet octamère. Pour reconstituer le nucléosome, la séquence Widom601 a été utilisée. Cette séquence a été identifiée dans un ensemble d'ADN synthétisés chimiquement de manière aléatoire pour sa très haute affinité pour les octamères d'histones (Lowary et Widom, 1998). Après définition de cette séquence de 147 paires de bases, des sites EcoRV placés aux extrémités 5' et 3' du fragment d'ADN ont permis la concaténation d'une vingtaine de fragments et leur insertion dans un vecteur navette (Makde et al., 2010).

Le protocole préparation de l'ADN et de reconstitution du nucléosome est présenté en annexe 8.7. Brièvement, les fragments de 147pb sont extraits des vecteurs par digestion avec l'enzyme de restriction EcoRV. Ces fragments sont ensuite séparés du vecteur par une précipitation au PEG. Après mélange de l'ADN purifié et des octamères d'histone avec un ratio 2-1 et 1-1 (protéine-ADN), le nucléosome est reconstitué par diminution progressive de la concentration saline.

La formation de l'octamère est vérifiée par dépôt d'un échantillon sur gel d'agarose natif. La présence des histones et de l'ADN dans le complexe est ensuite vérifiée par extraction des bandes correspondant au nucléosome formé et analyse sur gel SDS-page (Figure 66).

Le dépôt des échantillons sur gel d'agarose natif montre clairement la présence de deux espèces : (i) une première espèce (lignes 2 à 5) dont le déplacement électrophorétique est identique au déplacement de l'ADN purifié (ligne 1) et correspond à l'excès d'ADN non enroulé autour des protéines et (ii) une seconde espèce (lignes 2 à 5) dont la mobilité est réduite par un encombrement plus important et qui correspond au nucléosome reconstitué. Le dépôt des échantillons avant (lignes 2 et 4) et après centrifugation (lignes 3 et 5) confirme que le nucléosome reconstitué est soluble. Le dépôt sur gel d'acrylamide de la bande correspondant à cette seconde espèce confirme la présence des quatre protéines histones et de du fragment Widom 601.



**Figure 66 : Reconstitution du nucléosome à partir de l'octamère co-exprimé**

(A) Le nucléosome reconstitué est déposé sur gel d'agarose natif (0,8%) avant (lignes 2 et 4) et après (lignes 3 et 5) centrifugation à haute vitesse. (B) Le nucléosome reconstitué est déposé sur gel d'acrylamide (12,5%) et coloré à l'argent.

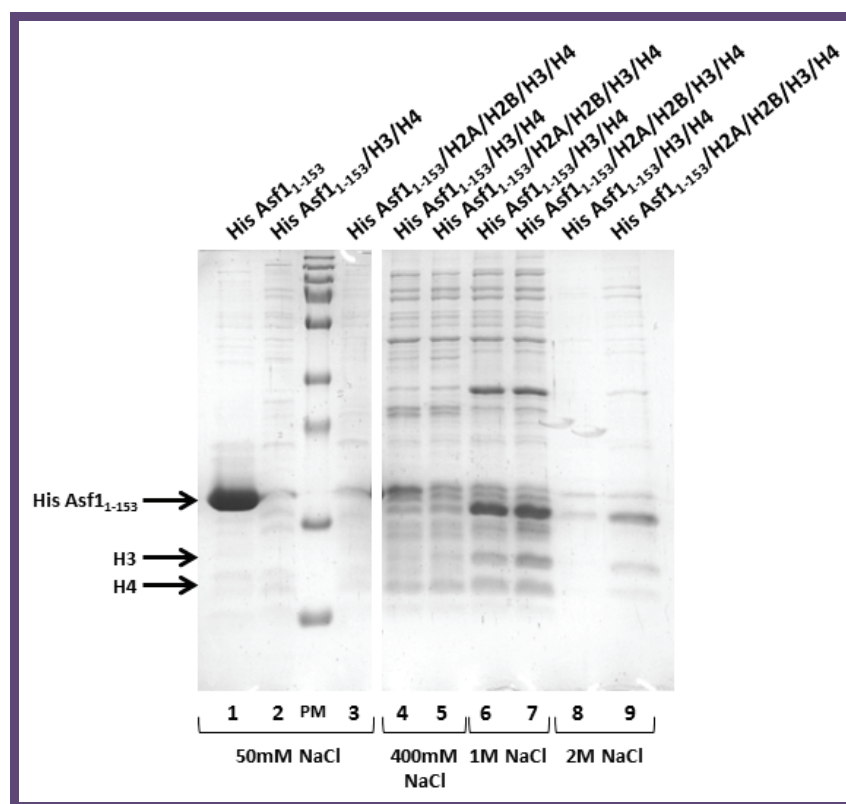
## 6.5. Etude de l'interaction entre des chaperonnes et les histones par co-expression

Les vecteurs créés pour la co-expression des histones ont été utilisés afin de tester l'interaction de Spt6 avec les paires et l'octamère d'histone. Afin de vérifier que ces tests permettaient bien d'observer ces interactions, ils ont été réalisés dans un premier temps avec la chaperonne d'histone Asf1, dont l'interaction avec la paire H3/H4 a déjà été caractérisée biochimiquement et structurellement (Natsume et al., 2007). De plus, la co-expression de cette chaperonne a déjà permis de solubiliser la paire H3/H4 produites dans les bactéries (Anderson et al., 2010).

### 6.5.1. Interaction de Asf1 avec la paire H3/H4

Le gène codant l'extrémité amino-terminale (construction 1-153) de la protéine Asf1 a été cloné dans le vecteur pNEA-tH permettant l'expression de la construction en fusion avec

une étiquette de purification poly-(His). Ce vecteur a ensuite été co-transformé avec les vecteurs (pnCS) codant soit pour la paire d'histone H3/H4, soit pour l'octamère H2A/H2B/H3/H4, les protéines entières ayant été utilisées. Quatre tampons de lyses ont été testés contenant : (i) 50mM NaCl, (ii) 400mM NaCl, (iii) 1M NaCl et (iv) 2M NaCl, tous contenant également 10mM Tris pH8. Ces différentes conditions salines ont été testées afin de couvrir les conditions de solubilité des protéines individuelles, Asf1 étant soluble à 50 et 400mM NaCl et la paire d'histone H3/H4 à 2M NaCl. Cependant un effet de solubilisation de ces différentes protéine est attendu dans le cas où les complexes seraient formés et notamment une solubilisation de la paire H3/H4 à plus faible concentration saline, qui serait alors plus proche des concentrations salines existant *in vivo*.



**Figure 67 : Co-expression du domaine amino-terminale de Asf1 et des histones**

*Le vecteur pnEA-tH codant pour la construction 1-153 de Asf1 a été co-transformé avec les vecteurs natif (pnCS) codant soit la paire d'histone H3/H4 (protéines entières) soit les quatre protéines histone (protéines entière). Les tests de co-expression ont été réalisés à quatre concentrations salines différentes.*

Un complexe entre Asf1 et la paire d'histone H3/H4 peut être formé lorsque le tampon de lyse contient 1M NaCl (Figure 67). En effet, à faible concentration saline (50mM NaCl), la protéine Asf1 est produite lorsqu'elle est exprimée seule (ligne 1), la co-expression avec les

paires et l'octamère d'histones conduisant à une diminution de la quantité de protéine visible sur le gel (lignes 2, 3, 4, et 5). Cette diminution pourrait notamment s'expliquer par une diminution de la solubilité de la protéine à ces concentrations salines lorsqu'elle interagit avec les histones. Par contre, ces tests réalisés à 1M NaCl font clairement apparaître un complexe entre Asf1 et H3/H4 (ligne 6 et 7), suggérant une diminution de la concentration saline nécessaire à la solubilisation de la paire H3/H4. Il semble également que dans le cas où Asf1 est co-exprimée avec l'octamère d'histone, la protéine séquestre la paire H3/H4 (ligne 7). La solubilité du complexe Asf1/H3/H4 à 2M est relativement faible, confirmant l'hypothèse d'une diminution de la salinité nécessaire à sa solubilisation. Aucune interaction aspécifique des histones avec la résine n'a pu être détectée au cours de ces tests (Figure 68, lignes 4, 5, 12 et 13). Il est à noter que lorsque Asf1 (1-153) est co-exprimée avec la paire H3/H4 dont les queues amino-terminales ont été enlevées, un complexe très fort est observé y compris à bas sel (notamment 50mM NaCl, l'optimum de solubilité étant à 400mM NaCl), montrant que la solubilité des histones entières H3 et H4 affecte aussi la formation de ce complexe (données non présentées).

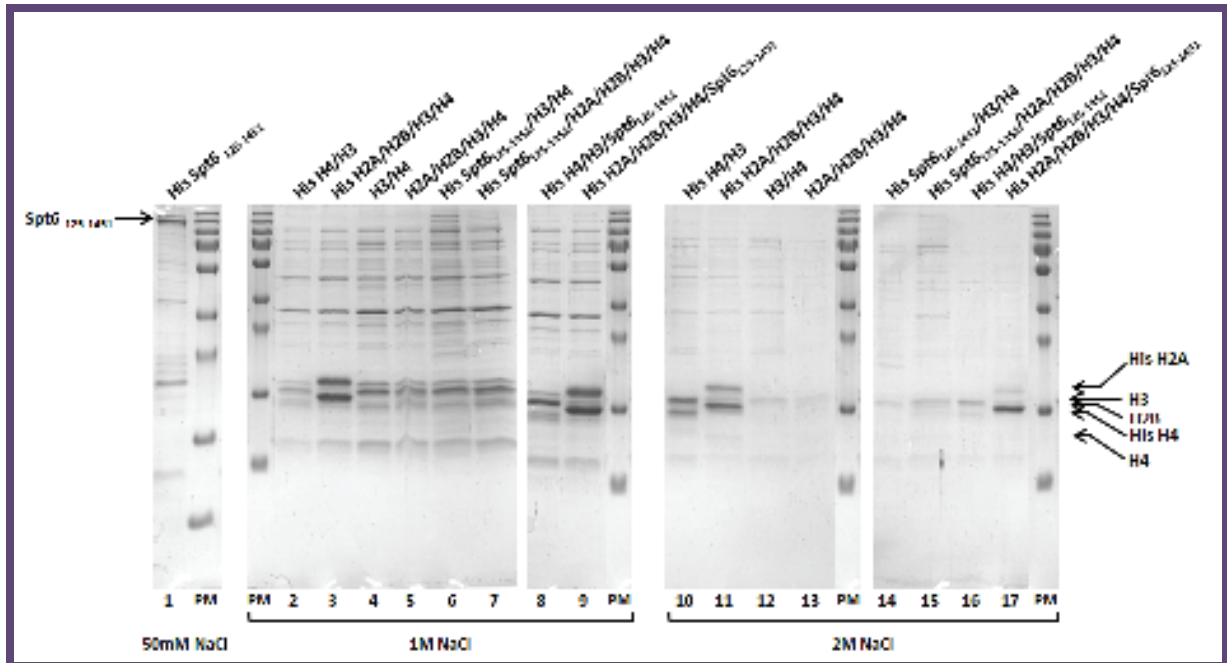
Ainsi, la co-expression d'une chaperonne d'histone, Asf1, avec la paire d'histone H3/H4 ou l'octamère d'histone permet bien de reconstituer le complexe ternaire Asf1/H3/H4. La formation du complexe lors de la co-expression avec l'octamère suggère une interaction préférentielle de la paire H3/H4 avec Asf1, cette interaction impliquant la zone d'interaction du second dimère H3/H4 dans l'octamère d'histone (Agez et al., 2007; Natsume et al., 2007).

### 6.5.2. Interaction de Spt6 avec les histones

Les tests de co-expression réalisés ont pour but (i) de tenter de vérifier l'interaction entre Spt6 et la paire d'histone H3/H4 et (ii) d'identifier la zone d'interaction au niveau de Spt6 dans l'optique de caractériser biochimiquement et structuralement ce complexe. Dans un premier temps, des tests d'expression ont été réalisés avec une construction longue de la protéine Spt6 de levure (construction 125-1451) afin de vérifier cette interaction. De la même manière que pour Asf1, quatre concentrations salines ont été testées, mais dans ce cas l'influence de la position de l'étiquette de purification a aussi été testée (Figure 68).

Ces tests d'expression ne permettent pas de mettre clairement en évidence des complexes entre Spt6 et les histones. En effet, la construction de la protéine Spt6 choisie ne permet pas d'obtenir de grandes quantités de protéine (ligne 1). Ainsi, aucun complexe ne

peut être détecté à faibles concentrations salines (50 et 400mM NaCl, gels non présentés). Spt6 reste faiblement produite à 1M NaCl (ligne 6 et 7) mais aucune bande correspondant aux protéines histones ne peut être détectée. Les taux d'expression sont très faibles à 2M NaCl, les bandes correspondants aux protéines histones sont notamment faibles, aucune bande ne correspondant à Spt6 ne pouvant être détectée.



**Figure 68 : Co-expression de la protéine Spt6 avec la paire d'histone H3/H4 et l'octamère d'histone.**

*La construction 125-1451 de la protéine Spt6 (portant l'étiquette de purification et native) a été co-transformée avec les vecteurs codant la paire d'histone H3/H4 et les quatre protéines histones.*

## 6.6. Conclusions

Nos travaux ont permis de créer un outil permettant la co-expression de l'octamère d'histone en nous appuyant sur les connaissances accumulées dans l'équipe sur cette technique (Fribourg et al., 2001; Romier et al., 2006; Perrakis et Romier, 2008). Ce complexe de près de 110kDa joue un rôle essentiel dans la régulation de la transcription, notamment par son interaction avec un très grand nombre de facteurs : facteurs épigénétiques, protéines chaperonnes et facteurs de remodelage de la chromatine. Cependant, l'étude des complexes formés entre ces facteurs et le nucléosome est compliquée par la difficulté que représente la production de l'octamère d'histone. En effet, les protocoles de reconstitution utilisés, bien que permettant l'obtention de très grandes quantités de protéines, sont longs et difficiles à mettre en œuvre puisqu'ils nécessitent (i) la production des protéines dans les bactéries sous forme

de corps d'inclusion, (ii) la purification à partir de ces corps d'inclusions et le refolding des protéines, (iii) et la formation des paires puis de l'octamère par dialyses successives (Luger et al., 1999). Ainsi, la mise en place de ce protocole peut nécessiter de nombreux jours de manipulations. L'outil de co-expression qui a été élaboré permet au contraire la production de l'octamère d'histone soluble, sa purification étant alors plus aisée. De plus, la capacité à reconstituer un nucléosome à partir des octamères produits par co-expression montre que ces octamères sont fonctionnels.

Par la réalisation de tests d'expression entre la chaperonne d'histone Asf1 et les vecteurs de co-expression, nous avons également démontré que cet outil permettait de tester l'interaction des protéines histones avec des interactants et notamment avec leur protéine chaperonne. Des tests réalisés avec la protéine Spt6 n'ont par contre pas permis de mettre en évidence une interaction claire entre la protéine et les paires ou l'octamère d'histone. Ces travaux restent cependant à poursuivre. Tout d'abord, la réalisation de tests de co-expression avec des constructions de la protéine Spt6 plus soluble que la construction utilisée permettrait d'observer plus aisément ce complexe s'il est formé et de définir la zone d'interaction. En effet des études récentes suggèrent que Spt6 interagit via son domaine HtH avec le nucléosome (McDonald et al., 2010) la présence de ce domaine localisé en amont de la zone d'interaction avec Iws1 ayant été confirmée par la résolution de la structure du domaine central de Spt6 de levure. Ainsi, il sera également nécessaire de tester l'interaction entre Spt6 et le nucléosome, notamment par la réalisation de gels natifs après mélange des protéines. Ces expériences pourront être réalisées avec plusieurs constructions solubles de Spt6, permettant également la définition de la zone d'interaction.

Ainsi, bien qu'aucune interaction entre Spt6 et la paire H3/H4 ou l'octamère n'ait pu être observée lors de la réalisation des premiers tests de co-expression, différentes expériences sont encore nécessaires afin de comprendre si ce complexe existe et, si oui, comment il est formé afin de résoudre sa structure.



# **Chapitre 7**

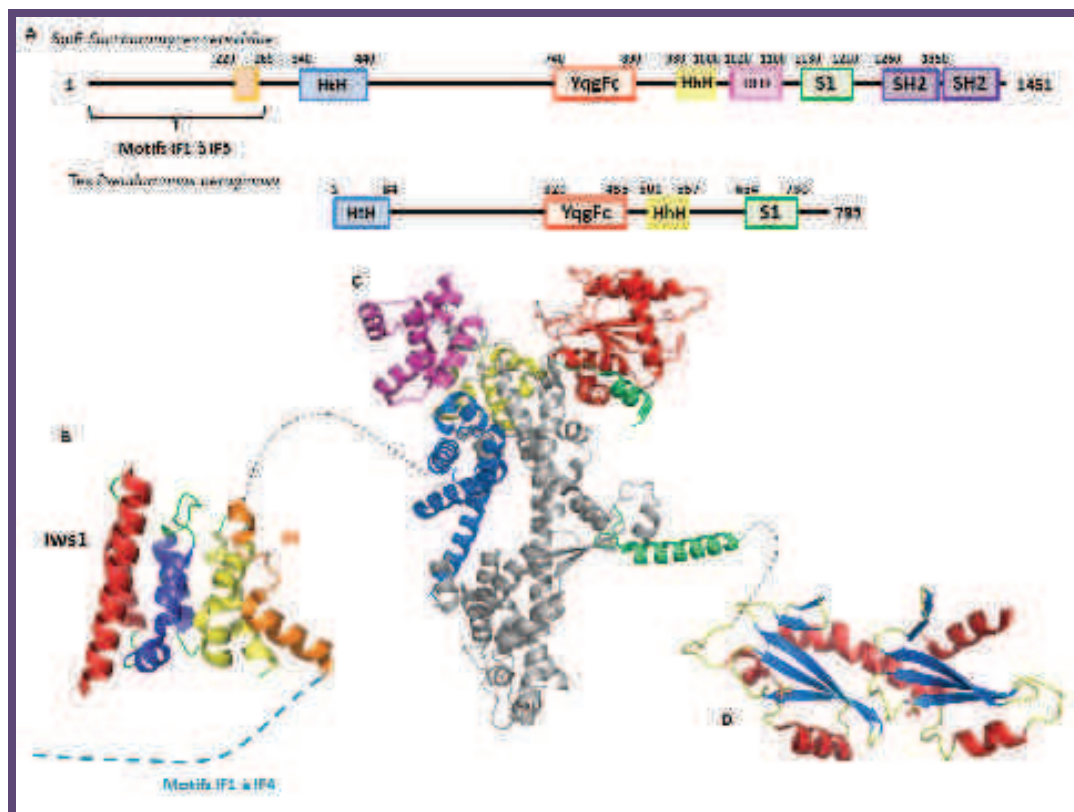
## **Conclusions générales et perspectives**





La protéine Spt6 et le complexe qu'elle forme avec Iws1 jouent un rôle important au cours de la transcription eucaryote. En effet, ce complexe protéique intervient à la fois dans les mécanismes de transcription et leur régulation, dans les mécanismes de modulation de la structure de la chromatine, et dans les mécanismes d'export et de maturation des ARN messagers. Lorsque ce travail a été initié, aucune information structurale sur cette protéine ni le complexe qu'elle forme avec Iws1 n'avait été obtenues et seule une vision parcellaire des domaines structuraux présents dans Spt6 et Iws1 était connue.

Les travaux réalisés au cours de cette thèse ont eu pour but de caractériser biochimiquement, structuralement et fonctionnellement la protéine Spt6, le complexe Spt6/Iws1 et leur interaction avec d'autres partenaires. Les études menées ces quatre dernières années ont permis de compléter la vision structurale de Spt6 mais également du complexe Spt6/Iws1 (Figure 69).



**Figure 69 : Organisation de la protéine Spt6**

(A) Les domaines fonctionnels des protéines Spt6 de levure et de la protéine Tex de *Pseudomonas aeruginosa* identifiés sont représentés, (B-C-D) informations structurales disponibles sur la protéine Spt6 : (B) structure du complexe ecSpt6/Iws1, le fragment de Spt6 est coloré en orange, la partie amino-terminale de Spt6 étant par des pointillés bleus (C) structure du domaine central de la protéine Spt6, (D) structure du domaine carboxy-terminal de alSpt6.

Nos travaux ont ainsi permis de caractériser biochimiquement, structuralement et fonctionnellement les complexes Spt6/Iws1 d'*Encephalotozoon cuniculi* et de levure (Figure 69B). Notre étude a également permis d'identifier une forte homologie structurale entre Iws1 et le domaine amino-terminal du facteur d'élongation de la transcription TFIIS. De plus, en nous basant sur une homologie de séquence entre Spt6, Spt8 et Med13, nous avons caractérisé précisément les complexes Spt8/TFIIS et Med13/TFIIS, qui avaient été décrit précédemment et qui ont été impliqués dans la formation de PIC transcriptionnellement actifs (Wery et al., 2004; Guglielmi et al., 2007; Kim et al., 2007). Nos travaux nous ont également poussés à analyser plus précisément le domaine amino-terminal de Spt6, décrit jusque là comme une région fortement chargée négativement, et à identifier cinq motifs hydrophobes, le dernier de ces motifs étant également impliqué dans la liaison à Iws1. Nous avons ainsi identifié un complexe potentiel formé entre deux motifs hydrophobes de Spt6 et le domaine amino-terminal de TFIIS. Alors que les homologies structurales et de séquences suggéraient que le mode de liaison entre les différents motifs hydrophobes identifiés sur Spt6, Spt8 et Med13 et les protéines Iws1 et TFIIS étaient similaires, nos premières caractérisations structurales montrent plutôt que le mode de liaison des complexes contenant TFIIS est différent de celui contenant Iws1.

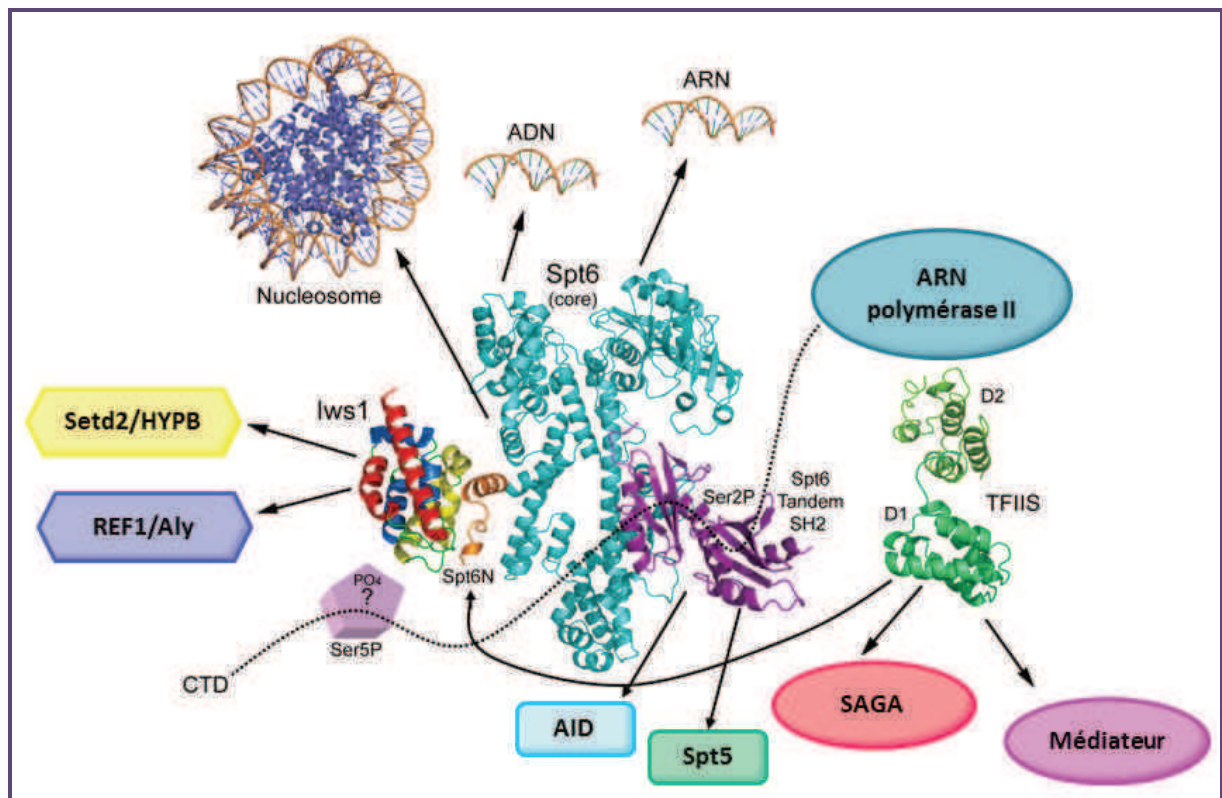
Nous avons également caractérisé le domaine carboxy-terminal de Spt6 (Figure 69D). Au début de ces travaux, le rôle de ce domaine dans le recrutement de Spt6 au cours de l'élongation de la transcription avait été décrit dans les cellules de souris, notamment l'implication du domaine SH2 contenu dans cette région de la protéine. La résolution par cristallographie de la structure du domaine carboxy-terminal entier de Spt6 a mis en évidence la présence d'un second domaine SH2 non canonique, formant ainsi un tandem de domaines SH2. L'analyse de la structure de ce domaine nous a conduit à proposer un mode de reconnaissance par le domaine carboxy-terminal de Spt6 de la plus grande sous-unité de l'ARN polymérase II. Cependant les tests d'interaction réalisés n'ont pas permis de mettre clairement en évidence une interaction directe avec des peptides phosphorylés du CTD de cette enzyme, alors que le tandem de domaines SH2 retient l'ARN polymérase présente dans un extrait de levure et que sa délétion chez la levure conduit à l'apparition de phénotypes qui confirment le rôle essentiel de ce domaine de la protéine.

Le rôle de chaperonne d'histone de Spt6 a été le premier rôle décrit de la protéine. Cependant, au début de ces travaux, aucune caractérisation biochimique ou structurale d'un

complexe entre les histones et Spt6 n'avait été réalisée. Afin de caractériser cette interaction, nous avons dans un premier temps utilisé la technique de co-expression afin de développer des outils pour co-exprimer les paires et l'octamère d'histone. Ces vecteurs de co-expression ont par la suite été utilisés afin de tester l'interaction de Spt6 avec ces deux complexes. Cependant, cette approche ne nous a pas permis de mettre en évidence une interaction entre Spt6 et les paires ou l'octamère d'histones.

Nos travaux et ceux publiés par d'autres équipes ont ainsi permis de compléter la vision structurale et fonctionnelle de Spt6 et du complexe Spt6/Iws1. En effet, la structure du domaine central de Spt6 a récemment été résolue (Figure 69C) (Close et al., 2011). Cette structure confirme l'homologie décrite précédemment avec la protéine bactérienne Tex (Johnson et al., 2008). Ces structures ont notamment révélé la présence de plusieurs domaines de liaison aux acides nucléiques, leurs liaisons aux acides nucléiques en termes moléculaires restant encore à caractériser.

Ainsi, bien qu'un grand nombre de données aient été publiées sur le rôle et la structure de Spt6 et du complexe Spt6/Iws1, un grand nombre de questions restent en suspens. En effet, différentes interactions ont été décrites qui ne sont à l'heure actuelle pas encore caractérisées. La Figure 70 récapitule l'ensemble de ces interactions et montre l'étendue des études structurales, biochimiques et fonctionnelles qui restent à effectuer sur Spt6 sachant que de nouvelles interactions seront très probablement découvertes dans le futur.



**Figure 70 : Représentation schématiques des différentes interactions mises en place par Spt6 et le complexe Spt6/Iws1**

Le complexe Spt6/Iws1 a été impliqué dans le recrutement, au cours de la transcription, de facteurs épigénétiques (Setd2), d'export (REF1/Aly) et de maturation (Rrp6). Les premiers tests que nous avons réalisés ne nous ont pas permis de mettre en évidence des complexes entre ces trois protéines et Spt6/Iws1. Cependant, différentes approches doivent encore être testées afin de caractériser ces complexes. Il est notamment possible que ces interactions ne soient mises en place que chez les eucaryotes supérieurs. Ces caractérisations et la résolution de la structure de ces complexes devraient permettre de comprendre comment ces facteurs sont recrutés au cours de la transcription et notamment quelles régions de Spt6 et Iws1 sont nécessaires à ces interactions.

L'interaction de Spt6 avec le nucléosome a été décrite récemment (McDonald et al., 2010). En effet, une interaction entre le nucléosome et un fragment de Spt6, contenant la zone d'interaction de Iws1 et le domaine HtH, a été démontrée par des expériences de gel retard. Cette interaction n'apparaît a priori possible qu'en l'absence de Iws1 et nécessite la présence de la protéine Nh6p de la famille des protéines HMG (High Mobility Group), cette protéine étant également nécessaire à l'activité de chaperonne du complexe FACT. Cette interaction entre Spt6 et le nucléosome n'ayant pour l'instant pas été caractérisée structuralement, un

projet important va être d'étudier ce complexe par cristallographie et/ou SAXS. Cette caractérisation devrait notamment être facilitée par l'utilisation de notre système de co-expression des histones, qui devrait faciliter la production d'octamères d'histones. L'analyse de la structure d'un complexe formé par Spt6 et le nucléosome devrait notamment permettre de comprendre le rôle précis de Spt6 dans la recondensation de la chromatine au cours de la transcription.

L'identification de plusieurs domaines de liaison aux acides nucléiques dans le domaine central de Spt6 suggère également que Spt6 pourrait être capable d'interagir avec des molécules d'ADN ou d'ARN, ce qui n'a pour l'instant pas été démontré. La présence de ces domaines de liaison pourrait notamment expliquer l'implication de Spt6 dans la régulation de l'expression du gène *SER3* de levure par l'ARN non codant *SRG1*. Une fonction similaire est retrouvée dans l'homologue bactérien Tex qui, chez *Staphylococcus aureus*, est impliqué dans la régulation de l'expression des gènes de virulence de cet agent pathogène par son interaction avec un ARN non codant (groupe de Pascal Romby, IBMC, données non publiées). Ainsi, l'identification des molécules d'ARN interagissant avec Spt6 permettrait de comprendre l'implication de cette protéine dans ces mécanismes de régulation, et la résolution de la structure de complexes Spt6/ARN, mais également Tex/ARN, permettrait de comprendre l'implication de ces deux protéines dans ces mécanismes transcriptionnels à la fois chez les procaryotes et les eucaryotes. Afin de comprendre ces différents aspects, une collaboration avec le Dr. Pascale Romby a été entreprise.

Le domaine carboxy-terminal de Spt6 est nécessaire à l'interaction de la protéine avec l'ARN polymérase II et à la croissance des levures. Alors que nos données montrent que l'interaction de ce domaine avec le CTD de l'ARN polymérase II nécessite des déterminants qu'il reste à définir, ce domaine de Spt6 a également été impliqué dans l'interaction avec d'autres partenaires. Au cours de cette thèse, l'interaction du tandem SH2 de Spt6 avec les protéines Spt5 et la cytidine déaminase AID a été testée, sans permettre d'identifier des interactions fortes. L'étude de ces complexes va encore nécessiter la réalisation de différents tests afin de caractériser ces complexes. De plus, Spt6 semble également participer à la transcription réalisée par l'ARN polymérase I, une interaction entre le domaine aminoterminal de Spt6 et la protéine Rpa43 de cette machinerie transcriptionnelle ayant récemment été décrite par double hybride. Cette interaction, qui n'est pour l'instant pas caractérisée

précisément, pourrait indiquer que Spt6 joue un rôle plus large dans la transcription que celui actuellement connu.

Si l'interaction entre Spt6 et le CTD de l'ARN polymérase II semble maintenant mieux caractérisée, notamment grâce aux études réalisées ces quatre dernières années, l'interaction entre Iws1 et l'ARN polymérase II n'est que peu comprise. En effet, une interaction avec le CTD de l'ARN polymérase II a été proposée sur la base d'études réalisées chez la levure, nos données structurales suggérant qu'une surface spécifique de Iws1 est dédiée à cette fonction. Cependant, aucune interaction directe n'a pour l'instant été démontrée. Cet aspect reste à étudier et l'identification de la sous-unité du complexe de transcription qui permet le recrutement de Iws1 en l'absence de Spt6 permettrait de comprendre le mode d'action de cette protéine au cours de la transcription, cette protéine étant notamment nécessaire à la régulation de l'expression de certains gènes inductibles de la levure.

Une étude que nous avons également entreprise est la caractérisation structurale du complexe Spt8/TFIIS, qui découle de l'observation des homologues de structure et de séquence avec les protéines du complexe Spt6/Iws1. La résolution de la structure de ce complexe va notamment permettre de comprendre quel est le mode de liaison entre ces deux protéines et de le comparer au mode de liaison que nous avons observé dans le cas du complexe Spt6/Iws1. De plus, la résolution de la structure du complexe Spt8/TFIIS permettra de comprendre le mode de recrutement de TFIIS par SAGA et le Médiateur au cours de l'initiation de la transcription.

Ainsi, bien que de nombreuses informations soient maintenant disponibles concernant l'organisation et la fonction de Spt6 et du complexe Spt6/Iws1, un très grand nombre de questions concernant le rôle de ces protéines restent en suspens. Les données actuellement disponibles montrent que ces deux protéines participent (i) à la transcription, en augmentant les taux de transcription par l'ARN polymérase II, (ii) à la régulation de la transcription, notamment en modulant la structure de la chromatine et le dépôt de marques épigénétiques, et (iii) aux mécanismes qui sont couplés à la transcription : la maturation du transcrit et son export. Un nombre très importants de facteurs sont requis pour la mise en place de ces fonctions, Spt6 et Iws1 semblant jouer un rôle important dans ces mécanismes par leur participation au recrutement des différents facteurs impliqués.

Les études biochimiques, structurales et fonctionnelles qu'il reste à entreprendre vont notamment avoir pour but la caractérisation des différents complexes formés par Spt6 et Iws1 avec d'autres effecteurs des mécanismes nucléaires et vont permettre d'aider à la compréhension du rôle de ces complexes. Ces études devraient notamment permettre de comprendre la régulation fine de la transcription à laquelle Spt6 et Iws1 participent. Au cours de ces études, il sera important de tenir compte des différences fonctionnelles entre les protéines Spt6 et Iws1 de levure et les mammifère. L'obtention d'une vision globale du rôle de ces protéines va ainsi nécessiter l'analyse de ces différents systèmes et leur comparaison.

Les organismes eucaryotes mettent en œuvre un mécanisme extrêmement complexe afin d'aboutir à la production des protéines. La transcription, qui constitue la première étape de ce mécanisme, est un parfait exemple de cette complexité. En effet, un nombre très important de protéines sont recrutées pour la mise en place de la transcription, de sa régulation, et des différents mécanismes associés. Les protéines Spt6 et Iws1, qui ont été étudiées au cours de cette thèse jouent un rôle essentiel dans le couplage de ces différents mécanismes, notamment par leur implication dans le recrutement d'autres facteurs. Les nombreuses études qu'il reste à développer pour comprendre le rôle de Spt6 et Iws1 en termes moléculaires et fonctionnels vont permettre de mieux appréhender la régulation fine de ces mécanismes, justifiant ainsi le travail à venir sur le complexe Spt6/Iws1. Clairement, les résultats obtenus au cours de cette thèse, ainsi que ceux de nos compétiteurs, constituent la base sur laquelle appuyer ces futures études.





# **Chapitre 8**

## **Annexes**



## 8.1. Protocoles de biologie moléculaire

Le clonage a pour but d'insérer le gène codant pour une protéine d'intérêt dans un vecteur dans le but d'exprimer cette protéine dans un organisme hôte. Le Chapitre 2 a pour but de décrire rapidement les différentes étapes nécessaires à l'obtention des vecteurs d'expression, le but de cette annexe étant de présenter de manière plus détaillée les diverses techniques employées.

### 8.1.1. La réaction PCR

L'amplification du gène à partir soit d'un génome, soit d'un vecteur de départ est réalisée par réaction de polymérisation en chaîne. Cette réaction est catalysée par une polymérase, l'enzyme Phusion (Finnzymes), qui va ajouter des dNTP par complémentarité à la matrice à partir des amorces sens et anti-sens. La composition du milieu réactionnel est présentée dans le Tableau 16.

Composition d'une réaction de PCR réalisée avec l'enzyme Phusion dans un volume réactionnel de 50 $\mu$ L.

	Volume réactionnel de 50 $\mu$ L
Matrice (50ng/ $\mu$ L)	1 $\mu$ L
Amorces sens et anti-sens (100 $\mu$ M)	0,5 $\mu$ L de chaque
Tampon de l'enzyme Phusion (10X)	10 $\mu$ L
Désoxyribonucléotides (dNTP) (2mM)	5 $\mu$ L
Enzyme Phusion (5U/ $\mu$ L)	0,5 $\mu$ L
Eau milliQ	qsp. 50 $\mu$ L

**Tableau 16 : Mélange réactionnel de la PCR**

Cette réaction est réalisée dans un thermocycleur, les cycles de la réaction sont présentés dans le Tableau 17.

Cycle			Conditions
Dénaturation		Ouverture complète des matrices doubles brins	1 minute, 95°C
Répétitions n cycles	Dénaturation	Dissociation des brins synthétisés de la matrice	10 secondes, 95°C
	Hybridation	Hybridation des amorces sens et anti-sens à la matrice	10 secondes, entre 55 et 58°C en fonction de la température d'hybridation définie lors de la comande des amorces
	Elongation	Synthèse du brin d'ADN complémentaire à la matrice	Temps fonction de la longueur du brin à synthétiser : 15 secondes / 1000nt, 72°C
Elongation finale		L'enzyme complète les fragments	10 minutes, 72°C
Conservation finale		L'échantillon est conservé	Temps infini, 4°C

**Tableau 17 : Cycles de la réaction de PCR.**

### 8.1.2. Clonage en vecteur navette par restriction/ligation

Une fois la réaction de PCR réalisée, les fragments amplifiés sont purifiés sur gel, excisés du gel puis extrait du gel avec le kit Nucleospin® Extract II de Macherey Nagel. Les fragments obtenus sont alors digérés avec les enzymes de restrictions dont les sites de reconnaissance ont été placés sur les amorces sens et anti-sens. Pour ces réactions, la totalité des fragments purifiés est digérée dans un volume de 20µL avec 2U de chaque enzyme, la réaction étant placée à 37°C.

Après une nouvelle purification de ces fragments sur colonne Nucleospin® Extract II de Macherey Nagel afin d'éliminer les sels et les enzymes présentes, une réaction de ligation est mise en place. Dans un volume réactionnel de 10µL, 5µL de fragments digérés sont mélangés à 0,5µL de vecteur ouvert (environ 10ng), au tampon de la ligase et à la ligase. Le vecteur est préparé indépendamment par digestion avec les enzymes de restriction et est déphosphorylé avec 1U d'enzyme SAP (Shrimp Alkaline Phosphatase) afin d'éviter sa recircularisation. La réaction de ligation est laissée à température ambiante au minimum 1h puis est transformée dans une souche bactérienne. Ces transformations étant réalisées dans le but de réaliser des clonages, des souches DH5α sont généralement utilisées. Le lendemain, un certain nombre de colonies sont mises en cultures dans un milieu liquide 2LB contenant l'antibiotique dilué au 1000<sup>ème</sup>, ces cultures étant placées sous agitation à 37°C pendant 16h. A cette étape, la présence de l'insert dans les vecteurs contenus dans les bactéries peut être vérifiée par PCR sur colonies. Le lendemain, les plasmides contenus dans ces bactéries sont

extraits, la présence de l'insert est vérifiée par digestion du vecteur avec les enzymes de restriction utilisées lors du clonage et l'insert est séquencé.

### 8.1.3. Sous clonage en vecteur d'expression

Une fois la séquence de l'insert vérifiée par séquençage, il est ré-extrait du vecteur navette et inséré dans les vecteurs d'expression de destination. Deux techniques ont été principalement utilisées, dont les protocoles sont décrits ici.

#### 8.1.3.1. Sous-clonage par restriction / ligation

Environ 1µg de vecteur navette est digéré avec les enzymes de restrictions qui ont été utilisées pour le clonage en vecteur navette. Cette réaction est réalisée dans un volume réactionnel de 10µL, 2U de chaque enzyme étant utilisées. Le produit de la digestion est alors déposé sur gel d'agarose et la bande correspondant à l'insert est découpée, l'insert étant alors élué de l'agarose avec le kit Nucleospin® Extract II de Macherey Nagel. Après purification, une réaction de ligation est mise en place entre 5µL de fragment et 0,5µL du vecteur d'expression ouvert et déphosphorylé. Après transformation dans des bactéries (généralement DH5α), mise en culture liquide et extraction des plasmides, la présence de l'insert est vérifiée par digestion avec les enzymes utilisées lors du clonage. Les inserts insérés dans les vecteurs d'expression ne sont plus reséquencés, leur séquence ayant déjà été vérifiée dans les vecteurs navettes.

#### 8.1.3.2. Sous-clonage par recombinaison homologue

Dans le cas certains cas, la recombinaison homologue a été utilisée afin de transférer l'insert d'un vecteur navette vers le vecteur d'expression, notamment dans le cas de la production de bacmides. Cette réaction de recombinaison est réalisée par incubation du vecteur navette (vecteur d'entrée) et du vecteur de destination (100ng de chaque vecteur) avec une enzyme LR clonase (2U) qui va réaliser la réaction d'échange des brins. Lors de la transformation du produit de la réaction, le gène *ccdB* présent dans le vecteur de destination non recombiné va permettre la sélection des vecteurs de destination contenant le gène d'intérêt. Des colonies sont alors mises en culture liquide et le lendemain, les plasmides sont extraits des bactéries. La présence de l'insert est alors vérifiée par PCR ou digestion avec des enzymes de restriction si des sites de reconnaissance ont été placés sur les amorces de la PCR. Là encore, les inserts ne sont pas reséquencés.

### 8.1.4. Les souches bactériennes

Au cours de ce travail de thèse, différentes souches bactériennes ont été utilisées, le choix de l'utilisation de l'une de ces souches dépend des caractéristiques de ces différentes souches, qui sont présentées dans le Tableau 18.

Souche	Utilisation	Génotype	Caractéristiques
<b>DH5α</b>	Clonage	F- $\phi 80lacZ\Delta M15 \Delta(lacZYA-argF)$ U169 <i>recA1 endA1 hsdR<sub>17</sub> (r<sub>k</sub><sup>-</sup>, m<sub>k</sub><sup>+</sup>) phoA supE44 <math>\lambda</math>- thi-1 gyrA96 relA1</i>	Sélection des clones par le système Blanc / Bleu, inactivation des endonucléases intracellulaires et du système de recombinaison, production d'un ADN méthylé.
<b>DH10Bac</b>	Génération de bacmides	F- <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$ $\phi 80lacZ\Delta M15 \Delta lacX74 recA1 endA1 araD139 \Delta (ara, leu)7697 galU galK \lambda$ - <i>rpsL nupG</i> /pMON14272 / pMON7124	Sélection des clones par le système Blanc / Bleu, inactivation des endonucléases intracellulaires et du système de recombinaison, contient un bacmide (pMON14272) et un plasmide helper (pMON7124)
<b>BL21(DE3)</b>	Expression	F- <i>ompT hsdS<sub>B</sub>(rB<sup>-</sup>, mB<sup>-</sup>) gal dcm</i> (DE3)	Bactérie lysogène qui a intégré l'ADN du phage $\lambda$ DE3, permettant de produire la polymérase T7, mutation des protéases intracellulaires
<b>BL21(DE3) pRARE2</b>	Expression	F- <i>ompT hsdS<sub>B</sub>(rB<sup>-</sup>, mB<sup>-</sup>) gal dcm</i> (DE3)	Caractéristiques identiques aux bactéries BL21(DE3), contient le plasmide pRARE2 qui code pour les ARN de transfert rare chez la bactérie.

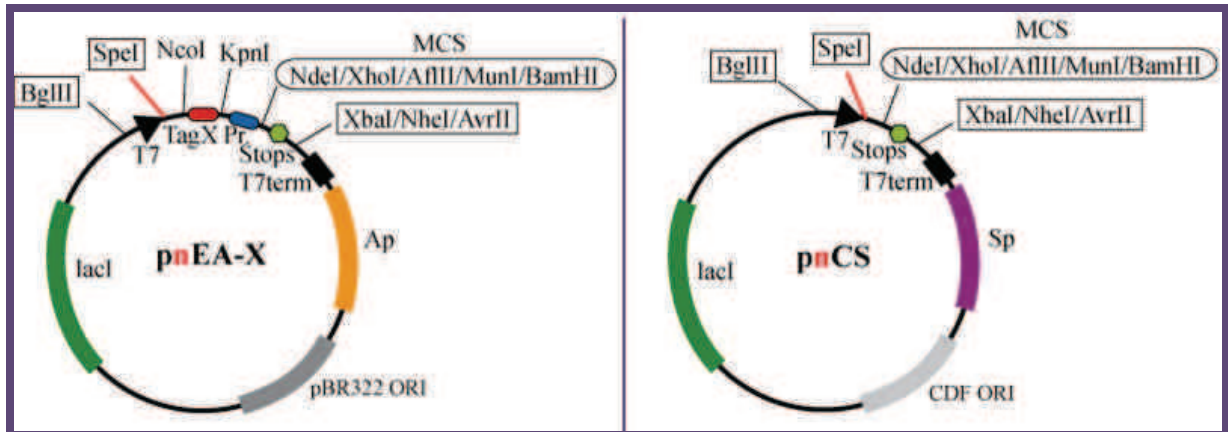
**Tableau 18 : Souches bactériennes utilisées au cours de cette étude.**

### 8.1.5. Les vecteurs de co-expression bactériens

Les vecteurs d'expression bactériens vont permettre l'expression des protéines d'intérêt dans les bactéries. Pour cela, les séquences nécessaires à la production de la protéine dans cet hôte sont placées sur le vecteur d'expression de part et d'autre de la cassette de clonage. Ainsi, en amont du gène, le promoteur de la polymérase T7 est placé afin d'assurer la transcription du gène, ainsi qu'un site de fixation des ribosomes qui permet la traduction de la protéine. En aval du gène, des codons stop vont permettre l'arrêt de la traduction et une séquence terminatrice (T7term) va permettre le décrochage de la polymérase T7.

Les deux vecteurs présentés dans la Figure 71, ont principalement été utilisés au cours de cette thèse (Diebold et al., 2011). Ces vecteurs permettent la co-expression de complexes protéiques, le vecteur pnEA-X permettant la production d'une protéine portant une étiquette de purification et le vecteur pnCS permettant l'expression d'une protéine native. De plus les sites

de restriction qui entourent qui sont présents dans ce vecteur permettent la concaténation des gènes qu'ils contiennent.



**Figure 71 : Carte des deux principaux vecteurs d'expression bactériens utilisés.**

*Les vecteurs dont les cartes sont présentées ici ont principalement été utilisés au cours de cette thèse. Ils permettent l'expression de complexes protéiques mais aussi la concaténation de gènes qu'ils contiennent.*

### 8.1.6. Vecteurs baculovirus

Les vecteurs baculovirus utilisés au cours de cette thèse permettent l'utilisation de la technique du Bac to Bac pour la génération des bacmides dans la bactérie. Ces vecteurs ont été modifiés par la plateforme de Biologie Structurale du Département à partir des vecteurs créés par Imre Berger (Berger et al., 2004). Ces vecteurs possèdent donc toutes les caractéristiques de vecteurs bactériens (origine de répllication et gène de résistance à un antibiotique) ainsi que des sites permettant la transposition du gène placé sous le contrôle du promoteur polyédrique dans le génome du virus. Le gène de la protéine d'intérêt a été inséré dans ces vecteurs par recombinaison homologe de type LR. Deux vecteurs ont principalement été utilisés dans l'optique de réaliser des expériences de co-expression : le vecteur pFC-0GW qui permet l'expression de protéines natives et le vecteur pFC-HGW qui permet l'expression de protéines portant une étiquette de purification poly-(His).



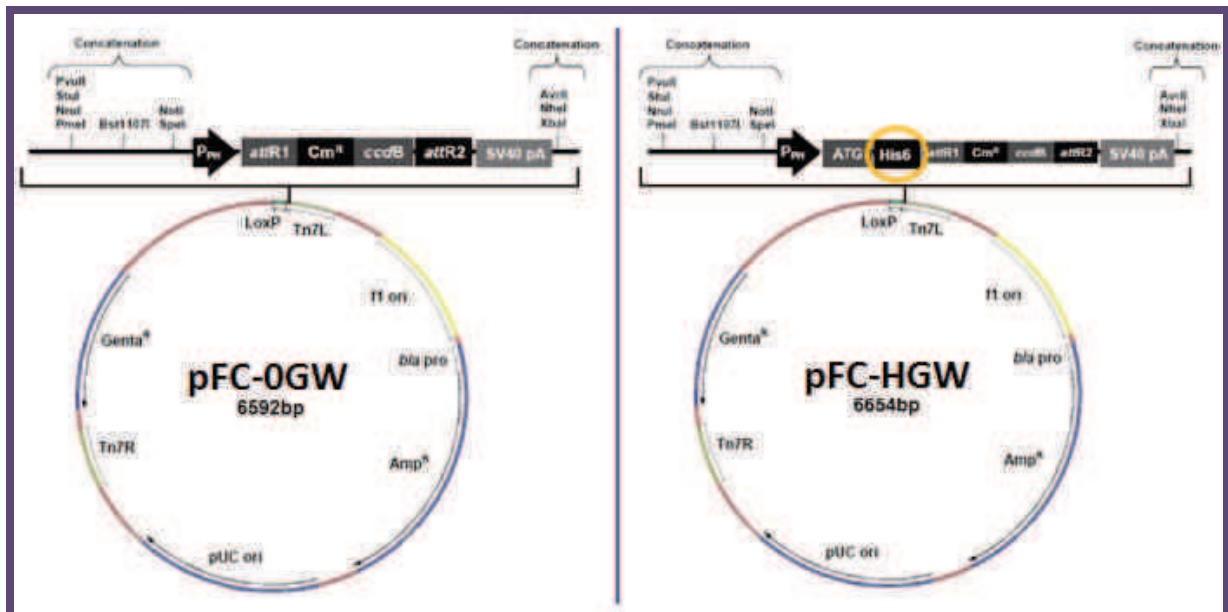


Figure 72 : carte des deux vecteurs baculovirus utilisés

*Les séquences présentes dans les vecteurs baculovirus vont permettre la transposition du gène d'intérêt dans le génome du virus.*

## 8.2. Données supplémentaires de la publication 1

## Supplementary Table 1.

Data collection, phasing and refinement statistics for *A. locustae* Spt6 tandem SH2

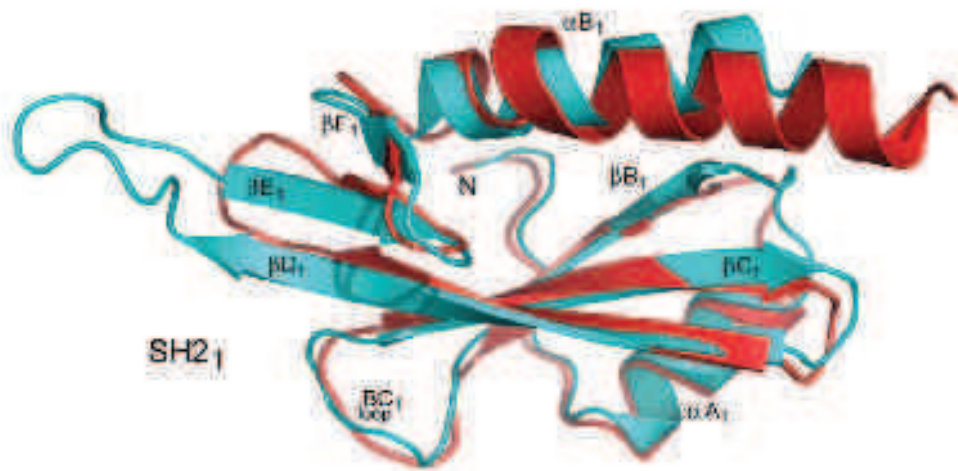
	Native	KAu(CN) <sub>2</sub>	(CH <sub>3</sub> ) <sub>3</sub> PbAc
<b>Data collection</b>			
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	59.10,49.78,59.53	59.84,50.56,59.51	59.29,50.20,59.48
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	32-2.20	32-2.45	32-2.45
	(2.28-2.20) <sup>a</sup>	(2.54-2.45) <sup>b</sup>	(2.54-2.45) <sup>b</sup>
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	4.2 (26.0)	3.5 (15.7)	4.3 (18.7)
<i>I</i> / $\sigma$ <i>I</i>	52.9 (7.3)	33.2 (9.8)	46.7 (11.5)
Completeness (%)	98.9 (99.9)	95.7 (95.0)	99.7 (99.9)
Redundancy	6.6 (6.7)	3.7 (3.4)	5.6 (5.2)
<b>Refinement</b>			
Resolution (Å)	32-2.20		
No. reflections	8803		
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.225 / 0.266		
No. atoms			
Protein	1457		
Ligand/ion	11		
Water	20		
<i>B</i> -factors			
Protein	29.35		
Ligand/ion	48.10		
Water	31.63		
R.m.s deviations			
Bond lengths (Å)	0.016		
Bond angles (°)	1.627		

<sup>a</sup> Values in parentheses are for the highest-resolution shell.

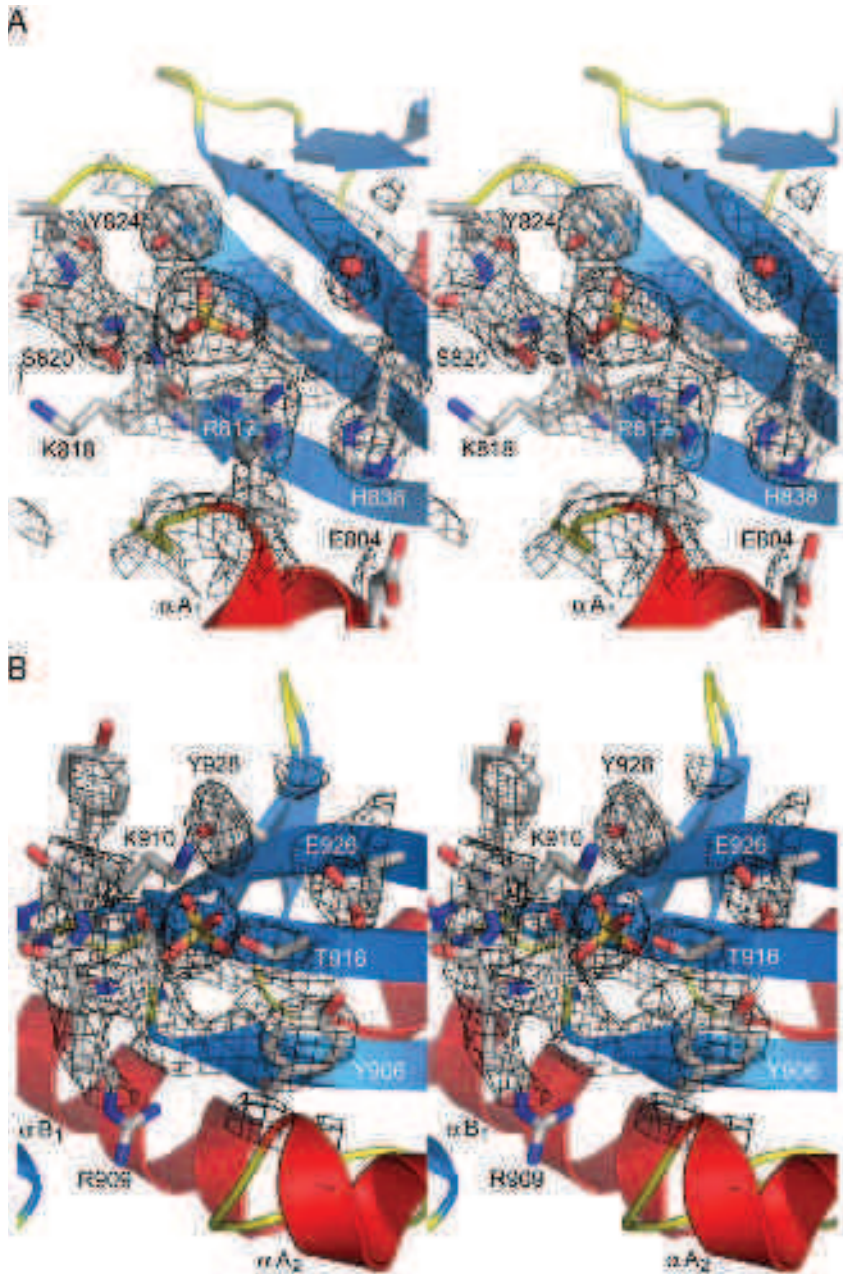
Supplementary Table 2.

*S. cerevisiae* strains.

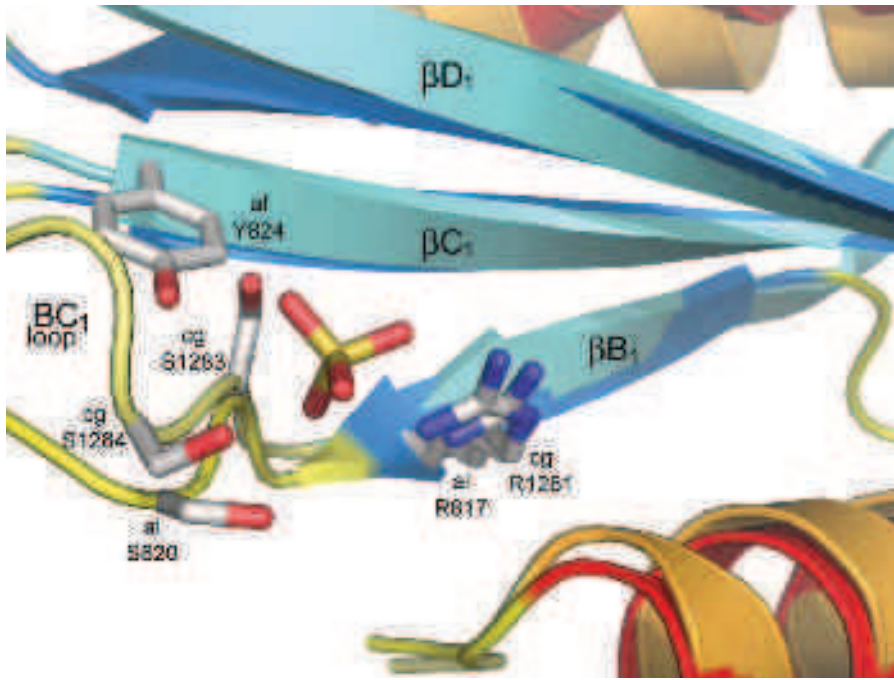
Strain	Genotype
FY2808	<i>MATa/MATa his3Δ200/” leu2Δ1/” lys2-128δ/” trp1Δ163/” ura3-52/” KanMx-GAL1pr-FLO8-HIS3/”</i>
FY2809	<i>MATa his3Δ200 leu2Δ1 lys2-128δ trp1Δ163 ura3-52 KanMx-GAL1pr-FLO8-HIS3</i>
FY2810	<i>MATa his3Δ200 leu2Δ1 lys2-128δ trp1Δ163 ura3-52 KanMx-GAL1pr-FLO8-HIS3 SPT6::TAP-NatMx</i>
FY2811	<i>MATa his3Δ200 leu2Δ1 lys2-128δ trp1Δ163 ura3-52 KanMx-GAL1pr-FLO8-HIS3 spt6ΔTandem::TAP-NatMx</i> (Last 201 codons of <i>SPT6</i> deleted and replaced with TAP tag)
FY2812	<i>MATa his3Δ200 leu2Δ1 lys2-128δ trp1Δ163 ura3-52 KanMx-GAL1pr-FLO8-HIS3 spt6ΔSH2<sub>2</sub>::TAP-NatMx</i> (Last 101 codons of <i>SPT6</i> deleted and replaced with TAP tag)
FY653	<i>MATa his4-912δ lys2-128δ ura3-52 leu2Δ1</i>
FY2796	<i>MATa/MATa his3Δ200/” leu2Δ1/” lys2-128δ/” trp1Δ163/” ura3-52/” KanMx-GAL1pr-FLO8-HIS3/” SPT6::TAP-NatMx/SPT6</i> (Note: Diploid strain with one copy of wild-type <i>SPT6</i> and one copy of TAP-tagged <i>SPT6</i> allele. )
FY2797	<i>MATa/MATa his3Δ200/” leu2Δ1/” lys2-128δ/” trp1Δ163/” ura3-52/” KanMx-GAL1pr-FLO8-HIS3/” spt6 ΔTandem::TAP-NatMx/SPT6</i> (Note: last 201 codons of <i>SPT6</i> deleted and replaced with TAP tag. Diploid strain with one copy of wild-type <i>SPT6</i> and one copy of truncated <i>spt6</i> allele. )
FY2798	<i>MATa/MATa his3Δ200/” leu2Δ1/” lys2-128δ/” trp1Δ163/” ura3-52/” KanMx-GAL1pr-FLO8-HIS3/” spt6ΔSH2<sub>2</sub>::TAP-NatMx/SPT6</i> (Note: last 101 codons of <i>SPT6</i> deleted and replaced with TAP tag. Diploid strain with one copy of wild-type <i>SPT6</i> and one copy of truncated <i>spt6</i> allele.)



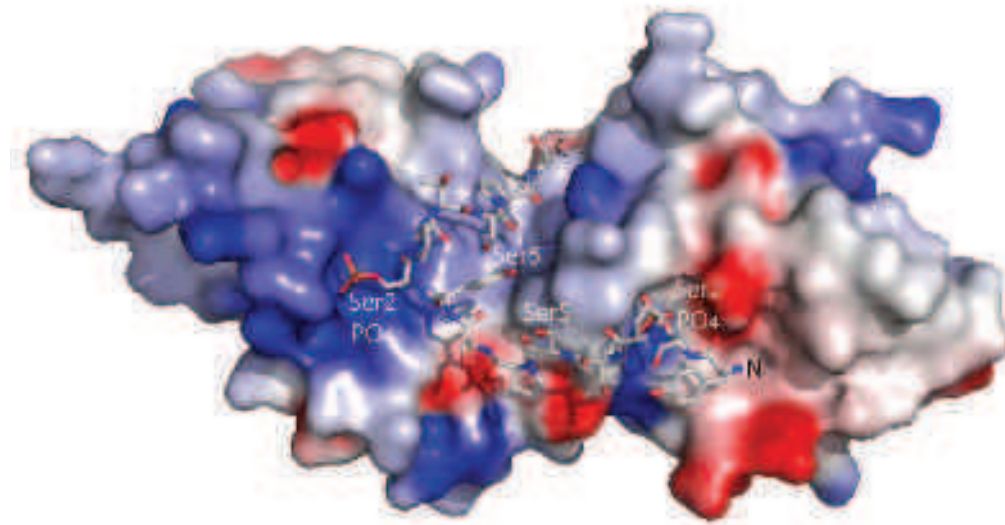
Supplementary Figure 1. Superposition of the SH2<sub>1</sub> domains from *A. locustae* and *C. glabrata*. The SH2<sub>1</sub> domains from *Antonospora locustae* (red) and *Candida glabrata* (cyan) have been superposed and are represented as ribbons. Secondary structure elements are labeled according to the nomenclature of SH2 domains. Major structural differences include: (i) extended  $\beta D_1$  and  $\beta E_1$  strands in the case of the *C. glabrata* SH2<sub>1</sub> and (ii) a kink in the  $\alpha B_1$  helix of the *A. locustae* SH2<sub>1</sub> which is most likely imposed by the interaction between the SH2<sub>1</sub> and SH2<sub>2</sub> domains and which induces a different path for the two  $\alpha B_1$  helices.



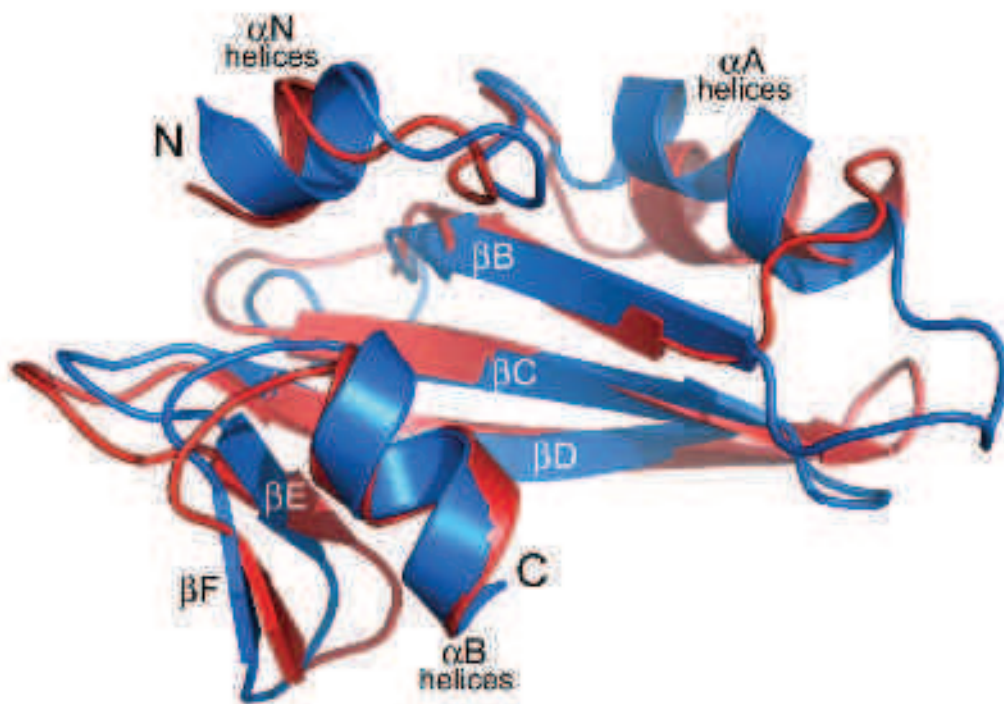
Supplementary Figure 2. Stereo views of the phosphate binding pockets of (A) the SH2<sub>1</sub> and (B) the SH2<sub>2</sub> domains. The domains are represented as ribbons with the exception of the central part of the BC loops which are represented as sticks. The bound sulfate ions and the residues coordinating these ions are shown as sticks and labeled. Water molecules are shown as spheres. The 2Fo-Fc electron density surrounding the sulfate ions is shown, contoured at 1  $\sigma$ .



Supplementary Figure 3. Superposition of the phosphate binding pockets of the SH2<sub>1</sub> domains from *A. locustae* and *C. glabrata*. The two phosphate binding pockets of the SH2<sub>1</sub> domains from *Antonospora locustae* (al) and *Candida glabrata* (cg) have been superposed and the residues putatively involved in phosphate binding are shown as sticks and labeled. The bound sulfate ion observed in the *A. locustae* structure of the Spt6 tandem SH2 is shown. Strict correspondence is observed for arginines alR817 and cgR1281 and serines alS820 and cgS1284. This is not the case for the *A. locustae* tyrosine alY824 which coordinates a sulfate ion in the tandem SH2 structure and which is not conserved throughout evolution. However, the serine cgS1283 in the *C. glabrata* BC<sub>1</sub> loop, which is conserved throughout evolution except in *A. locustae*, is perfectly positioned to replace alY824 for coordinating a phosphate ion. This positioning is notably due to a different conformation of the BC<sub>1</sub> loop which is one residue longer in *C. glabrata* - and in most Spt6 proteins - than in *A. locustae*.



Supplementary Figure 4. Tentative modeling of the interaction of the tandem SH2 with two consecutive Ser2-phosphorylated CTD repeats. In this model, the phosphorylated Ser2 of the first repeat binds to the phosphoresidue binding pocket of the SH2<sub>2</sub> domain. The repeat goes in an extended conformation towards the phosphoresidue binding pocket of the SH2<sub>1</sub> domain where the phosphorylated Ser2 of the second repeat binds. The rest of the second repeat is recognized by the groove between both SH2 domains where specific binding might occur.



Supplementary Figure 5. Superposition of the SH2<sub>1</sub> and SH2<sub>2</sub> domains of *A. locustae* Spt6 tandem SH2. The SH2<sub>1</sub> (red) and SH2<sub>2</sub> (blue) domains of the tandem SH2 from *Antonosporea locustae* Spt6 have been superposed and are shown as ribbons. For the superposition, the long  $\alpha$ B<sub>1</sub> helix connecting both domains has been divided in three parts: the N-terminal part forms the normal  $\alpha$ B<sub>1</sub> helix of the SH2<sub>1</sub>, the central part has been removed, and the C-terminal part has been considered as a small  $\alpha$ N helix of the SH2<sub>2</sub>. Upon superposition, the  $\alpha$ B helices but also the  $\alpha$ N helices superpose nicely, suggesting that the tandem SH2 has been created through duplication of an ancestral SH2 domain followed by the evolution of the SH2<sub>2</sub> domain into a non canonical SH2 motif.



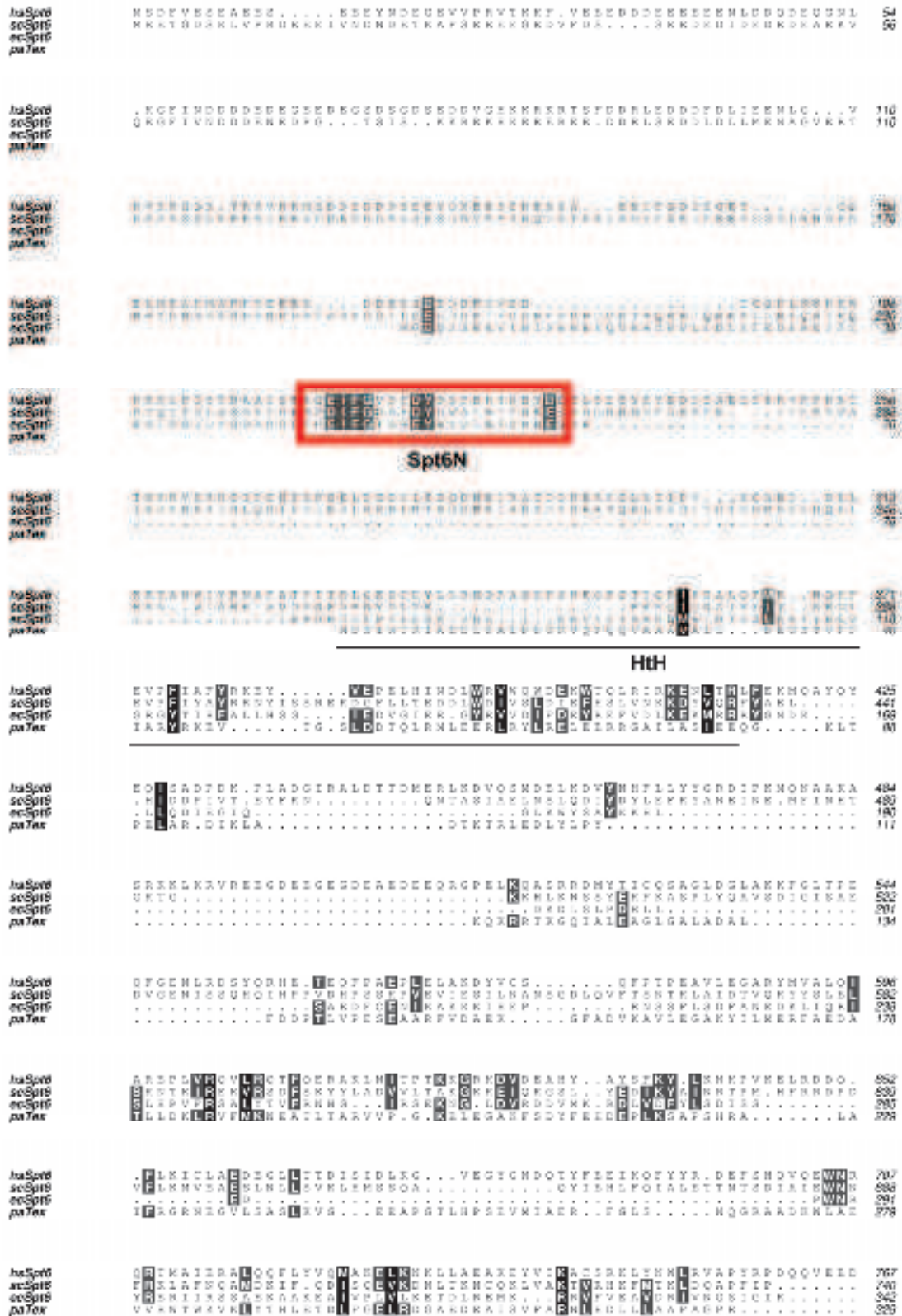
### 8.3. Préparation des extraits de levure pour la réalisation d'expériences de pulldown

Les extraits de levure préparés à partir de ce protocole ont été utilisés lors de la réalisation de pulldown pour identifier les partenaires de nos protéines d'intérêt.

- Des cultures de 100mL de milieu YPD de levures sauvages sont inoculées et mises sous agitation sur la nuit à 30°C.
- Le lendemain, lorsqu'une densité optique à 600nm de 2 est atteinte, la culture est centrifugée (5000 rpm, 5 minutes à 4°C) puis les cellules sont lavées à deux reprises avec de l'eau ultra-pure froide.
- Les cellules sont lavées dans 3mL de tampon A (10mM K-HEPES pH 7,9, 10mM KCl, 1,5mM MgCl<sub>2</sub>, 0,5mM DTT, 0,5mM PMSF, 2 mM Benzamidine, 1μM Leupeptin, 2μM Pepstatin A, 4μM Chymostatin et 2,6μM Aprotonin 500000KIE/ml) puis resuspendues dans 1mL de ce même tampon.
- 750μL de billes de verres stérilisées sont ajoutées et le mélange est vortexé à haute vitesse pendant 30 secondes quatre fois en les laissant reposer sur glace entre chaque étape. L'échantillon est alors re-centrifugé pour décanter les billes de verre (5000 rpm pendant 5 minutes) et le surnageant est transféré dans un eppendorf.
- 900μL de KCl 2M sont ajoutés à l'extrait qui est alors centrifugé pendant 2 minutes à pleine vitesse (4°C),
- Le surnageant est transféré dans un tube en polycarbonate et est centrifugé 20 minutes, 56000rpm à 4°C
- La phase liquide (contenue entre le culot et une couche supérieure lipidique) est mise à dialyser pendant 1 heure contre un tampon A contenant 20% de glycérol.
- Les échantillons sont aliquotés, congelés dans l'azote liquide et placés à -80°C.

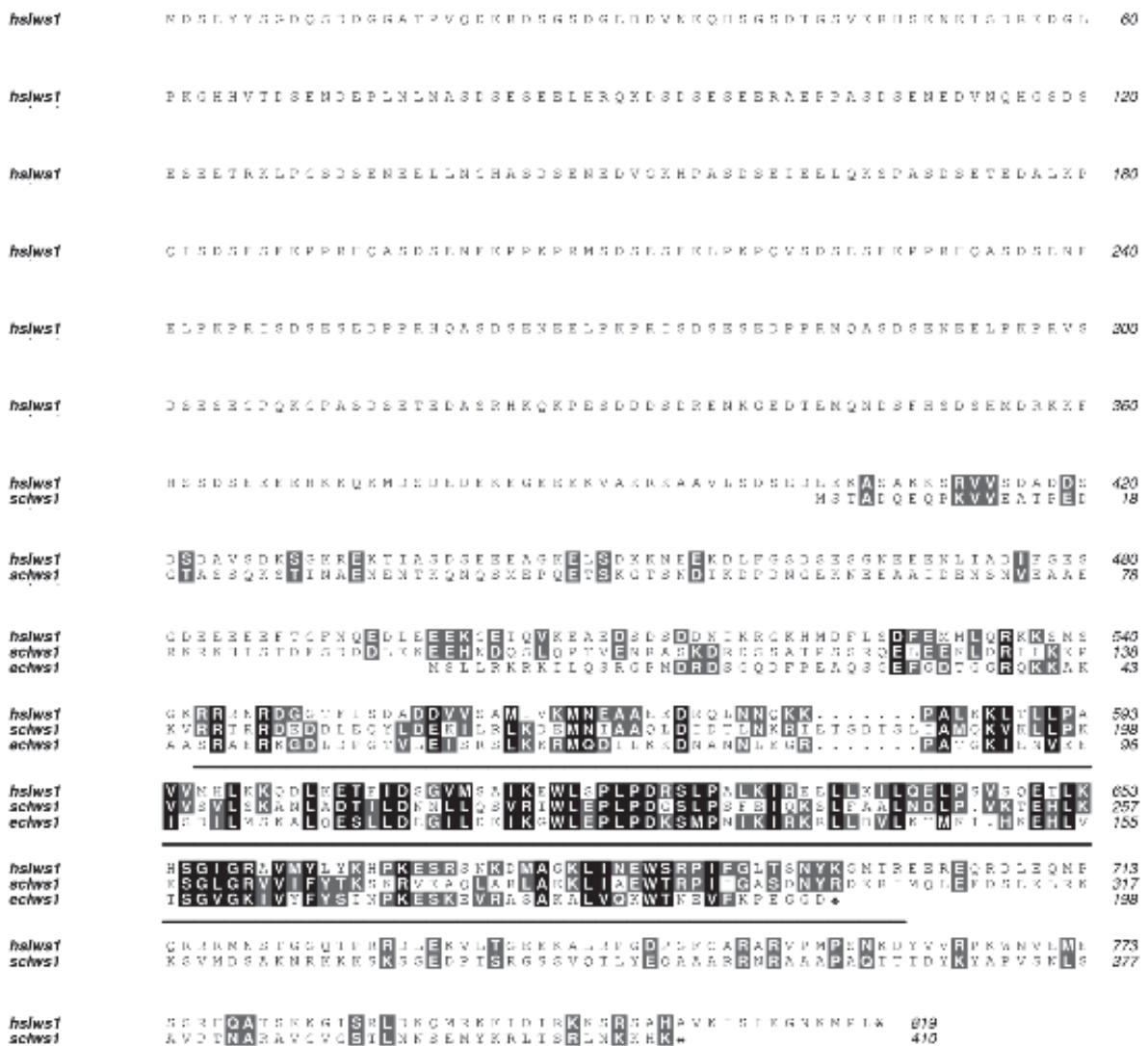
### 8.4. Données supplémentaires de la publication 2

A



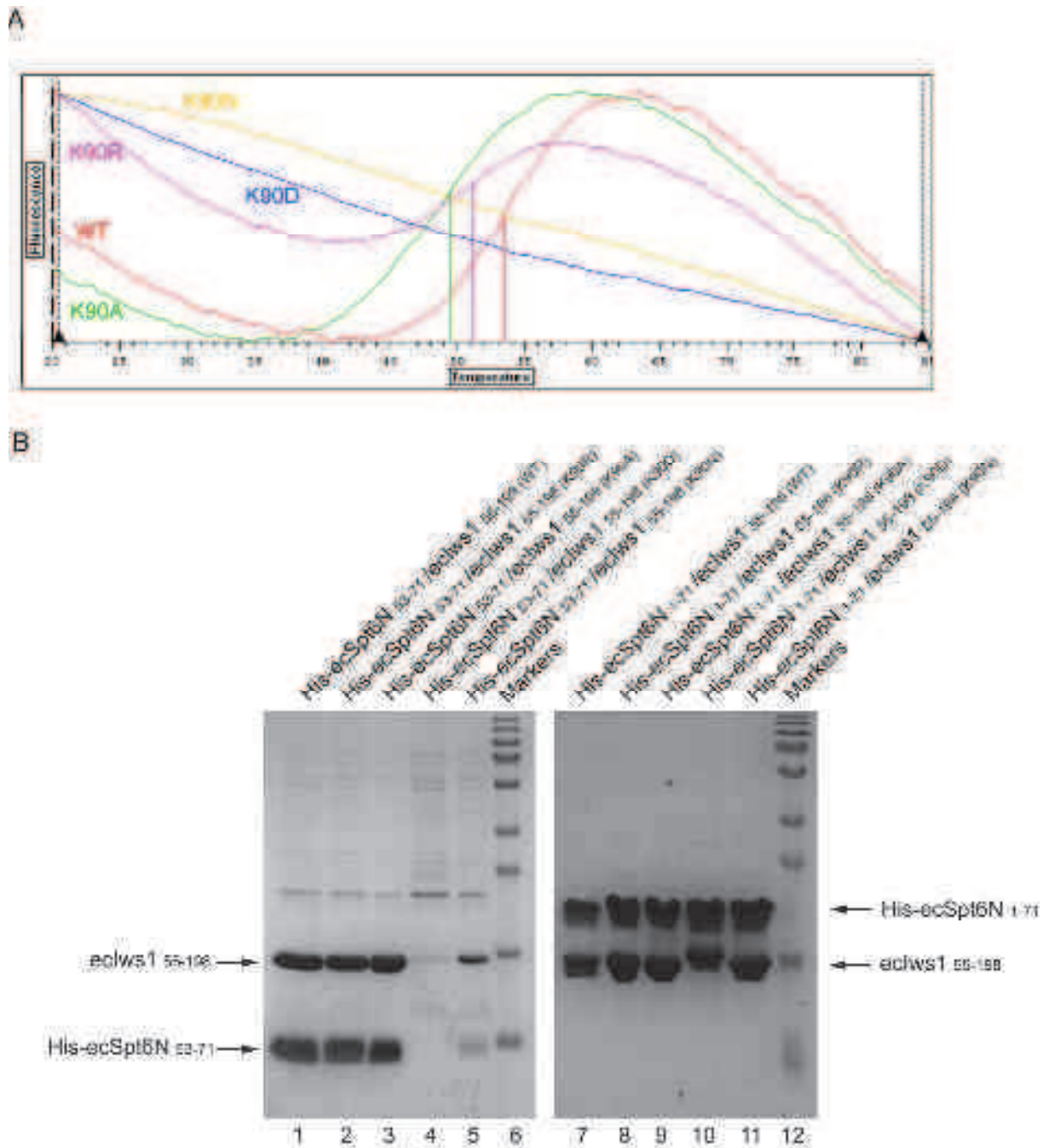


B



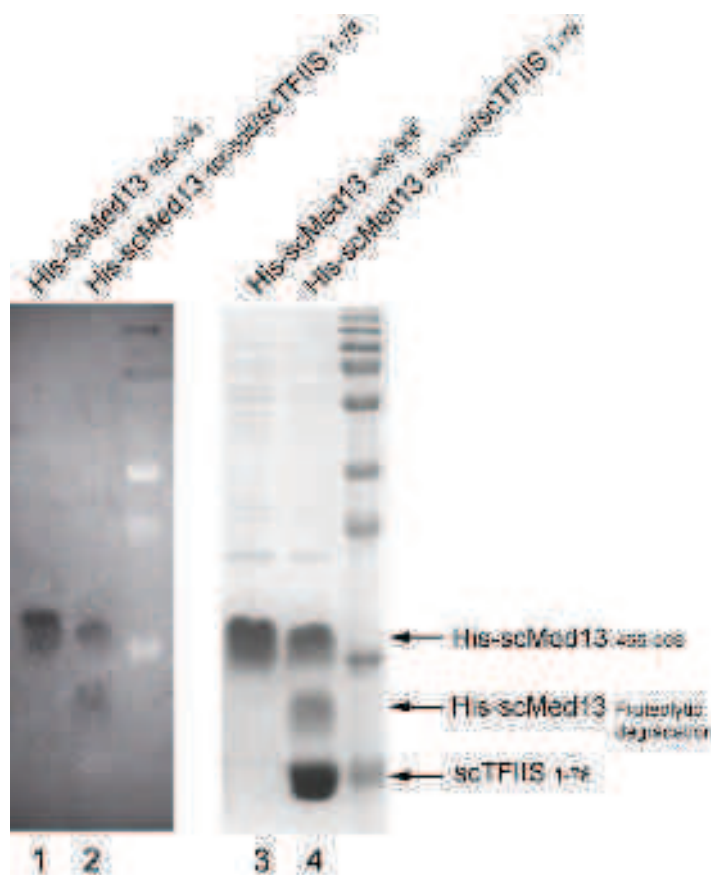
Supplementary Figure 1. Multiple alignments of full-length Spt6, Tex and Iws1 proteins.

*ec*, *Encephalitozoon cuniculi*; *sc*, *Saccharomyces cerevisiae*; *hs*, *Homo sapiens*; *pa*, *Pseudomonas aeruginosa*. Sequence similarities are indicated by shading. End of sequences are marked with a ‘\*’. (A) Multiple sequence alignment of *ec*, *sc*, *hs* Spt6 and Spt6 bacterial paralog *pa*Tex. The specific conserved region at the N-terminus of Spt6 (Spt6N) is indicated with a red box. The structural domains of Tex as well as the Spt6-specific tandem SH2 domains are indicated underneath the sequences. (B) Multiple sequence alignment of *ec*, *sc*, *hs* Iws1. The region sufficient for the essential function of Iws1 in yeast is underlined.



Supplementary Figure 2. Effect of K90 mutants on *E. cucuruli* Iws1 stability and Spt6N binding.

(A) Biophysical characterization of *E. cucuruli* Iws1 wild-type (WT) and K90 mutants using the Thermofluor technology. For each protein, the fluorescence observed upon binding of the SYPRO orange dye to the protein is plotted versus the temperature. (B) Effect of K90 mutations on formation of the complexes between *ecIws1*<sub>55-198</sub> and *ecSpt6N*<sub>53-71</sub> or *ecSpt6N*<sub>1-71</sub>. The K90D mutant migrates more slowly (lane 10).



Supplementary Figure 3. A C-terminal degraded fragment of *scMed13* is solubilised upon binding to TFIIS.

Upon co-expression of His-*scMed13*<sub>466-508</sub> with *scTFIIS*<sub>1-78</sub>, a small protein is observed, bound to the affinity resin, which does not migrate at the expected molecular weight of His-*scMed13*<sub>466-508</sub> and *scTFIIS*<sub>1-78</sub> (right panel; Coomassie-stained SDS gel). A western blot analysis (left panel) using an anti-polyhistidine antibody shows that this protein is his-tagged and corresponds most likely to a C-terminal degradation of His-*scMed13*<sub>466-508</sub> which is solubilised through its interaction with *scTFIIS*<sub>1-78</sub>.

Supplementary Table I. Data collection and refinement statistics for *E. cuniculi* Iws1<sub>55-198</sub>

<b>Data collection statistics</b>				
<i>Data sets</i>	Native	Se-Met $\lambda 1$ peak	Se-Met $\lambda 2$ inflexion	Se-Met $\lambda 3$ remote
<i>Wavelength (Å)</i>	0.933	0.9792	0.9795	0.9753
<i>Space group</i>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
<b>Cell constants</b>				
<i>a (Å)</i>	33.7	41.3	41.4	41.3
<i>b (Å)</i>	128.8	218.3	218.5	218.6
<i>c (Å)</i>	33.6	71.7	72.0	72.0
<i>β (°)</i>	101.2	102.7	102.7	102.8
<i>Resolution (Å)</i>	64-2.25	30-2.7	30-3.0	30-3.0
<i>Reflections (measured/unique)</i>	60417 13209	245211 68237	167714 49682	136789 47418
<i>Redundancy (overall/last shell)</i>	4.6/4.4	3.6/3.6	3.4/3.4	2.9/2.9
<i>Completeness (%) (overall/last shell)</i>	99.2/99.3	99.7/99.9	99.7/100	98.4/100
<i>R<sub>sym</sub><sup>§</sup> (%) (overall/last shell)</i>	5.9/12.3	7.5/18.2	7.3/19.1	8.1/21.0
<i>I/σ(I) (overall/last shell)</i>	25.9/15.1	17.9/7.6	16.9/6.2	15.8/6.4
<b>Refinement statistics</b>				
<i>Resolution (Å)</i>	64-2.25			
<i>Number of protein atoms</i>	2268			
<i>Number of water molecules/ions</i>	35			
<i>Number of reflections (F&gt;0) (all/test set)</i>	12534 650			
<i>R-factor<sup>#</sup> (%)</i>	20.4			
<i>R-free<sup>#</sup> (%)</i>	26.5			
<b>Deviations from ideal geometry</b>				
<i>Bonds (Å)</i>	0.019			
<i>Angles (°)</i>	1.780			
<i>Mean temperature factors (Å<sup>2</sup>)</i>	15.79			

<sup>§</sup>  $R_{sym} = \sum |I - \langle I \rangle| / \sum I$ , where  $I$  is the observed intensity and  $\langle I \rangle$  is the average intensity for multiple measurements.

<sup>#</sup>  $R = \sum_{hkl} |F_{obs} - F_{calc}| / \sum_{hkl} F_{obs}$ , where  $F_{obs}$  and  $F_{calc}$  are the observed and calculated structure factor amplitudes for reflection  $hkl$ , applied to the work (R-factor) and test (R-free) sets.

Supplementary Table II. Data collection and refinement statistics for *E. cuniculi* Iws1<sub>55-198</sub>/Spt6N complexes

<b>Data collection statistics</b>			
<i>Data sets</i>	ecSpt6N <sub>53-71</sub> ecIws1 <sub>55-198</sub> Form1	ecSpt6N <sub>53-71</sub> ecIws1 <sub>55-198</sub> Form2	ecSpt6N <sub>34-71</sub> ecIws1 <sub>55-198</sub>
<i>Wavelength (Å)</i>	0.9762	0.9334	0.977
<i>Space group</i>	P4 <sub>1</sub> 2 <sub>1</sub> 2	P6 <sub>5</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
<b>Cell constants</b>			
<i>a (Å)</i>	52.3	112.6	27.4
<i>b (Å)</i>	52.3	112.6	56.1
<i>c (Å)</i>	151.1	51.4	116.4
<i>β (°)</i>	90	120	90
<i>Resolution (Å)</i>	50-1.95	100-2.1	58-1.75
<i>Reflections (measured/unique)</i>	199644 16077	234782 21935	85689 19008
<i>Redundancy (overall/last shell)</i>	12.4/11.4	10.7/10.2	4.5/4.6
<i>Completeness (%) (overall/last shell)</i>	99.4/99.6	99.9/100	99.4/99.9
<i>R<sub>sym</sub><sup>S</sup> (%) (overall/last shell)</i>	4.1/17.1	8.5/28.8	4.2/16.6
<i>I/σ(I) (overall/last shell)</i>	78.8/17.0	33.6/10.5	44.6/9.2
<b>Refinement statistics</b>			
<i>Resolution (Å)</i>	50-1.95	100-2.1	58-1.75
<i>Number of protein atoms</i>	1269	2474	1300
<i>Number of water molecules/ions</i>	133	153	170
<i>Number of reflections (F&gt;0) (all/test set)</i>	15204 803	20780 1124	17980 973
<i>R-factor<sup>#</sup> (%)</i>	21.0	20.0	19.4
<i>R-free<sup>#</sup> (%)</i>	23.6	24.7	23.6
<b>Deviations from ideal geometry</b>			
<i>Bonds (Å)</i>	0.017	0.017	0.018
<i>Angles (°)</i>	1.49	1.57	1.61
<i>Mean temperature factors (Å<sup>2</sup>)</i>	16.15	11.63	11.89

<sup>S</sup>  $R_{sym} = \frac{\sum |I - \langle I \rangle|}{\sum I}$ , where *I* is the observed intensity and  $\langle I \rangle$  is the average intensity for multiple measurements.

<sup>#</sup>  $R = \frac{\sum_{hkl} |F_{obs} - F_{calc}|}{\sum_{hkl} F_{obs}}$ , where *F*<sub>obs</sub> and *F*<sub>calc</sub> are the observed and calculated structure factor amplitudes for reflection *hkl*, applied to the work (R-factor) and test (R-free) sets.



Supplementary Table III. Yeast strains.

<b>Strain</b>	<b>Genotype</b>
FY2801	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, FLAG-SPT6)</i>
FY2802	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, spt6-50)</i>
FY2803	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, FLAG-spt6-Y255A,W257A)</i>
FY2804	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, FLAG0spt6-Y255A)</i>
FY2805	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1,FLAG- spt6-W257A)</i>
FY2806	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, FLAG-spt6-I248A,F249A)</i>
FY2807	<i>MAT<math>\alpha</math> his4-912<math>\delta</math> lys2-128<math>\delta</math> leu2<math>\Delta</math>1 ura3-52 trp1<math>\Delta</math>63 spt6<math>\Delta</math>::LEU2 (CEN, TRP1, FLAG-spt6-G250A,G252A)</i>

## 8.5. Purification des complexes de TFIIS et méthylation des lysines

Le protocole de méthylation des lysines a été employé dans le cadre de la cristallisation des complexes Spt8/TFIIS, Spt6 IF3/TFIIS et Spt6 IF5/TFIIS. Cette réaction de méthylation est réalisée par une substitution chimique des hydrogènes liés à l'amine terminale de la lysine par des groupements méthyles. Ce protocole a été adapté de la publication (Walter et al., 2006) afin de pouvoir réaliser la réaction entre les étapes de chromatographie d'affinité et de coupure par la protéase. Ainsi, les étapes de lyse et de chromatographie d'affinité sont identiques au protocole standard, le tampon de lyse utilisé contenant 50mM NaCl, 50mM HEPES pH 7,5. Après lavage de la résine, la quantité de protéine purifiée est estimée et la réaction de réaction de méthylation est réalisée.

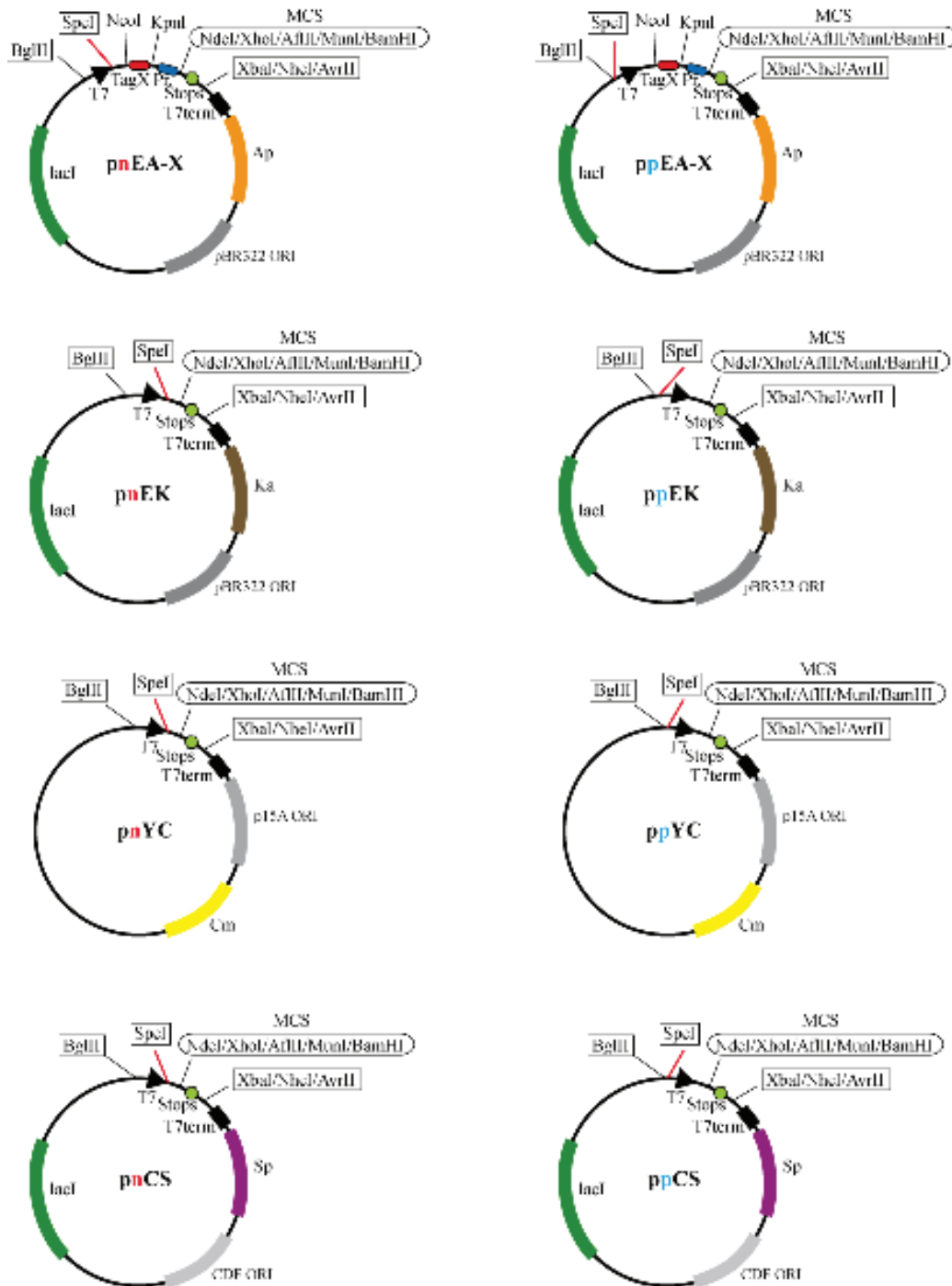
- 20 $\mu$ L de diméthylamine borane (1M) et 40 $\mu$ L de formaldéhyde (1M) sont ajoutés par milligramme de protéine estimée. La réaction est placée à 4°C pendant deux heures.
- 20 $\mu$ L de diméthylamine borane (1M) et 40 $\mu$ L de formaldéhyde (1M) sont à nouveau ajoutés dans les mêmes proportions. La réaction est placée à 4°C pendant deux heures.
- 10 $\mu$ L de diméthylamine borane (10 $\mu$ L) sont alors ajoutés et la réaction est laissée pendant la nuit à 4°C.
- Le lendemain, la résine est lavée avec une solution tampon contenant 50mM NaCl, 20mM Tris pH7,5 afin de stopper la réaction.
- La thrombine est alors ajoutée et, le lendemain, l'échantillon est injecté sur une colonne de chromatographie d'exclusion.

## 8.6. Données supplémentaires de la publication numéro 3

**Deciphering correct strategies for multiprotein complex assembly  
by co-expression: Application to complexes as large as the histone  
octamer**

Marie-Laure Diebold, Sébastien Fribourg, Michael Koch, Thibaud Metzger and Christophe  
Romier

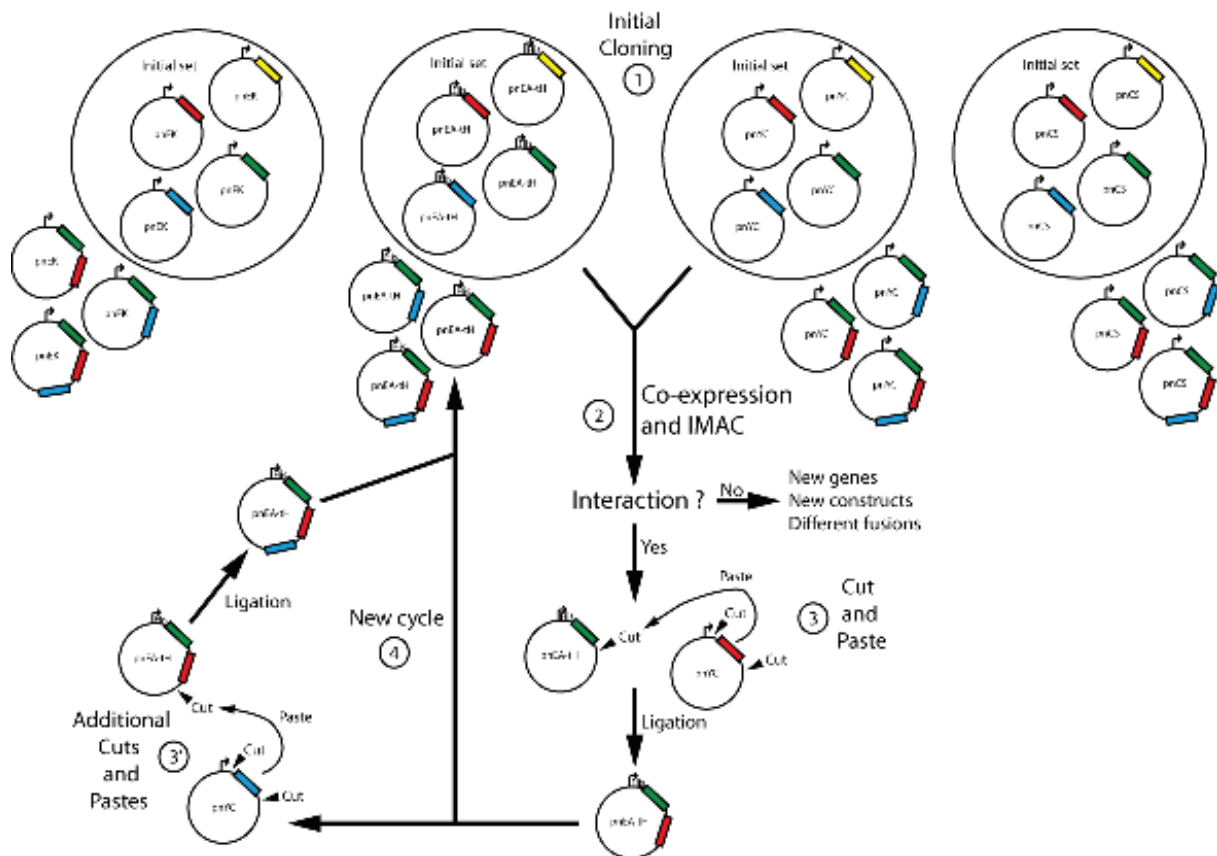
**Supplementary Data**



**Supplementary Figure 1. Detailed features of the pET-MCN and pET-MCP vector series.**

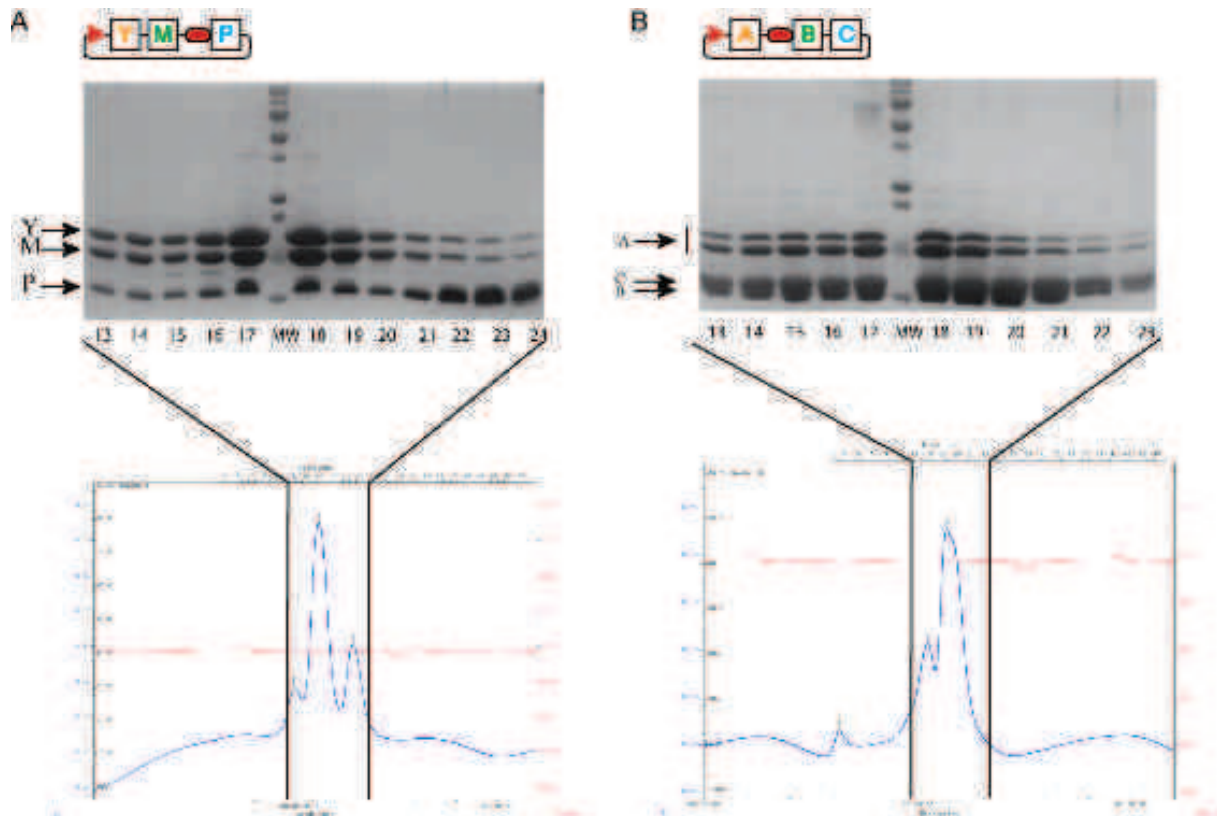
Vectors of the pET-MCN (left column) and pET-MCP (right column) series are depicted. Nomenclature for the vectors are the following: p (first position), plasmid; n or p (second position), MCN or MCP, respectively; E (third position), based on either pET15b or pET28b; Y (third position), based on the pACYC-11b vector; C (third position), based on the pCDF-DuET vector (Novagen); A, K, C, S (fourth position), Ampicillin, Kanamycin, Chloramphenicol, Spectinomycin, respectively; X denotes the various tags available.

Replication origins, lacI gene (coding for the lac repressor) and antibiotic resistance genes are shown as coloured segments. Position of the T7 promoter (black arrow), sequence encoding purification tag (red ellipse; N-terminal position in this case), sequence encoding the protease cleavage site (blue ellipse), stop codons (green dot) and T7 terminator (black segment) are displayed. The multi-cloning restriction sites (MCS; *NdeI*, *XhoI*, *AflIII*, *MunI*, *BamHI*) are surrounded by an ellipse. Note that some vectors have a Gateway compatible version instead of restriction sites. The *BglIII*, *SpeI*, *XbaI*, *NheI* and *AvrII* concatenation sites are boxed. Note the different position of the *SpeI* site between the pET-MCN and pET-MCP vectors that allow the creation of either polycistronic or multiple cassettes vectors. The *NcoI*, *KpnI* and *NdeI* restriction sites are used for facilitating the exchange of the sequences encoding the tag and the protease cleavage site. Further detailed information on these vectors and their use are provided in the Supplementary Website ([http://igbmc.fr/Romier\\_Multi-Expression](http://igbmc.fr/Romier_Multi-Expression)).



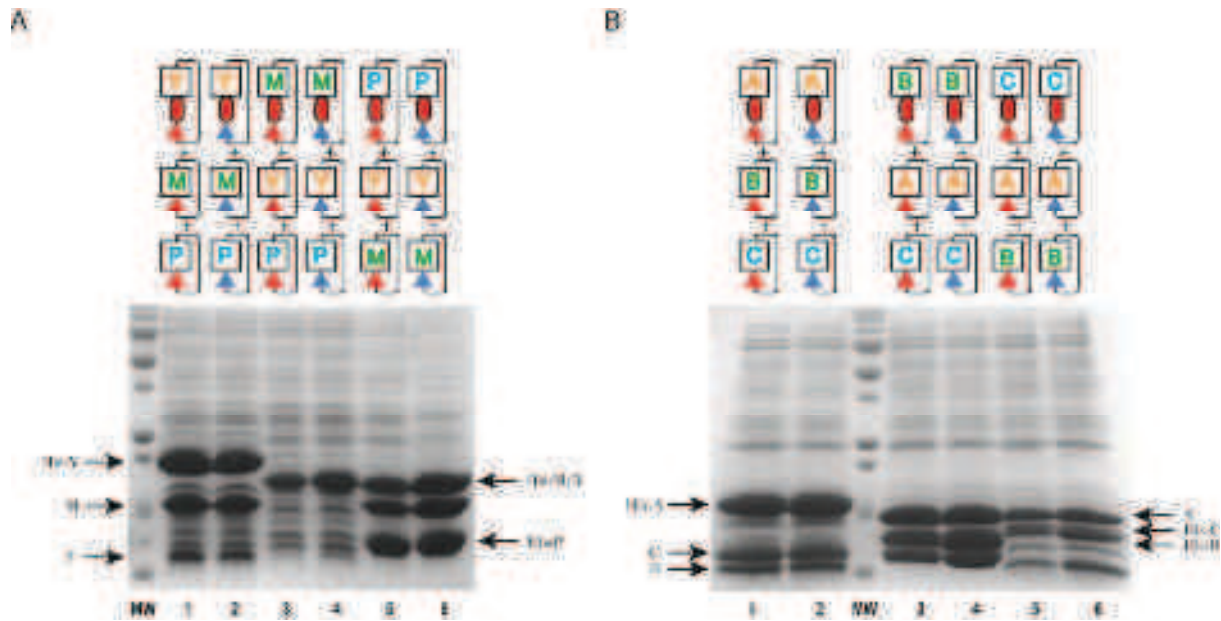
**Supplementary Figure 2. Extended complex reconstitution strategy using the pET-MCN and pET-MCP vector series.**

The strategy depicted extends the previously published strategy (see text) by making use of the newly developed pnCS and pnEK vectors. Although the strategy depicts the use of vectors from the pET-MCN series, it is fully compatible with the vectors of the pET-MCP series and can, in fact, be used with a combination of pET-MCN and pET-MCP vectors (Supplementary Website: [http://igbmc.fr/Romier\\_Multi-Expression](http://igbmc.fr/Romier_Multi-Expression)). A main advantage of this strategy is to accelerate complex reconstitution by testing co-expression of at least four different proteins in a single experiment (through the use of the four different vectors), thus enabling several Cuts and Pastes steps in a single reconstitution cycle.



**Supplementary Figure 3. Large scale purification of the YMP and NFY complexes.**

(A) Analytical gel filtration profile and fraction analysis of large-scale purified Y14/Mago/Pym (YMP) complex. The vector used for multi-expression is depicted as in Figure 2. The major peak corresponds to the ternary complex and the small peak to the excess of Pym that is bearing the affinity purification tag. Obviously, tagging the smallest subunit clearly simplified purification. (B) Analytical gel filtration profile and fraction analysis of large-scale purified NFYA/NFYB/NFYC (NFY) complex. The major peak consists in two species that co-purify: the ternary NFY complex and the NFYB/NFYC sub-complex that cannot be separated by other chromatographic techniques. Placement of the polyhistidine tag at the NFYA C-terminus finally resolved the issue (data not shown). The NFYA protein appears as two bands since proteolytic degradation occurs during large scale purification. Use of constructs shorter at the C-terminus prevents degradation (data not shown).



**Supplementary Figure 4. Reconstitution of the YMP and NFY complexes with three independent vectors.**

SDS-PAGE analysis of soluble (A) YMP and (B) NFY complexes retained on affinity resins. The different combinations of vectors used are depicted above each lane as in Figs. 2 and 3. The results obtained are sometimes different from what was obtained with the pET-MCP series for ternary complex assembly using two or a single vector, showing that the observed differences of stoichiometry between the pET-MCN and pET-MCP series are not only due to the production of a single mRNA versus multiple mRNAs.



## 8.7. Reconstitution du nucléosome à partir de l'octamère d'histone co-exprimé

La reconstitution du nucléosome va nécessiter la purification de l'octamère d'histone d'une part et la préparation de l'ADN d'autre part. L'octamère est purifié à partir de six litres de cultures bactériennes, le protocole comprenant une étape de purification par affinité sur une résine cobalt (Talon, Clontech) suivi d'une étape de gel filtration qui va permettre de séparer les différentes espèces purifiées et notamment l'octamère d'histone du la paire H2A/H2B (publication 3).

### 8.7.1. Préparation de l'ADN Widom 601

- 160µg de vecteur contenant les séquences Widom 601 sont digérés sur la nuit dans un volume réactionnel de 300µL avec 40 unités d'enzyme EcoRV en présence de BSA.
- Le lendemain, 10% de PEG 8000 et 0,5M de NaCl (déterminé par rapport au volume réactionnel final) sont ajoutés. Le mélange est laissé sur glace pendant 30 minutes avant d'être centrifugé à 16000rpm pendant 20 minutes. Cette étape permet de précipité les fragments d'ADN de grande taille, dans notre cas le vecteur ouvert (la concentration de PEG 8000 utilisée a été déterminée expérimentalement).
- Les fragments d'ADN de 147pb présents dans le surnageant sont précipités à l'isopropanol (par ajout de 0,8 fois le volume dans lequel se trouve l'ADN d'isopropanol 100% à -20°C), le culot étant alors lavé à l'éthanol.
- Le culot d'ADN est alors resuspendue dans un tampon contenant 2M NaCl, 10mM Tris pH8 identique au tampon des histones.

### 8.7.2. Reconstitution des nucléosomes

- L'ADN purifié et les octamères d'histones produits sont mélangés avec des ratios molaires 2 : 1 et 1 : 1 (protéines : ADN) et laissés sur glace pendant 30 minutes.

- Le mélange est alors placé dans un tube de dialyse (Slide-A-Lyzer, Pierce) la membrane ayant une limite d'exclusion de 7kDa. Le tube est placé dans 200mL d'un tampon contenant 2M NaCl, 10mM Tris pH8.
- Une pompe péristaltique est utilisée afin d'ajouter lentement un tampon contenant 10mM Tris pH8 jusqu'à atteindre une concentration finale de 250mM NaCl. (Ainsi, 1400mL de tampon sont ajoutés, le débit calculé étant de 40mL par heure).
- Lorsque le volume de tampon a été ajouté (au bout de 36 heures environ) les échantillons sont centrifugés, afin d'éliminer la fraction insoluble, et déposés sur gel.

## 8.8. Publication numéro 4

Michael Koch, Marie-Laure  
Diebold, Jean Cavarelli and  
Christophe Romier\*

IGBMC (Institut de Génétique et Biologie  
Moléculaire et Cellulaire), Département de  
Biologie et Génétique Structurales, UDS,  
CNRS, INSERM, 1 Rue Laurent Fries, BP 10142,  
67404 Illkirch CEDEX, France

Correspondence e-mail: romier@igbmc.fr

Received 26 August 2009

Accepted 8 December 2009

## Crystallization and preliminary crystallographic analysis of eukaryotic transcription and mRNA export factor Iws1 from *Encephalitozoon cuniculi*

Transcription elongation by eukaryotic RNA polymerase II requires the coupling of mRNA synthesis and mRNA processing and export. The essential protein Iws1 is at the interface of these processes through its interaction with histone chaperone and elongation factor Spt6 as well as with complexes involved in mRNA processing and export. Upon crystallization of the evolutionarily conserved domain of Iws1 from *Encephalitozoon cuniculi*, four different crystal forms were obtained. Three of the crystal forms belonged to space group  $P2_1$  and one belonged to space group  $P222_1$ . Preliminary X-ray crystallographic analysis of one of the crystal forms allowed the collection of data to 2.5 Å resolution.

### 1. Introduction

During eukaryotic transcription by RNA polymerase II (RNAPII), mRNA synthesis is tightly linked to chromatin modulation and to mRNA processing and export (Sims *et al.*, 2004). The essential protein Iws1 (interacts with Spt6; yeast Spn1) is at the interface of these different processes. Firstly, Iws1 forms a complex with Spt6 (Krogan *et al.*, 2002; Lindstrom *et al.*, 2003; Yoh *et al.*, 2007), a protein which is both a putative histone chaperone (Bortvin & Winston, 1996; Kaplan *et al.*, 2003; Adkins & Tyler, 2006) and an elongation factor of RNAPII (Endoh *et al.*, 2004; Yoh *et al.*, 2007; Ardehali *et al.*, 2009). Secondly, Iws1 also interacts with the mRNA processing and export factor REF1/Aly, which then recruits the mRNA surveillance-factor exosome (Yoh *et al.*, 2007). The Spt6–Iws1 complex travels with elongating RNAPII through binding of the Spt6 SH2 domain to the hyperphosphorylated C-terminal domain (CTD) of the polymerase. Accordingly, depletion of Iws1 or mutation within Spt6 SH2 leads to the accumulation of bulk poly(A)<sup>+</sup> RNAs in the nucleus (Yoh *et al.*, 2007). Furthermore, mammalian Iws1 has also been shown to recruit the HYPB/Setd2 histone methyltransferase to the elongating polymerase. This recruitment is required for H3K36 trimethylation (Yoh *et al.*, 2008). Thus, formation of the Spt6–Iws1 complex provides an effective means of coupling transcription, epigenetics mechanisms and mRNA processing and export.

In yeast, Iws1 has also been shown to have a role that is independent of Spt6. Indeed, yeast Iws1 affects transcription regulation at the post-recruitment-regulated promoter CYC1, being recruited constitutively in an Spt6-independent manner to this promoter and preventing loading of the chromatin remodelling factor Swi/Snf (Fischbeck *et al.*, 2002; Zhang *et al.*, 2008). Interestingly, this recruitment also requires a hyperphosphorylated RNAPII CTD. However, the interaction between Iws1 and its different partners remains unknown in molecular terms.

In yeast, the evolutionarily conserved region of Iws1 has been shown to be sufficient and necessary for growth (Fischbeck *et al.*, 2002). Surprisingly, the C-terminal domain of this region is homologous in sequence to the N-terminal domains of the transcriptional effectors TFIIS, elongin A and Med26 (Ling *et al.*, 2006), suggesting that these proteins may act in related transcriptional pathways and



© 2010 International Union of Crystallography  
All rights reserved

may possibly share common interaction partners. Interestingly, the sequence identity between the homologous regions of human Iws1 and human TFIIS is 18% and should be compared with the 20% identity for the same regions of yeast and human TFIIS. This suggests that the structures of the TFIIS regions from yeast and mouse (PDB code 1eo0, Booth *et al.*, 2000; PDB code 1wjt, M. Yoneyama, N. Tochio, S. Koshiba, M. Inoue, T. Kigawa & S. Yokoyama, unpublished work) could be used for determination of the Iws1 structure by molecular replacement.

Determining the structure of Iws1 is an important requirement for elucidating the role of Iws1 in the various mechanisms that it contributes to. We are studying the Iws1 protein from the yeast-related eukaryotic parasite *Encephalitozoon cuniculi*. A main characteristic of this organism is the smaller size of its proteins compared with their eukaryotic orthologues, a feature that simplifies structural characterization. Despite their smaller size, these proteins retain most of the features of their orthologues (Romier *et al.*, 2007). For instance, the evolutionarily conserved domain of Iws1 from *E. cuniculi* is 28% identical to its yeast and human orthologues, whereas 38% identity is observed between the yeast and human domains. Once structural knowledge has been gained on the *E. cuniculi* proteins, it can then be used for functional studies of the yeast and human proteins (Romier *et al.*, 2007). Here, we present the cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of the evolutionarily conserved *E. cuniculi* Iws1 domain.

## 2. Material and methods

### 2.1. Cloning, expression and purification

The Iws1 gene from *E. cuniculi* (GenBank CAD26349.1) was amplified by PCR from genomic DNA using the following oligonucleotides: ecIws1-N, 5'-GGA TAT CCA TAT GTC ATT ATT ACG AAA ACG GAA A-3', and ecIws1-C, 5'-CGC GGA TCC TCA ATC TCC ACC TTC TGG CTT-3'. The amplified full-length construct (amino-acid residues 1–198; ecIws1<sub>1–198</sub>) was cloned into the pNEA-tH vector (Romier *et al.*, 2006) using *NdeI* and *BamHI* restriction sites. The vector provides an N-terminal fusion composed of a hexahistidine tag and a thrombin-cleavage site. Another shorter construct was made using the oligonucleotide ecIws1- $\Delta$ N, 5'-GGA TAT CCA TAT GGA CCC TGG GAC TGT TCT GG-3'. This construct encompasses the evolutionarily conserved region of ecIws1 (amino-acid residues 55–198; ecIws1<sub>55–198</sub>). The ecIws1<sub>55–198</sub> construct was also cloned into the pNEA-tH vector.

For expression, plasmids encoding ecIws1<sub>1–198</sub> and ecIws1<sub>55–198</sub> were co-transformed with the pRare vector (Novagen) in *Escherichia coli* BL21 (DE3) cells (Novagen). 6 l auto-induction (Studier, 2005) cultures containing 100  $\mu\text{g ml}^{-1}$  ampicillin and 35  $\mu\text{g ml}^{-1}$  chloramphenicol were inoculated with co-transformants gathered on Petri dishes and the cells were grown at 310 K to an optical density of 0.4 at 600 nm. The temperature was then lowered to 298 K to allow induced growth overnight. Cells were harvested by centrifugation at 4300g, resuspended in buffer A (10 mM Tris pH 8.0, 400 mM NaCl) and lysed by ultrasonication. The soluble fraction was recovered by high-speed centrifugation at 29 000g and was mixed with 1.5 ml Talon resin (Clontech). After 1 h incubation at 277 K, the supernatant was removed and the resin was washed extensively with buffer A. The resin was then resuspended in 2 ml buffer A and left overnight at 277 K in the presence of 50 U bovine thrombin (Sigma) to cleave off the hexahistidine tag. This cleavage yielded additional residues at the N-terminus of the proteins: Gly-Ser from the thrombin site followed by His-Met from the *NdeI* cloning site.

The supernatants were recovered and applied onto a Hiload 16/60 Superdex 75 gel-filtration column (GE Healthcare) equilibrated with buffer A containing 2 mM DTT. Protein purity was analyzed by SDS-PAGE and fractions with appropriate purity were pooled and concentrated with Centripreps and Centricons (Millipore) to a final concentration of 20 mg ml<sup>-1</sup> as assessed by the Bio-Rad protein assay (Bio-Rad). The monodispersity of the purified sample was confirmed by dynamic light scattering (DLS; DynaPro, Protein Solutions).

### 2.2. Crystallization

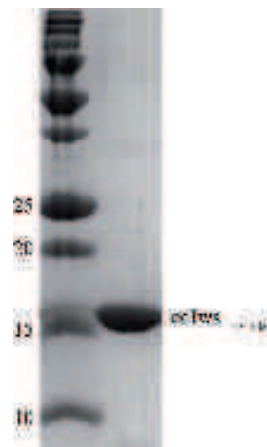
Initial crystallization trials for ecIws1<sub>55–198</sub> were performed in 96-well sitting-drop plates (Innovaplate) using a Cartesian Honeybee 8+1 robot (Genomics Solutions). The commercial Classics, JCSG+, PACT (Qiagen) and Wizard (Emerald BioSystems) screens were used for screening at 277 and 290 K. For each condition, 200 nl protein solution was mixed with 200 nl reservoir solution and equilibrated against 50  $\mu\text{l}$  reservoir solution. Crystal improvement was carried out using hanging drops (2  $\mu\text{l}$  protein solution mixed with 2  $\mu\text{l}$  reservoir solution) in 24-well plates (Greiner).

### 2.3. Data collection

All crystals were initially tested for diffraction on ESRF beamline ID14-2. A large range of cryoprotectants were tested for each crystal form. In all cases, the best results were obtained when the mother liquor was supplemented with 20% (m/v) PEG 200. Data collection at high resolution was carried out using crystals belonging to form IV. A full data set was collected from a single crystal on ESRF beamline ID23-1, which was equipped with an ADSC Q315 detector. Data were collected at a wavelength of 0.97530 Å with 1° oscillations over a range of 180°. The data were processed and scaled using the *HKL-2000* program package (Otwinowski & Minor, 1997).

## 3. Results and discussion

Initial expression and purification experiments using the full-length *E. cuniculi* Iws1 protein ecIws1<sub>1–198</sub> showed that it was very rapidly N-terminally degraded. We therefore made a C-terminal construct of this protein that corresponded to the evolutionarily conserved region of Iws1 (ecIws1<sub>55–198</sub>). This domain turned out to be stable throughout purification by affinity and size-exclusion chromato-



**Figure 1**  
Purified ecIws1<sub>55–198</sub> protein. Molecular-mass markers are shown in the left lane and their corresponding masses are given in kDa.

**Table 1**  
Crystal forms of eCIws<sub>155–198</sub>.

Crystal form	Form I	Form II	Form III	Form IV
Crystallization condition	0.1 M sodium citrate pH 5.0, 20% PEG 4000, 8% 2-propanol		0.1 M MES pH 6.6, 20% PEG 8000, 0.5% ethyl acetate	
Space group	$P2_1$	$P222_1$	$P2_1$	$P2_1$
Crystal system	Monoclinic	Orthorhombic	Monoclinic	Monoclinic
Unit-cell parameters				
$a$ (Å)	43.0	82.5	33.7	41.9
$b$ (Å)	125.4	53.1	128.8	220.0
$c$ (Å)	55.7	35.1	33.6	71.8
$\beta$ (°)	91.7	90	101.2	102.4
Molecules per ASU <sup>†</sup>	4	1	2	8
Matthews coefficient <sup>†</sup> (Å <sup>3</sup> Da <sup>-1</sup> )	2.25	2.30	2.14	2.42
Solvent content (%)	45.3	46.6	42.6	49.2
Maximum diffraction (Å)	2.00	2.20	2.30	2.50

<sup>†</sup> The number of molecules per asymmetric unit was chosen as the most probable solution obtained from the analysis of Kantardjieff & Rupp (2003) and was used for calculation of the most probable Matthews coefficient.

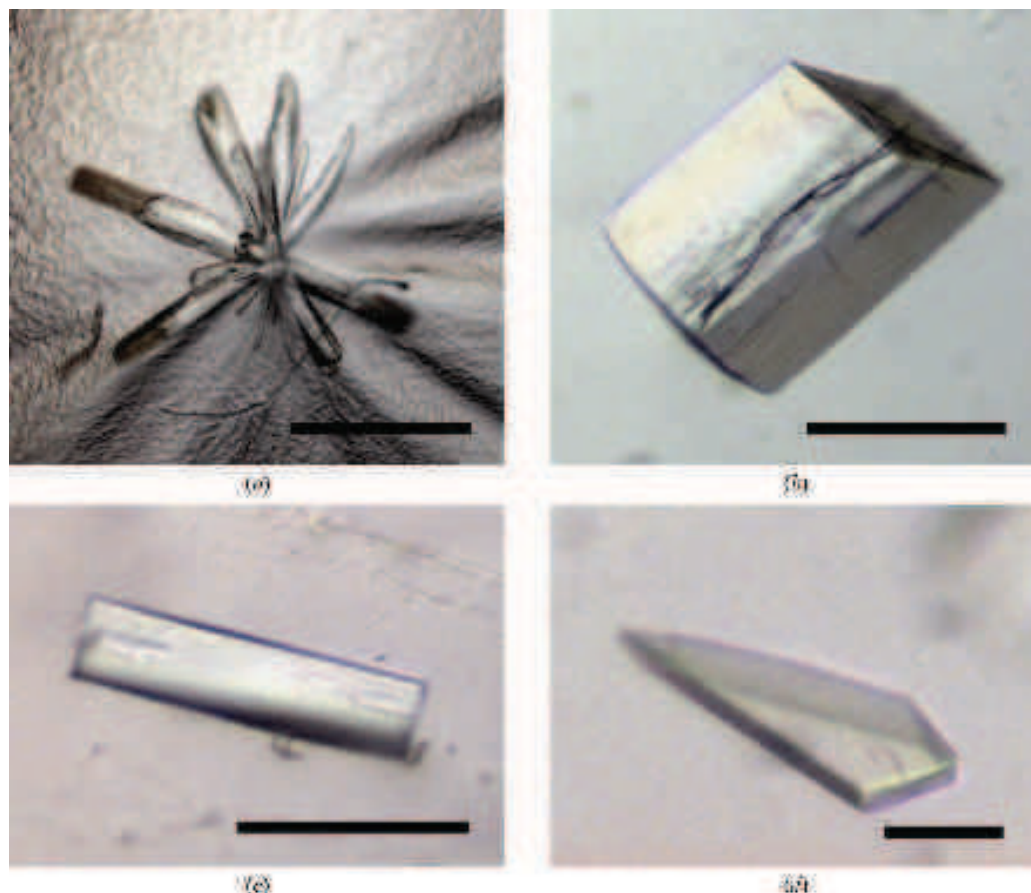
graphy and yielded 20 mg of protein that was about 95% pure per litre of culture (Fig. 1).

The eCIws<sub>155–198</sub> protein could readily be crystallized. Crystals were further refined by varying the pH and the crystallizing agent concentration, yielding an optimal condition composed of 0.1 M sodium citrate pH 5.0, 20% (*m/v*) PEG 4000 and 8% 2-propanol. At 290 K three-dimensional needles with approximate dimensions of

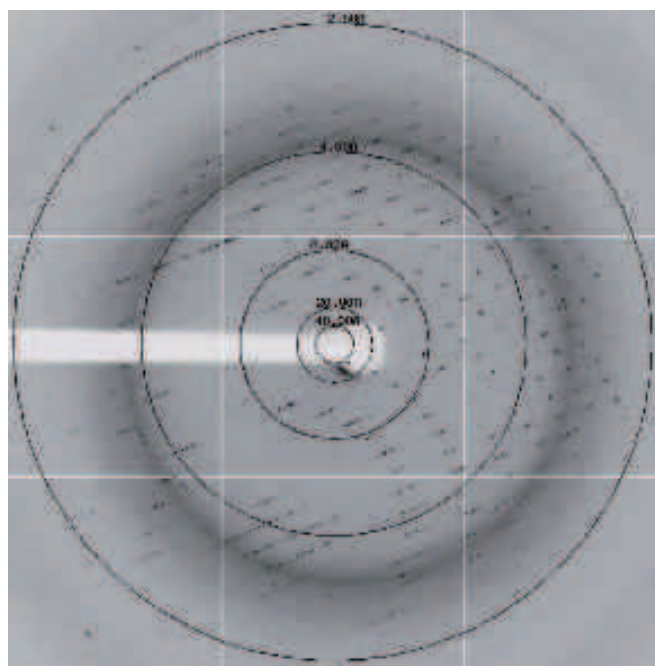
100 × 20 × 20 μm were observed (crystal form I; Fig. 2*a*). At 297 K two optically different crystal forms were observed in the same droplets: cuboid-shaped crystals with approximate dimensions of 200 × 100 × 100 μm (crystal form II; Fig. 2*b*) and needles with approximate dimensions of 150 × 40 × 40 μm (crystal form III; Fig. 2*c*). All crystals appeared within 1 d and continued to grow for a few days. A major problem that was encountered with all these crystals was that they cracked immediately after removing the cover slip, most likely owing to rapid evaporation of the 2-propanol present in the crystallization condition.

Further inspection of the crystallization trials revealed the presence of plate-shaped crystals in a different condition that was refined to 0.1 M MES pH 6.6, 20% (*m/v*) PEG 8000 and 0.5% ethyl acetate. These crystals grew at 290 and 297 K within two weeks to reach approximate dimensions of 300 × 150 × 20 μm (crystal form IV; Fig. 2*d*). Although these crystals were very fragile, they could be handled more easily than the other crystal forms obtained previously.

The crystal-cracking problems observed with crystals forms I, II and III yielded poor diffraction patterns, notably multiple lattices. Furthermore, although most crystals showed initial diffraction to 2.0–2.3 Å resolution, anisotropic diffraction and rapid crystal decay were commonly encountered. These problems allowed characterization of the different crystal forms but prevented the collection of complete data sets. Crystal forms I and III both belonged to the monoclinic space group  $P2_1$ , but differed in their unit-cell parameters (Table 1). Calculation of the most probable Matthews coefficients (Matthews, 1968; Kantardjieff & Rupp, 2003) showed that crystal forms I and III



**Figure 2**  
Different crystal forms obtained upon crystallization of eCIws<sub>155–198</sub>. (a) Form I (space group  $P2_1$ ). (b) Form II (space group  $P222_1$ ). (c) Form III (space group  $P2_1$ ). (d) Form IV (space group  $P2_1$ ). The black bars represent 100 μm.



**Figure 3** Diffraction pattern obtained with crystal form IV (space group  $P2_1$ ). The resolution rings displayed are at 40.0, 20.0, 8.0, 4.0 and 2.5 Å.

are most likely to contain four and two molecules per asymmetric unit, respectively. Crystal form II belonged to the orthorhombic space group  $P222_1$  and calculation of the Matthews coefficient suggested the presence of one molecule per asymmetric unit (Table 1).

In contrast, the fourth crystal form yielded crystals that could be used for data collection (a representative diffraction pattern is presented in Fig. 3). These crystals also belonged to the monoclinic space group  $P2_1$  but were different from crystal forms I and III and were characterized by a long  $b$  axis of 220 Å (Table 1). Diffraction data with good statistics were obtained in the resolution range 30–2.5 Å (Table 2). Calculation of the most probable Matthews coefficient suggested the presence of more than four molecules per asymmetric unit (Table 1). A self-rotation analysis showed the presence of two clear twofold axes but could not help to resolve the exact number of molecules per asymmetric unit.

Owing to the partial sequence homology between Iws1 and TFIIS, structure determination was attempted using the NMR structures of yeast and mouse TFIIS domain I available in the Protein Data Bank (PDB codes 1eo0 and 1wjt, respectively). However, no satisfactory solution was found. Therefore, we are now concentrating on

**Table 2**

Experimental X-ray data for crystal form IV.

Values in parentheses are for the highest resolution shell.

Resolution range (Å)	30.0–2.50 (2.59–2.50)
Total observations	130808
Unique reflections	43576 (4385)
Multiplicity	3.0 (2.7)
Completeness (%)	99.3 (99.7)
$R_{\text{merge}}$ (%)	6.2 (36.4)
$I/\sigma(I)$	20.6 (3.5)

obtaining experimental phases for the structure determination of eClws1<sub>55–198</sub>.

This work was supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Université de Strasbourg (UDS) as well as the European Commission SPINE2-Complexes project (contract No. LSHG-CT-2006-031220).

## References

- Adkins, M. W. & Tyler, J. K. (2006). *Mol. Cell*, **21**, 405–416.
- Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W. & Lis, J. T. (2009). *EMBO J.* **28**, 1067–1077.
- Booth, V., Koth, C. M., Edwards, A. M. & Arrowsmith, C. H. (2000). *J. Biol. Chem.* **275**, 31266–31268.
- Bortvin, A. & Winston, F. (1996). *Science*, **272**, 1473–1476.
- Endoh, M. *et al.* (2004). *Mol. Cell. Biol.* **24**, 3324–3336.
- Fischbeck, J. A., Kraemer, S. M. & Stargell, L. A. (2002). *Genetics*, **162**, 1605–1616.
- Kantardjiev, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kaplan, C. D., Laprade, L. & Winston, F. (2003). *Science*, **301**, 1096–1099.
- Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S. & Greenblatt, J. F. (2002). *Mol. Cell. Biol.* **22**, 6979–6992.
- Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R. III & Hartzog, G. A. (2003). *Mol. Cell. Biol.* **23**, 1368–1378.
- Ling, Y., Smith, A. J. & Morgan, G. T. (2006). *Nucleic Acids Res.* **34**, 2219–2229.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Romier, C. *et al.* (2006). *Acta Cryst.* **D62**, 1232–1242.
- Romier, C., James, N., Birck, C., Cavarelli, J., Vivarès, C., Collart, M. A. & Moras, D. (2007). *J. Mol. Biol.* **368**, 1292–1306.
- Sims, R. J. III, Belotserkovskaya, R. & Reinberg, D. (2004). *Genes Dev.* **18**, 2437–2468.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 225–232.
- Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. (2007). *Genes Dev.* **21**, 160–174.
- Yoh, S. M., Lucas, J. S. & Jones, K. A. (2008). *Genes Dev.* **22**, 3422–3434.
- Zhang, L., Fletcher, A. G., Cheung, V., Winston, F. & Stargell, L. A. (2008). *Mol. Cell. Biol.* **28**, 1393–1403.



# **Chapitre 9**

## **Références bibliographiques**





- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., et Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section D* 58, 1948-1954.
- Adkins, M. W., et Tyler, J. K. (2006). Transcriptional Activators Are Dispensable for Transcription in the Absence of Spt6-Mediated Chromatin Reassembly of Promoter Regions. *Mol Cell* 21, 405-416.
- Agez, M., Chen, J., Guerois, R., van Heijenoort, C., Thuret, J.-Y., Mann, C., et Ochsenbein, F. (2007). Structure of the Histone Chaperone Asf1 Bound to the Histone H3 C-Terminal Helix and Functional Insights. *Structure* 15, 191-199.
- Anderson, M., Huh, J. H., Ngo, T., Lee, A., Hernandez, G., Pang, J., Perkins, J., et Dutnall, R. N. (2010). Co-expression as a convenient method for the production and purification of core histones in bacteria. *Protein Expression and Purification* 72, 194-204.
- Andrulis, E. D., Guzmán, E., Döring, P., Werner, J., et Lis, J. T. (2000). High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes *in vivo*: roles in promoter proximal pausing and transcription elongation. *Genes & Development* 14, 2635 -2649.
- Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., et Lis, J. T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* 420, 837-841.
- Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., et Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. *EMBO J* 28, 1067-1077.
- Avvakumov, N., Nourani, A., et Côté, J. (2011). Histone Chaperones: Modulators of Chromatin Marks. *Molecular Cell* 41, 502-514.
- Bai, L., et Morozov, A. V. (2010). Gene regulation by nucleosome positioning. *Trends in Genetics* 26, 476-483.
- Baniahmad, C., Nawaz, Z., Baniahmad, A., Gleeson, M. A., Tsai, M. J., et O'Malley, B. W. (1995). Enhancement of human estrogen receptor activity by SPT6: a potential coactivator. *Molecular Endocrinology* 9, 34 -43.
- Basehoar, A. D., Zanton, S. J., et Pugh, B. F. (2004). Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell* 116, 699-709.
- Baumann, M., Pontiller, J., et Ernst, W. (2010). Structure and Basal Transcription Complex of RNA Polymerase II Core Promoters in the Mammalian Genome: An Overview. *Molecular Biotechnology* 45, 241-247.
- Becker, R., Loll, B., et Meinhart, A. (2008). Snapshots of the RNA Processing Factor SCAF8 Bound to Different Phosphorylated Forms of the Carboxyl-terminal Domain of RNA Polymerase II. *Journal of Biological Chemistry* 283, 22659 -22669.

- Beckouët, F., Mariotte-Labarre, S., Peyroche, G., Nogi, Y., et Thuriaux, P. (2011). Rpa43 and its partners in the yeast RNA polymerase I transcription complex. *FEBS Letters* 585, 3355-3359.
- Berger, I., Fitzgerald, D. J., et Richmond, T. J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nat Biotech* 22, 1583-1587.
- Besse, S., Vigneron, M., Pichard, E., et Puvion-Dutilleul, F. (1995). Synthesis and maturation of viral transcripts in herpes simplex virus type 1 infected HeLa cells: the role of interchromatin granules. *Gene Expression* 4, 143-161.
- Bhaumik, S. R. (2011). Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1809, 97-108.
- De Boer, H. A., Comstock, L. J., Hui, A., Wong, E., et Vasser, M. (1983). A hybrid promoter and portable Shine-Dalgarno regions of *Escherichia coli*. *Biochem. Soc. Symp.* 48, 233-244.
- Booth, V., Koth, C. M., Edwards, A. M., et Arrowsmith, C. H. (2000). Structure of a Conserved Domain Common to the Transcription Factors TFIIS, Elongin A, and CRSP70. *Journal of Biological Chemistry* 275, 31266 -31268.
- Bortvin, A., et Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* 272, 1473-1476.
- Brès, V., Yoh, S. M., et Jones, K. A. (2008). The multi-tasking P-TEFb complex. *Current Opinion in Cell Biology* 20, 334-340.
- Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell* 36, 541-546.
- Buratowski, S. (2003). The CTD code. *Nat Struct Mol Biol* 10, 679-680.
- Burley, S. K., et Petsko, G. A. (1986). Amino-aromatic interactions in proteins. *FEBS Letters* 203, 139-143.
- Cai, G., Imasaki, T., Yamada, K., Cardelli, F., Takagi, Y., et Asturias, F. J. (2010). Mediator Head module structure and functional interactions. *Nat Struct Mol Biol* 17, 273-279.
- Chaban, Y., Ezeokonkwo, C., Chung, W.-H., Zhang, F., Kornberg, R. D., Maier-Davis, B., Lorch, Y., et Asturias, F. J. (2008). Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat Struct Mol Biol* 15, 1272-1277.
- Chang, B., Chen, Y., Zhao, Y., et Bruick, R. K. (2007). JMJD6 Is a Histone Arginine Demethylase. *Science* 318, 444-447.
- Chapman, R. D., Heidemann, M., Hintermair, C., et Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. *Trends in Genetics* 24, 289-296.

- Chen, H.-T., Warfield, L., et Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat Struct Mol Biol* *14*, 696-703.
- Cheung, A. C. M., et Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* *471*, 249-253.
- Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R., et Winston, F. (2008). Chromatin- and Transcription-Related Factors Repress Transcription from within Coding Regions throughout the *Saccharomyces cerevisiae* Genome. *PLoS Biol* *6*, e277.
- Chiang, P.-W., Wang, S., Smithivas, P., Song, W.-J., Ramamoorthy, S., Hillman, J., Puett, S., Van Keuren, M. L., Crombez, E., Kumar, A., et al. (1996). Identification and Analysis of the Human and Murine Putative Chromatin Structure Regulator SUPT6H and Supt6h. *Genomics* *34*, 328-333.
- Chung, C.-wa (2007). The use of biophysical methods increases success in obtaining liganded crystal structures. *Acta Crystallographica Section D* *63*, 62-71.
- Clapier, C. R., et Cairns, B. R. (2009). The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* *78*, 273-304.
- Clark-Adams, C. D., et Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* *7*, 679 -686.
- Clark-Adams, C. D., Norris, D., Osley, M. A., Fassler, J. S., et Winston, F. (1988). Changes in histone gene dosage alter transcription in yeast. *Genes & Development* *2*, 150 -159.
- Close, D., Johnson, S. J., Sdano, M. A., McDonald, S. M., Robinson, H., Formosa, T., et Hill, C. P. (2011). Crystal Structures of the *S. cerevisiae* Spt6 Core and C-Terminal Tandem SH2 Domain. *Journal of Molecular Biology* *408*, 697-713.
- Collaborative (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* *50*, 760-763.
- Collins, R. E., Northrop, J. P., Horton, J. R., Lee, D. Y., Zhang, X., Stallcup, M. R., et Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol* *15*, 245-250.
- Conaway, R. C., et Conaway, J. W. (2011). Function and regulation of the Mediator complex. *Current Opinion in Genetics & Development* *21*, 225-230.
- Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. *Current Opinion in Genetics & Development* *14*, 218-226.
- Cramer, P., Bushnell, D. A., et Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. *Science* *292*, 1863-1876.

- Das, C., Tyler, J. K., et Churchill, M. E. A. (2010). The histone shuffle: histone chaperones in an energetic dance. *Trends in Biochemical Sciences* 35, 476-489.
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35, W375 -W383.
- Dengl, S., Mayer, A., Sun, M., et Cramer, P. (2009). Structure and *in vivo* Requirement of the Yeast Spt6 SH2 Domain. *Journal of Molecular Biology* 389, 211-225.
- Denis, C. L. (1984). Identification of new genes involved in the regulation of yeast alcohol dehydrogenase II. *Genetics* 108, 833 -844.
- Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T., et Romier, C. (2011). Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. *Journal of Structural Biology* 175, 178-188.
- Diebold, M.-L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., et Romier, C. (2010a). The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. *EMBO J* 29, 3979-3991.
- Diebold, M.-L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., et Romier, C. (2010b). Noncanonical Tandem SH2 Enables Interaction of Elongation Factor Spt6 with RNA Polymerase II. *Journal of Biological Chemistry* 285, 38389 -38398.
- Diribarne, G., et Bensaude, O. (2009). 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biology* 6, 122-128.
- Egloff, S., et Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends in Genetics* 24, 280-288.
- Egloff, S., O'Reilly, D., Chapman, R. D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., et Murphy, S. (2007). Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression. *Science* 318, 1777 -1779.
- Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., et Murphy, S. (2012). Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. *Molecular Cell*. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276511008562>.
- Emsley, P., et Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* 60, 2126-2132.
- Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.-K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., et al. (2004). Human Spt6 Stimulates Transcription Elongation by RNA Polymerase II *In vitro*. *Molecular and Cellular Biology* 24, 3324 -3336.

- English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. A., et Tyler, J. K. (2006). Structural Basis for the Histone Chaperone Activity of Asf1. *Cell* *127*, 495-508.
- Erdel, F., Krug, J., Längst, G., et Rippe, K. (2011). Targeting chromatin remodelers: Signals and search mechanisms. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* *1809*, 497-508.
- Fassler, J. S., et Winston, F. (1988). Isolation and Analysis of a Novel Class of Suppressor of Ty Insertion Mutations in *Saccharomyces cerevisiae*. *Genetics* *118*, 203 -212.
- Felsenfeld, G., et Groudine, M. (2003). Controlling the double helix. *Nature* *421*, 448-453.
- Fischbeck, J. A., Kraemer, S. M., et Stargell, L. A. (2002). SPN1, a Conserved Gene Identified by Suppression of a Postrecruitment-Defective Yeast TATA-Binding Protein Mutant. *Genetics* *162*, 1605 -1616.
- Fribourg, S., Romier, C., Werten, S., Gangloff, Y.-G., Poterszman, A., et Moras, D. (2001). Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli*. *Journal of Molecular Biology* *306*, 363-373.
- Gangloff, Y.-G., Romier, C., Thuault, S., Werten, S., et Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. *Trends in Biochemical Sciences* *26*, 250-257.
- Gao, Q., Hua, J., Kimura, R., Headd, J. J., Fu, X.-yuan, et Chin, Y. E. (2004). Identification of the Linker-SH2 Domain of STAT as the Origin of the SH2 Domain Using Two-dimensional Structural Alignment. *Molecular & Cellular Proteomics* *3*, 704 -714.
- García-Oliver, E., García-Molinero, V., et Rodríguez-Navarro, S. (2012). mRNA export and gene expression: The SAGA–TREX-2 connection. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Available at: <http://www.sciencedirect.com/science/article/pii/S1874939911002173>.
- Gavin, A.-C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* *415*, 141-147.
- Gershenson, N. I., et Ioshikhes, I. P. (2004). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* *21*, 1295 -1300.
- Goddard, T. D., et Kneller, D. G. SPARKY 3 (University of California, San Francisco).
- Guerry, P., et Herrmann, T. (2012). Protein NMR Techniques. Dans *Methods in Molecular Biology*. (Springer New York), p. 429-451. Available at: [http://dx.doi.org/10.1007/978-1-61779-480-3\\_22](http://dx.doi.org/10.1007/978-1-61779-480-3_22).
- Guglielmi, B., Soutourina, J., Esnault, C., et Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation *in vivo*. *Proceedings of the National Academy of Sciences* *104*, 16062 -16067.

- Güntert, P. (2004). Automated NMR Structure Calculation With CYANA. Dans *Protein NMR Techniques Methods in Molecular Biology.*, p. 353 -378. Available at: <http://www.springerprotocols.com/Abstract/doi/10.1385/1-59259-809-9:353>.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* *11*, 394-403.
- Hainer, S. J., Pruneski, J. A., Mitchell, R. D., Monteverde, R. M., et Martens, J. A. (2011). Intergenic transcription causes repression by directing nucleosome assembly. *Genes & Development* *25*, 29 -40.
- Happel, A. M., Swanson, M. S., et Winston, F. (1991). The SNF2, SNF5 and SNF6 Genes Are Required for Ty Transcription in *Saccharomyces cerevisiae*. *Genetics* *128*, 69 -77.
- Hartzog, G. A., Wada, T., Handa, H., et Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes & Development* *12*, 357 -369.
- Hirschhorn, J. N., Brown, S. A., Clark, C. D., et Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes & Development* *6*, 2288 -2298.
- Holm, L., et Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research* *38*, W545 -W549.
- Huang, H., Li, L., Wu, C., Schibli, D., Colwill, K., Ma, S., Li, C., Roy, P., Ho, K., Songyang, Z., et al. (2008). Defining the Specificity Space of the Human Src Homology 2 Domain. *Molecular & Cellular Proteomics* *7*, 768 -784.
- Ivanovska, I., Jacques, P.-É., Rando, O. J., Robert, F., et Winston, F. (2011). Control of Chromatin Structure by Spt6: Different Consequences in Coding and Regulatory Regions. *Molecular and Cellular Biology* *31*, 531 -541.
- Izzo, A., et Schneider, R. (2010). Chatting histone modifications in mammals. *Briefings in Functional Genomics* *9*, 429 -443.
- Jaehning, J. A. (2010). The Paf1 complex: Platform or player in RNA polymerase II transcription? *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* *1799*, 379-388.
- Jensen, M. M., Christensen, M. S., Bonven, B., et Jensen, T. H. (2008). Requirements for chromatin reassembly during transcriptional downregulation of a heat shock gene in *Saccharomyces cerevisiae*. *FEBS Journal* *275*, 2956-2964.
- John MacLennan, A., et Shaw, G. (1993). A yeast SH2 domain. *Trends in Biochemical Sciences* *18*, 464-465.
- Johnson, S. J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S. L., et Hill, C. P. (2008). Crystal Structure and RNA Binding of the Tex Protein from *Pseudomonas aeruginosa*. *Journal of Molecular Biology* *377*, 1460-1473.

- Juven-Gershon, T., et Kadonaga, J. T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339, 225-229.
- Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W., et Kadonaga, J. T. (2008). The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology* 20, 253-259.
- Kaplan, C. A., Laprade, L., et Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096-1099.
- Kaplan, C. D., Holland, M. J., et Winston, F. (2005). Interaction between Transcription Elongation Factors and mRNA 3'-End Formation at the *Saccharomyces cerevisiae* GAL10-GAL7 Locus. *Journal of Biological Chemistry* 280, 913 -922.
- Kaplan, C. D., Morris, J. R., Wu, C.-ting, et Winston, F. (2000). Spt5 and Spt6 are associated with active transcription and have characteristics of general elongation factors in *D. melanogaster*. *Genes & Development* 14, 2623 -2634.
- Kasten, M. M., Clapier, C. R., et Cairns, B. R. (2011). SnapShot: Chromatin Remodeling:SWI/SNF. *Cell* 144, 310-310.e1.
- Kim, B., Nesvizhskii, A. I., Rani, P. G., Hahn, S., Aebersold, R., et Ranish, J. A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences* 104, 16068 -16073.
- Kim, M., Ahn, S.-H., Krogan, N. J., Greenblatt, J. F., et Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23, 354-364.
- Klopf, E., Paskova, L., Sole, C., Mas, G., Petryshyn, A., Posas, F., Wintersberger, U., Ammerer, G., et Schuller, C. (2009). Cooperation between the INO80 Complex and Histone Chaperones Determines Adaptation of Stress Gene Transcription in the Yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 29, 4994-5007.
- Koch, M., Diebold, M.-L., Cavarelli, J., et Romier, C. (2010). Crystallization and preliminary crystallographic analysis of eukaryotic transcription and mRNA export factor Iws1 from *Encephalitozoon cuniculi*. *Acta Crystallographica Section F* 66, 207-210.
- Kohler, A., et Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* 8, 761-773.
- Kouzarides, T. (2007). Chromatin Modifications and Their Function. *Cell* 128, 693-705.
- Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., et Greenblatt, J. F. (2002). RNA Polymerase II Elongation Factors of *Saccharomyces cerevisiae*: a Targeted Proteomics Approach. *Mol. Cell. Biol.* 22, 6979-6992.



- Kuriyan, J., et Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* *26*, 259-288.
- Kutach, A. K., et Kadonaga, J. T. (2000). The Downstream Promoter Element DPE Appears To Be as Widely Used as the TATA Box in *Drosophila* Core Promoters. *Molecular and Cellular Biology* *20*, 4754 -4764.
- Laurent, B. C., Treitel, M. A., et Carlson, M. (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proceedings of the National Academy of Sciences* *88*, 2687 -2691.
- Lemmon, M. A., et Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell* *141*, 1117-1134.
- Levine, M. (2011). Paused RNA Polymerase II as a Developmental Checkpoint. *Cell* *145*, 502-511.
- Li, B., Carey, M., et Workman, J. L. (2007). The Role of Chromatin during Transcription. *Cell* *128*, 707-719.
- Li, J., et Gilmour, D. S. (2011). Promoter proximal pausing and the control of gene expression. *Current Opinion in Genetics & Development* *21*, 231-235.
- Li, L., Ye, H., Guo, H., et Yin, Y. (2010). Arabidopsis IWS1 interacts with transcription factor BES1 and is involved in plant steroid hormone brassinosteroid regulated gene expression. *Proceedings of the National Academy of Sciences* *107*, 3918 -3923.
- Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R., et Hartzog, G. A. (2003). Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Molecular and Cellular Biology* *23*, 1368 -1378.
- Ling, Y., Smith, A. J., et Morgan, G. T. (2006). A sequence motif conserved in diverse nuclear proteins identifies a protein interaction domain utilised for nuclear targeting by human TFIIS. *Nucleic Acids Research* *34*, 2219 -2229.
- Lis, J. (1998). Promoter-associated Pausing in Promoter Architecture and Postinitiation Transcriptional Regulation. *Cold Spring Harbor Symposia on Quantitative Biology* *63*, 347 -356.
- Liu, B. A., Jablonowski, K., Raina, M., Arcé, M., Pawson, T., et Nash, P. D. (2006). The Human and Mouse Complement of SH2 Domain Proteins—Establishing the Boundaries of Phosphotyrosine Signaling. *Molecular Cell* *22*, 851-868.
- Liu, B. A., Jablonowski, K., Shah, E. E., Engelmann, B. W., Jones, R. B., et Nash, P. D. (2010). SH2 Domains Recognize Contextual Peptide Sequence Information to Determine Selectivity. *Molecular & Cellular Proteomics* *9*, 2391 -2404.
- Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., et Shi, Y. (2011). Solution Structure of Tandem SH2 Domains from Spt6 Protein and Their Binding to

- the Phosphorylated RNA Polymerase II C-terminal Domain. *Journal of Biological Chemistry* 286, 29218 -29226.
- Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H., Ranish, J., Shokat, K. M., et Hahn, S. (2009). Phosphorylation of the Transcription Elongation Factor Spt5 by Yeast Bur1 Kinase Stimulates Recruitment of the PAF Complex. *Molecular and Cellular Biology* 29, 4852 -4863.
- Liu, Z., Zhou, Z., Chen, G., et Bao, S. (2007). A putative transcriptional elongation factor hIws1 is essential for mammalian cell proliferation. *Biochemical and Biophysical Research Communications* 353, 47-53.
- Lowary, P. ., et Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology* 276, 19-42.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., et Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* 389, 251-260.
- Luger, K., Rechsteiner, T. J., et Richmond, T. J. (1999). Preparation of nucleosome core particle from recombinant histones. Dans *Chromatin* (Academic Press), p. 3-19. Available at: <http://www.sciencedirect.com/science/article/pii/S0076687999040033>.
- Machida, K., et Mayer, B. J. (2005). The SH2 domain: versatile signaling module and pharmaceutical target. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1747, 1-25.
- Maciejewski, P. M., Peterson, F. C., Anderson, P. J., et Brooks, C. L. (1995). Mutation of Serine 90 to Glutamic Acid Mimics Phosphorylation of Bovine Prolactin. *Journal of Biological Chemistry* 270, 27661 -27665.
- Makde, R. D., England, J. R., Yennawar, H. P., et Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 467, 562-566.
- Malagón, F., et Aguilera, A. (1996). Differential intrachromosomal hyper-recombination phenotype of spt4 and spt6 mutants of *S. cerevisiae*. *Current Genetics* 30, 101-106.
- Malik, S., et Roeder, R. G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends in Biochemical Sciences* 30, 256-263.
- Maniatis, T., et Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., et Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO J* 30, 1302-1310.
- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., et Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17, 1272-1278.

- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., et al., Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* *40*, 658-674.
- McDonald, S. M., Close, D., Xin, H., Formosa, T., et al., Hill, C. P. (2010). Structure and Biological Importance of the Spn1-Spt6 Interaction, and Its Regulatory Role in Nucleosome Binding. *Molecular Cell* *40*, 725-735.
- Millevoi, S., et al., Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research* *38*, 2757-2774.
- Mohibullah, N., et al., Hahn, S. (2008). Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes & Development* *22*, 2994-3006.
- Murshudov, G. N., Vagin, A. A., et al., Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* *53*, 240-255.
- Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., et al., Senda, T. (2007). Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature* *446*, 338-341.
- Nechaev, S., et al., Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* *1809*, 34-45.
- Neigeborn, L., Rubin, K., et al., Carlson, M. (1986). Suppressors of Snf2 mutations restore invertase derepression and cause temperature-sensitive lethality in yeast. *Genetics* *112*, 741-753.
- Okazaki, I.-mi, Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T., et al. (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proceedings of the National Academy of Sciences* *108*, 7920-7925.
- Orphanides, G., Lagrange, T., et al., Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes & Development* *10*, 2657-2683.
- Otwinowski, Z., et al., Minor, W. (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. *Dans Macromolecular Crystallography Part A* (Academic Press), p. 307-326. Available at: <http://www.sciencedirect.com/science/article/pii/S007668799776066X>.
- Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., et al. (2001). High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *Journal of Biomolecular Screening* *6*, 429-440.

- Perales, R., et Bentley, D. (2009). « Cotranscriptionality »: The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Molecular Cell* 36, 178-191.
- Perrakis, A., et Romier, C. (2008). Structural Proteomics. Dans *Methods in Molecular Biology*<sup>TM</sup>. (Humana Press), p. 247-256. Available at: [http://dx.doi.org/10.1007/978-1-60327-058-8\\_15](http://dx.doi.org/10.1007/978-1-60327-058-8_15).
- Peyretailade, E., El Alaoui, H., Diogon, M., Polonais, V., Parisot, N., Biron, D. G., Peyret, P., et Delbac, F. (2011). Extreme reduction and compaction of microsporidian genomes. *Research in Microbiology* 162, 598-606.
- Phatnani, H. P., et Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development* 20, 2922 -2936.
- Portela, A., et Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotech* 28, 1057-1068.
- Pruneski, J. A., Hainer, S. J., Petrov, K. O., et Martens, J. A. (2011). The Paf1 Complex Represses SER3 Transcription in *Saccharomyces cerevisiae* by Facilitating Intergenic Transcription-Dependent Nucleosome Occupancy of the SER3 Promoter. *Eukaryotic Cell* 10, 1283 -1294.
- Pujari, V., Radebaugh, C. A., Chodaparambil, J. V., Muthurajan, U. M., Almeida, A. R., Fischbeck, J. A., Luger, K., et Stargell, L. A. (2010). The Transcription Factor Spn1 Regulates Gene Expression via a Highly Conserved Novel Structural Motif. *Journal of Molecular Biology* 404, 1-15.
- Ramakrishnan, V. (1997). Histone structure and the organization of the nucleosome. *Annual Review of Biophysics and Biomolecular Structure* 26, 83-112.
- Reinberg, D., et Sims, R. J. (2006). de FACTo Nucleosome Dynamics. *Journal of Biological Chemistry* 281, 23297 -23301.
- Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P. H. N., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., et al. (2006). Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. *Acta Crystallographica Section D* 62, 1232-1242.
- Romier, C., James, N., Birck, C., Cavarelli, J., Vivarès, C., Collart, M. A., et Moras, D. (2007). Crystal Structure, Biochemical and Genetic Characterization of Yeast and *E. cuniculi* TAFII5 N-terminal Domain: Implications for TFIID Assembly. *Journal of Molecular Biology* 368, 1292-1306.
- Sarma, K., et Reinberg, D. (2005). Histone variants meet their match. *Nat Rev Mol Cell Biol* 6, 139-149.
- Saunders, A., Core, L. J., et Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7, 557-567.

- Schomaker, V., et Trueblood, K. N. (1968). On the rigid-body motion of molecules in crystals. *Acta Crystallographica Section B* *24*, 63-76.
- Segal, E., et Widom, J. (2009). What controls nucleosome positions? *Trends in Genetics* *25*, 335-343.
- Seraphin, B., et Rosbash, M. (1989). Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* *59*, 349-358.
- Shen, Y., Delaglio, F., Cornilescu, G., et Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* *44*, 213-223.
- Shuaib, M., Ouararhni, K., Dimitrov, S., et Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. *Proceedings of the National Academy of Sciences* *107*, 1349 -1354.
- Sikorski, T. W., et Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology* *21*, 344-351.
- Sims, R. J., Belotserkovskaya, R., et Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes & Development* *18*, 2437 -2468.
- Smale, S. T., et Kadonaga, J. T. (2003). The RNA Polymerase II core promoter. *Annu. Rev. Biochem.* *72*, 449-479.
- Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., et Werner, M. (2011). Direct Interaction of RNA Polymerase II and Mediator Required for Transcription *in vivo*. *Science* *331*, 1451 -1454.
- Strahl, B. D., et Allis, C. D. (2000). The language of covalent histone modifications. *Nature* *403*, 41-45.
- Studier F. William (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* *41*, 207-234.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J., et Dubendorff, J. W. (1990). [6] Use of T7 RNA polymerase to direct expression of cloned genes. Dans *Gene Expression Technology* (Academic Press), p. 60-89. Available at: <http://www.sciencedirect.com/science/article/pii/007668799085008C>.
- Subtil-Rodríguez, A., et Reyes, J. C. (2011). To cross or not to cross the nucleosome, that is the elongation question.. *RNA Biol* *8*, 389-393.
- Sun, M., Larivière, L., Dengl, S., Mayer, A., et Cramer, P. (2010). A Tandem SH2 Domain in Transcription Elongation Factor Spt6 Binds the Phosphorylated RNA Polymerase II C-terminal Repeat Domain (CTD). *Journal of Biological Chemistry* *285*, 41597 - 41603.

- Swanson, M. S., Carlson, M., et Winston, F. (1990). SPT6, an essential gene that affects transcription in *Saccharomyces cerevisiae*, encodes a nuclear protein with an extremely acidic amino terminus. *Molecular and Cellular Biology* *10*, 4935-4941.
- Taatjes, D. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in Biochemical Sciences* *35*, 315-322.
- Talbert, P. B., et Henikoff, S. (2010). Histone variants - ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* *11*, 264-275.
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., et Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* *14*, 1025-1040.
- Thebault, P., Boutin, G., Bhat, W., Rufiange, A., Martens, J., et Nourani, A. (2011). Transcription Regulation by the Noncoding RNA SRG1 Requires Spt2-Dependent Chromatin Deposition in the Wake of RNA Polymerase II. *Molecular and Cellular Biology* *31*, 1288-1300.
- Thomas, M. C., et Chiang, C.-M. (2006). The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology* *41*, 105-178.
- Travers, A., Caserta, M., Churcher, M., Hiriart, E., et Di Mauro, E. (2009). Nucleosome positioning-what do we really know? *Mol. Biosyst.* *5*, 1582-1592.
- Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., et Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. *Journal of Structural Biology* *172*, 45-54.
- Vasudevan, D., Chua, E. Y. D., et Davey, C. A. (2010). Crystal Structures of Nucleosome Core Particles Containing the '601' Strong Positioning Sequence. *Journal of Molecular Biology* *403*, 1-10.
- Verdaasdonk, J. S., et Bloom, K. (2011). Centromeres: unique chromatin structures that drive chromosome segregation. *Nat Rev Mol Cell Biol* *12*, 320-332.
- Vijayachandran, L. S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., et al. (2011). Robots, pipelines, polyproteins: Enabling multiprotein expression in prokaryotic and eukaryotic cells. *Journal of Structural Biology* *175*, 198-208.
- Vonrhein, C., Blanc, E., Roversi, P., et Bricogne, G. (2007). Automated Structure Solution With autoSHARP. Dans *Macromolecular Crystallography Protocols Methods in Molecular Biology*, p. 215-230. Available at: <http://www.springerprotocols.com/Abstract/doi/10.1385/1-59745-266-1:215>.
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development* *12*, 343-356.

- Walter, T. S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., et Grimes, J. M. (2006). Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* *14*, 1617-1622.
- Wery, M., Shematorova, E., Van Driessche, B., Vandehaute, J., Thuriaux, P., et Van Mullem, V. (2004). Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J* *23*, 4232-4242.
- White, A. E., Burch, B. D., Yang, X.-cui, Gasdaska, P. Y., Dominski, Z., Marzluff, W. F., et Duronio, R. J. (2011). Drosophila histone locus bodies form by hierarchical recruitment of components. *The Journal of Cell Biology* *193*, 677 -694.
- Widiez, T., El Kafafi, E. S., Girin, T., Berr, A., Ruffel, S., Krouk, G., Vayssières, A., Shen, W.-H., Coruzzi, G. M., Gojon, A., et al. (2011). HIGH NITROGEN INSENSITIVE 9 (HNI9)-mediated systemic repression of root NO<sub>3</sub><sup>-</sup> uptake is associated with changes in histone methylation. *Proceedings of the National Academy of Sciences* *108*, 13329 -13334.
- Winkler, M., aus dem Siepen, T., et Stamminger, T. (2000). Functional Interaction between Pleiotropic Transactivator pUL69 of Human Cytomegalovirus and the Human Homolog of Yeast Chromatin Regulatory Protein SPT6. *Journal of Virology* *74*, 8053 -8064.
- Winston, F., Chaleff, D. T., Valent, B., et Fink, G. R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* *107*, 179 -197.
- Wu, P.-Y. J., Ruhlmann, C., Winston, F., et Schultz, P. (2004). Molecular Architecture of the *S. cerevisiae* SAGA Complex. *Molecular Cell* *15*, 199-208.
- Yamaguchi, Y., Narita, T., Inukai, N., Wada, T., et Handa, H. (2001). SPT Genes: Key Players in the Regulation of Transcription, Chromatin Structure and Other Cellular Processes. *Journal of Biochemistry* *129*, 185 -191.
- Yoh, S. M., Cho, H., Pickle, L., Evans, R. M., et Jones, K. A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes & Development* *21*, 160 -174.
- Yoh, S. M., Lucas, J. S., et Jones, K. A. (2008). The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & Development* *22*, 3422 -3434.
- Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D., et Mellor, J. (2008). Roles for Ctk1 and Spt6 in Regulating the Different Methylation States of Histone H3 Lysine 36. *Molecular and Cellular Biology* *28*, 4915 -4926.
- Zhang, L., Fletcher, A. G. L., Cheung, V., Winston, F., et Stargell, L. A. (2008). Spn1 Regulates the Recruitment of Spt6 and the Swi/Snf Complex during Transcriptional Activation by RNA Polymerase II. *Molecular and Cellular Biology* *28*, 1393 -1403.

Zlatanova, J., et Thakar, A. (2008). H2A.Z: View from the Top. *Structure* *16*, 166-179.

Zobeck, K. L., Buckley, M. S., Zipfel, W. R., et Lis, J. T. (2010). Recruitment Timing and Dynamics of Transcription Factors at the Hsp70 Loci in Living Cells. *Molecular Cell* *40*, 965-975.





## **Chapitre 10**

### **Liste des figures et des tableaux**



## Liste des figures

FIGURE 1 : DIFFERENTS ELEMENTS CONSTITUANTS LES PROMOTEURS DES GENES. ....	23
FIGURE 2 : ARCHITECTURE DES COMPLEXES CO-ACTIVATEURS DE LA TRANSCRIPTION SAGA ET MEDIATEUR. ....	24
FIGURE 3 : ASSEMBLAGE SEQUENTIEL DU COMPLEXE DE PRE-INITIATION DE LA TRANSCRIPTION AU NIVEAU DES PROMOTEURS .....	27
FIGURE 4 : STRUCTURE DE L'ARN POLYMERASE II DE LEVURE .....	28
FIGURE 5 : ORGANISATION DE L'ARN POLYMERASE II DANS LE COMPLEXE DE PRE-INITIATION .....	29
FIGURE 6 : PASSAGE EN PHASE D'ELONGATION PRODUCTIVE DE LA TRANSCRIPTION .....	31
FIGURE 7 : PHOSPHORYLATION DU CTD DE L'ARN POLYMERASE II AU COURS DE LA TRANSCRIPTION.....	35
FIGURE 8 : MECANISME D'EXPORT DES ARN MESSAGERS .....	37
FIGURE 9 : INTERCONNEXION ENTRE LES MECANISMES DE TRANSCRIPTION, MATURATION, EXPORT ET SURVEILLANCE DES ARN MESSAGERS .....	38
FIGURE 10 : ORGANISATION DE LA CHROMATINE DANS LE NOYAU DES CELLULES EUCARYOTES. ....	39
FIGURE 11 : ARCHITECTURE ET STRUCTURE D'UNE PROTEINE HISTONE, FORMATION DE LA PAIRE H3/H4. ....	40
FIGURE 12 : STRUCTURE TRIDIMENSIONNELLE DU NUCLEOSOME .....	42
FIGURE 13 : ORGANISATION DE L'HETERO- ET L'EUCHROMATINE DANS LES NOYAUX DES CELLULES .....	43
FIGURE 14 : LES DIFFERENTS SITES DE MODIFICATION DES HISTONES CANONIQUES .....	45
FIGURE 15 : MODIFICATIONS DES HISTONES SUR LA PHASE CODANTE DES GENES AU COURS DE LA TRANSCRIPTION .....	47
FIGURE 16 : ARCHITECTURE GLOBALE DES VARIANTS D'HISTONES .....	48
FIGURE 17 : ARCHITECTURE DES SOUS-UNITES CATALYTIQUES DES QUATRE FAMILLES DE COMPLEXE DE REMODELAGE DE LA CHROMATINE ET MECANISME DE REMODELAGE.....	50
FIGURE 18 : IMPLICATION DES CHAPERONNES D'HISTONES FACT ET ASF1 AU COURS DE LA REPLICATION .....	52
FIGURE 19 : STRUCTURE DU COMPLEXE CHAPERONNE D'HISTONE ASF1/H3/H4.....	54
FIGURE 20 : ROLE DE CHAPERONNE D'HISTONE DE LA PROTEINE SPT6 .....	60
FIGURE 21 : COLOCALISATION DE SPT6 AVEC L'ARN POLYMERASE II .....	62
FIGURE 22 : IMPLICATION DE SPT6 ET IWS1 DANS LA TRANSCRIPTION, LA REGULATION EPIGENETIQUE, L'EXPORT DES ARNm ET LA SURVEILLANCE.....	66
FIGURE 23 : ARCHITECTURE DES PROTEINES SPT6 ET IWS1 .....	69
FIGURE 24 : PRINCIPE DE LA REACTION DE TOPO TA-CLONING .....	76
FIGURE 25 : PRINCIPE DE LA REACTION LR D'UN CLONAGE GATEWAY.....	80
FIGURE 26 : PROCEDURE DE CONCATENATION DE DEUX GENES DANS UN VECTEUR D'EXPRESSION. ....	81
FIGURE 27 : GENERATION D'UN BACMIDE PAR LA METHODE « BAC TO BAC » .....	85
FIGURE 28 : LE SYSTEME D'EXPRESSION BACTERIEN PET. ....	86
FIGURE 29 : REPRESSION DE L'OPERON LACTOSE EN PRESENCE DE GLUCOSE OU REPRESSION DES CATABOLITES. ....	87
FIGURE 30 : LES DIFFERENTES ETAPES DES TESTS DE CO-EXPRESSION, (ADAPTE DE VIJAYACHANDRAN ET AL., 2011) .....	91
FIGURE 31 : INTERPRETATION DES RESULTATS DES TESTS D'EXPRESSION. ....	95
FIGURE 32 : MONTAGE D'UNE EXPERIENCE DE MICROCALORIMETRIE. ....	107

FIGURE 33 : RESULTATS TYPIQUES D'UNE EXPERIENCE DE MICROCALORIMETRIE.....	108
FIGURE 34 : PRINCIPE DE L'EXPERIENCE DE THERMOFLUOR. ....	109
FIGURE 35 : DIAGRAMME DE PHASE A DEUX DIMENSIONS .....	114
FIGURE 36 : SCHEMA DES DISPOSITIFS DE GOUTTE ASSISE ET DE GOUTTE SUSPENDUE.....	115
FIGURE 37 : ETAT DES SPINS NUCLEAIRES EN ABSENCE ET EN PRESENCE D'UN CHAMP MAGNETIQUE. ....	123
FIGURE 38 : MESURE ET EXPLOITATION DU PHENOMENE DE RELAXATION .....	123
FIGURE 39 : STRATEGIE D'ASSIGNEMENT SEQUENTIEL .....	127
FIGURE 40 : REPLIEMENT DU DOMAINE SH2 DE LA PROTEINE SRC ET INTERACTION AU PEPTIDE PHOSPHORYLE. ....	134
FIGURE 41 : INTERACTION ENTRE LA PHOSPHOTYROSINE ET LA POCHES DE FIXATION DU DOMAINE SH2 DE LA PROTEINE SRC. ....	136
FIGURE 42 : CRISTAUX DU DOMAINE CARBOXY-TERMINAL DE LA PROTEINE SPT6 D'ANTONOSPORA LOCUSTAE .....	139
FIGURE 43 : CO-EXPRESSION DES VECTEURS GST-CTD AVEC LES PROTEINES Pcf11 ET SPT6.....	154
FIGURE 44 : ALIGNEMENT DE SEQUENCES DE L'EXTREMITÉ CARBOXY-TERMINAL DE SPT5 (ADAPTE DE PFAM).....	156
FIGURE 45 : INTERACTION SPT6/SPT5 - EXPERIENCES DE PULLDOWN .....	157
FIGURE 46 : INTERACTION SPT6/SPT5 – TESTS DE CO-EXPRESSION .....	158
FIGURE 47 : ALIGNEMENT DES SEQUENCES ET SUPERPOSITION DES STRUCTURES DE DOMAINES SH2 DE LA PROTEINE SPT6.....	161
FIGURE 48 : DIFFERENTS MODE DE LIAISON ENTRE LE TANDEM SH2 DE SPT6 ET LE CTD PHOSPHORYLE DE L'ARN POLYMERASE II. ....	165
FIGURE 49 : ORGANISATION DE LA PROTEINE IWS1 .....	172
FIGURE 50 : FORMES CRISTALLINES OBTENUES LORS DE LA RESOLUTION DE LA STRUCTURE DU COMPLEXE SPT6/IWS1 PAR DIFFRACTION DES RAYONS X.....	172
FIGURE 51 : CO-EXPRESSION DE SPT6 AVEC DES MUTANTS DE SURFACE DE IWS1.....	188
FIGURE 52 : TESTS DE L'INTERACTION IWS1/ARN POLYMERASE II – EXPERIENCES DE PULLDOWN. ....	191
FIGURE 53 : INTERACTION ENTRE IWS1 ET DES PEPTIDES PHOSPHORYLES DU CTD DE L'ARN POLYMERASE II – EXPERIENCES DE THERMOFLUOR. ....	192
FIGURE 54 : ALIGNEMENT DES SEQUENCES ET DES STRUCTURES DES PROTEINES IWS1 DONT LES STRUCTURES ONT ETE RESOLUES... ..	195
FIGURE 55 : ALIGNEMENT DES SEQUENCES ET DES STRUCTURES DES COMPLEXES SPT6/IWS1 DONT LES STRUCTURES ONT ETE RESOLUES.....	196
FIGURE 56 : MOTIFS HYDROPHOBES DE L'EXTREMITÉ AMINO-TERMINALE DE SPT6 .....	205
FIGURE 57 : TEST DE L'INTERACTION DE IWS1 AVEC LES MOTIFS HYDROPHOBES DE L'EXTREMITÉ AMINO-TERMINALE DE SPT6 .....	207
FIGURE 58 : INTERACTION DE TFIIS AVEC LES MOTIFS HYDROPHOBES DE L'EXTREMITÉ AMINO-TERMINALE DE SPT6. ....	209
FIGURE 59 : SPECIFICITE DES COMPLEXES SPT6/TFIIS ET SPT6/IWS1 - INTERACTION DE TFIIS ET IWS1 AVEC UN FRAGMENT MUTE DE SPT6.....	210
FIGURE 60 : CO-EXPRESSION DU COMPLEXE SPT6 IF5/IWS1/TFIIS.....	212
FIGURE 61 : CARACTERISATION BIOCHIMIQUE DES COMPLEXES SPT6 IF3/TFIIS ET SPT6 IF5/TFIIS. ....	214
FIGURE 62 : ANNOTATION DU SPECTRE 15N-HSQC DU COMPLEXE SPT8/TFIIS. ....	217
FIGURE 63 : MODELE DE LA STRUCTURE DU DOMAINE AMINO-TERMINAL DE TFIIS RESOLUE PAR RMN .....	218
FIGURE 64 : RELAXATION T1/T2 DU COMPLEXE SPT8/TFIIS.....	219
FIGURE 65 : PHENOTYPAGE DES DIFFERENTS MUTANTS DE SPT6 .....	222

FIGURE 66 : RECONSTITUTION DU NUCLEOSOME A PARTIR DE L'OCTAMERE CO-EXPRIME.....	241
FIGURE 67 : CO-EXPRESSION DU DOMAINE AMINO-TERMINALE DE ASF1 ET DES HISTONES .....	242
FIGURE 68 : CO-EXPRESSION DE LA PROTEINE SPT6 AVEC LA PAIRE D'HISTONE H3/H4 ET L'OCTAMERE D'HISTONE. ....	244
FIGURE 69 : ORGANISATION DE LA PROTEINE SPT6.....	249
FIGURE 70 : REPRESENTATION SCHEMATIQUES DES DIFFERENTES INTERACTIONS MISES EN PLACE PAR SPT6 ET LE COMPLEXE SPT6/Iws1 .....	252
FIGURE 71 : CARTE DES DEUX PRINCIPAUX VECTEURS D'EXPRESSION BACTERIENS UTILISES. ....	263
FIGURE 72 : CARTE DES DEUX VECTEURS BACULOVIRUS UTILISES .....	264



## Liste des tableaux

TABLEAU 1 : GENES DE LA FAMILLE SPT (ADAPTE DE YAMAGUCHI ET AL., 2001).....	56
TABLEAU 2: LES DIFFERENTS VECTEURS NAVETTES EMPLOYES ET LEURS PRINCIPALES CARACTERISTIQUES. ....	75
TABLEAU 3 : LES VECTEURS DE CO-EXPRESSION EMPLOYES. ....	78
TABLEAU 4 : COMPOSITION DES DIFFERENTS MILIEUX DE CULTURE UTILISES. ....	88
TABLEAU 5 : LES DIFFERENTS TYPES DE MILIEU EMPLOYES POUR LA REALISATION DES TESTS D'EXPRESSION .....	93
TABLEAU 6 : PRINCIPALES ETIQUETTES DE PURIFICATION UTILISEES.....	101
TABLEAU 7 : SPECTRES TRIDIMENSIONNELLES MESURES POUR L'ATTRIBUTION TRIDIMENSIONNELLE DES SPECTRES RMN .....	126
TABLEAU 8 : DIFFERENTES FAMILLES DE FAMILLES DE DOMAINES SH2 ET LEURS CARACTERISTIQUES. ....	135
TABLEAU 9 : DIFFERENTES CONSTRUCTIONS DE L'EXTREMITÉ CARBOXY-TERMINALE DE L'ARN POLYMERASE.....	153
TABLEAU 10 : VALEURS DE DEVIATIONS STANDARD ENTRE LES STRUCTURES DES DOMAINES SH2. ....	160
TABLEAU 11 : MESURE DES CONSTANTES DE DISSOCIATION ENTRE LES TANDEMS SH2 ET DE PEPTIDES PHOSPHORYLES MIMANT LE CTD DE L'ARN POLYMERASE II ( $\mu\text{M}$ ). ....	163
TABLEAU 12 : SEQUENCE DES PEPTIDES UTILISES POUR LES EXPERIENCES DE MICROCALORIMETRIE ET THERMOFLUOR ENTRE IWS1 ET LE CTD DE L'ARN POLYMERASE II. ....	192
TABLEAU 13 : PHENOTYPES ASSOCIES AUX MUTATIONS DE SPT6 ET IWS1. ....	198
TABLEAU 14 : CARACTERISATIONS FONCTIONNELLES DE L'EXTREMITÉ AMINO-TERMINALE DE SPT6 – MUTATIONS PONCTUELLES INTRODUITES. ....	221
TABLEAU 15 : PRINCIPAUX PHENOTYPES CAUSES PAR LA MUTATION DES MOTIFS HYDROPHOBES DE L'EXTREMITÉ AMINO-TERMINALE DE SPT6.....	222
TABLEAU 16 : MELANGE REACTIONNEL DE LA PCR.....	259
TABLEAU 17 : CYCLES DE LA REACTION DE PCR.....	260
TABLEAU 18 : SOUCHES BACTERIENNES UTILISEES AU COURS DE CETTE ETUDE. ....	262



## Etude biochimique, structurale et fonctionnelle du complexe chaperonne d'histone/facteur d'élongation Spt6/Iws1

Les ARN messagers (ARNm) fonctionnels sont produits au cours d'un mécanisme complexe qui allie la transcription, la maturation du transcrit et son export. De plus, ces différentes machineries vont devoir faire face à la structure compacte de la chromatine, nécessitant une activité de décondensation/recondensation de la chromatine qui est notamment régulée par les mécanismes épigénétiques. Un très grand nombre de facteurs sont donc requis pour la production des ARNm fonctionnels. Parmi ces facteurs, les protéines Spt6 et Iws1 sont impliquées dans le mécanisme général de la transcription, dans la modulation de la structure de la chromatine et la maturation et l'export des ARNm. Ces travaux de thèse ont permis de caractériser biochimiquement, structuralement et fonctionnellement ces deux protéines, leur complexe et leur interaction avec d'autres effecteurs de la transcription. Ces travaux ont notamment permis de comprendre en termes moléculaires et fonctionnels (i) comment Spt6 est recrutée par l'ARN polymérase II au cours de la transcription et (ii) comment le complexe Spt6/Iws1 est formé. Ils ont également permis d'identifier de nouveaux interactants potentiels de Spt6, et notamment le facteur d'élongation de la transcription TFIIS. Ces travaux ont ainsi permis de révéler le rôle essentiel et extrêmement complexe joué par Spt6 et Iws1 lors de la production d'un ARNm, mais également de permettre l'étude future de leur interaction avec d'autres facteurs transcriptionnels.

*Mots clés : Transcription, ARN polymérase II, Elongation, Epigénétique, mRNA export, Histones*

Production of functional messenger RNA (mRNA) requires a complex mechanism that couples transcription with maturation and export of the mRNA. In addition to this mechanism, chromatin needs to be unwound to allow the transcription machinery access the DNA, this unwinding being also highly regulated. Thus, production of a functional mRNA requires a huge number of factors implicated in these different processes. Among these proteins Spt6 and Iws1 are participating in the mechanism of transcription, chromatin unwinding, and maturation and export of the mRNA. The work carried out during this thesis has enabled the biochemical, structural and functional characterization of these proteins, their complex and their interaction with other effectors of transcription. This work has specifically enabled the molecular and functional characterization (i) of the recruitment of Spt6 by RNA polymerase II and (ii) of the formation of the Spt6/Iws1 complex. Moreover, this work has identified putative new partners of Spt6, notably the elongation factor TFIIS. Thus, our work has highlighted the essential and complex role of Spt6 and Iws1 during the production of functional mRNA, and has also enabled future studies of the complexes formed by these two proteins with other transcriptional factors.

*Keywords: Transcription, RNA polymerase II, Elongation, Epigenetic, mRNA export, Histones*