



# THÈSE DE DOCTORAT DE L'UNIVERSITÉ BORDEAUX 2 ET DE L'UNIVERSITÉ DE SARAGOSSE

**Mention :** Sciences, Technologie et Santé

**Option :** Biochimie

**Présentée et soutenue publiquement :**

*Le 18 novembre de 2010*

**Présentée par:**

*María MORENO MORCILLO*

## **Etude structurale du complexe CstF et de son homologue chez la levure CF IA, deux facteurs indispensables pour la maturation 3' des pré- ARN messagers**

### **Membres du Jury**

Christophe CULLIN .....	Président du jury
Juan HERMOSO DOMINGUEZ .....	Rapporteur
Domenico LIBRI .....	Rapporteur
Eric ENNIFAR .....	Examineur
Marta MARTINEZ JULVEZ .....	Directrice de thèse
Sébastien FRIBOURG .....	Directeur de thèse





## Remerciements

L'aboutissement de ces travaux n'est pas uniquement résultat d'un travail individuel, beaucoup de gens ont participé d'une manière ou d'une autre à cette thèse permettant qu'elle devienne un jour réalité. Je tiens donc à remercier toutes ces personnes en espérant qu'elles ont gardé les mêmes bons souvenirs que moi j'ai conservé de chacun d'entre eux.

Je voudrais commencer en remerciant chaleureusement les rapporteurs et les examinateurs de cette thèse car ils ont investi beaucoup de temps pour lire attentivement le manuscrit. Le Pr. Christophe Cullin pour avoir accepté de présider ce jury, les Dr. Juan A. Hermoso et Dr. Domenico Libri pour avoir fait le déplacement et accepté de rapporter ce travail. Enfin, le Dr. Eric Ennifar pour avoir pris le soin d'examiner en détail mon travail. J'ai eu la chance d'avoir d'un jury compréhensif qui m'a aidé énormément faisant de la soutenance une discussion intéressante et très agréable. Merci également pour vos encouragements pour que je continue dans cette voie.

Je tiens à remercier les directeurs de l'institut qui m'ont accueilli pendant mes 3 ans de thèse et mon année de stage, les Drs. Jean-Jacques Toulmé, Ivan Huc et Léon Ghosez, codirecteurs et ancien codirecteur de l'institut. Mes travaux ont été réalisés au sein de l'unité U869 dirigé également par le Dr. Jean-Jacques Toulmé qui a été d'un grand soutien.

Un grand merci aussi à la fondation « La Caixa » et à l'ARC (« l'Association pour la recherche sur le cancer ») qui m'ont donné la possibilité en me finançant de réaliser cette thèse dans les meilleurs conditions.

Ce travail de thèse a été réalisé en cotutelle entre l'Université de Bordeaux 2 et l'Université de Saragosse. Dans cette dernière, je voudrais remercier l'implication de ma directrice de thèse, la Dr. Marta Martínez Júlvez, qui m'a aidé autant que la distance le permettait. Le travail de laboratoire a été effectué dans l'équipe de Cristallographie des Complexes Macromoléculaires, à l'Institut Européen de Chimie et de Biologie (IECB) sous la direction du Dr. Sébastien Fribourg. Mes plus profonds remerciements sont adressés à Sébastien, pour m'avoir accepté dans son équipe et pour m'avoir montré ce domaine de la recherche. Merci pour tout ce que tu m'as apporté ces 4 dernières années, tes conseils, ton expérience et ton aide. Dans des nombreuses occasions nous avons discuté des erreurs que nous avons commises ensemble, cependant, il faut dire aujourd'hui que tout ce qui a été réussi a été aussi un travail d'équipe, tu a toujours été présent et tu as cru en moi dès le début, merci donc pour ta confiance, ça m'a fait évoluer énormément professionnellement et personnellement.

Mes remerciements sont adressés également aux gens qui on participé à ce projet. A l'IECB j'ai eu la chance de collaborer avec Cameron Mackereth qui a pris le temps de m'initier au monde de la spectrométrie RMN, merci pour ta patience et ta bonne humeur. Je tiens à remercier Hélène Dumay-Odelot, pour son soutien, sa disponibilité et son aide précieuse, Carmelo Di Primo, d'avoir partagé son savoir sur la SPR, Brice Kauffmann pour son aide en cristallographie et Katel Bathany pour les expériences réalisés en spectrométrie de masse. En dehors de l'IECB, je remercie Pierre Legrand, toujours disponible pour m'aider dans la résolution des structures cristallographiques depuis le synchrotron SOLEIL et le synchrotron ESRF (à Grenoble) où nous avons collecté la plupart des données cristallographiques.

Je tiens à exprimer toute ma gratitude à tous les gens qui sont passés dans l'équipe. Je voudrais tout d'abord remercier Noelio d'avoir été là depuis le début, je n'oublierai jamais ce que tu as fait pour moi. Merci à Estefano, avec qui j'ai partagé tous nos soucis de thésards et plein de délires au labo, et merci aux autres anciens membres Fred garçon, Cécile, Lynn, Anna. Merci Lionelio pour ta bonne humeur et pour nos discussion scientifiques (et moins scientifiques...). Un grand merci aux filles, Cécile et Natacha, ensemble nous avons fait du labo un endroit où l'on a envie d'arriver le matin, j'ai adoré les fous rires et nos conversations pendant les petites pauses. Et j'ai laissé délibérément pour la fin ma grande amie Freda, depuis le début tu as été beaucoup plus qu'une collègue, toujours là pour moi, au boulot et en dehors. Cette thèse n'aurait pas été possible sans toi, tu as eu la patience de corriger chaque mot, d'abord au niveau du français et ensuite en te servant de ton énorme qualité scientifique et je te remercie infiniment. Je ne sais pas si j'ai mérité ton soutien inconditionnel, mais c'est ça qui m'a fait tenir debout quand je n'en pouvais plus. Réaliser cette thèse valait déjà la peine du simple fait de t'avoir rencontré.

Je suis heureuse de partir de l'IECB en ayant rencontré de gens formidable avec qui je partage une profonde amitié. Parmi eux, ma « communauté hispanique ». *Gracias Rosina por tu tiempo, llegaste tarde pero entraste fuerte...* Marisa et Jone qui ont su m'écouter depuis le premier jour et avec qui j'ai partagé des innombrables moments incroyables, *nunca podré agradecer todo lo que habéis hecho por mí, compartir esto con vosotras ha sido un regalo, creo que las palabras aquí se quedarían cortas*. Je remercie également Margot qui a été une des premières à me faire sentir chez moi à Bordeaux. Et un grand merci à Vanou, tu as été à mes côtés du début à la fin, il y a eu des moments très durs mais ils ont été plus faciles à surmonter grâce à toi, donc merci pour tous nos bons moments.

A l'IECB le nombre de gens que je veux remercier est trop important pour citer tout le monde. J'ai partagé des repas, des apéros, des discussions, des soirées et des moments inoubliables avec des gens qui ont été ou qui se trouvent toujours dans l'institut et ils ont fait que ces quatre années de thèse passent très vite, avec plein des bons souvenirs... Merci à tous les chimistes, aux « Quideau's » (Marion, Gaëlle, Céline, Tony, Rémy, Anna, les deux Mélanies, Séverine, Delphine, Anne-Laure...), Oscar (*Oscarito... no pierdas esa sonrisaca permanente que te caracteriza*), Tracy, Elisabeth, Simon et tous ce qui m'ont accompagné ces années dans la « salle de pause », merci pour votre bonne humeur. Merci Denis pour ton soutien, à Karine et Ilyass, à Martin Teichmann et son groupe pour leurs conseils, aux gens du premier (Samir, Yoann, Sara, Benoit, Patricia, Clothilde, Julien, Sabrina...), au groupe de RMN (Vania, Anna, Olivier et Cécile et Axelle particulièrement), à Stéphane Chaignepain, Wissamito, Fred Guillem, Sébastien Buchoux... Je voulais remercier également l'administration, spécialement Annie, Sandra, Stéphanie, Kati et Patricia.

Un grand merci aux gens de bordeaux, externes à l'IECB, ça a été grâce à vous que j'ai pu m'évader dans les moments difficiles remplissant ainsi ma thèse de plein de bons souvenirs. Merci spécialement à Ady, *por ver y dejarme ver lo mejor de mí, mil gracias a « los vecinos » (Daniel y Cris) a los que quiero un resto, gracias a Caro (y Pilucho) por vuestro apoyo y por las innumerables conversaciones detrás de una taza té, a mi gente de flamenco que me ha dado tanto*. Merci à tous ces gens avec qui j'ai passé des soirées incroyables, Benji (même si pour tes voisins c'était moins incroyable), James, Emilie, Fred, Astrid... *Y gracias a los que han sabido transmitir todo su apoyo más allá de las fronteras : a los albaceteños, con una mención especial a Cris, mi amiga del alma, fer, tere, rodri, barby, carletes, los gemelos y el resto de la panda, a Blanca, Palomator, a Fiona « zanahoria », a mi truchita madrileña que siempre me comprende, gracias a Oscar que me animó a comenzar esta aventura, a mi hermanita de Almería que ha tenido que aguantar muchas de mis « crisis existenciales », a mis murcianicos (Ana y Diego, que estuvieron presentes más de lo nunca imaginaron, a Juanu, Pablo, Javichu...).*

*La palabra « gracias » queda demasiado corta para demostrar lo que le debo a mi familia. Sobre todo en momentos como estos me doy cuenta de la inmensa suerte que tengo de poder contar con ellos, y lo mejor es que siempre fue así y que siempre lo será. Mi mayor placer al recibir este título es ver en vuestra cara lo orgullosos que os sentís de mí, pues sabed que este escrito también es vuestro, que la mayoría de las cosas que me han permitido llegar hasta aquí me las habéis dado o enseñado vosotros. Soy yo la que se siente orgullosa de pertenecer a esta familia tan maravillosa. Gracias por vuestro apoyo tan incondicional.*

Enfin, je tiens à remercier la personne qui m'a "supportée" le plus pendant tout ce temps, qui a été présente dans les meilleurs et les pires moments (j'en ai eu pas mal...) de cette thèse. Thomas, je n'ai jamais rencontré quelqu'un capable de faire tout ce que tu as fait pour moi. Ta seule présence me rassure et je me sens protégée. J'ai eu la chance de t'avoir à mes côtés et de compter sur ton aide, non seulement dans notre vie personnelle, mais aussi (et les gens de mon groupe le savent...!) dans ma vie professionnelle. Tu as toujours su me donner une priorité sur tout le reste, quoi qu'il arrive. Je ne pourrais jamais te remercier suffisamment, tu dirais que c'est très « cursi », mais je pense vraiment que le monde devrait être rempli des personnes comme toi pour qu'il en vaille plus la peine. *Mil gracias MPCM.*

# Table de matières

<b>CHAPITRE I : INTRODUCTION</b> .....	<b>7</b>
A. PROCESSUS DE MATURATION DE L'EXTREMITÉ 3' DES ARN MESSAGERS.....	7
<b>1. Contexte scientifique du processus de maturation 3' des pré-ARNms</b> .....	<b>8</b>
a.) Les premiers pas dans la connaissance du processus.....	8
b.) Importance de la modification post-transcriptionnelle sur l'extrémité 3' du pré-ARNm.....	9
c.) Étapes de la maturation 3'.....	12
d.) Éléments en cis impliqués dans le processus de maturation 3'.....	14
e.) Protéines impliquées dans la modification sur l'extrémité 3' du pré-ARNm.....	19
f.) Implication de la maturation 3' dans différentes maladies.....	48
g.) Maturation 3' des pré-ARNms d'histones.....	50
h.) Maturation chez les procaryotes.....	52
<b>2. Connexion entre la maturation 3' et d'autres processus</b> .....	<b>53</b>
a.) Rôle du CTD.....	54
b.) Connexion entre le processus de maturation 3' et de transcription.....	56
c.) Connexion entre la modification post-transcriptionnelle en 3' et l'épissage.....	59
d.) Connexion entre la formation de la queue poly(A) et l'export des ARNms vers le cytoplasme... ..	61
e.) Connexion entre la formation de la queue poly(A) et la traduction.....	64
B. OBJECTIFS.....	65
<b>CHAPITRE II : RESULTATS OBTENUS POUR LE COMPLEXE CSTF</b> .....	<b>67</b>
A. ETUDE STRUCTURALE DU DOMAINE N-TERMINAL DE LA SOUS-UNITE CSTF-50.....	67
<b>1. Etude de la région CstF-50 (1-92) de <i>Drosophila melanogaster</i></b> .....	<b>68</b>
a.) Identification du domaine pour la réalisation des travaux.....	68
b.) Expression et purification de CstF-50 (1-92) de drosophile.....	70
c.) Etude structurale de CstF-50 dm (1 - 92).....	71
<b>2. Etude structurale de CstF-50 (1-65) de drosophile</b> .....	<b>72</b>
a.) Clonage, expression et purification de la protéine CstF-50dm (1-65).....	72
b.) Comparaison des constructions CstF-50dm (1-92) et CstF-50dm (1-65).....	74
c.) Cristallisation de CstF-50dm (1-65).....	75
d.) Résolution de la structure de CstF-50dm (1-65).....	76
e.) Le domaine N-terminal de CstF-50 forme un homodimère conservé.....	78
f.) Conservation de la surface du dimère de CstF-50dm (1-65).....	82
g.) Etudes de l'interaction de CstF-50 avec le CTD de l'ARN pol II.....	82
<b>3. Conclusions et perspectives</b> .....	<b>93</b>
B. INTERACTIONS DANS LE COMPLEXE CSTF.....	95
<b>1. Etude des domaines impliqués dans l'interaction entre CstF-77 et CstF-64 ..</b> <b>95</b>	
a.) Travaux préliminaires : étude du domaine « hinged » de CstF-64.....	95
b.) Identification des régions impliquées dans l'interaction entre CstF-77 et CstF-64.....	97
c.) Analyse du complexe CstF-77hs (626-717) / CstF-64hs (1-211).....	99
d.) Le complexe CstF-77hs (626-717) / CstF-64hs (1-211) interagit avec un d'ARN riche en G/U....	102
e.) Essais de cristallisation du complexe CstF-77hs (626-717) / CstF-64hs (1-211) / ARN.....	103
<b>2. Conclusions et perspectives</b> .....	<b>107</b>

<b>CHAPITRE III : RESULTATS OBTENUS POUR LE COMPLEXE CF IA .....</b>	<b>110</b>
A. ETAT DE L'ART : COMPARAISON DES COMPLEXES CSTF ET CF IA.....	110
B. ANALYSE STRUCTURALE ET FONCTIONNELLE DU COMPLEXE CF IA.....	112
1. <i>Etude structurale des domaines d'interaction des sous-unités Rna15p et Rna14p</i> .....	<b>112</b>
a.) Etudes précédentes réalisées sur l'interaction entre Rna14p et Rna15p.....	112
b.) Identification des régions de Rna14p et de Rna15p impliquées dans l'interaction .....	112
c.) Etudes préliminaires : purification et essais de cristallisation des complexes formés entre différentes régions de Rna14p et Rna15p .....	117
d.) Structure des domaines d'interaction des protéines Rna14p et Rna15p .....	121
e.) L'association hinge/monkeytail est nécessaire pour l'assemblage de Rna15p au complexe CF IA	127
f.) Conclusions et perspectives pour l'interaction Rna15p/Rna14p.....	129
2. <i>Reconstitution du complexe CF IA</i> .....	<b>132</b>
a.) Etudes préliminaires : purification du complexe natif .....	133
b.) Reconstitution du complexe CF IA 'in vitro': tests d'expression.....	134
c.) Pcf11p interagit avec toutes les sous-unités du complexe CF IA ? .....	135
d.) Purification du complexe CF IA à partir des protéines recombinantes .....	136
e.) Conclusion et perspectives pour la reconstitution 'in vitro' du complexe CF IA .....	138
<b>CHAPITRE IV : DISCUSSION .....</b>	<b>141</b>
A. MODELES STRUCTURAUX DES COMPLEXES CSTF ET CF IA.....	141
B. INTERPRETATION ET ANALYSE DES MODELES STRUCTURAUX : MODELE FONCTIONNEL .....	147
<b>CHAPITRE V : CONCLUSIONS ET PERSPECTIVES.....</b>	<b>157</b>
<b>CAPÍTULO V : CONCLUSIONES Y PERSPECTIVAS .....</b>	<b>161</b>
<b>CHAPITRE VI : MATERIELS ET METHODES.....</b>	<b>165</b>
A. PROTOCOLES DE BIOLOGIE MOLECULAIRE .....	165
1. <i>Le clonage</i> .....	<b>165</b>
a.) Préparation de l'insert par PCR.....	165
b.) Préparation du vecteur .....	167
c.) Ligation de l'insert et le vecteur .....	168
d.) Préparation de l'ADN plasmidique.....	169
e.) Application des techniques du clonage pour la génération de mutants .....	169
f.) Un cas particulier : les plasmides bicistroniques .....	171
2. <i>Systèmes d'expression</i> .....	<b>172</b>
a.) Expression chez <i>Escherichia coli</i> .....	172
b.) Un autre système d'expression : TNT, traduction 'in vitro' .....	174
3. <i>Etudes d'interaction</i> .....	<b>175</b>
a.) Western Blot .....	175
b.) Double hybride.....	177
B. PROTOCOLES DE BIOCHIMIE .....	181

<b>1. Tests d'expression .....</b>	<b>181</b>
<b>2. Production de protéines à grande échelle .....</b>	<b>182</b>
a.) Production de protéines dans des conditions natives .....	182
b.) Production de protéines marquées avec du Se pour des études cristallographiques .....	182
c.) Production de protéines marquées avec du <sup>15</sup> N, <sup>13</sup> C pour des études en RMN .....	184
<b>3. Purification de protéines .....</b>	<b>184</b>
<b>4. Purification de l'ARN .....</b>	<b>187</b>
<b>C. TECHNIQUES BIOPHYSIQUES.....</b>	<b>190</b>
<b>1. Cristallographie aux rayons X .....</b>	<b>190</b>
a.) Cristallo-génèse .....	190
b.) Diffraction aux rayons X .....	193
<b>BIBLIOGRAPHIE .....</b>	<b>207</b>
<b>ANNEXES.....</b>	<b>233</b>

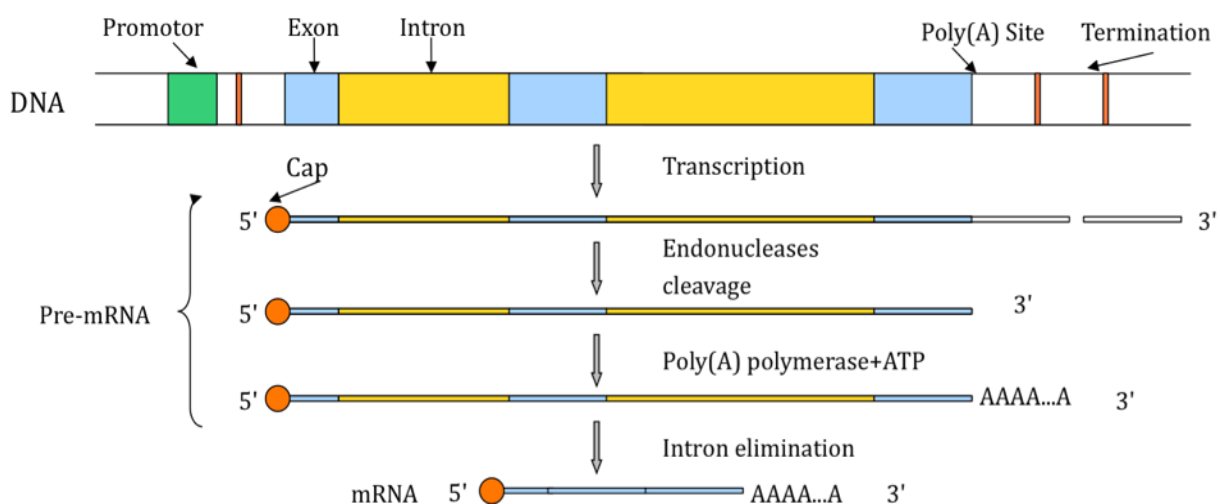


## Chapitre I : Introduction

### A. Processus de maturation de l'extrémité 3' des ARN messagers

Chez les eucaryotes, les précurseurs des ARN messagers (pré-ARNms) sont transcrits dans le noyau à partir de l'ADN grâce à l'ARN polymérase II. Ces pré-ARNms doivent subir des modifications avant leur export vers le cytoplasme où ils seront traduits en protéines. Ces modifications post-transcriptionnelles régulent l'expression génique spatio-temporelle des protéines.

Il existe trois étapes pour la formation de l'ARNm mature (*fig. 1*). D'abord se produit une **modification sur l'extrémité 5'** qui aboutira à la formation de la coiffe (Shatkin 1976). Cette modification implique le clivage de l'extrémité 5'-triphosphate de l'ARN précurseur par une triphosphatase (Spencer et al. 1978). Ensuite, une guanosine triphosphate est ajoutée grâce à l'action d'une guanylyl-transférase qui permet la formation d'un pont 5'-5' triphosphate (Venkatesan et al. 1980). La coiffe est enfin constituée grâce à la méthylation du carbone 7 sur la guanosine ajoutée et sur les 2'-hydroxyles des deux premiers riboses (Ensinger et al. 1975) (Barbosa and Moss 1978; Barbosa and Moss 1978). La modification sur l'extrémité 5' empêche la dégradation des ARNms par action des exonucléases 5' → 3' augmentant leur durée de vie (Furuichi et al. 1977) (Shimotohno et al. 1977). Elle intervient aussi dans l'export des ARNms vers le cytoplasme (Rottman et al. 1974) et joue un rôle dans le recrutement du ribosome au moment de la traduction (Both et al. 1975).



**Figure 1.** Etapes de la maturation des ARN pré-messagers. Les exons sont colorés en bleu, les introns en jaune. Le site poly(A) correspond au point de clivage et polyadénylation. La coiffe (« cap » en anglais) est représentée par un cercle rouge à l'extrémité 5' et la queue poly(A) par une série de « A » à l'extrémité 3'.

Pendant la **maturation 3'**, le pré-ARNm est clivé et une queue de poly(A) (poly Adénines), composée de 150 à 200 résidus adényliques chez les mammifères et de 55 à 90 chez la levure, est ajoutée à l'extrémité 3' (Brown and Sachs 1998) (Martin and Keller 1998).

La dernière modification post-transcriptionnelle est l'**épissage** durant laquelle les introns sont éliminés. L'épissage est assuré par un ensemble de ribonucléoprotéines (snRNP pour small nuclear ribonucleoprotein) appelé **splicéosome** (Maniatis and Reed 1987). Le splicéosome reconnaît des séquences nucléotidiques conservées situées dans les jonctions entre les exons et les introns et des séquences internes dans les introns appelés « boîtes de branchement ». Généralement, l'épissage est effectué en deux étapes. Tout d'abord une attaque nucléophile du 2'-OH du ribose de l'adénosine de la « boîte de branchement » a lieu sur le phosphate de la jonction exon-intron en 5'. Après cette coupure, le 3'-OH libéré au niveau de l'exon situé en amont établit une liaison avec le phosphate de la jonction intron-exon en aval. De cette façon, les produits obtenus dans la réaction sont d'une part les deux exons ligaturés et d'autre part, l'intron intermédiaire (Sharp 1985; Padgett et al. 1986) (Ruby and Abelson 1991) (Green 1991).

Les trois types de modification sont essentielles à la maturation des pré-ARNms et sont très connectées entre elles et avec le processus de transcription (*plus en détail dans le Chapitre I.A.2*). Les travaux de cette thèse sont centrés sur la modification post-transcriptionnelle des ARNms sur son extrémité 3'.

## ***1. Contexte scientifique du processus de maturation 3' des pré-ARNms***

### **a.) Les premiers pas dans la connaissance du processus**

L'identification d'une activité **poly(A) polymerase** (PAP) a été mise en évidence en **1960** (Edmonds and Abrams 1960) mais ce n'est que 10 ans plus tard que son rôle dans la maturation 3' a été révélé. Plus tard, la queue de poly(A) a été identifiée comme étant une modification post-transcriptionnelle de l'extrémité 3' des ARNms (Edmonds et al. 1971) (Darnell et al. 1971). Des études postérieures ont démontré qu'une étape de **clivage** de l'ARN précédait la formation de la **queue poly(A)** (Nevins and Darnell 1978). Cette découverte a permis de trouver la différence fondamentale entre le processus chez la bactérie et celui chez les mammifères. En effet, chez les bactéries, l'extrémité 3' des ARNms matures correspond exactement au site de terminaison de la transcription contrairement aux mammifères où un clivage et l'addition d'une queue de poly(A) ont lieu sur cette extrémité.

A partir de **1970-1980**, avec l'apparition des nouvelles technologies, la communauté scientifique prend conscience de la complexité et de l'importance de ce processus dans la régulation génique. Des nombreuses études ont été réalisées avec pour but d'élucider les détails du mécanisme de maturation 3'. Les premières évidences sur la fonction de la queue poly(A) se sont manifestées grâce à des expériences qui montraient la **stabilisation** par la queue de poly(A) des ARNms injectés



dans des œufs de grenouille (Huez et al. 1978). En effet, la polyadénylation sert à la protection des ARNm des nucléases. Postérieurement, d'autres rôles l'ont été attribués, comme sa participation dans la **traduction** (Sachs and Davis 1989), dans l'**export** des ARNm vers le cytoplasme (Huang and Carmichael 1996) ou dans l'**épissage** (Berget 1995).

Le premier élément de séquence d'ARNm identifiée dans la régulation du processus est **AAUAAA** (Proudfoot and Brownlee 1976). Des postérieures études ont démontré l'importance de cet région et sa conservation à travers les espèces (Fitzgerald and Shenk 1981) (Montell et al. 1983) (Wickens and Stephenson 1984). Dans les années 80, différents groupes de recherche ont travaillé sur la caractérisation biochimique des composants de la maturation 3' (Manley 1983) (Moore and Sharp 1984 ; Moore and Sharp 1985) et ces études ont mis en évidence que le clivage et l'addition d'adénines sont deux réactions couplées qui ont lieu grâce à la participation d'un grand complexe protéique.

Pendant les années qui ont suivi, les efforts se sont focalisés sur les détails du mécanisme et l'assignation des fonctions individuelles de chaque facteur. Un grand nombre des protéines et des éléments de séquence impliqués, leurs interactions et les connexions avec d'autres étapes de la maturation et de la transcription sont maintenant connues, et seront présentées ultérieurement.

#### **b.) Importance de la modification post-transcriptionnelle sur l'extrémité 3' du pré-ARNm**

La maturation 3' joue un rôle primordial dans l'expression génique. La queue poly(A) d'un ARNm mature est essentielle pour sa stabilité, sa translocation vers le cytoplasme et sa traduction (Bernstein and Ross 1989; Sachs 1990). De cette façon, les transcrits qui ne sont pas traités dans leur extrémité 3' vont être dégradés ou retenus dans le noyau. En outre, la maturation connecte les machineries impliquées dans les processus de transcription, modification 5' et épissage des ARNm. Elle participe aussi à la régulation de la quantité et la qualité de l'expression dans différents processus biologiques à travers la sélection de signaux alternatifs de polyadénylation.

Premièrement, la maturation 3' est importante pour la **stabilité de l'ARNm** (Bernstein and Ross 1989). Les niveaux d'ARNm dans le cytoplasme dépendent d'un équilibre entre la synthèse des ARNs, leur maturation, leur export et leur dégradation dans le cytoplasme. Il a été démontré que la protection de la région 3' par la queue poly(A) et par la protéine PABP (**poly(A) binding protein**) prévient l'ARN de sa dégradation (Bernstein and Ross 1989). En effet, dans le cytoplasme, l'ARNm est dégradé à partir de l'extrémité 3' par action de ribonucléases. Il a été montré que l'excision de la queue de poly(A) précède à la dégradation de l'ARNm (Sachs and Wahle 1993). Cette dégradation peut commencer pendant le processus de traduction par un mécanisme dépendent de la déadénylation. Dans ce mécanisme, la queue de poly(A) est d'abord raccourci par action de poly(A) ribonucléases (**PARN** chez les mammifères) (Korner and Wahle 1997; Korner et al. 1998).

Ensuite, la coiffe serait éliminée (Couttet et al. 1997) et l'ARNm reste accessible pour des exonucléases 5'→3' ou 3'→5' (Guhaniyogi and Brewer 2001).

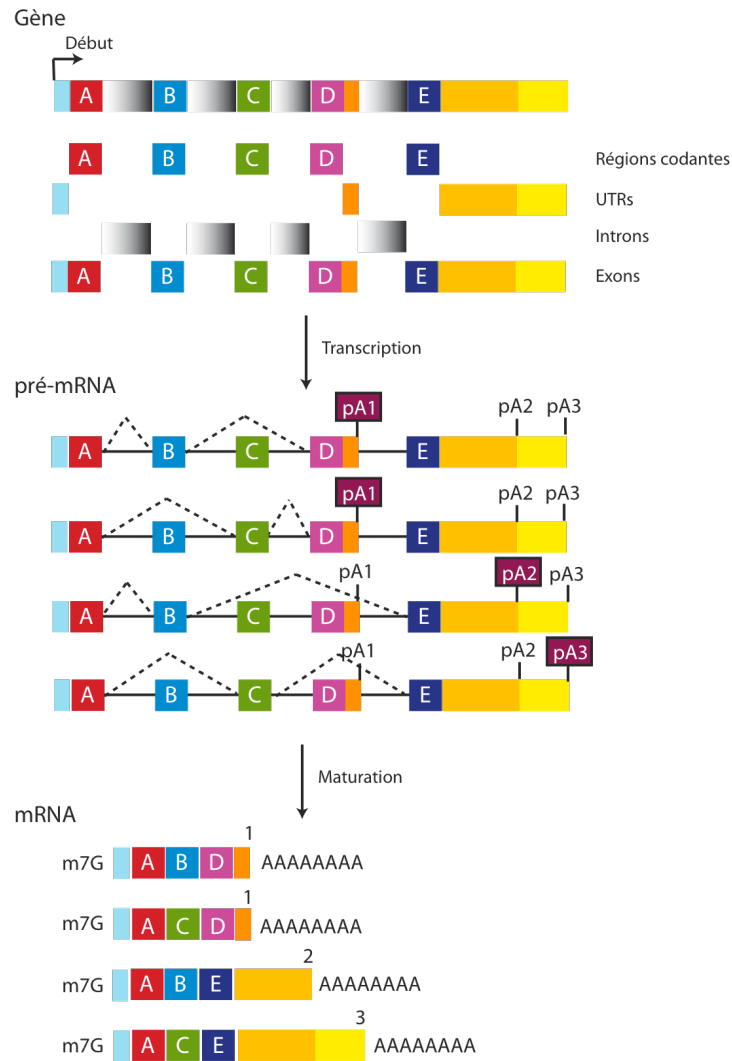
D'un autre côté, la maturation 3' promeut le **transport des ARNms du noyau vers le cytoplasme** (Huang and Carmichael 1996). Il a été démontré que la modification sur l'extrémité 3' est nécessaire et suffisante pour l'export (Dower and Rosbash 2002). De plus, Pab1p et Nab2p, des facteurs qui interviennent dans le contrôle de la queue poly(A) chez la levure, contribuent aussi à l'export des ARNms (Hammell et al. 2002) (Hector et al. 2002) (Vinciguerra and Stutz 2004) (*voir aussi Chapitre I.A.2.d*). En effet, le ratio entre l'ARNm présent dans le cytoplasme et celui présent dans le noyau diminue de 10 fois quand le pré-ARNm est clivé mais ne contient pas la queue poly(A). Ceci indique que l'ARNm est retenu dans le noyau en absence de la queue d'adénines (Huang and Carmichael 1996). A l'appui de cette théorie, il a été montré que des mutations sur les éléments (en *cis* et en *trans*) impliqués dans la maturation 3' inhibent l'export des ARNms matures vers le cytoplasme (Hammell et al. 2002).

Chez la levure, il a été observé que la présence de la queue poly(A) est suffisante pour l'initiation de la **traduction**, la présence en plus de la coiffe sur l'extrémité 5' assurant une activité optimale (Preiss and Hentze 1998). Des mutations réalisées sur la protéine **Pab1p** (Poly(A) binding protein) ont apporté les premières évidences *in vivo* de l'implication de cette protéine et de la queue de poly(A) dans l'initiation de la traduction (Sachs and Davis 1989). D'autre part, les effets sur la traduction des pré-ARNms de différentes longueurs de queue poly(A) ont été comparés dans des extraits de réticulocytes. Il a été prouvé que la queue stimulait l'initiation de la traduction. De cette façon, un lien entre les deux processus a été clairement établi (Munroe and Jacobson 1990).

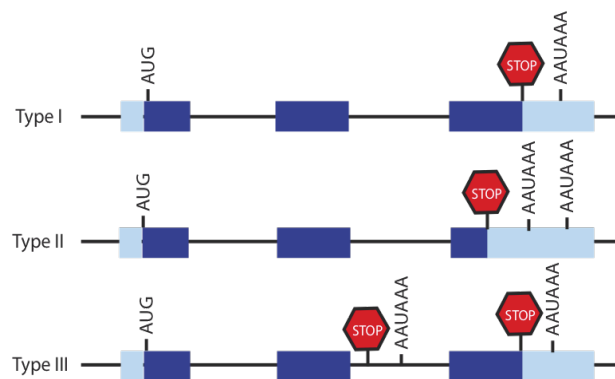
Actuellement, il est indiscutable qu'une **polyadénylation alternative** du pré-ARNm peut conférer des propriétés structurales et fonctionnelles différentes aux protéines. De ce fait, un pré-ARNm pourrait produire des ARNms matures avec différentes séquences 3' UTR importantes dans la régulation du transcrit (*fig. 2*). Ainsi, les conséquences d'une polyadénylation alternative pourraient être la génération de protéines distinctes à partir du même gène ou la formation des ARNms avec une stabilité différentielle qui modifie sa demi-vie (Licatalosi and Darnell 2010).

Il existe dans la littérature de nombreux exemples d'utilisation de différentes **PASs** (**polyadénylation signals**) chez les mammifères, de façon tissu spécifique ou en réponse à une stimulation extracellulaire (Neilson and Sandberg 2010) (Edwalds-Gilbert G 1997). Trois types de polyadénylation ont été décrits (*fig. 3*) (Lutz 2008):

- type 1 : un seul site de polyadénylation et l'épissage n'est pas affecté,
- type 2 : plusieurs sites de poly(A) dans la région 3' UTR et l'épissage n'est pas affecté,
- type 3 : au moins une polyadénylation alternative couplée à un épissage alternatif.



**Figure 2. Polyadénylation alternative des pré-ARNms** (adaptée de (Licatalosi and Darnell 2010)). Sur la figure, un exemple d'un gène transcrit en pré-ARNm et qui est lui-même traduit en différentes isoformes d'ARNm grâce au phénomène de polyadénylation alternative. Ici, il existe une sélection alternative de sites poly(A) (pA) ce qui génère des isoformes d'ARNm avec un 3' UTR plus ou moins long. La coiffe est représentée par m7G.



**Figure 3. Adaptée de (Lutz 2008). Types de polyadénylation.**

Finalement, l'importance de la modification en 3' vient aussi de l'interaction et de l'influence de ce processus sur **l'épissage et la transcription**. Ainsi, le complexe protéique impliqué dans la maturation 3' interagit avec des facteurs qui interviennent dans ces deux processus, en particulier, avec le **CTD** (carboxy terminal domain) de la sous-unité majeure de l'ARN polymérase II, Rpb1 (*ce point sera développé dans la section A.2 de ce chapitre*).

### c.) **Étapes de la maturation 3'**

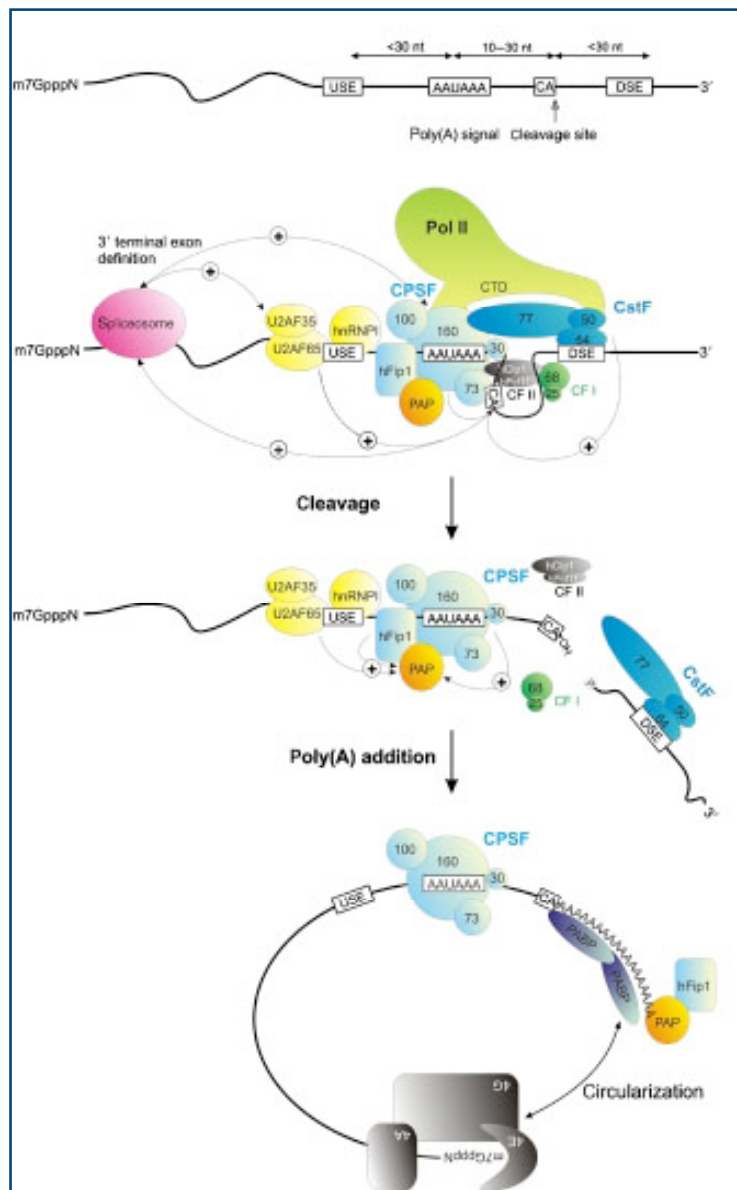
La maturation 3' des ARN pré-messagers est un processus nucléaire se déroulant en deux étapes successives et obligatoires, un **clivage endonucléolitique**, où les précurseurs des ARNm perdent un fragment non codant de leur extrémité 3', suivi de la polymérisation des adénosines phosphate (AMP) qui permet la formation de la **queue poly(A)**.

Le **clivage** à l'extrémité 3' est une étape de l'expression génique commune à tous les eucaryotes. Avant que le clivage n'ait lieu, les facteurs protéiques qui participent à la réaction doivent s'assembler et interagir avec le pré-ARNm de façon coopérative. Il n'est pas clair si les interactions entre les différents complexes précèdent à l'assemblage avec l'ARN ou si les deux types de liaison ont lieu en même temps. La reconnaissance des éléments en *cis* du pré-ARNm par les facteurs déterminera le site de coupure (Edmonds 2002).

Chez les **mammifères**, l'ARN est clivé au niveau d'un site endonucléolitique **CA** (cytosine-adénosine) et approximativement 10 adénosines phosphate sont ajoutées par l'enzyme **PAP** (**p**oly(**A**) **p**olymerase) grâce à l'utilisation d'ATP (Wahle 1991). Le fragment 3' clivé est dégradé rapidement par l'exonucléase **Xrn2** (5'-3' exoribonucléase 2) (West et al. 2004). Deux des complexes les mieux caractérisés sont **CPSF** (**c**leavage and **p**olyadenylation **s**pecificity **f**actor) et **CstF** (**c**leavage **s**timulation **f**actor). D'abord, CPSF va reconnaître de façon spécifique le signal poly(A) pour former un complexe protéine-ARN relativement instable (Gilmartin and Nevins 1989). Cette interaction est stabilisée par le complexe protéique CstF qui interagit au même temps avec une région adjacente au site de clivage riche en G/U (Weiss et al. 1991). De plus, un troisième facteur et la présence de PAP sont nécessaires pour la réaction de clivage (Wahle and Keller 1996).

Dans la deuxième étape du processus, plus de 200 adénosines sont additionnées à l'extrémité 3' pour former la **queue poly(A)**. L'enzyme PAP, qui possède une faible affinité pour l'ARN, est recrutée grâce à son interaction avec la sous-unité hFip1 de CPSF (Kaufmann et al. 2004). CPSF est indispensable à cette étape ce qui suggère l'existence d'une étroite relation entre les deux phases de la maturation 3' (Gilmartin and Nevins 1989) (Weiss et al. 1991) (Wahle and Keller 1996). En plus de cette interaction, PAP a besoin de **PABP** (**p**oly(**A**) **b**inding **p**rotein) qui rejoint le complexe de polyadénylation quand une queue d'au moins 10 résidus adénylés a été formée sur le pré-ARNm. De cette façon, PAP permet l'addition des adénines et PABP contrôle la longueur de la queue. Seule l'association des trois éléments, PAP, PABP et ARN, est efficace (Kerwitz et al. 2003).

La polyadénylation processive continue à une vitesse approximative de 25 nucléotides par seconde jusqu'à ce que la queue poly(A) atteigne une longueur de **150-200 nucléotides**. À partir de ce moment, la réaction processive se termine, et une élongation lente et distributive persiste (Bienroth et al. 1993). L'interaction de la protéine PABP avec ses partenaires à cet instant est caractérisée par un taux élevé d'association-dissociation et elle est finalement remplacée par la PABP cytoplasmique (**PABPC**) au moment de l'**export des ARNm**. PABPC interagit alors avec le facteur d'initiation de la translation **eIF4G**. Cette interaction entraîne la formation d'une particule ribonucléoprotéique pseudocirculaire qui pourra être traduite (Mangus et al. 2003). Le schéma des étapes du processus est représenté dans la *figure 4* (adaptée de (Danckwardt et al. 2008)).



**Figure 4.** Adaptée de (Danckwardt et al. 2008). Étapes du processus de maturation 3' des ARNm.

Chez la levure, le principe du processus est le même mais les acteurs intervenants sont différents. Dans un premier temps, au moins cinq protéines vont reconnaître le pré-ARNm. Le complexe **CF I** (cleavage factor I) lie des séquences situées en amont du site de clivage et **CPF** (cleavage polyadenylation factor) va permettre une identification plus précise de ce site. Quand la machinerie est fixée sur l'ARN, elle se stabilise via des interactions protéine-protéine afin de procéder aux réactions de clivage et polyadénylation (Zhao et al. 1999). Le fragment 3' clivé est dégradé par une 5'-3' exonucléase appelée **Rat1**, une enzyme qui participe à la terminaison de la transcription (Kim et al. 2004).

L'enzyme qui permet l'addition des adénines chez la levure est la **Pap1p** (poly(A) polymerase). La régulation de cette synthèse nécessite surtout la présence de la protéine **Fip1p** (Factor interacting with Pap1p) (Preker et al. 1995). Dans un modèle concernant ce processus, il a été proposé que Pap1p soit maintenue dans un état inhibé par Fip1p au moment du clivage. Une fois la réaction de clivage réalisée, la protéine Pap1p serait activée grâce à un mécanisme mal défini qui requiert des interactions à l'extrémité C-terminale de Fip1p avec d'autres facteurs de la machinerie (Helmling et al. 2001). La queue atteint une longueur de 55 à 90 résidus adényliques chez la levure (Martin and Keller 1998).

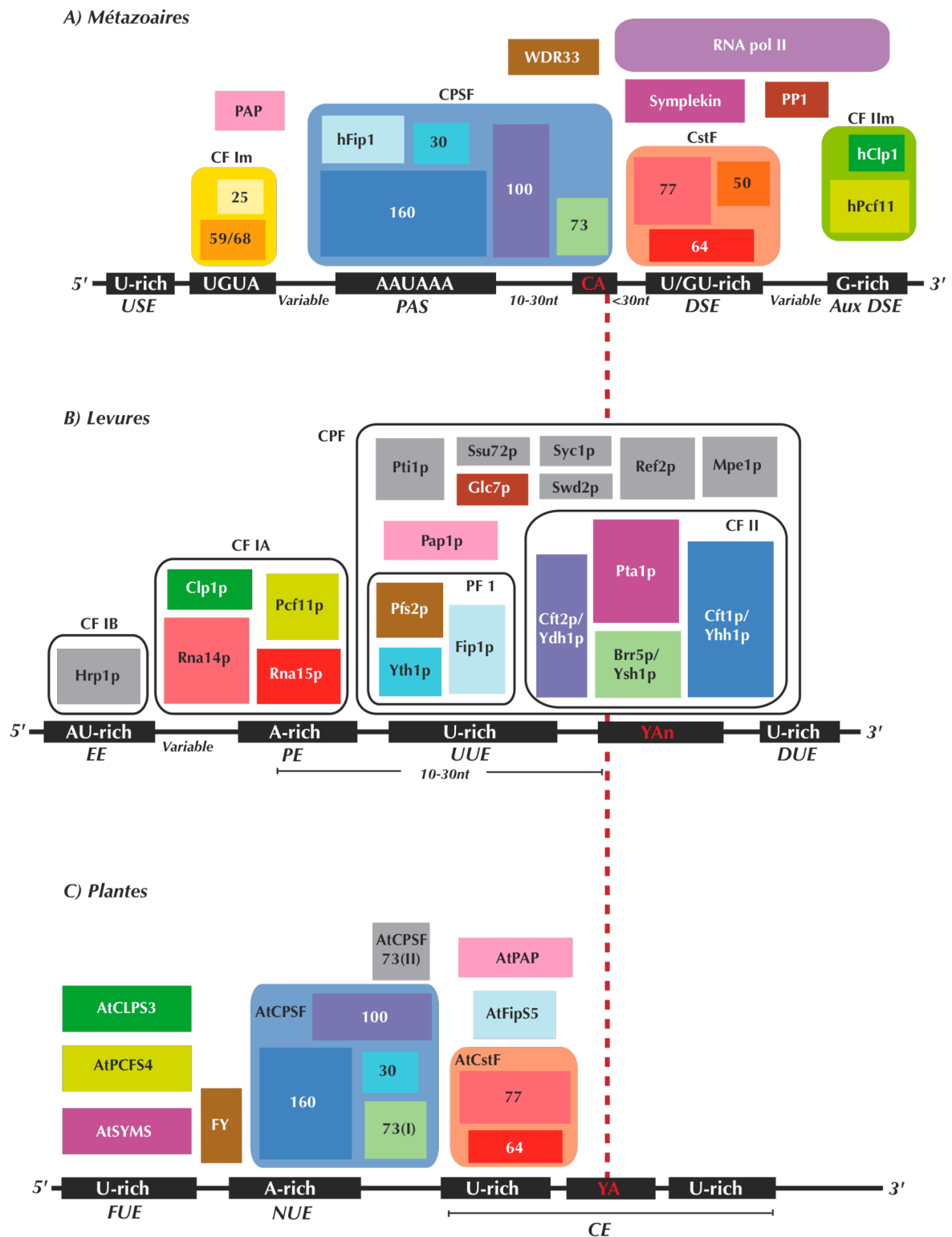
De cette façon, de nombreux facteurs sont essentiels pour l'aboutissement des réactions de clivage et polyadénylation. Les caractéristiques principales de l'ensemble des facteurs protéiques et des séquences de pré-ARNm impliqués seront présentées ultérieurement.

#### **d.) Éléments en cis impliqués dans le processus de maturation 3'**

Les réactions de clivage et de polyadénylation sont régulées par des éléments de séquence présents dans une région non-traduite. Ces éléments existent dans presque tous les pré-ARNms qui vont ultérieurement subir une polyadénylation, mais ils ne sont pas présents dans les pré-ARNms d'histones où la polyadénylation n'a pas lieu après le clivage (*voir section I. A.1.g*).

La *figure 5* correspond à un schéma incluant la globalité des facteurs connus pour leur implication dans la maturation 3' et les éléments en *cis* auxquels ils se lient. Les interactions entre les facteurs n'ont pas pu être toutes respectées sur cette image mais des figures postérieures les décrivent en détail (*figs. 10 et 11 ; tableaux 2 et 3*).

Certains des éléments en *cis* sont similaires chez les mammifères, les levures et les plantes, mais pour d'autres, on trouve des différences significatives au niveau des séquences et de leur localisation. D'autre part, les séquences chez la levure sont moins conservées que celles trouvées chez les mammifères et elles peuvent apparaître dupliquées. Pour cette raison, je vais présenter séparément les éléments *cis* des différentes espèces.



**Figure 5.** Schéma représentant les facteurs et les éléments en cis impliqués dans la maturation 3' des ARN pré-messagers chez les métazoaires (A), les levures (B) et les plantes (C). Les couleurs des facteurs ont été conservées pour les homologues dans les différentes espèces. Les facteurs spécifiques sont colorés en gris. La position des protéines par rapport à l'ARN met en évidence les interactions entre celles-ci et les éléments de séquence. Les signaux de poly(A) sont représentés par des carrés noirs et le site de clivage est montré par un pointillé rouge.

*d.1.) Éléments en cis chez les mammifères*

Chez les mammifères, il a été montré l'existence de trois éléments de séquence qui déterminent le site de polyadénylation et de deux auxiliaires qui régulent la réaction. Les trois premiers incluent : **AAUAAA** nommée PAS (polyadénylation signal), le **site de clivage, CA**, et une région riche en G/U appelée **DSE** (downstream element) (Zhao et al. 1999). Des deux **éléments auxiliaires**, un se trouve en aval et l'autre en amont du site de clivage.

- **PAS (polyadenylation signals)**: cet élément a été identifié en 1976 (Proudfoot and Brownlee 1976). Sa séquence est hautement conservée et des mutations dans cette région provoquent des troubles importants dans le processus de maturation 3' (exemple en (Wickens and Stephenson 1984)). Non seulement la séquence PAS est conservée, mais sa distance par rapport au site de clivage reste également constante à travers les espèces. Cette distance varie entre **10 et 30 bases en 5'** du site CA (*fig. 5*).

Selon les dernières études réalisées, seulement un 4% des ARNms ne contiennent pas une séquence signal de polyadénylation (Tian et al. 2005). Pour la plupart des ARNms (70%) cette séquence correspond à l'hexamère **AAUAAA**, ou dans certains cas, à une variation de cette séquence à la deuxième position qui donne comme résultat un hexamère **AUUAAA** (Manley 1988). En général, AAUAAA est employée plus souvent que d'autres variantes. Des analyses bioinformatiques ont été réalisées afin d'étudier pourquoi certains PAS étaient utilisés de façon préférentielle dans la cellule. Ces études ont corroboré d'autres précédentes et soutiennent que l'apparition des PAS modifiés permettrait une **polyadénylation alternative ou tissu spécifique** (Hook and Kellems 1988) (Wallace et al. 1999) (Neilson and Sandberg 2010).

Cependant, au moins 20% des pré-ARNms, ne contiennent pas la séquence A(A/U)UAAA. Dans ces cas, le mécanisme de reconnaissance des signaux poly(A) est moins connu. Des éléments auxiliaires situés en 3' et en 5' du site de clivage devraient compenser l'absence de cette séquence. Ces éléments faciliteraient la stabilisation du complexe de polyadénylation mettant en place des sites alternatifs d'ancrage. Ainsi, il a été montré que la séquence **UGUAN**, reconnue par le facteur CFIm, pourrait déterminer le site de clivage en absence de l'élément AAUAAA (Venkataraman et al. 2005). Des études récentes réalisées à partir du pré-ARNm de MCR4 (melanocortin 4 receptor) ont montré que des mutations sur l'hexamère AAUAAA n'ont pas d'effets importants dans la maturation 3' de cet ARN et que le site de polyadénylation nécessite seulement le **DSE** et une séquence **riche en A** en 5' (Nunes et al. 2010). Ces études démontrent que le caractère essentiel du PAS ne constitue pas une généralité.

- **DSE** : sa séquence est beaucoup moins conservée que pour le PAS. Deux types de séquence ont été observées, une riche en **GU** (YGUGUUUYY, où Y=pyrimidine) (McLauchlan et al.



1985) et une autre riche en U (UUUUU) (Chou et al. 1994). Cet élément est situé à moins de **30 bases en 3'** du site de clivage.

- **Le site de clivage** : sa position reste constante entre le PAS et le DSE (Chen et al. 1995) mais les séquences situées immédiatement à côté ne sont pas strictement conservées (MacDonald et al. 1994). En ce qui concerne la séquence, elle consiste, pour la plupart de pré-ARNms, en une adénosine précédée d'une cytosine (**CA** sur la *figure 5*) (Sheets et al. 1990).
- **Éléments auxiliaires** : un de ces éléments se trouve en **amont du site PAS** et contient souvent une séquence **riche en U** (UUUU, UGUA, UAUUA) (Hu et al. 2005). Les éléments **en aval du site de clivage** sont moins connus et moins conservés en séquence et en distance. Généralement une séquence **riche en G** est constatée. La présence d'autres éléments auxiliaires a également été observée. La fonction de ces éléments n'est pas très bien définie, mais ils pourraient être impliqués dans la reconnaissance des facteurs de polyadénylation pour le site de clivage et dans l'efficacité du processus.

#### *d.2.) Éléments en cis chez la levure*

Quatre éléments ont été identifiés dans les pré-ARNms de levures : une région riche en AU appelée **EE** (Efficiency Element), une autre riche en A, **PE** (Positioning Element), le **site de clivage** et finalement des **éléments riches en U** encadrant ce dernier.

- **EE** : cette séquence riche en AU est localisée toujours en amont du site de clivage mais à une distance variable. Des études de mutagenèses ont montré que la séquence **UAUAUA** est la plus utilisée dans les pré-ARNms de levures, la première et la cinquième position de U étant les plus critiques pour son rôle dans l'activation de la maturation 3' (Irniger and Braus 1994).
- **PE** : est une séquence riche en A située en aval de EE et à, approximativement, 10-30 nucléotides en amont du site de clivage. Sa position et sa séquence (habituellement AAAAAA ou AAUAAA (Guo and Sherman 1996)) accomplissent un rôle essentiel dans le processus. Le nom de PE (Positioning Element) est dû au fait que la plupart de ses mutations sont capables d'altérer le positionnement de la polyadénylation mais pas son efficacité (Guo and Sherman 1995) (Wahle and Ruegsegger 1999).
- **Le site de clivage** : à la différence de ce que l'on trouve chez les mammifères, le site est composé généralement d'une séquence commençant par une pyrimidine suivie de multiples adénosines (sur la *figure 5* **Y(A)<sub>n</sub>**). La coupure a lieu sur une position en 3' juste avant une adénosine (Heidmann, 1994 #110).

- **Éléments riches en U** : ce type de séquences a été observé aussi bien en amont (**UUE**) qu'en aval (**DUE**) du site de clivage (Graber et al. 1999). Des mutations effectuées dans ces éléments réduisent l'efficacité de la réaction de clivage (Dichtl and Keller 2001).

*d.3.) Éléments en cis chez les plantes*

Des études réalisées à partir des mutations génétiques ont révélé trois éléments principaux de séquence qui ne sont que peu conservés chez les plantes (*fig. 5*). Le premier, **FUE** (**f**ar-**u**pstream **U**-rich **e**lement) est riche en U et situé en amont du site de clivage, à une distance d'environ 100 nucléotides. Il existe aussi une séquence riche en A nommée **NUE** (**n**ear-**u**pstream **e**lement), de plus ou moins 7 bases de longueur. La séquence NUE se trouve plus près du site de clivage, à approximativement 10-40 nucléotides. Le site de clivage comporte un dinucléotide pyrimidine-adénosine entouré de régions de 5-10 nucléotides riches en U (**CE** ou **c**leavage **e**lement) (Li and Hunt 1995) (Loke et al. 2005) (Hunt et al. 2008).

Si on fait une **comparaison** du signal de polyadénylation des trois groupes, on observe qu'un **signal minimal tripartite**, composé d'une séquence riche en A, un élément riche en U et un site de clivage, est commun à tous les eucaryotes. Chez les mammifères, l'élément PAS est hautement conservé, il est présent en une seule copie et absolument nécessaire dans le processus de maturation 3'. Chez la levure, le motif riche en A donne la position du site de clivage et il est fréquemment dupliqué. En ce qui concerne les séquences riches en U ou UA, elles collaborent avec les signaux riches en A. Chez les mammifères, ce type de séquences apparaissent normalement en une seule copie en aval du site de clivage et sont essentielles. Quand elles se trouvent en amont du CA c'est pour stimuler la réaction de clivage (Zhao et al. 1999). Chez les levures et les plantes, ce type de signal est fréquemment redondant et il est situé en amont du site de clivage et, dans la plupart de cas, en amont aussi de la séquence riche en A. Le clivage a lieu de façon préférentielle sur un site CA chez les mammifères et sur un site Y(A)<sub>n</sub> chez les levures et les plantes. Des signaux en amont de ce site n'ont pas été clairement déterminés pour ces organismes.

L'identification des éléments de séquence sur le site de clivage et polyadénylation et les effets observés sur la maturation 3' après leurs mutations, ont permis d'imaginer l'existence de facteurs protéiques qui pourraient, par interaction avec les éléments en *cis*, réguler le processus.

e.) **Protéines impliquées dans la modification post-transcriptionnelle sur l'extrémité 3' du pré-ARNm**

Comme mentionné ci-dessus, seulement deux réactions sont nécessaires dans la modification 3', le clivage et l'addition de la queue poly(A). Malgré cela, nous connaissons aujourd'hui plus de **15** facteurs impliqués directement dans le processus chez les mammifères, autour de **20** chez la levure et **13** chez les plantes. La plupart des facteurs sont conservés entre les différentes espèces. Les détails sur la nécessité d'une telle complexité, les rôles des protéines impliquées et leurs interactions sont à présent le centre de nombreuses recherches.

L'ensemble des protéines impliquées dans la maturation 3' chez les **mammifères** est organisé en différents sous-complexes : **CPSF** (c̄leavage and polyadenylation s̄pecificity f̄actor), **CstF** (c̄leavage s̄timulation f̄actor), **CFIm** (c̄leavage f̄actor I) et **CFIIm** (c̄leavage f̄actor II). En plus, d'autres facteurs participent aussi dans le processus et ils semblent être étroitement liés à ces complexes, notamment **PAP** (poly (A) p̄olymerase), **PABP** (poly(A) p̄olymerase b̄inding p̄rotein), la **symplektine** et le **CTD** (c̄arboxy-t̄erminal d̄omain) de l'**ARN polymerase II** (fig. 5A). Tous ces facteurs sont essentiels dans la réaction de clivage *in vitro* sauf PABP, tandis que seul CPSF, PAP et PABP sont nécessaires pour la formation de la queue de poly(A).

Les sous-complexes chez la **levure** incluent **CPF** (c̄leavage and polyadenylation f̄actor), **CF IA** (c̄leavage f̄actor IA) et **CF IB** (c̄leavage f̄actor IB). CPF peut être divisé en **CF II** (c̄leavage f̄actor II) et **PF I** (p̄olyadenylation f̄actor I) (Chen and Moore 1992). CF II contient des sous-unités homologues à celles présentes dans le complexe CPSF de mammifères à l'exception de Yth1p et Fip1p, CPSF-30 et hFip1 chez les métazoaires, qui font partie du complexe PF I (fig. 5B). Deux des protéines de CF IA de levures correspondraient au CstF-77 et CstF-64 de mammifères, cependant pour CstF-50 un homologue n'a pas été trouvé chez *Saccharomyces cerevisiae*. Quelques facteurs apparaissent uniquement chez la levure, certains d'entre eux ont des homologues chez les mammifères, mais leur implication dans la maturation 3' n'a pas été prouvée. CF IA, CF IB et CF II sont requis pour la réaction de clivage *in vitro* et CPF, CF IA, CF IB et Pabp1 pour la polyadénylation. Le reste des facteurs joue un rôle dans la régulation du processus (Mandel et al. 2008).

Ces dernières années, les études réalisées chez les **plantes** ciblent la caractérisation des gènes et la détermination de la fonction des homologues des sous-unités impliquées dans la maturation 3' de mammifères et de levures. Un grand nombre de facteurs présents chez d'autres eucaryotes ont été identifiés dans ces organismes. Comme exception, nous pourrions citer l'absence des orthologues de CFIm59/68 et Hrp1. Cependant, pour la plupart des protéines, la similarité de séquence est importante et ceci suggère une conservation de la fonction de ces facteurs dans le processus (Hunt et al. 2008). Chez les plantes, trois complexes ont été identifiés. Le premier complexe est équivalent au complexe CPSF et inclut AtCPSF100, AtCPSF160, AtCPSF73-I, AtCPSF73-II, AtCPSF30,

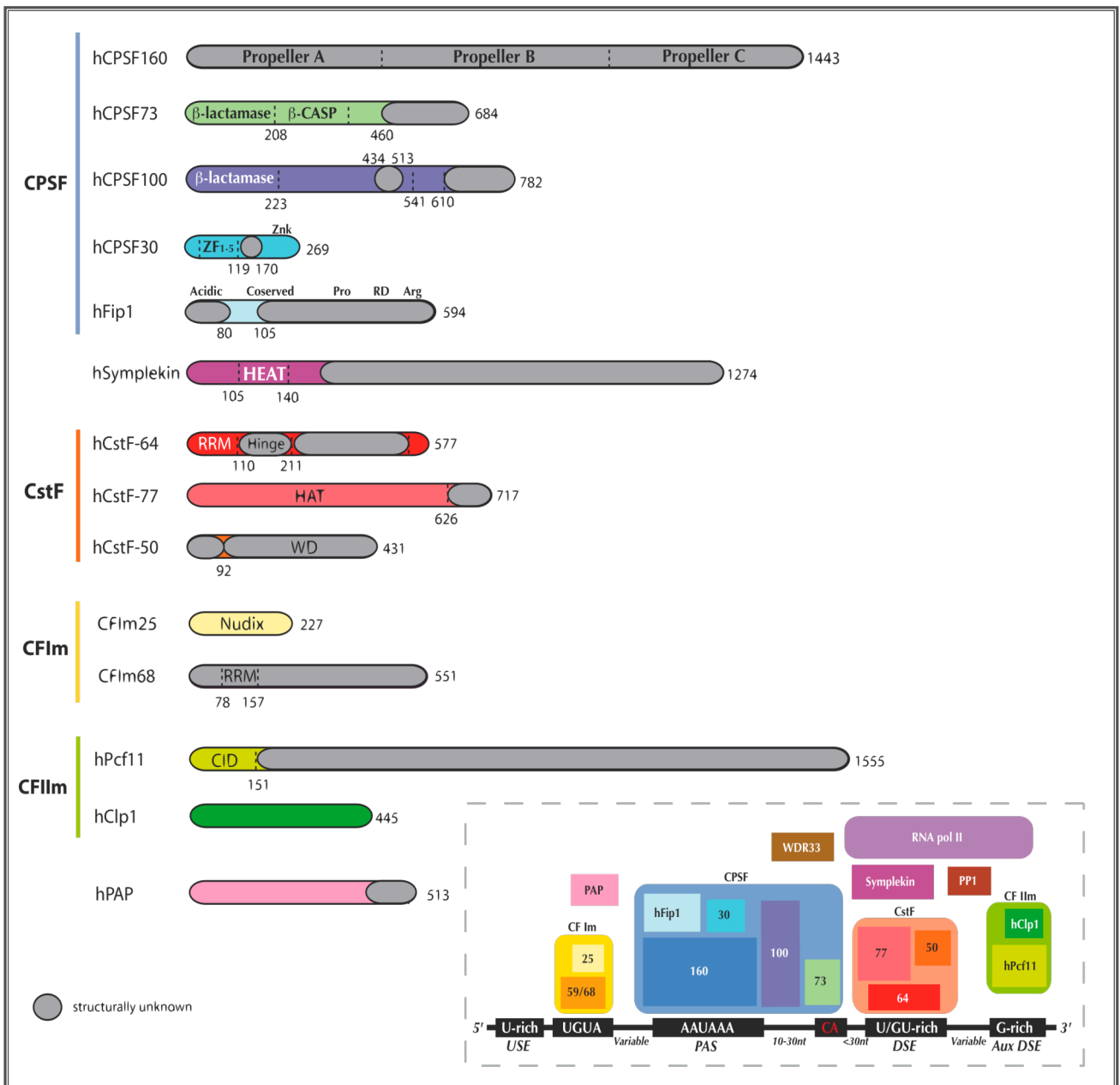
AtPAPS2, symplekine et FY. Le deuxième ensemble contient les facteurs PCFS (homologues de Pcf11p) et un des orthologues de Clp1p (CLPS). Enfin, le dernier complexe est composé des protéines AtFIPS, PAPS, PABN, CFIS et du reste des orthologues CLPS (fig. 5C).

Les principales caractéristiques des différents facteurs sont résumées ci-dessous. Au cours de ma thèse, j'ai travaillé principalement sur le complexe **CstF** et son homologue chez la levure **CF IA**, pour cette raison, ces complexes seront présentés plus en détail par rapport aux autres facteurs dans la section suivante. Sur le **tableau 1**, les noms des principales protéines du processus maturation 3' chez les mammifères et leurs homologues connus chez *Saccharomyces cerevisiae* et chez *Arabidopsis thaliana* ont été indiqués. Les complexes dans lesquels ces protéines sont incluses ont également été marqués.

<i>Homo sapiens</i>		<i>Saccharomyces cerevisiae</i>		<i>Arabidopsis thaliana</i>	
Complexe	Protéine	Complexe	Protéine	Complexe	Protéine
CPSF	CPSF-160	CPF (CF II)	Cft1/Yhh1p	AtCPSF	AtCPSF160
CPSF	CPSF-73	CPF (CF II)	Brr5p/Ysh1p	AtCPSF	AtCPSF73-I/II*
CPSF	CPSF-100	CPF (CF II)	Cft2p/Ydh1p	AtCPSF	AtCPSF100
CPSF	CPSF-30	CPF (PF I)	Yth1p	AtCPSF	AtCPSF30
CPSF	hFip1	CPF (PF I)	Fip1p	-	AtFIPS3/5*
-	WDR33	CPF (PF I)	Pfs2p	-	AtFY
-	Symplekine	CPF (CF II)	Pta1	-	AtSYM1/2/5*
CstF	CstF-77	CFIA	Rna14p	CSTF	CSTF-77
CstF	CstF-64	CFIA	Rna15p	CSTF	CSTF-64
CstF	CstF-50	-	-	-	CSTF-50
CF Im	CF Im -25	-	-	-	CFIS1/2*
CF Im	CF Im -68	-	-	-	-
CF II m	hPcf11	CFIA	Pcf11p	-	PCFS1/4/5*
CF II m	hClp1	CFIA	Clp1p	-	CLPS3/5*
-	PAP	CPF	Pap1	-	AtPAPS1/2/3/4*
-	PABP	-	Pab1p	-	AtPABN1/2/3*
-	Pol II CTD	-	Pol II CTD	-	-
-	PP1	CPF	Glc7p	-	-

**Tableau 1.** Facteurs impliqués directement dans le processus de maturation 3' des ARNm. Les homologues chez *S. cerevisiae* et *A. thaliana* de chaque protéine humaine sont indiqués. – indique l'absence d'un homologue ou d'un complexe dans lequel la protéine est incluse. \*Pour certains facteurs de plantes, différents gènes ont été identifiés. Ceux-ci seront représentés par des numéros séparés du symbole /. Ainsi, AtFIPS3/5 correspond à AtFIPS3 et AtFIPS5.

Sur la **figure 6**, une représentation schématique de facteurs impliqués dans le processus chez les mammifères est montrée. Les domaines colorés en gris correspondent aux régions des protéines dont la structure n'a pas été déterminée. Ainsi, nous disposons d'un **modèle structural** pour certains facteurs, pour d'autres, seulement l'architecture d'un ou plusieurs domaines et il y a des protéines dont la structure reste entièrement inconnue.



**Figure 6.** Représentation schématique de facteurs impliqués dans le processus chez les mammifères. Les domaines principaux de chaque protéine sont indiqués. Les parties colorées en gris correspondent aux régions dont la structure est inconnue. Un rappel de la figura 5 pour le code couleur est inclus en bas à droite.

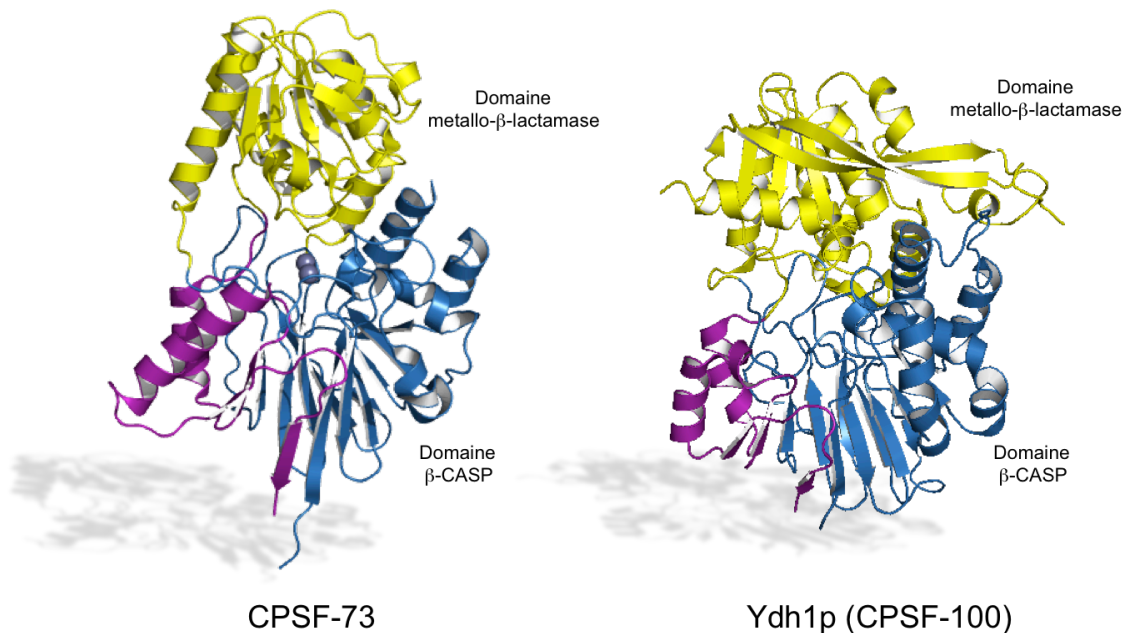
*e.1.) Caractéristiques principales de protéines impliquées dans la maturation 3' des ARNm*

- **CPSF-160 (Cft1/Yhh1p, AtCPSF160)\*** : est une protéine essentielle requise dans les deux réactions de la maturation 3' et dans la connexion avec la terminaison de la transcription (Proudfoot 2004) (Dichtl et al. 2002) (McCracken et al. 1997). CPSF-160 interagit directement avec l'élément PAS grâce à une forte affinité pour la séquence AAUAAA et une plus faible pour les séquences qui diffèrent de celle-ci (Moore et al. 1988). Cependant, l'union CPSF-PAS dépend aussi du complexe CstF via l'interaction entre CstF-77 et CPSF-160 (Murthy and Manley 1995). Chez la levure, **Yhh1p** interagit avec l'ARNm sur un site autour de la région de clivage riche en A et non sur le PE (l'équivalent du PAS de mammifères) (Dichtl et al. 2002). La délétion de ce facteur provoque l'abolition du clivage et l'absence de formation de la queue. La structure de cette protéine n'a pas été résolue, mais **16  $\beta$ -propeller** ont été identifiés à partir de l'analyse de sa séquence (Neuwald and Poleksic 2000). La structure d'une protéine qui partage cet homologie de séquence, **DBP1** (**D**NA **d**amage-**b**inding protein **1**), a été résolue et peut donner une idée de la possible architecture de CPSF-160 (Li et al. 2006).
- **CPSF-73 (Brr5p/Ysh1p ; AtCPSF73-I/II)** : cette sous-unité de CPSF est indispensable pour la viabilité cellulaire (Chanfreau et al. 1996). CPSF-73 contient un domaine **metallo- $\beta$ -lactamase** dans la région N-terminale et un domaine  $\beta$ -CASP, faisant ainsi partie de la superfamille de protéines  **$\beta$ -CASP** (Aravind 1999) (Callebaut et al. 2002). La région la plus conservée de la protéine est précisément ce dernier domaine (Jenny et al. 1996). La famille de  $\beta$ -CASP a été tout d'abord connue pour son implication dans l'interaction et le traitement des acides nucléiques. Ces observations suggéraient une implication possible de CPSF-73 dans le clivage du pré-ARNm. Des études structurales et biochimiques ont depuis donné des preuves évidentes sur l'activité **endonucléase** de cette protéine démontrant son implication dans la réaction de clivage (Mandel et al. 2006) (*fig. 7*) (Ryan et al. 2004). Des expériences réalisées à partir de mutations ponctuelles sur CPSF-73 et CPSF-100 ont montré que des résidus conservés des deux facteurs sont nécessaires pour la réaction de clivage. Les deux protéines sont ainsi requises pour la réaction (Kolev et al. 2008).
- **CPSF-100 (Cft2p/Ydh1p ; AtCPSF100)** : cette protéine est cruciale pour la viabilité cellulaire et essentielle dans les deux étapes de la polyadénylation (Kyburz et al. 2003) (Preker et al. 1997). Elle est aussi un membre de la famille  **$\beta$ -CASP** mais aucune action catalytique ne lui a été attribuée. Ainsi, même si la structure de ses domaines metallo- $\beta$ -lactamase et  $\beta$ -CASP a été résolue, sa fonction reste inconnue (Mandel et al. 2006)

---

\* Les noms entre parenthèses correspondent aux homologues de mammifères, en première position pour *Saccharomyces cerevisiae* et ensuite pour *Arabidopsis thaliana*.

(Callebaut et al. 2002) (*fig. 7*). Seule son interaction avec d'autres protéines impliquées dans le processus, CPSF-73, hClp1 et le CTD de l'ARN pol II, a été déterminée (Dominski et al. 2005) (de Vries et al. 2000) (McCracken et al. 1997). De plus, sa liaison avec le pré-ARNm dans une région proche au site de clivage a été observée (Dichtl and Keller 2001).



**Figure 7.** Modèle structural des protéines CPSF-73 humaine et l'homologue de CPSF-100 chez la levure (*Ydh1p*). Le domaine metallo-β-lactamase des deux protéines est marqué en jaune et le domaine β-CASP en bleu.

- **CPSF-30 (Yth1p ; AtCPSF30)** : intervient dans les deux étapes de la formation de la queue poly(A). CPSF-30 contient 5 motifs doigts de zinc (**ZF1-ZF5**) plus, chez les métazoaires seulement, une articulation de zinc sur l'extrémité C-terminale (« **Zn knuckle** »). Ces motifs permettent à la protéine de fixer l'ARN. Ainsi, il a été démontré l'interaction de **Yth1p** avec des régions de l'ARN riches en U (UUE et DUE) (Barabino et al. 2000). Les doigts de zinc sont aussi les responsables des interactions protéine-protéine. Cette caractéristique est important compte tenu de l'implication de CPSF-30 dans la coordination d'autres facteurs présents dans la maturation 3'. L'association CPSF-30/Yth1p aux protéines hFip1/Fip1p et Ysh1p (chez la levure) a été mise en évidence (Kaufmann et al. 2004) (Preker et al. 1995) (Takahashi et al. 2003). De plus, CPSF-30 interagit avec la protéine **NS1** du virus influenza. La formation du complexe CPSF-30/NS1 empêche la liaison de CPSF à l'ARN inhibant, par conséquent, le clivage et la polyadénylation des ARNms de l'hôte. De cette façon, NS1 produit une rétention sélective des ARNms hôtes dans le noyau et non des ARNs viraux (Nemeroff et al. 1998). La structure

cristallographique de la zone C-terminale de NS1 en complexe avec la région de CPSF-30 contenant les doigts de zinc 2 et 3 a été résolue (Das et al. 2008). L'architecture d'autres régions de la protéine est aussi disponible sur la PDB, (Protein Data Bank, code :2d9n) mais aucune publication n'est liée à cette structure. AtCPSF30 diffère de CPSF-30 vu qu'il possède, en plus de sa capacité de reconnaissance de l'ARN, une activité endonucléase qui n'a pas été révélée chez les mammifères ou la levure (Delaney et al. 2006) (Addepalli and Hunt 2007).

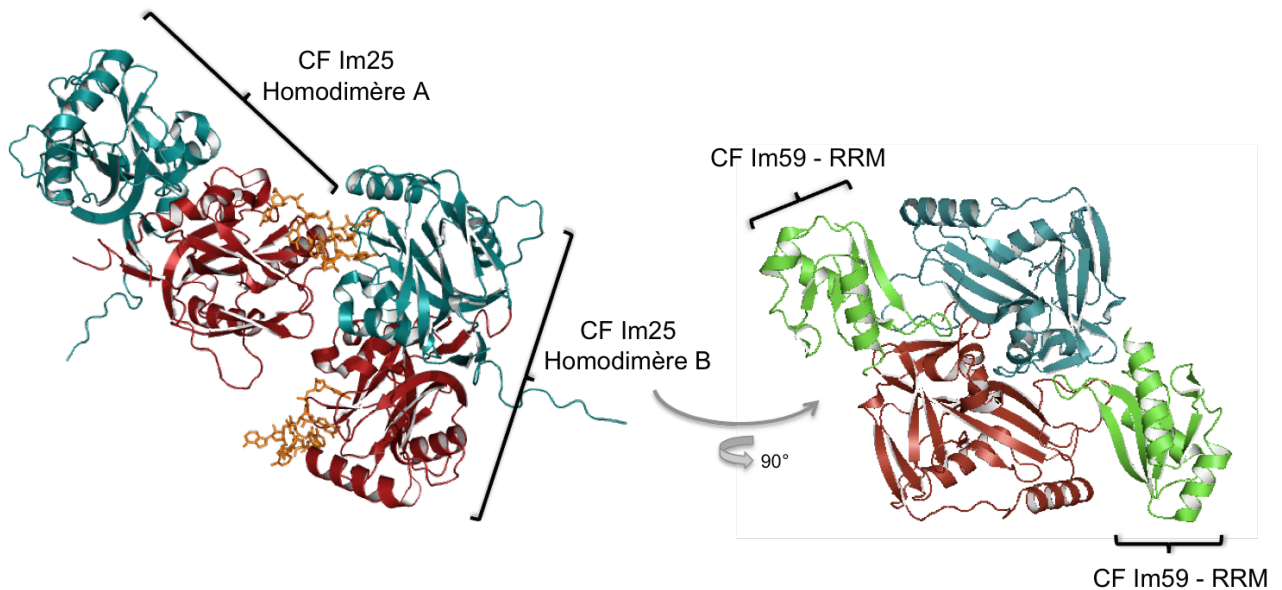
- **hFip1 (Fip1p ; AtFIPS3/5)** : Fip1p doit son nom à son interaction avec **Pap1p** (Factor interacting with Pap1p) (Preker et al. 1995). hFip1 contient un segment acide, proche de la région N-terminale, nécessaire pour l'interaction avec PAP. Cette région est suivie d'une zone très conservée qui lui permet d'interagir avec d'autres protéines de la maturation 3', CPSF-160, CPSF-30 et CstF-77 (Preker et al. 1995). Sa région C-terminale inclut un segment riche en Pro, un autre riche en Arg et Asp impliqué dans l'interaction avec la séquence d'ARNm (RD sur la *figure 6*) et un dernier riche en Arg. La région RD n'existe pas chez la levure (Kaufmann et al. 2004) (Mandel et al. 2008). La principale fonction de hFip1 est de rapprocher PAP au site de polyadénylation. Chez Arabidopsis, un des orthologues de ce facteur est capable d'interagir en plus avec l'ARN (Forbes et al. 2006).
- **WDR33 (Pfs2p ; AtFY)** : l'implication de WDR33 dans la maturation 3' a été découverte récemment (Shi et al. 2009) tandis que l'importance de son homologue Pfs2p, une sous-unité du complexe CPF, dans les deux réactions de la formation de la queue poly(A) était déjà connue (Ohnacker et al. 2000). WDR33/Pfs2p comporte des motifs répétés **WD40** (7 régions de 40 aminoacides riches en Trp et Asp). Pfs2p interagit, entre autres, avec Ysh1p, Ydh1p et Fip1p (Ohnacker et al. 2000) (Kyburz et al. 2003). Ces trois protéines possèdent des homologues chez les mammifères. Il serait intéressant d'étudier si ces interactions sont conservées pour WDR33 dans ces organismes.
- **Symplekine (Pta1p ; AtSYM1/2/5)** : la Symplekine et Pta1p partagent une faible homologie de séquence, néanmoins, elles ont été considérées comme homologues grâce à une région commune (Takagaki and Manley 2000). La participation de Pta1p dans le clivage et la polyadénylation a été démontrée (Zhao et al. 1999). La structure cristallographique de la région N-terminale de la Symplekine de *D. melanogaster* a été déterminée (Kennedy et al. 2009). La structure de cette région montre la présence de cinq répétitions **HEAT** (Huntington, Elongation Factor 3, PR65/A, TOR) avec une boucle (« loop »), entre la quatrième et la cinquième répétition, dont la séquence est très conservée à travers les espèces. Il a été observé que la boucle du domaine HEAT apporte à la Symplekine la capacité d'interagir avec un grand nombre de partenaires comme Syc1p, Swd2p, Ssu72p, Pti1p, Glc7p, Fip1p, Cft2p/Ydh1p, Brr5p/Ysh1p, Clp1p et Pcf11p (Zhelkovsky et al. 2006) (Nedeja et al. 2003) (Ghazy et al. 2009) (*voir tableau 2*).



Symplekine/Pta1p joue donc un rôle important dans la connexion des différents facteurs impliqués dans la maturation 3'. Chez les plantes, AtSYM fait partie du complexe CPSF (Herr et al. 2006).

- **CstF** : est composé de 3 protéines nommées en fonction de leurs poids moléculaires, **CstF-77**, **CstF-64** et **CstF-50**. Les deux premières ont un homologue de séquence et fonctionnel chez la levure, **Rna14p** et **Rna15p** respectivement. Pour CstF-50, aucun homologue n'a été identifié chez la levure. Ce complexe constitue le cœur de cette thèse, pour cette raison, ce point sera abordé plus en détail dans la section suivante.
- **CF Im** : ce facteur apparaît seulement chez les eucaryotes supérieurs où il est nécessaire pour la réaction de clivage (Takagaki et al. 1989). Il a été montré que sa déphosphorylation par une Ser/Thr phosphatase abolit cette réaction (Ryan 2007). CF Im a été pensé fonctionner comme un hétérodimère composé d'une sous-unité de **25kDa** et d'une deuxième de **59**, **68** ou **72kDa** (Mandel et al. 2008). Cependant, les données structurales disponibles montrent l'organisation dimérique de la protéine CF Im-25 (Tresaugues et al. 2008) (Coseno et al. 2008) (Yang et al. 2010). Les trois grandes sous-unités contiennent la même séquence en acides aminés, mais elles sont transcrites à partir de différents gènes. **CF Im25** interagit avec PAP et PABP (Dettwiler et al. 2004) et avec hClp1 (de Vries et al. 2000). La région N-terminale de **CF Im68** est composée d'un **RBD** (RNA Binding Domain) qui est également indispensable pour l'interaction avec la sous-unité de 25kDa (Dettwiler et al. 2004). De plus, sa région C-terminale est impliquée dans la liaison avec la machinerie d'épissage (Rappsilber et al. 2002). La fonction principale de CF Im est d'apporter un site additionnel de reconnaissance de l'ARNm pour aider à la définition d'un site de polyadénylation adéquat. Il va interagir avec une séquence qui généralement contient **UGUAA** situé en amont du PAS (Brown and Gilmartin 2003). Chez Arabidopsis, l'homologue correspondant à CF Im25 a été identifié, CFIS1/2, mais on ne lui connaît pas de partenaire.

La structure de l'homodimère formé par **CF Im25** est connue (Tresaugues et al. 2008) (Coseno et al. 2008). L'architecture de cette sous-unité est typique de protéines Nudix, mais l'activité hydrolase caractéristique de cette famille protéique n'a pas été détectée pour CF Im25 (McLennan 2006). Une partie de mes travaux de thèse s'est intéressé au complexe CF Im25/CF Im68-ARN. L'obtention du complexe '*in vitro*' seul et en interaction avec l'ARN a permis la réalisation des essais de cristallisation. Cependant, avant la découverte des conditions adéquates pour la cristallisation, la structure de CF Im25 avec des séquences d'ARN contenant UGUA a été résolue dans d'autres groupes de recherche (Yang et al. 2010) (*fig. 8*). De plus, la structure cristallographique du complexe CF Im25/CF Im59 a été également déposée dans la PDB (Protein Data Bank ; code : 3N9U). Nous avons donc mis de côté ces essais et les résultats préliminaires obtenus ne seront pas présentés ici.

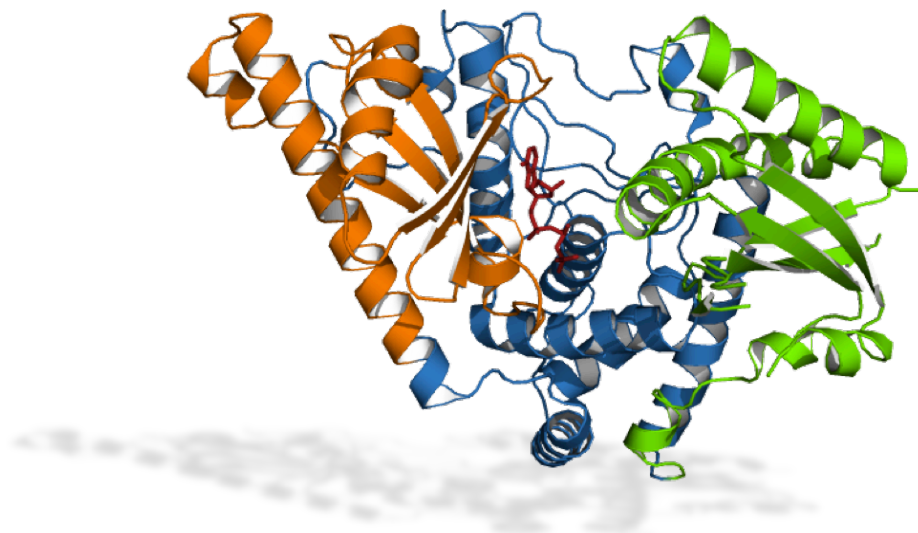


**Figure 8.** Structures obtenues pour la protéine CF Im25. A gauche, deux unités asymétriques, chacune comportant un homodimère. L'ARN (UGUAAA), coloré en orange sur la figure, interagit avec un des monomères d'une des unités asymétriques et avec l'autre monomère de l'unité asymétrique située à proximité (Yang et al. 2010). A droite, représentation du modèle structural de CF Im25 en complexe avec le domaine RRM de CF Im59 (PDB : 3N9U).

- CF IIm** : de la même façon que CF Im, CF IIm intervient seulement dans la réaction de clivage (Ruegsegger et al. 1996). Ce complexe contient deux protéines, **hPcf11** et **hClp1**, qui avaient déjà été identifiées chez la levure mais appartenant au complexe CF IA avec Rna14p et Rna15p (*plus en détail dans la section suivante*) (de Vries et al. 2000). **hPcf11** a été identifié à partir des extraits nucléaires avec un poids moléculaire apparent de 140 à 200 kDa. Son homologue chez la levure correspond à une version beaucoup plus courte de cette protéine où seulement la partie N-terminale est présente (de Vries et al. 2000). Les deux protéines, hPcf11 et Pcf11p, ont en commun une région qui permet leur interaction avec le CTD de l'ARN pol II (*plus détaillé dans la section I.A.1.e.3*). La fonction de la région C-terminale de hPcf11 n'a pas été révélée à présent. Cette protéine accomplit un rôle crucial dans la terminaison de la transcription et est requise pour une dégradation efficace du segment 3' restant après le clivage (West and Proudfoot 2008). **hClp1** interagit avec CPSF-100 et CFIm25 (de Vries et al. 2000) (*fig. 10*). Cette protéine, aussi bien que ses homologues, contient un motif **Walker A** généralement associé à l'union de l'ATP/GTP (Walker et al. 1982) (Noble et al. 2007). Un rôle physiologique **ARN-kinase**, importante dans le processus d'épissage des tRNA, a été attribué à hClp1. Il a été démontré que hClp1 phosphoryle l'extrémité 5' de l'exon 3' permettant la future liaison des exons par une enzyme inconnue (Weitzer and Martinez 2007). Chez les plantes, les homologues de hPcf11 et de hClp1 sont les seuls composants d'un complexe et ils ont été nommés

PCFS1/4/5 et CLP3. Deux des isoformes de PCFS contiennent le domaine d'interaction avec le CTD de l'ARN pol II (**CID**) retrouvé chez leurs homologues (Xing et al. 2008). Une interaction entre PCFS4 et l'orthologue de Pfs2p, FY a été montrée (Simpson et al. 2003). Pour CLPS, l'existence d'une interaction avec CPSF-30 a été mise en évidence, mais cette protéine n'est pas associée à CF Im comme dans le cas des mammifères (Hunt et al. 2008).

- **PAP ou poly(A) polymerase (Pap1p ; AtPAPS1/2/3/4)** : chez les mammifères PAP est une enzyme requise pour les deux réactions, clivage et polyadénylation, en revanche, chez la levure Pap1p est nécessaire seulement pour la polyadénylation (Edmonds 2002). Les structures des protéines originaires de levures et de mammifères ont été résolues (Bard et al. 2000) (Martin et al. 2000) (Martin et al. 2004). À partir des structures, la présence de trois domaines différents a été révélée (*fig. 9*). Le site actif est localisé dans une fissure formée entre le domaine N-terminal et le domaine C-terminal. Deux ions,  $Mg^{2+}$  et  $Mn^{2+}$ , coordonnés par le domaine N-terminal, sont nécessaires pour la catalyse. Le **domaine C-terminal**, riche en Ser/Thr, peut être phosphorylé et ceci joue un rôle potentiel dans la régulation du processus. Cette région est aussi impliquée dans l'interaction avec hFip1 (Fip1p) et CPSF-160 (Yhh1p) (Kaufmann et al. 2004) (Preker et al. 1995) (Martin et al. 2000). De plus, PAP contient un domaine **RBD** capable d'interagir avec l'ARN et deux signaux de localisation nucléaire (**NLSs**) essentiels pour la polyadénylation (Raabe et al. 1994). Chez les plantes, un des quatre orthologues de PAP, AtPAPS3, semble être une enzyme cytoplasmique et joue un rôle important pendant le développement du pollen (Hunt et al. 2008). De façon similaire, il existe une enzyme PAP cytoplasmique de mammifères spécifique aux testicules (Kashiwabara et al. 2000).



**Figure 9.** Modèle structural de la PAP humaine (Martin et al. 2000). La région colorée en orange représente le domaine catalytique, celle en bleu est le domaine central et en vert, le domaine C-terminal. Un analogue de l'ATP, indiqué en rouge, est lié au site actif de l'enzyme.

- **PABP (Pab1p/Nab2p ; AtPABN1/2/3)** : cette protéine est nécessaire pour une polyadénylation correcte et efficace. L'interaction de PABP avec l'ARNm augmente l'efficacité de la formation de la queue (Kerwitz et al. 2003). Il existe cinq PABPs chez l'humaine et huit chez Arabidopsis. On peut distinguer entre **PABPs nucléaires et cytoplasmiques**. Les premières (PABPNs) sont nécessaires pour la synthèse de la queue d'adénines et le control de sa longueur (Amrani et al. 1997) (Minvielle-Sebastia et al. 1997). PABPN1 interagit directement avec la queue poly(A) naissante, avec les premières 11-14 adénines, et cette union est maintenue jusqu'à la fin de la formation d'une queue de 200-300 bases chez les mammifères (Meyer et al. 2002) (Keller et al. 2000). PABPN1 stimule la synthèse processive des adénines par la PAP (Bienroth et al. 1993) (Wahle 1995). Les PABPs nucléaires contiennent un seul **RRM (RNA Recognition Motif)** et un domaine carboxy-terminal riche en Arg important aussi pour la reconnaissance de l'ARN (Nemeth et al. 1995) (Wahle and Ruegsegger 1999). Chez la levure, l'homologue de PABPN provient du gène **NAB2** (Anderson et al. 1993). Une mutation sur la protéine **Nab2p** promeut l'hyperpolyadénylation. Ceci signifierait qu'elle joue un rôle dans le contrôle de la longueur de la queue (Hector et al. 2002). Un deuxième homologue de PABPNs chez la levure est la protéine **Pab2p**. Il a été démontré récemment que cette protéine interagit avec la queue poly(A) des pré-ARNsno (« small nucleolar RNA ») et promeut sa dégradation par recrutement de l'exosome (Lemay et al. 2010). Pab2p interagit également avec l'ARNm (Lemieux and Bachand 2009), cependant, dans ces cas-là le transcrit n'est pas dégradé par l'exosome et traverse la membrane nucléaire pour être traduit. Il est possible que l'ARNm soit alors transporté dans le cytoplasme grâce à un mécanisme de compétition cinétique entre les processus d'export et la dégradation (Libri 2010).

Les **PABPCs**, contiennent quatre RRMes connectés au domaine carboxy-terminal par un « linker » riche en Pro et Met (Adam et al. 1986). Les structures de la région C-terminale de PABPC1 et des deux motifs RRMes en complexe avec 11 adénines ont été résolues par RMN (Kozlov et al. 2001) et par cristallographie (Deo et al. 1999), respectivement. La région C-terminale est importante pour son rôle dans la régulation de la polyadénylation, la déadénylation, l'initiation de la traduction et la terminaison de la traduction (Mangus et al. 2003). Rna15p est la seule protéine connue dans la maturation 3' qui interagit directement avec **Pab1p** et elle pourrait, probablement, participer à son recrutement (Amrani et al. 1997).

- **Le domaine carboxy-terminal (CTD) de Rbp1, la sous-unité plus grande de la ARN polymerase II** : il est composé d'un heptapeptide répété **52** fois chez l'humain et **26** fois chez la levure. La séquence de chaque répétition est très conservée, généralement **YSPTSPS**. Le CTD est considéré comme un facteur de la maturation 3' vu que sa délétion provoque l'inhibition de la polyadénylation (McCracken et al. 1997). Son rôle dans la

connexion entre les processus de transcription, de maturation 3' et d'épissage sera exposé plus en détail ultérieurement (*voir Chapitre I.2.a*).

- **hSsu72p (Ssu72p)** : Ssu72p est une protéine nécessaire pour le clivage. Elle fait partie du complexe CPF (He et al. 2003) (Dichtl et al. 2002). Ssu72p est une **phosphatase** spécifique de la Ser<sub>5</sub> du CTD (Krishnamurthy et al. 2004). La déphosphorylation de la Ser<sub>5</sub> est une activité requise pour la formation de la coiffe en 5'. Par conséquent, l'inhibition de l'activité phosphatase de Ssu72p provoque une accumulation d'ARN pol II phosphorylée en Ser<sub>5</sub> *in vivo* (Reyes-Reyes and Hampsey 2007). Chez l'humain, la découverte de l'implication de cette protéine dans le processus de maturation 3' a été plus tardive (St-Pierre et al. 2005). Cependant, la structure du complexe ternaire formé par la région N-terminale de la Symplekine, hSsu72 et un heptapeptide du CTD phosphorylé en Ser<sub>5</sub> a été résolue très récemment (Xiang et al. 2010).
- **D'autres facteurs présents chez la levure** : d'autres protéines ont été identifiées chez la levure. Certaines d'entre elles ont des homologues chez les mammifères, mais leur rôle dans la maturation 3' n'a pas été décrit.

**Hrp1p ou Nab4p** est le seul composant du facteur **CF IB**. Cette protéine intervient dans le clivage où elle n'est pas essentielle, mais joue un rôle important dans la sélection du site de poly(A) (Minvielle-Sebastia et al. 1998). Elle contient deux **RRMs** qui reconnaissent de façon spécifique l'élément de séquence EE riche en U-U/A (Perez-Canadillas 2006). A partir d'une étude réalisée par SELEX, une préférence de cette protéine pour la séquence (UA)<sub>3</sub> a été observée (Valentini et al. 1999). Hrp1p fait partie de la famille des **hnRNP** (heterogeneous nuclear ribonucleoprotein).

**Pti1p** fait partie du complexe CPF associé au **snoRNA** (Nedea et al. 2003). Elle n'est pas indispensable dans la formation de la queue poly(A), mais des défauts dans cette protéine affectent la sélection du site de clivage et la longueur de la queue poly(A). Cette protéine partage une homologie de séquence avec CstF-64 et Rna15p (Skaar and Greenleaf 2002).

**Swd2p** est aussi intégré dans le complexe CPF (Gavin et al. 2002) (Nedea et al. 2003). La présence de 7 répétitions de **WD40** a été prédite à partir de sa séquence. Elle n'a pas un rôle essentiel dans la polyadénylation, mais participe à la stabilisation d'autres protéines et à la connexion avec la transcription (Dichtl et al. 2004). L'implication dans le processus de son homologue chez l'humain (**WDR82**) n'a pas été démontrée jusqu'à présent (Lee and Skalnik 2005).

**Mpe1p**, un membre du complexe CPF, est requis pour les deux réactions de la maturation 3' (Vo et al. 2001). Son homologue chez les mammifères a été découvert récemment (Shi et al. 2009) et correspond à la protéine **Rbbp6**. Rbbp6 contient des domaines additionnels par rapport à Mpe1. Elle a été originalement identifiée pour son

implication dans l'apoptose, le cycle cellulaire et la régulation de p53 (Simons et al. 1997). Il est possible que cette protéine ait un rôle dans l'interaction entre ces processus et la maturation 3' (Shi et al. 2009).

**Syc1p** est un composant du CPF qui garde une homologie de séquence importante avec la région C-terminale de Brr5p/Ysh1p. Son rôle n'est pas bien défini, mais elle pourrait agir comme régulateur négatif du processus de maturation par compétition avec Brr5p/Ysh1p (Zhelkovsky et al. 2006) (Mandel et al. 2008).

**Glc7p** est une des **phosphatases** participant dans la maturation 3' (Gavin et al. 2002). Elle joue un rôle dans la réaction de polyadénylation (Garcia-Gimeno et al. 2003) (He and Moore 2005). Pta1p pourrait être une de ses cibles pour la déphosphorylation (He and Moore 2005). L'homologue de Glc7p correspond à la phosphatase **PP1** (Shi et al. 2009).

**Ref2p** est aussi inclus dans le complexe CPF (Nedea et al. 2003). La participation de ce facteur dans la maturation 3' des snoRNAs a été démontrée (Dheur et al. 2003).

En 2009, le groupe de J. Manley a utilisé la technique du « RNA-tagging » pour la purification de l'ensemble des protéines impliquées dans le processus de maturation 3' chez l'humain (Shi et al. 2009). Pour l'expérience, trois copies de la tige-boucle liant spécifiquement la protéine MS2 de bactériophage sont ajoutées à l'extrémité 5' des pré-ARNms L3 d'adénovirus et SV40 late (SVL). Les ARNs liés à MBP-MS2 (maltose binding protein-MS2) ont été incubés avec un extrait nucléaire HeLa, sous des conditions adéquates pour la réaction de polyadénylation (avec ATP), afin de permettre l'association des complexes impliqués dans la maturation 3' avec l'ARN d'étude. A partir de ces réactions, la méthode de fractionnement sur gradient de glycérol suivi d'une purification ont permis l'indentification d'approximativement **85** protéines. Dans l'ensemble de protéines, le cœur est composé de facteurs participant directement dans la maturation 3' et environ 50 protéines, n'intervenant pas directement, s'associent aux facteurs principaux et lient ce mécanisme à d'autres processus comme l'épissage, la transcription, la traduction ou la réponse aux endommagements de l'ADN. La plupart des protéines connues pour leur implication dans le processus de maturation 3' étaient présentes, cependant, il a été constaté, de manière remarquable, l'absence de facteurs importants comme hClp1 ou les faibles niveaux stœchiométriques de hPcf11. Ainsi, ces deux facteurs ont du être ajoutés pour rendre la machinerie purifiée fonctionnelle. Malgré ceci, ces travaux ont permis l'indentification des nouveaux éléments en '*trans*' dont ces homologues chez la levure étaient connus (**WDR33**, **Rbbp6**, **PP1**). De plus, des analyses réalisées en **microscopie électronique** dans la même étude, ont montré que le cœur protéique de la machinerie acquiert une forme de **rein** et atteint une longueur de **250Å**. Ceci a donné les premiers indices sur l'architecture générale de la machinerie.

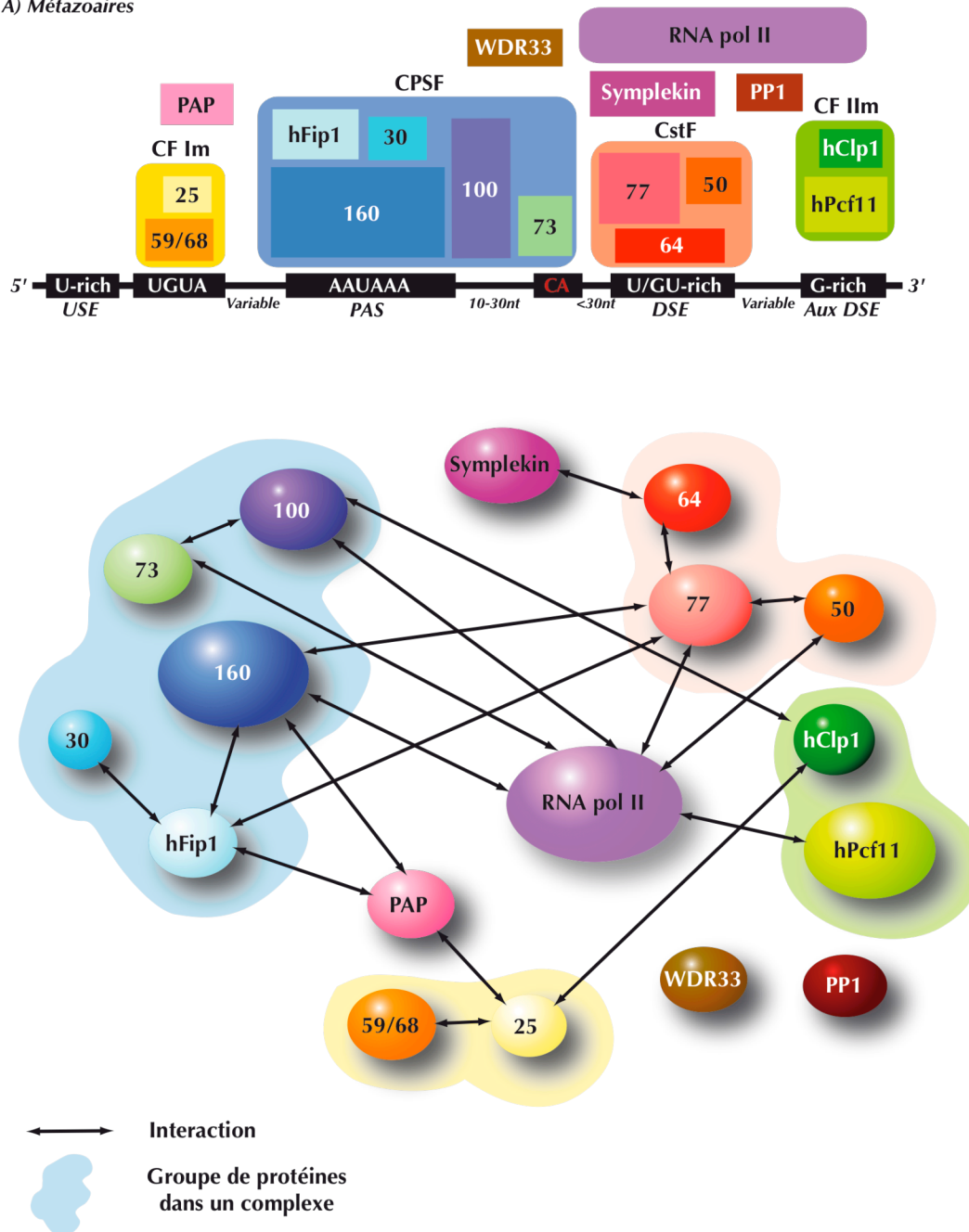
Les caractéristiques des facteurs présentés les rendent indispensables pour accomplir leur fonction, mais sans leurs interactions la réalisation du processus serait impossible. Les protéines qui participent à la maturation 3' forment un réseau d'interactions et vont permettre de réguler le processus. De plus, ces interactions permettent la connexion entre les deux étapes de la polyadénylation ainsi qu'entre la polyadénylation et d'autres événements ayant lieu dans la cellule. Des avancés sur la connaissance de ces interactions ont été obtenues ces dernières années ; deux tableaux montrant ces connexions ainsi que des schémas représentatifs font l'objet des figures suivantes (*tableaux 2 et 3 et figures 10 et 11*). Un grand nombre d'interactions n'ont certainement pas encore été identifiées. Malgré cela, ce processus apparaît déjà d'une grande complexité et ses capacités de régulation sont très nombreuses. De plus, il faut considérer le fait que quelques interactions sont exclusives, ainsi, dans certains cas, une protéine utilise la même région pour interagir avec deux facteurs différents et l'association avec l'un des deux empêche l'union à l'autre. Ceci est une façon importante de réguler le processus.

CPSF-73		1, 6												4	
CPSF-100	1, 6		15	15							2		14	4	
CPSF-30		15			7		9								
CPSF-160		15			7		13						13, 17	4	
hFip1p			7	7			7			3			7, 16		
Symplekin								5							
CstF-77			9	13	7		5	5, 12	5, 12					4	
CstF-64						5	5, 12								
CstF-50							5, 12		5					4, 8	
CFIm 25					3						10	2		10	
CFIm 68										10					
CFIm hClp1		2								2					
CFIm hPcf11														11	
PAP		14		13, 17	7, 16					10					
Rpb1 (Pol II)	4	4		4			4		4, 8				11		
	CPSF-73	CPSF-100	CPSF-30	CPSF-160	hFip1p	Symplekin	CstF-77	CstF-64	CstF-50	CFIm 25	CFIm 68	CFIm hClp1	CFIm hPcf11	PAP	Rpb1 (Pol II)

**Tableau 2.** Interactions connues entre les protéines de la machinerie de la maturation 3'. La couleur verte indique l'existence d'interaction directe entre deux protéines. La couleur saumon représente des interactions qui ont seulement été démontrés chez les plantes mais qui pourraient également exister chez les mammifères. Les numéros correspondent aux publications où les interactions apparaissent : 1. (Kolev et al. 2008) ; 2. (de Vries et al. 2000) ; 3. (Forbes et al. 2006) ; 4. (McCracken et al. 1997) ; 5. (Takagaki and Manley 2000) ; 6. (Dominski et al. 2005) ; 7. (Kaufmann et al. 2004) ; 8. (Fong and Bentley 2001) ; 9. (Bell and Hunt 2010) ; 10. (Dettwiler et al. 2004) ; 11. (Zhang et al. 2005) ; 12. (Takagaki and Manley 1994) ; 13. (Murthy and Manley 1995) ; 14. (Elliott et al. 2003) ; 15. (Xu et al. 2006) ; 16. (Meinke et al. 2008) ; 17. (Murthy and Manley 1992).

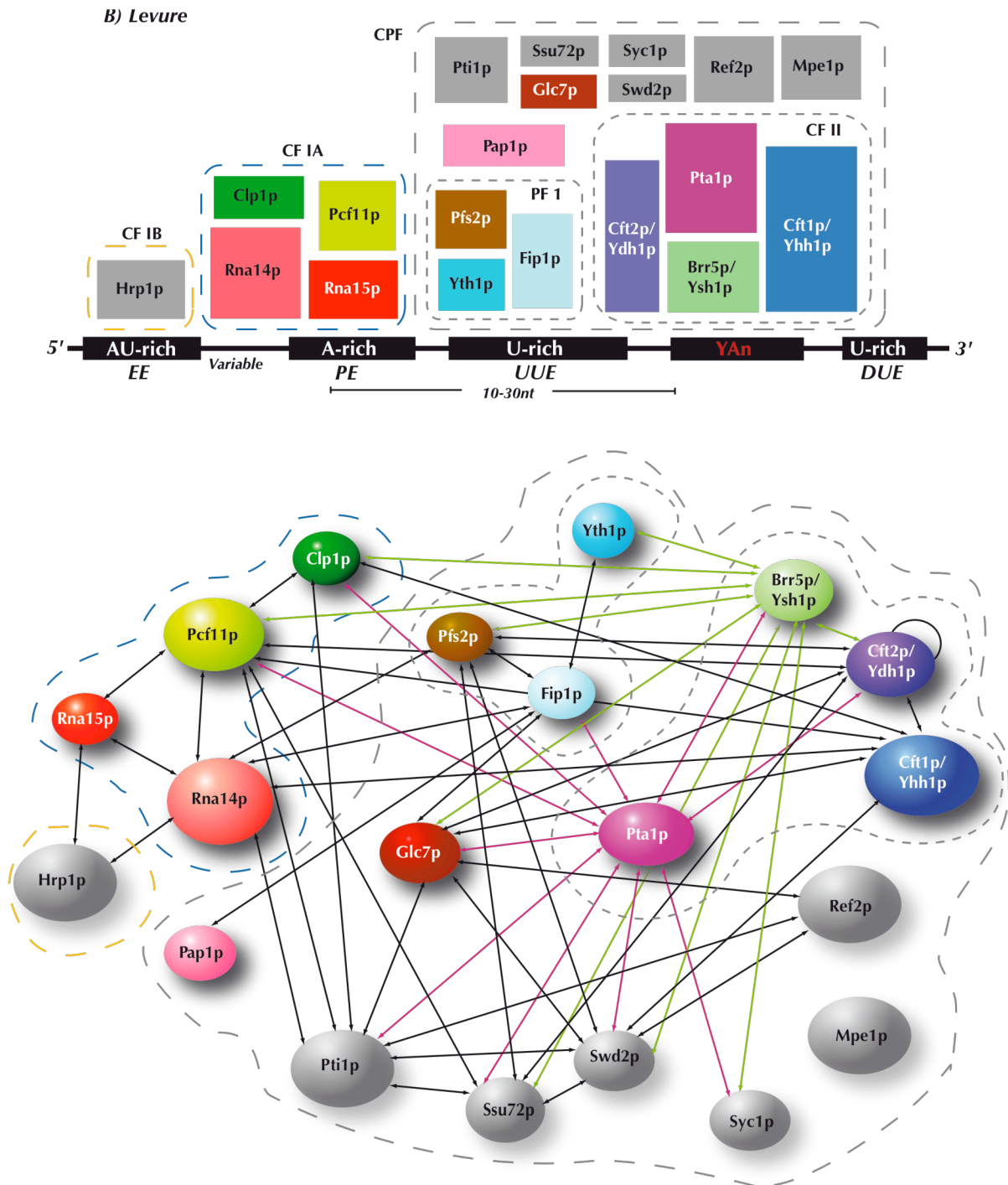


## A) Métazoaires



**Figure 10.** Interactions existantes entre les facteurs impliqués dans la maturation 3' chez les *mammifères*. Au dessus : rappel de la figure 5 où sont représentés les facteurs impliqués dans la maturation 3', En dessous, les interactions entre protéines sont représentées par des flèches. Les protéines ont été groupées dans leurs complexes correspondants.





**Figure 11.** Représentation des interactions connues chez la levure. Les flèches indiquent les capacités d'association entre protéines. Pour Brr5p/Ysh1p et Pta1p, leurs interactions sont marquées avec leurs couleurs respectives pour simplifier la lecture du schéma. Les pointillés délimitent le groupement de facteurs intégrant un complexe.

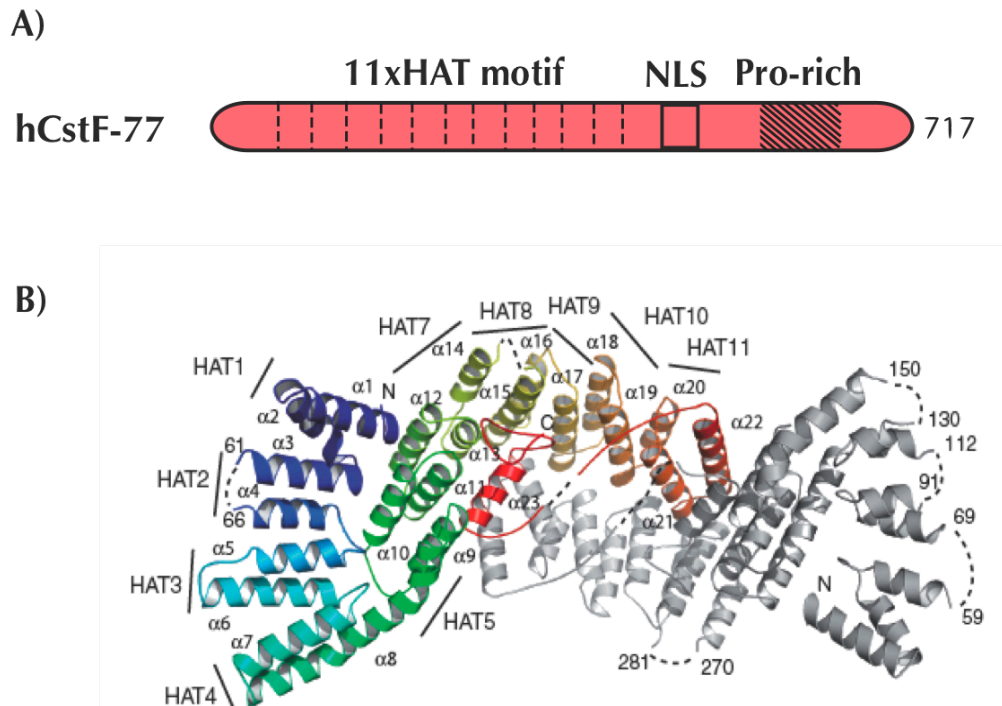
e.2.) Le complexe CstF

Le complexe CstF (Cleavage stimulation Factor) est constitué de trois protéines nommées CstF-77, CstF-64 et CstF-50 (Takagaki et al. 1990) (Gilmartin and Nevins 1991) en fonction de leur poids moléculaire respectif, 77, 64 et 50kDa. Ce complexe interagit, via la sous-unité de 64kDa, avec une séquence riche en G et U en aval du site de clivage et stabilise le complexe CPSF/ARN préformé. Un résumé des caractéristiques principales de chaque sous-unité est présenté ci-dessous.

- **CstF-77** : étant la sous-unité centrale de CstF maintient le complexe formé. CstF-77 est requis pour la réaction de clivage. Il a été mis en évidence que des mutations sur l'homologue de CstF-77 chez *Drosophila melanogaster*, *Suppressor-of-forked* ou *Su(f)*, impliquent des changements dans l'expression génique et l'utilisation de sites de polyadénylation alternatifs (O'Hare 1995) (Juge et al. 2000).

Cette protéine est principalement constituée d'un nombre variable (12 chez les mammifères) de séquences répétées appelées **HAT** (half a TPR). Ce type de séquences est similaire au motif TPR (Tetratricopeptide Repeat) (Preker and Keller 1998), un motif généralement associé aux interactions protéine-protéine. Des études biochimiques et la résolution de sa structure ont permis d'identifier deux sous-domaines dans le domaine HAT, **HAT-N** (du résidu 1 au 140 qui inclut les motifs HAT 1-5) et **HAT-C** (de l'acide aminé 241 au 549 comprenant les motifs du 6 au 12) (Bai et al. 2007) (Legrand et al. 2007) (*fig. 12*). Les deux structures résolues de CstF-77, l'une d'entre elles au sein du laboratoire, révèlent l'organisation en **homodimère** de son domaine HAT-C. L'architecture de CstF-77 corrobore, donc, les expériences réalisées en double hybride (Bai et al. 2007) et « far western blot » (Takagaki and Manley 2000), et aussi les résultats de microscopie électronique obtenus à partir de son homologue chez la levure, Rna14p (Noble et al. 2004). L'ensemble de ces données suggère que CstF-77 pourrait fonctionner comme un dimère dans une étape essentielle du processus de maturation.

Le domaine HAT est suivi d'une **région dite riche en Pro**. Des expériences en « far western » montrent la capacité de ce segment de la protéine d'interagir avec le domaine « **Hinge** » de **CstF-64** et le **WD-40 de CstF-50** (Takagaki and Manley 2000). L'interaction de CstF-77 et CstF-64 a été postérieurement confirmée par « pull-down » et ultracentrifugation analytique (Bai et al. 2007). De la même manière, la formation d'un hétérodimère entre leurs homologues chez la levure, Rna14p et Rna15p, a été également démontrée (Minvielle-Sebastia et al. 1994) (Gross and Moore 2001) (Noble et al. 2004). Ainsi, CstF-77 jouerait un rôle principal dans l'assemblage du complexe CstF. Par ailleurs, CstF-77 est la seule des trois protéines du complexe contenant un domaine **NLS** (Nuclear Localization Signal). Il a été proposé un modèle dans lequel CstF-77 permettrait l'entrée de CstF-64 et de Cst-50 dans le noyau via son NLS. Cette théorie suggère que la formation du complexe CstF a lieu dans le cytoplasme, avant l'internalisation. Ainsi, les trois protéines traverseraient en même temps la membrane nucléaire (Hockert et al. 2010).



**Figure 12.** *CstF-77.* A) Organisation générale de *CstF-77*. B) Structure de la protéine (adaptée de (Legrand et al. 2007)). Sur la figure seulement un des monomères est coloré de bleu à rouge de l'extrémité N-terminale à la C-terminale. Les motifs HAT sont indiqués. Le deuxième monomère est coloré en gris. Les pointillés représentent des régions désordonnées.

On peut affirmer que *CstF-77* joue un rôle de « pont », non seulement dans le complexe *CstF*, mais aussi avec d'autres facteurs qui interviennent dans le clivage et la formation de la queue poly(A). Ainsi, son domaine HAT-C est impliqué dans l'interaction avec le deuxième  $\beta$ -hélix («  $\beta$ -propeller ») de **CPSF-160**, la même région responsable de la reconnaissance du PAS (Murthy and Manley 1995) (Bai et al. 2007). Son association à **hFip1** (Kaufmann et al. 2004) et au **CTD de l'ARN polymérase II** a été également démontrée, néanmoins, elle interagit avec le CTD de façon moins efficace que *CstF-50* (McCracken et al. 1997). L'homologue de *CstF-77* chez *A. thaliana* (**AtCstF77**), s'associe aux protéines *AtCstF64* et *AtCPSF160* (Yao et al. 2002) (Hunt et al. 2008) et, à différence de son orthologue chez les mammifères, une interaction entre le domaine C-terminal de *AtCstF77* et **AtCPSF30** a été décrite (Hunt et al. 2008). De plus, ce même domaine possède la capacité d'interagir avec l'ARN, une caractéristique de *CstF-77* qui a été seulement identifiée chez les plantes.

La région C-terminale de cette sous-unité peut être phosphorylée sur la Ser 691, néanmoins, la fonction possible de cette phosphorylation reste inconnue (Ryan and Bauer 2008).

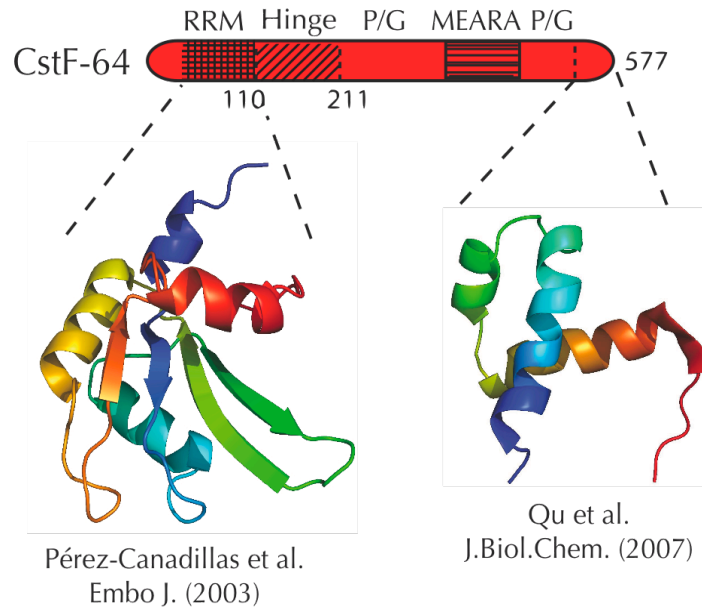
Par ailleurs, des analyses bioinformatiques et de biologie moléculaire ont démontré la présence d'une **variante plus courte de CstF-77**. Cette isoforme contient seulement les 103 premiers acides aminés et est formée grâce à l'existence d'un site de poly(A) intronique. Il n'a pas été défini si cette variante est fonctionnelle ou si elle arrive à traverser la membrane du noyau, mais elle pourrait jouer un rôle dans une régulation tissu-spécifique (Pan et al. 2006). Des transcrits de taille variable ont aussi été observés pour son homologue chez la levure, Rna14p (Minvielle-Sebastia et al. 1991).

La résolution de la structure de la version intégrale de cette protéine (à exception de l'extrémité C-terminale) nous a permis d'avancer dans la connaissance du mécanisme général de maturation 3'. Cependant, elle a aussi soulevé de nouvelles questions. Son organisation en dimère suggère une complexité majeure du processus, cependant, l'origine de cette complexité est aujourd'hui l'un des grands mystères. La résolution des structures de cette protéine interagissant avec d'autres facteurs, comme ces partenaires dans le complexe ou avec CPSF-160, pourrait aider à la compréhension du mécanisme.

- **CstF-64** : son existence a été connue rapidement grâce à son interaction avec l'ARN (Wilusz and Shenk 1988). Cette protéine contient un domaine **RRM** ou **RNA Recognition Motif** suffisant pour l'interaction avec une séquence d'ARN de préférence riche en **G/U** (Takagaki and Manley 1997). L'organisation du domaine RRM en complexe avec l'ARN a été déterminée par **Résonance Magnétique Nucléaire (RMN)** (*fig. 13*). La structure montre la présence d'un site d'union spécifique **UU** dans la séquence d'ARN. Ces résultats ont été corroborés par des études réalisées *a posteriori* (Perez Canadillas and Varani 2003) (Deka et al. 2005). Néanmoins, l'observation d'une surface d'interaction très flexible expliquerait la formation de complexes stables avec différents types de séquences toutes riches en **GU**. Des analyses bioinformatiques ont montré l'existence d'une **séquence DSE divisée en deux parties**, une riche en **UG** à 5-10 nucléotides en aval du site de clivage, et une autre riche en **U** à 15-25 nucléotides. À partir de ces résultats, un modèle a été proposé où CstF-64 interagirait avec les deux éléments séparément et séquentiellement. Dans cette hypothèse, la région riche en **UG** contribuerait au déplacement de l'hélice C (située à l'extrémité C-terminale du RRM) permettant l'exposition des feuillets  $\beta$  et l'interaction avec l'élément le plus éloigné riche en **U** (Salisbury et al. 2006).

Le RRM de CstF-64 est suivi de la région « **hinge** » (charnière) qui compte 100 résidus très conservés. Cette partie de la protéine est impliquée dans l'interaction avec **CstF-77** et avec la **symplektine** (Hatton et al. 2000) (Takagaki and Manley 2000). De ce fait, il est possible qu'il y ait une compétition entre ces deux protéines pour le domaine « hinge » de CstF-64 (Takagaki and Manley 2000). Ce domaine est aussi requis pour l'internalisation de la protéine dans le noyau, suggérant que son interaction avec CstF-77 est importante pour la localisation nucléaire (Hockert et al. 2010).

Son **domaine C-terminal** est aussi conservé. Sa structure montre une organisation en trois hélices alpha (Qu et al. 2007) (*fig. 13*). Cette région semble être responsable des interactions avec des facteurs de transcription et joue un rôle dans la régulation de ce processus (Proudfoot 2004).



**Figure 13.** Représentation schématique des domaines faisant partie de la protéine CstF-64. Sur la partie inférieure de la figure, les modèles structuraux des régions indiquées sont représentés dans un gradient de couleur, de bleu au rouge, de la région N-terminale à la C-terminale.

Les eucaryotes supérieurs possèdent une région caractéristique située entre l'extrémité C-terminale et le « hinge ». Cette région est composée d'un segment **riche en Pro/Gly** suivi de 12 répétitions d'un motif pentapeptide, **MEARA(A/G)** qui forment potentiellement une hélice  $\alpha$ . La longueur et la composition de cette région est variable entre les espèces et sa fonction reste inconnue (Colgan and Manley 1997).

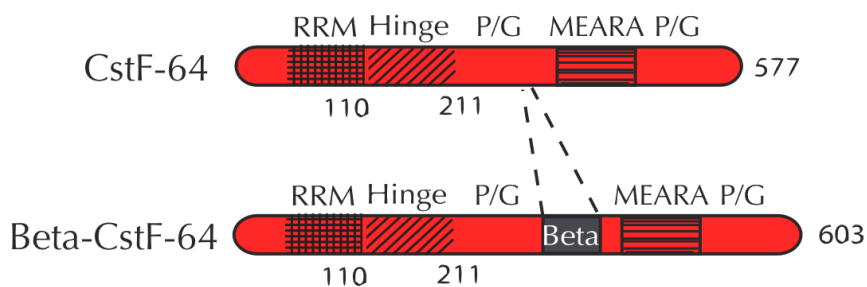
De plus, des modifications dans CstF-64 affectent la viabilité cellulaire. En effet, des mutations qui provoquent une altération dans la progression du cycle cellulaire sont critiques dans le développement de tumeurs. Une réduction du taux de protéine CstF-64 entraîne un arrêt du cycle cellulaire et provoque l'apoptose (Danckwardt et al. 2008) (*voir aussi Chapitre I.A.1.f*).

Par ailleurs, CstF-64 est un important **régulateur de la polyadénylation tissu-spécifique**. Il existe une variante de CstF-64 nommée  **$\tau$ CstF**.  **$\tau$ CstF** contient une substitution Pro $\rightarrow$ Ser dans le domaine RRM, 9 répétitions de la séquence MEARA à la place de 12 et deux insertions qui ne sont pas présentes sur CstF-64. En conséquence, la protéine  **$\tau$ CstF** contient 39 acides aminés de plus par rapport à CstF-64 et elle atteint un poids moléculaire supérieur. Elle s'exprime de manière importante dans les cellules mâles germinales et participe à la polyadénylation spécifique de ce type cellulaire et dans le processus de spermatogenèse (Wallace et al. 1999). La substitution dans son

motif RRM pourrait engendrer la reconnaissance d'une séquence d'ARNm plus spécifique d'un tissu déterminé (Monarez et al. 2007). Ainsi, cette variante a été associée à certains problèmes neurologiques et d'infertilité masculine (Dass et al. 2002) (Dass et al. 2007) (*voir aussi I.A.1.f*). A l'opposé de ces résultats, cette variante a été actuellement caractérisée comme un composant général de la machinerie de polyadénylation, étant donné qu'elle co-purifie avec l'ensemble de protéines du processus de maturation (Shi et al. 2009) ; des études fonctionnelles, structurales, et de localisation tissulaire devraient nous permettre d'éclaircir cette question.

De plus, d'autres variantes de CstF-64 ont été identifiées (Shankarling et al. 2009). Parmi toutes les isoformes, la présence de **βCstF-64** pourrait être soulignée. Dans cette isoforme, la région riche en Pro/Gly est interrompue par l'insertion d'un autre domaine (*fig 14*). βCstF-64 contribue à la diversité protéomique du système nerveux. Deux explications ont été proposées pour un tel effet. Dans un premier modèle, βCstF-64 reconnaîtrait des séquences de l'ARNm spécifiques du système nerveux ; dans le deuxième, l'insertion du domaine βCstF-64 pourrait provoquer une altération structurale qui affecterait son interaction avec CPSF ou avec d'autres composants de la machinerie de poly(A) (Dass et al. 2002) (Shankarling et al. 2009).

La résolution des structures de la sous-unité CstF-64 entière et de ses différentes isoformes contribuera à la compréhension de sa fonction dans le processus de maturation 3' et, plus globalement, de son implication dans le développement de maladies et la viabilité cellulaire.

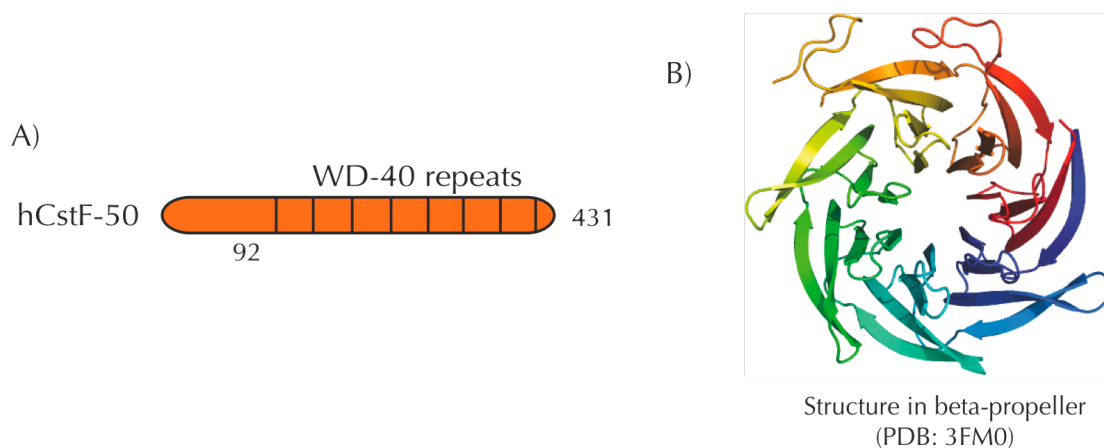


**Figure 14.** Différences entre **CstF-64** et **βCstF-64**. La région riche en Pro/Gly est interrompue par l'insertion d'un autre domaine dans la variante **βCstF-64**.



- **CstF-50** : est une protéine indispensable pour la réaction de clivage du pré-ARNm. Son homologue chez la levure n'a pas été identifié.

CstF-50 contient une **région N-terminale** et 7 motifs **WD-40** (fig. 15). Des expériences de « far Western blotting » ont montré l'interaction de CstF-50 sur elle-même. À partir de l'étude de mutants délétés de CstF-50, il a été observé que la zone **N-terminale** de la protéine permet la formation d'un **dimère** (Takagaki and Manley 2000). Ce domaine est, par ailleurs, impliqué dans l'interaction avec le **CTD de l'ARN pol II** (Fong and Bentley 2001). La délétion des premiers 95 acides aminés réduit l'interaction CstF-50/CTD suggérant que les motifs WD-40 ne sont pas essentiels pour cette association. L'activation du processus de maturation 3' du pré-ARNm par le CTD requière un contact entre le CTD et la sous-unité CstF-50. Ceci permet le couplage entre la terminaison de la transcription et la maturation 3'. L'interaction CstF-50/CTD participerait, de cette façon, au recrutement du complexe CstF à la machinerie de maturation 3'.



**Figure 15.** CstF-50. A) le schéma représentant les domaines de CstF-50 avec une région N-terminale suivis de sept répétitions du motif WD-40. B) on observe une structure typique en beta-propeller extraite à partir du PDB.

D'autres observations indiquent que la surexpression de la région N-terminale de CstF-50 a un effet de dominant négatif sur le processus. Cette surexpression inhibe la maturation 3' malgré la formation d'un complexe CstF intact. Il est probable que l'excès du domaine N-terminal de CstF-50 entre en compétition avec la protéine native pour son interaction avec le CTD et empêche ainsi le recrutement du complexe CstF dans la machinerie de la maturation (Fong and Bentley 2001).

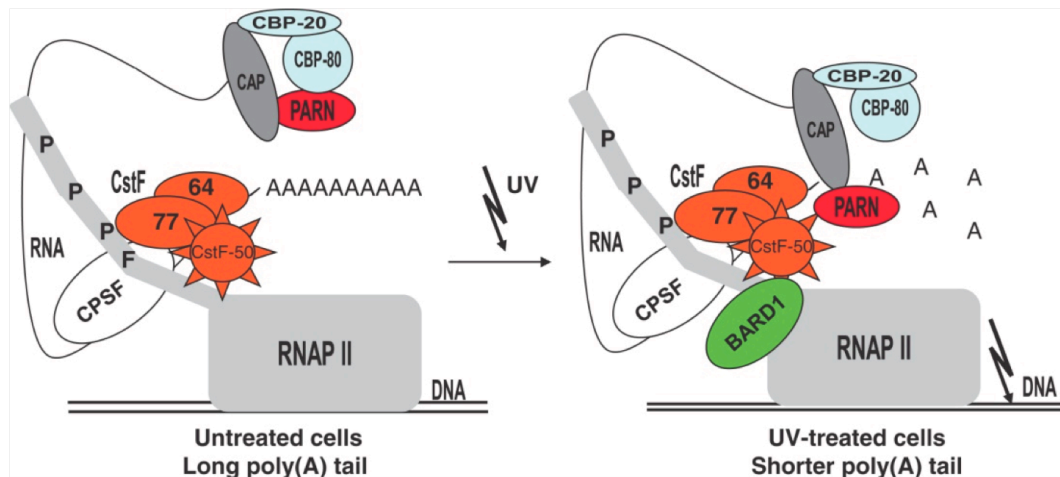
Pratiquement le reste de la séquence de CstF-50 est constitué de répétitions d'un motif potentiel d'interaction protéine-protéine appelé **WD-40** ou transducine. Des études structurales indiquent que les protéines de cette famille peuvent s'associer avec ses partenaires grâce à la dépression formée au centre de la structure en hélice- $\beta$ , ou «  $\beta$ -propeller » des motifs WD-40 (exemple en (Gaudet et al. 1996)) (fig. 15). En effet, la caractéristique principale des motifs WD-40 est la formation d'une architecture toroïdale entièrement  $\beta$ , avec 4-8 feuillets bêta disposés autour d'un axe central. Le

WD-40 est nécessaire pour son interaction avec **CstF-77** et la délétion des dernières répétitions du motif réduit l'association des deux (Takagaki and Manley 2000).

En outre, cette région participe à l'interaction avec **BARD1** (**BRCA1 Associated RING Domain 1**), un facteur de réplication/réparation de l'ADN (Kleiman and Manley 1999). **BRCA1** est une protéine associée aux cancers du sein et des ovaires (*voir aussi I.A.1.f*). L'association de BARD1 avec BRCA1 joue un rôle dans la maintenance de la stabilité du génome en réponse à des dommages produits sur l'ADN (Scully et al. 2004) (Irminger-Finger and Leung 2002). L'hétérodimère formé entre les deux protéines a une activité ubiquitine ligase (Brzovic et al. 2003). D'un autre côté, la participation de BARD1 dans le processus de transcription a été prouvée vu qu'elle s'associe à l'ARN pol II, à l'ARN helicase et à la protéine CtlP. BARD1 contient des domaines **BRCT** (**BRCA1 C-terminal repeats**) en tandem et une collection de **répétitions d'ankyrines** dans une région située plus proche de l'extrémité N-terminale. Les deux types de domaines sont connus pour leur implication dans des interactions protéine-protéine (Bork et al. 1997) (Mosavi et al. 2004).

L'interaction BARD1/CstF-50 s'effectue entre la région portant le WD-40 de CstF-50 et le « linker » situé entre la région BRCT et le domaine ankyrine de BARD1 (Edwards et al. 2008). L'assemblage **CstF-50/BARD1/BRCA1** provoque une **inhibition transitoire mais spécifique de la polyadénylation en réponse au dommage à l'ADN** (par traitement avec hydroxy-urée ou UV). De ce fait, il a été mis en évidence la **relation entre la maturation 3' des ARNms et la réparation de l'ADN/suppression de tumeurs** (Kleiman and Manley 2001) (Mirkin et al. 2008). L'utilisation de BARD1 comme un agent protecteur de tumeurs est envisagée par différents groupes de recherche (Danckwardt et al. 2008) (*voir I.A.1.f*).

Très récemment (Cevher et al. 2010), l'existence d'une interaction entre CstF-50 et **PARN** (nuclear poly(A)-specific ribonuclease), un facteur de déadénylation a été mise en lumière. Après un dommage à l'ADN, l'interaction PARN/CstF-50 aboutit à l'inhibition de la réaction de clivage couplée à une activation de la déadénylation. Lors du traitement aux UV, CstF-50 entre en compétition avec **CBP80** (cap binding protein 80) pour l'interaction avec PARN. Un modèle a été proposé selon lequel, en absence de dommage de l'ADN, CBP80 interagit avec PARN et inhibe son activité hydrolytique empêchant de cette façon la dégradation de l'ARNm. Après dommages à l'ADN, le complexe BRCA1/BARD1 est recruté pour inhiber le processus d'ubiquitinylation de l'ARN pol II. Le complexe **CstF/PARN** serait également formé pour assurer la dégradation des ARNms mal polyadénylés. Dans cette hypothèse, CBP80 serait dissociée de PARN pour laisser la place au complexe CstF-50/BARD1 (Cevher et al. 2010) (*fig. 16*).



**Figure 16.** Modèle du fonctionnement de la queue poly(A) après traitement de l'ADN par UV (adapté de (Cevher et al. 2010)). PARN interagit avec CstF à travers la sous-unité de 50kDa. Cette interaction active le processus de déadénylation. L'interaction de CstF-50 avec PARN et BARD1 met en évidence son rôle important dans le couplage maturation 3'-reparation de l'ADN.

CstF-50 interagit par ailleurs avec **Srm160** (the SR-related nuclear matrix protein of 160 kDa) mettant en évidence l'existence d'une nouvelle association entre le processus de maturation 3' des ARNm et la transcription par l'ARN Pol II (McCracken et al. 2003).

Une phosphorylation sur la **Tyr 367** de la sous-unité CstF-50 a été détectée dans deux lignées cellulaires de lymphome différentes, cependant, l'effet de cette phosphorylation est pour l'instant inconnu (Ryan 2007).

La connaissance de la structure de CstF-50 participerait à la compréhension des bases du processus de maturation 3' des ARNm. L'interaction avec le CTD est essentielle pour la relation entre la fin de la transcription et la polyadénylation. D'un autre côté, la connaissance de l'architecture du complexe avec BARD1 nous permettrait d'interpréter son rôle dans le développement de certains types de cancers.

### e.3) Le complexe CF IA

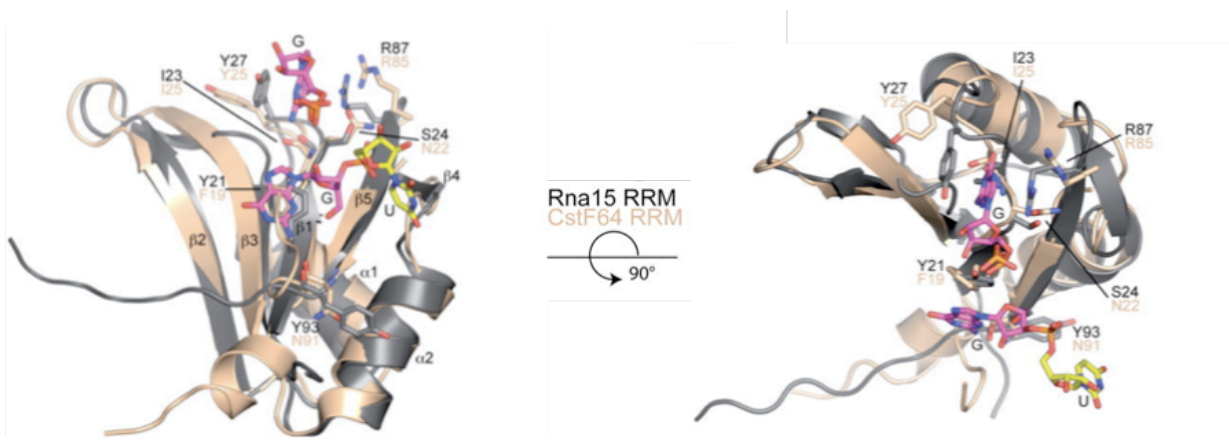
À la différence de CstF, ce complexe est composé de 4 facteurs, les homologues de CstF-77 et CstF-64, **Rna14p** et **Rna15p** respectivement (Minvielle-Sebastia et al. 1991) (Minvielle-Sebastia et al. 1994), et deux protéines supplémentaires, **Pcf11p** (Amrani et al. 1997) et **Clp1p** (Minvielle-Sebastia et al. 1997) dont leurs homologues chez les mammifères, hPcf11 et hClp1, sont associées dans le complexe CF IIm.

Rna14p et Rna15p, à différence de leurs homologues CstF-77 et CstF-64, sont impliquées dans les deux réactions, clivage et polyadénylation (Minvielle-Sebastia et al. 1994). Des mutation dans leurs gènes correspondants, donnent comme résultat la formation de transcrits où la queue poly(A) est absente. Ces transcrits sont rapidement dégradés dans le noyau (Torchet et al. 2002) (Minvielle-

Sebastia et al. 1994). Les deux protéines ont été localisées dans le noyau (Bonneaud et al. 1994), mais Rna14p a aussi été détectée dans la mitochondrie (Rouillard et al. 2000).

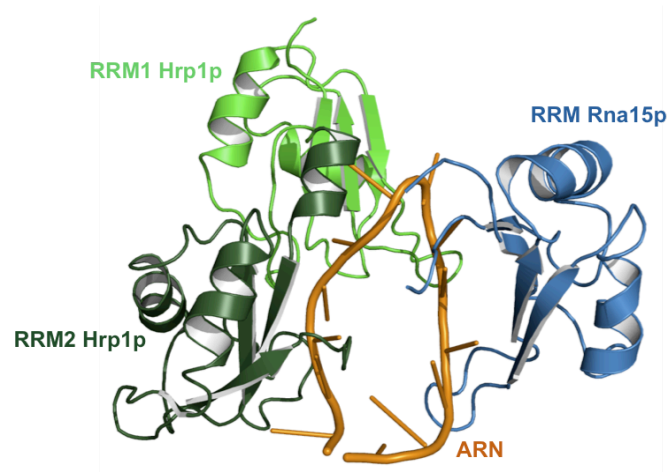
L'association directe de Rna14p et Rna15p a été mise en évidence par des nombreuses études biochimiques (Kessler et al. 1996) (Gross and Moore 2001) (Noble et al. 2004) (Legrand et al. 2007). Les techniques de microscopie électronique et d'ultracentrifugation analytique ont mis en évidence la formation d'un hétérotétramère compatible avec la présence de 2 copies de Rna14p et de Rna15p (Noble et al. 2004).

La protéine **Rna15p** possède une région d'interaction avec l'ARN, un domaine **RRM**, contenant les deux motifs classiques **RNP**. Le site de liaison à l'ARNm diffère par rapport à son homologue chez les mammifères, ainsi, CF IA est localisé en amont du CPF pendant que le complexe CstF est situé en aval du CPSF (Birse et al. 1998). Initialement, il a été montré que la protéine Rna15p, purifiée à partir de *Escherichia coli* en absence d'autres facteurs, reconnaissait une séquence **riche en U** (Minvielle-Sebastia et al. 1994). Récemment, la structure du domaine RRM de Rna15p a été résolue en interaction avec des séquences d'ARN **riches en GU**. Des études d'interaction en RMN montrent une préférence claire de ce motif pour ce type de séquences (Pancevac et al. 2010) (code PDB :2X1A). Le modèle structural présente un motif de reconnaissance de l'ARN similaire à celui trouvé pour CstF-64, la superposition des deux modèles montrant la forte conservation de ce domaine (*fig. 17*). Selon ces résultats, des séquences riches en U ou G/U autour du site de poly(A) constitueraient les éléments en *cis* d'interaction aussi bien pour CstF que pour CF IA. Une autre caractéristique commune aux deux protéines observée à partir de cette structure, est la capacité probable des résidus de la région C-terminale du RRM, de refermer le site d'interaction de l'ARN en absence de celui-ci (Perez Canadillas and Varani 2003) (Pancevac et al. 2010).



**Figure 17.** Adaptée de (Pancevac, 2010 #416). Superposition des modèles structuraux de Rna15p (16-111)-GUUGU et le domaine RRM de son homologue chez les mammifère, CstF-64 (PDB :1P1T). La protéine Rna15p est colorée en gris, CstF-64 est en beige.

Contrairement à ces études, d'autres travaux suggèrent que le RRM de Rna15p est capable d'interagir spécifiquement avec la séquence PE de l'ARNm **riche en A**, mais uniquement en présence d'autres protéines, spécialement de **Rna14p** et de **Hrp1p** (Gross and Moore 2001). Cette étude a été réalisée à partir des expériences de « cross-linking » entre Rna15p et un ARN *GAL7* 3'-UTR avec différentes délétions et mutations. En soutien de cette hypothèse, la résolution récente de la structure par RMN (Résonance Magnétique Nucléaire) de Hrp1p en complexe avec le motif RRM de Rna15p et un fragment d'ARN (UAUAUAUAAUAAU), montre la capacité de Rna15p de reconnaître des séquences riches en A correspondantes au PE (Leeper et al. 2010) (*fig. 18*).



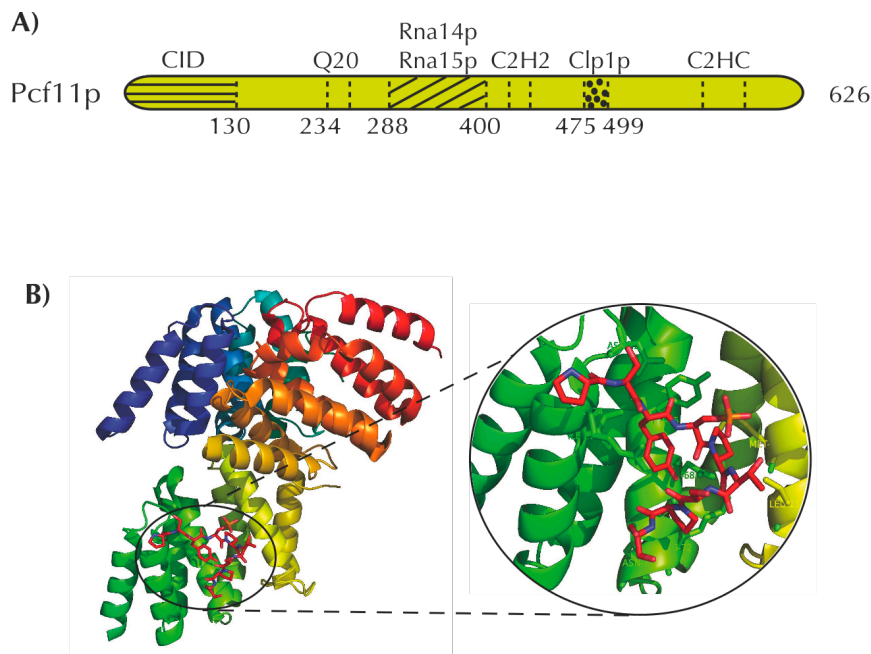
**Figure 18.** Modèle structural du complexe ternaire Rna15p/Hrp1p/ARN (Leeper et al. 2010). La protéine Rna15p reconnaît la séquence d'ARN AAUAAU. De plus, des contacts entre les deux sous-unités de CF I, Rna15p et Hrp1p, sont établis au moins en présence d'ARN (code PDB: 2km8).

La **région « Hinge »** identifiée en CstF-64 est aussi présente en Rna15p, et, dans les deux cas, elle sert à l'interaction avec CstF-77/Rna14p, cependant, les détails de cette interaction restent inconnus. En ce qui concerne la **partie C-terminale** de Rna15p, son implication dans l'interaction avec Pcf11p a été démontrée par des expériences de « pull down » (Qu et al. 2007).

Le 24% de la séquence de **Rna14p** et de CstF-77 est identique (Takagaki and Manley 1994). Rna14p, comme son homologue chez l'homme, présente des motifs **HAT** impliqués dans des interactions protéine-protéine et joue un rôle important dans la connexion des différents facteurs de la maturation. A partir de la structure de son homologue CstF-77 (Bai et al. 2007) (Legrand et al. 2007) et des études précédentes (Noble et al. 2004), cette protéine serait vraisemblablement organisée en dimère. La **région C-terminale** de Rna14p est moins conservée. Cependant, son importance a été mise en évidence lorsque des mutations des résidus de cette zone de la protéine provoquaient un phénotype thermosensible chez la levure (Bloch et al. 1978).

Rna14p interagit avec un CTD de l'ARN pol II préférentiellement phosphorylé (Sadowski et al. 2003). De plus, elle permet l'association des trois complexes qui participent dans la maturation 3' chez les levures, CF IA, CF IB et CPF via l'interaction avec plusieurs facteurs (*voir tableau 3 et fig. 11*).

**Pcf11p** contient un domaine d'interaction avec le CTD de l'ARN pol II (**CID** ou CTD interacting domain) composé de 130 résidus dans sa région N-terminale. L'interaction se produit avec un CTD phosphorylé sur la Ser<sub>2</sub> de façon préférentielle (Sadowski et al. 2003) (Barilla et al. 2001). La structure du complexe CTD/Pcf11p a été résolue (Meinhart and Cramer 2004) (*fig.19*). Des mutations dans le CID entraînent une terminaison de la transcription erronée sans que la polyadénylation ne soit affectée (Noble et al. 2005) (Zhang et al. 2005). De plus, le CID peut interagir avec l'ARN. Cette liaison affecte également à la terminaison de la transcription (Zhang et al. 2005). En effet, les zones du CID impliquées dans l'interaction avec l'ARN et avec le CTD se chevauchent et ceci engendre une compétition entre les deux pour leur association. Cette compétition pourrait jouer un rôle dans la libération des facteurs de la maturation 3' de la machinerie de la pol II (Hollingworth et al. 2006).

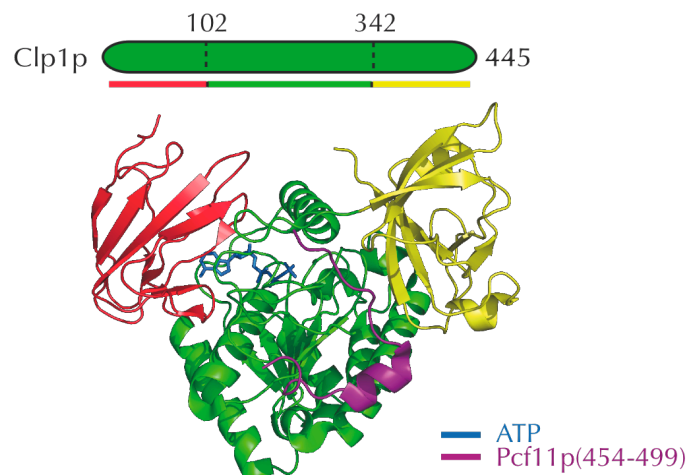


**Figure 19.** A) Représentation des domaines de la protéine **Pcf11p** chez la levure. B) A gauche, le modèle structural de **Pcf11p** (avec 3 unités asymétriques) interagissant avec le **CTD** (X répétition du motif YSPYSYS). A droite, un détail de l'interaction. La protéine est colorée avec un gradient de bleu à rouge, de la région N-terminale à la C-terminale.

Le CID est suivi d'une extension de 20 résidus de glutamine **Q<sub>20</sub>** (Pcf11p 234-253) et d'une région postérieure impliquée dans l'interaction avec Rna14p/Rna15p (Gross and Moore 2001). La zone de Pcf11p comprise entre les acides aminés 477 à 499 est suffisante pour l'interaction avec Clp1p et est entourée de **doigts de zinc** (C<sub>2</sub>H<sub>2</sub> et C<sub>2</sub>HC sur la figure) (Noble et al. 2007). Pcf11p interagit avec d'autres protéines, Mpe1p, Swd2p, Ssu72p, Pti1p, Cft2p/Ydh1p, Brr5p/Ysh1p, Cft1p/Yhh1p et Pta1p mais les régions impliquées dans ces interactions sont pour le moment inconnues (Kyburz et al. 2003) (Uetz et al. 2000) (Nedea et al. 2003) (Vo et al. 2001) (He et al. 2003) (Skaar and Greenleaf 2002) (Dichtl et al. 2004) (*tableau 3 et figure 11*).



Clp1p comporte un **motif Walker A** (Walker et al. 1982) typique de protéines qui interagissent avec l'ATP. Clp1p est associée au complexe CF IA via Pcf11p, mais elle n'interagit pas avec Rna14p et Rna15p. La **structure de Clp1p** a été résolue en complexe avec l'ATP et avec une région minimale de Pcf11p indispensable pour l'interaction (Noble et al. 2007) (*fig. 20*). Les régions impliquées dans la liaison avec l'ATP forment un motif « **P-loop** » (ou « phosphate binding loop »)/**Walker A**. Cette structure montre une forte similitude par rapport à d'autres ATPases ; malgré cela, une telle activité n'a pas pu être démontrée pour cette protéine. Dans le modèle proposé, l'association CF IA-CPF serait nécessaire pour stimuler l'activité ATPase de Clp1p (Noble et al. 2007). La structure a permis aussi d'établir les bases moléculaires de l'interaction Clp1p-Pcf11p. Le site de liaison à Pcf11p est localisé dans le domaine central de Clp1p.



**Figure 20.** Structure du facteur Clp1p. La région N-terminale est colorée en rouge, la C-terminale en jaune et le reste de la protéine en vert. Pcf11p (454-499) est représenté en violet et l'ATP en bleu. Ces deux derniers, interagissent avec la région intermédiaire de Clp1p.

#### *e.4) Le complexe CstF chez les plantes*

Les détails du complexe chez *Arabidopsis* restent moins connus. Néanmoins, certaines caractéristiques des facteurs impliqués sont similaires à celles du complexe CstF de mammifères. Par exemple, AtCstF-64 contient aussi un domaine RRM qui lie le même type de séquences de pré-ARNm. La différence la plus importante entre le complexe CstF chez les mammifères et chez les plantes est basée sur les relations entre leurs partenaires. L'interaction entre AtCstF-77 et AtCstF-64 a été mise en évidence, cependant, aucune preuve de l'interaction AtCstF-50/AtCstF-77 n'a été trouvée. D'ailleurs, il n'est pas sûr qu'AtCstF-50 fasse partie d'un complexe CstF comparable à celui de mammifères. Il a été montré qu'elle interagit avec CPSF, AtCstF-64 et PAPS, suggérant l'existence d'une fonction complètement différente de cette protéine chez plantes (Yao et al. 2002).

**f.) Implication de la maturation 3' dans différentes maladies**

Un certain nombre de maladies trouvent leur origine dans l'altération directe d'un facteur de la maturation 3' des ARNm et reflètent ainsi l'importance de la précision de la machinerie de polyadénylation. Des variations dans les facteurs impliqués, ainsi que dans les séquences de pré-ARNm reconnues par ceux-ci, dérèglent le processus et entraînent, dans certains cas, l'apparition des maladies. D'un point de vue médical, autant la perte que le gain d'une fonction ont un impact important. En effet, des troubles dans l'équilibre général du processus, l'excès d'un facteur, la carence d'un autre, leurs altérations ou des modifications dans la séquence du pré-ARNm, peuvent provoquer le développement des maladies.

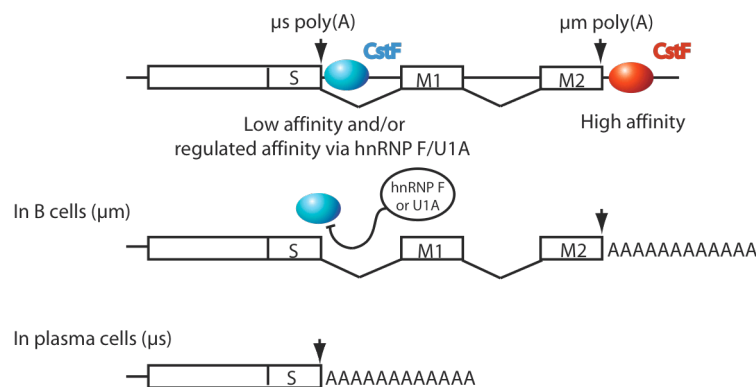
Dans cette section, des exemples des pathologies humaines produites par une modification du processus de maturation 3' sont présentés. Je me suis focalisée sur celles qui sont en relation avec les facteurs constituant l'objet principal de cette thèse, cependant beaucoup d'autres maladies y sont connectées.

Dans une section antérieure j'ai cité l'existence d'une interaction entre la sous-unité CstF-50 et la protéine **BARD1**. En effet, l'interaction CstF/BARD1/BRCA1 promeut une inhibition spécifique de la polyadénylation en réponse aux dommages à l'ADN, mettant en évidence la présence d'une relation entre la maturation 3' de l'ARNm et la réparation de l'ADN/suppression de tumeurs (Kleiman and Manley 2001). Différentes mutations de BARD1 ont été trouvées dans certains cas de **cancer du sein**, de **tumeurs utérines** et de **tumeurs ovariennes**. La mutation de BARD1 dans ces derniers est connue, il s'agit de la mutation **BARD1-Gln564His** (Irminger-Finger and Leung 2002) (Danckwardt et al. 2008). Cette mutation réduit son affinité pour CstF-50 et la rend inactive dans l'inhibition de la polyadénylation. L'utilisation de BARD1 comme agent protecteur contre l'apparition de tumeurs a été envisagée par plusieurs équipes de recherche.

D'autre part, il existe une variante de BARD1, dénommée délétion **ΔRIN**, qui contient une délétion/interruption des exons 2 à 6 codants pour une grande partie du domaine RING et la totalité du domaine ankyrine. Elle a été détectée dans le **cancer du sein**. La mutation est aussi présente dans les lignés cellulaires normales dérivées du sein, mais dans ces cas-là les niveaux d'expression de cette protéine mutée sont très bas. Il reste à déterminer si cette présence est due à la réponse aux dommages de l'ADN ou s'il existe une différence pendant la régulation du cycle cellulaire. En effet, BARD1ΔRIN peut entrer en compétition avec BARD1 pour l'interaction avec CstF-50. Cette compétition va inhiber la formation du complexe CstF/BARD1/BRCA1. Alternativement, il a été proposé un modèle dans lequel BARD1ΔRIN empêcherait l'entrée de CstF-50 dans le noyau (Tsuzuki et al. 2006).



Par ailleurs, **CstF-64** est un important **régulateur de la polyadénylation tissu-spécifique**. La sélection d'un site poly(A) alternatif représente un mécanisme régulateur important et conservé pour un contrôle spatial et temporel de l'expression. L'importance de la polyadénylation alternative a été corroborée quand il a été montré qu'à peu près la moitié des ARNm humains contiennent plusieurs sites de polyadénylation (Tian et al. 2005) (Yan and Marr 2005). La régulation de l'expression de la **chaîne lourde de l'IgM** pendant la différenciation de lymphocytes B constitue un des premiers exemples de ce type de phénomène. Cette régulation dépend des variations dans la concentration de la sous-unité CstF-64. Selon le modèle proposé (Takagaki and Manley 1997) (Danckwardt et al. 2008), pendant la différenciation de cellules B, une augmentation de CstF-64 permet une élévation du taux d'expression de la « forme sécrétée » de la chaîne lourde de l'IgM par rapport à la « forme membranaire ». Ceci se produit grâce à l'activation d'un site spécifique de poly(A) ( **$\mu$ s poly(A)**) situé plus en amont qu'un autre site appelé  **$\mu$ m poly(A)**. Ainsi, une quantité plus importante de CstF-64 favorise la production de la forme sécrétée (*fig. 21*). Comme représenté sur la figure, U1A et hnRNP F peuvent intervenir aussi dans le processus inhibant l'interaction de CstF-64 avec le site  $\mu$ s (Phillips et al. 2004) (Veraldi et al. 2001).



**Figure 21.** Adaptée de (Danckwardt et al. 2008). Dans l'exemple représenté, le site de plus forte affinité pour l'IgM de membrane ( $\mu$ m) est utilisé pour les cellules B quand la concentration de CstF-64 est faible ou quand une concentration élevée de hnRNP ou U1A inhibent l'union de CstF-64 au site ( $\mu$ s). Sur la figure, S correspond à l'exon pour l'IgM sécrétée et M1 et M2 à ceux de l'IgM de membrane.

En outre, une variante de cette protéine ( **$\tau$ CstF-64**) a été décrite dans la *section Chapitre I.A.I.e.I*. Cette isoforme, s'exprimant de façon préférentielle dans les cellules germinales et dans le cerveau, est associée à certains **problèmes neurologiques** ainsi qu'à des **problèmes d'infertilité masculine** (Dass et al. 2002). Il a été montré qu'une altération dans le gène codant pour  **$\tau$ CstF-64** entraîne une erreur dans la spermatogenèse.

De plus, il est observé que des modifications de CstF-64 peuvent affecter la **viabilité cellulaire**. En effet, les mutations qui provoquent une altération dans la progression du cycle cellulaire sont des événements critiques dans le **développement de tumeurs**. Dans un cas normal, le taux

d'expression de CstF-64 augmente cinq fois pendant le cycle cellulaire entre la phase  $G_0$  et S, de cette façon, il y aura une plus grande quantité de complexe CstF formé à cette étape (Martincic et al. 1998). L'augmentation du CstF entraîne une accumulation d'ARNms avec une queue poly(A) dans les cellules en sortant de la phase  $G_0$  pour rentrer dans la phase S (Getz et al. 1976). Une réduction de CstF-64 peut mener à un **arrêt du cycle cellulaire** dans la phase  $G_0/G_1$  et provoquer **l'apoptose** (Takagaki and Manley 1998) (Danckwardt et al. 2008).

Des travaux récents (Rozenblatt-Rosen et al. 2009) ont mis en évidence un autre lien entre CstF et la formation de tumeurs à travers l'association de **Cdc73** (la parafibromine) avec les complexes CstF et CPSF. Cdc73 est une protéine impliquée dans la suppression de tumeurs. Elle est inactivée dans le cas de **tumeurs héréditaires et sporadiques des parathyroïdes** et dans le **syndrome d'hyperparathyroïdisme associé aux tumeurs de la mâchoire (HPT- JT)**. Ces tumeurs seraient provoquées par une mutation de la protéine Cdc73 qui entraînerait une modification de la structure de la chromatine et/ou une altération dans la maturation 3' où CstF/CPSF agirait comme intermédiaires.

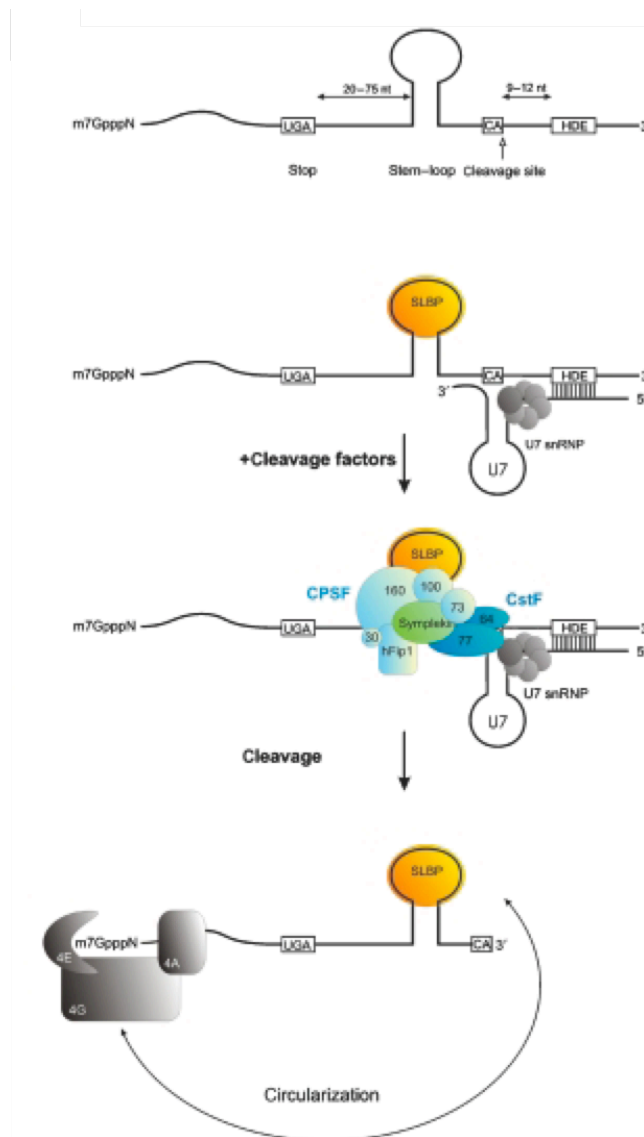
Finalement, des altérations dans les **éléments cis** peuvent provoquer également l'apparition de pathologies. Un exemple concernant ce sujet de thèse est une modification dans la région **DSE**. Le signal de poly(A) de la **prothrombine** humaine présente de façon générale, une architecture non usuelle dans ses séquences USE et DSE (la région d'interaction avec le facteur CstF-64). En effet, le pré-ARNm de la prothrombine contient un DSE pauvre en U en plus d'un site de clivage peu efficace. L'addition de résidus U sur le site DSE facilite son interaction avec CstF-64 ce qui provoque des troubles dans le processus de maturation 3' et dans l'expression de la prothrombine. Ce type de mutation a été observé chez des patients atteints de **thrombophilie** (Danckwardt et al. 2004; Danckwardt et al. 2006).

### **g.) Maturation 3' des pré-ARNms d'histones**

Chez les vertébrés, il existe deux familles de gènes d'histones. La première contient des gènes qui codent pour des histones appelées majeures ou **réplication-dépendantes**. L'expression de ces histones est massive et leur incorporation à l'ADN s'effectue uniquement au cours de la phase S du cycle cellulaire lors de la réplication d'ADN. Ce groupe de gènes produit des ARNms **non polyadénylés** (Schumperli 1988). La deuxième famille correspond aux gènes qui codent pour des histones qualifiées de histones de réparation ou **réplication indépendantes**. Leurs séquences diffèrent parfois très peu des histones majeures et leur expression s'effectue à un niveau basal tout au long du cycle cellulaire. Ce type de gènes est transcrit en ARNms polyadénylés (Wang et al. 1996) (Wells and Kedes 1985).

Pour la maturation 3' des ARNms **d'histones réplication-dépendantes**, des éléments de séquence et des facteurs en *trans* sont requis. Le signal de maturation 3' est composé d'une **tige-**

**boucle** (« stem-loop ») très conservée et localisée en amont du site de clivage et d'un segment riche en purines en aval appelé **HDE** (histone down-stream element) (Schumperli 1988) (Muller and Schumperli 1997). Le HDE est reconnu, par appariement de bases, par l'extrémité 5' de l'U7 snRNA (small nuclear RNA). Ce snRNA est incorporé à la machinerie comme composant d'une ribonucléoprotéine (RNP), **U7 snRNP** (Mowry and Steitz 1987) (Cotten et al. 1988). La tige-boucle interagit avec **SLBP** (stem-loop binding proteine) qui participe dans le processus et stabilise, avec le facteur **eIF4G**, la « pseudo-circularisation » de l'ARNm mature (Dominski and Marzluff 1999). Le clivage a lieu après un dinucléotide **CA** de la séquence du pré-ARNm, situé à une dizaine de nucléotides en amont du duplex ARN U7/HDE (Bond et al. 1991) (Scharl and Steitz 1994). Des travaux récents ont démontré que la protéine capable d'effectuer la réaction de clivage, **CPSF-73**, est au même temps l'exonucléase chargée de la dégradation de l'ARN clivé en 3' ou DPC (downstream cleavage product) (Yang et al. 2009) (*fig. 22*).



**Figure 22.** Adaptée de (Danckwardt et al. 2008). Etapes dans la maturation 3' d'histones.

La différence fondamentale entre la maturation 3' avec formation d'une queue poly(A) et la synthèse des ARNm matures d'histones a été établie. Vraisemblablement, la première se réalise en deux étapes et est couplée à d'autres processus de l'expression génique, tandis que la maturation des ARNm d'histones réplication-dépendantes requiert seulement une réaction de coupure endonucléolytique et semble être incompatible avec la réaction d'épissage et indépendante de la transcription (Pandey et al. 1990) (Adamson and Price 2003).

Malgré toutes ces variations, les deux mécanismes partagent certains des facteurs impliqués. Ainsi, le cœur catalytique de la réaction de clivage comportant les complexes **CPSF** et **CstF** est commun (Kolev and Steitz 2005). En effet, les cinq sous-unités du complexe **CPSF**, dont **CPSF-73** qui remplit la fonction d'endonucléase (Dominski et al. 2005), deux des trois protéines de **CstF** (**64kDa** et **77kDa**) et la **Symplekine** sont impliquées dans la réaction de clivage des pré-ARNm d'histones. Ces similarités suggèrent une origine commune.

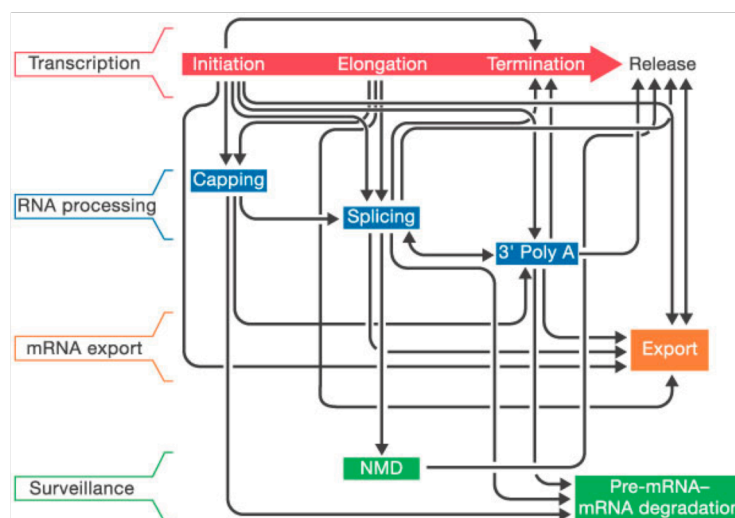
#### **h.) Maturation chez les procaryotes**

Deux types de PAPs ont été identifiées chez *Escherichia coli*. Aucune des deux ne présente une ressemblance de séquence par rapport à l'enzyme PAP d'eucaryotes (Cao and Sarkar 1992; Cao et al. 1996), mais **PAP1** contient un site catalytique nucléotide transférase (Martin and Keller 1996). Elle est capable d'ajouter une queue de poly(A) de moins de 20 nucléotides. Dans le modèle proposé, **PAP1** adényle tous les ARNs et des exoribonucléases sont chargées de raccourcir ou éliminer la queue formée quand cela est nécessaire. Ces enzymes ne sont pas essentielles pour le développement des bactéries, mais il a été observé que la polyadénylation promeut la dégradation des fragments d'ARNm ce qui permettrait un contrôle de la concentration intracellulaire des ARNs régulateurs (Zhao et al. 1999).

## 2. Connexion entre la maturation 3' et d'autres processus

Les étapes principales de la biogenèse des ARNm, transcription, formation de la coiffe, clivage/polyadénylation et épissage, peuvent être indépendamment reconstitués *in vitro*, mais dans le noyau, ces processus ont lieu au même temps et sont intimement connectés. L'interdépendance entre les étapes améliore la précision et l'efficacité du processus général (fig. 23).

La connexion entre la transcription et les étapes distinctes de la maturation peut être établie de différentes manières. La façon la plus simple est à travers les **interactions** entre les facteurs qui interviennent dans les différents processus. Un exemple clair à ce sujet est le rôle du domaine carboxy terminal (**CTD**) de l'ARN pol II qui interagit avec des nombreuses protéines impliquées dans la maturation (Phatnani and Greenleaf 2006). La connexion avec le processus de transcription peut être aussi du type **allostérique**, comme par exemple l'activation de l'enzyme guanylyltransférase par le CTD pendant la formation de la coiffe (Ho and Shuman 1999). Finalement, un **couplage cinétique** entre la transcription et d'autres processus a été observé. En effet, le taux de synthèse d'ARNm et les pauses effectuées par l'ARN pol II peuvent affecter l'épissage des pré-ARNms (Eperon et al. 1988) (Roberts et al. 1998) (Kornbliht et al. 2004).



**Figure 23.** Adaptée de (Maniatis and Reed 2002). Connexion entre les différents processus. À gauche on apprécie les étapes de l'expression génique. « Release » représente la libération des ARNm matures de la machinerie de transcription et « NMD » (*nonsense-mediated decay*) la dégradation des ARNs non-codants.

Les flèches indiquent les connexions physiques et/ou fonctionnelles entre les différents processus.

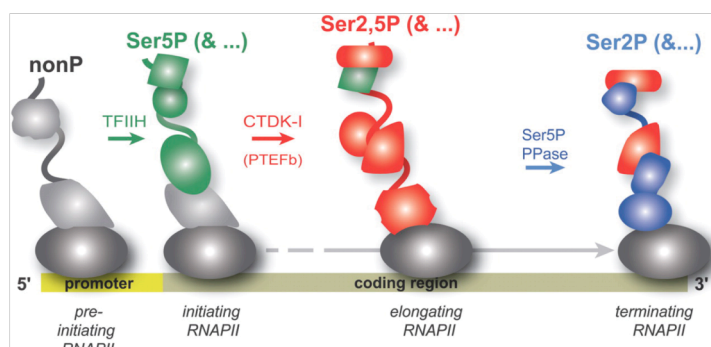
L'exécution des différents mécanismes de façon **co-transcriptionnelle**, permet un contrôle général du processus. Cette régulation prend forme lorsqu'il existe, par exemple, une compétition des deux facteurs pour un même site d'interaction de l'ARN ou d'une troisième protéine. Ce mécanisme permet de réguler l'ordre des événements (Perales and Bentley 2009). Le principal régulateur dans la globalité de la synthèse des ARNm matures est le CTD de l'ARN pol II. Cet élément donne à l'enzyme la capacité de participer aux différents processus « co-transcriptionnels ».

### a.) Rôle du CTD

L'ARN pol II est composée de 12 sous-unités appelées **Rpb1-Rpb12**. Les différentes sous-unités sont très conservées structurellement et fonctionnellement de levures aux mammifères (Cramer 2002). **Rpb1**, la plus grande sous-unité, contient un **domaine carboxy-terminal (CTD)** caractéristique (Corden et al. 1985). Le CTD fonctionne comme une zone flexible d'interaction pour les protéines qui entrent en contact avec l'ARN pol II. La longueur du CTD complètement déplié correspond à plusieurs fois le diamètre du cœur de la polymérase (Meinhart et al. 2005) ce qui confère un espace d'interaction très vaste pour l'union des multiples partenaires.

Le CTD est constitué chez les mammifères de **52** répétitions en tandem de l'heptapeptide **Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>**. Le numéro de répétitions varie entre les espèces. Chez la levure, le CTD est composé de 26 répétitions, il y a 32 chez les nématodes et 45 chez la drosophile. *In vitro*, le CTD intervient dans l'activation des trois étapes du processus de maturation (5', 3' et épissage). En outre, il a été montré *in vivo*, que sa délétion inhibait non seulement la terminaison de la transcription, mais aussi la maturation 3' et l'épissage (McCracken et al. 1997).

Les trois Ser faisant partie de l'heptapeptide sont sensibles à la **phosphorylation** (Zhao et al. 1999) (Egloff et al. 2007). Des kinases et phosphatases agissent sur les différentes positions d'une façon coordonnée durant les phases d'initiation, d'élongation, de terminaison de la transcription et de maturation des ARNm (Meinhart et al. 2005) (Phatnani and Greenleaf 2006) (*fig. 24*). Certaines de ces enzymes ont été identifiées. Ainsi, **Cdk7** (l'homologue de **Kin28** chez la levure) et **Cdk9** (**Ctk1** chez la levure) promeuvent la phosphorylation de la Ser<sub>5</sub> et la Ser<sub>2</sub> respectivement (Peterlin and Price 2006) (Komarnitsky et al. 2000). **Fcp1p** a été identifiée chez *S. cerevisiae* comme la phosphatase chargée de déphosphoryler préférentiellement la Ser<sub>2</sub> (Cho et al. 2001), tandis que **Ssu72p** chez la levure et **SCP** chez les mammifères déphosphorylent la Ser<sub>5</sub> (Krishnamurthy et al. 2004) (Yeo et al. 2003).



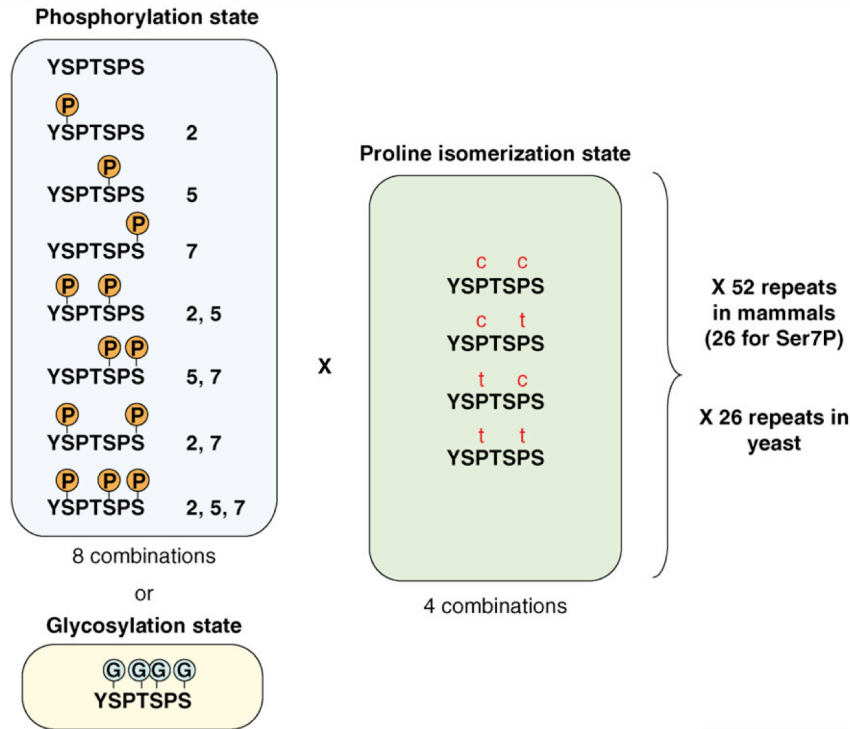
**Figure 24.** Adaptée de (Phatnani and Greenleaf 2006). Les différentes phosphorylations du CTD dictent quels facteurs vont s'associer à l'ARN pol II. L'ARN pol II est représentée par un ovale gris. Le CTD est coloré différemment dans les étapes distinctes pour montrer la phosphorylation différentielle : gris pour le CTD non-phosphorylé, vert pour la phosphorylation en Ser<sub>5</sub>, rouge pour Ser<sub>2</sub> et Ser<sub>5</sub> et bleu pour Ser<sub>2</sub>.

La combinaison de phosphorylations dynamiques sur la Ser<sub>2</sub>, Ser<sub>5</sub> ou la Ser<sub>7</sub> permet de réguler finement le mécanisme. Ainsi, il a été suggéré que les protéines participant à la maturation 3' vont reconnaître de façon préférentielle un CTD avec un groupe phosphate sur la Ser<sub>2</sub> (McCracken et al. 1997) (Barilla et al. 2001) (Maniatis and Reed 2002) (Proudfoot et al. 2002) (Kyburz et al. 2003). L'interaction mieux caractérisée entre une de ces protéines et le CTD correspond à l'association Pcf11p/CTD et, comme il a été décrit précédemment, la sous-unité **Pcf11p** de CF IA s'associe préférentiellement au CTD phosphorylé en Ser<sub>2</sub> (Licatalosi et al. 2002) (Noble et al. 2005). La phosphorylation en Ser<sub>5</sub> serait associée aux étapes initiales de l'expression génique étant donnée qu'elle promeut la libération du promoteur, la formation du complexe d'élongation et la maturation en 5' (Proudfoot 2004). Le rôle de Ser<sub>7</sub> a été observé dans le recrutement du complexe « integrator » aux snRNA (Egloff et al. 2007).

D'un autre côté, il a été décrit que **différentes régions** du CTD peuvent avoir des fonctions distinctes (Corden 1990). Bien que ce ne soit pas essentiel, certaines variations dans le peptide répété favorisent l'interaction avec différents facteurs. Ainsi, chez les mammifères, la région comprise entre les répétitions 27 et 52 inclut souvent une **K** (Lys) ou une **T** (Thr) à la place de Ser<sub>7</sub> et ceci aiderait à son association avec **CstF-50**. En outre, chez les vertébrés, le **motif C-terminal** de 10 acides aminés situé après la série d'heptapeptides, serait essentiel pour maintenir les niveaux nécessaires de transcription, de clivage du pré-ARNm et d'épissage (Fong et al. 2003).

Une autre caractéristique du CTD est sa capacité **d'isomérisation**. En effet, les deux Pro faisant partie de l'heptapeptide peuvent se trouver en conformation *cis* ou *trans* (Buratowski 2003). Des études effectués par RMN (Résonance Magnétique Nucléaire) sur l'interaction Pcf11p/CTD ont montré une préférence de Pcf11p pour un CTD en forme *trans* (Noble et al. 2005). Ces résultats ont donné les premières évidences sur la participation de l'isomérisation dans la régulation du processus.

Par conséquent, au moins trente deux états différents peuvent être établis pour une seule répétition du CTD (Buratowski 2003) (Egloff and Murphy 2008) (*fig. 25*). Le « code » du CTD détermine quelle étape de la transcription aura lieu à un moment précis. Ainsi, le rôle régulateur du CTD est crucial dans l'organisation générale du processus.



**Figure 25.** Le « code » du CTD (adaptée de (Egloff and Murphy 2008)). Les 64 phosphorylations potentielles combinées avec les 16 formes d'isomérisations génèrent un « code » qui rend possible la régulation des processus. Ce code permet une réduction des protéines interagissant au même temps avec le CTD.

### **b.) Connexion entre le processus de maturation 3' et de transcription**

Certains facteurs impliqués dans la maturation 3' sont retrouvés lors de la transcription. Il a été montré que des interactions entre l'ARN pol II et la machinerie de polyadénylation sont présentes même au moment de la reconnaissance du promoteur. De plus, l'étape de clivage peut altérer la terminaison de la transcription (Zhao et al. 1999). De cette manière, les connexions transcription-polyadénylation auraient lieu à tout moment de la transcription.

Dans la section antérieure, une description sur le rôle du **CTD** a été réalisée. De plus, nous avons mentionné sa capacité d'interagir avec divers facteurs de la polyadénylation et de stimuler les réactions de clivage et de formation de la queue poly(A) *in vivo* et *in vitro* (Hirose and Manley 1998) (McCracken et al. 1997). L'interaction de ce domaine de l'ARN pol II avec la sous-unité de 50kDa de **CstF** et avec la protéine **Pcf11p** du complexe CF IA de levures a été mise en évidence (Meinhart and Cramer 2004) (Meinhart et al. 2005) (Fong and Bentley 2001). Le complexe **CPSF** a aussi la capacité de s'associer au CTD. Des résultats obtenus à partir du complexe de levures montrent que Cft1p/Yhh1p (l'homologue de CPSF-160) lie un CTD phosphorylé. Il faut tenir en compte que CPSF peut lier le complexe CstF et la pol II à travers le même site d'interaction (Nag et al. 2007). Ceci suggère que la formation d'un complexe fonctionnelle CPSF/CstF peut être contrôlée par le déplacement de CPSF depuis l'ARN pol II vers CstF.



Dans les étapes initiales de la transcription, des facteurs intervenant dans la maturation 3' sont présents. Cependant, de façon inattendue, ils ne se retrouvent pas à l'extrémité 3' du transcrit. En effet, le **CPSF** humain et le complexe **CF IA** de la levure ont été localisés sur l'extrémité 5' bien avant les signaux de poly(A) (Dantonel et al. 1997) (Glover-Cutter et al. 2008) (Licatalosi et al. 2002). Ces facteurs seraient donc initialement recrutés par le **PIC** (preinitiation complex) ou le **TEC** (transcription elongation complex), par des interactions protéine-protéine, et transférés par la suite au pré-ARNm après que le site de poly(A) ait été transcrit.

Il a été montré que, pendant l'**initiation de la transcription**, le complexe CPSF peut établir des interactions avec **TFIID**, un facteur essentiel dans l'assemblage du complexe de pré-initiation (**PIC**) avec les promoteurs (Dantonel et al. 1997). **CPSF** est recruté par TFIID et quand la transcription commence, il se dissocie pour rejoindre l'ARN pol II. Cette interaction met en évidence la relation entre l'initiation de la transcription et la polyadénylation. De la même manière, le facteur **TFIIB**, qui participe à l'assemblage du PIC et au recrutement de la polymérase, interagit avec des protéines impliquées dans la maturation 3' (**CPSF-73** et **CstF-64**) (Wu et al. 1999) (Knaus et al. 1996) (Wang et al. 2010). Les dernières études à ce sujet ont montré que la phosphorylation de TFIIB sur sa Ser<sub>65</sub> est requise pour son interaction avec **CstF-64** (Wang et al. 2010). De cette façon, **CPSF-73** et **CstF-64** seraient présents au moment de l'initiation de la transcription. Ces expériences révèlent l'existence d'un rôle des promoteurs dans le recrutement de facteurs de la maturation 3' (Perales and Bentley 2009).

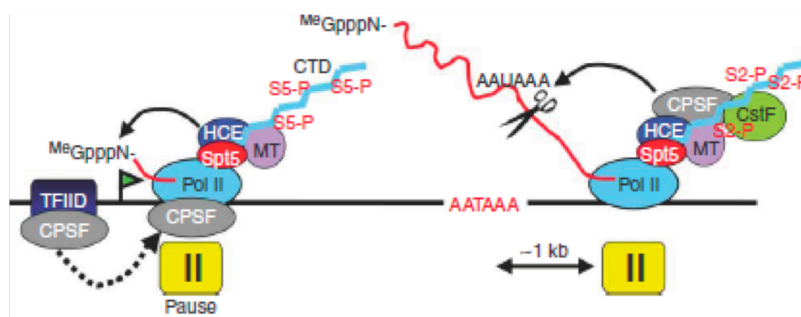
Chez *S. cerevisiae* la protéine **Rna15p** rejoint aussi le TEC avant le début de la maturation 3' du pré-ARNm. Elle entre en compétition avec **Npl3p** (nucleolar protein 3) pour l'association au même site du pré-ARNm. Npl3p est un homologue des protéines SR (Ser/Arg) d'eucaryotes supérieures et elle semble participer au transport des ARNms (Bucheli et al. 2007). La structure de son domaine d'interaction avec l'ARN, composé de deux motifs RRM, a été résolue (Deka et al. 2008) et la reconnaissance préférentielle d'une région riche en G/U a été démontrée par RMN. La structure récente du domaine équivalent de Rna15p (Pancevac et al. 2010) montre aussi une interaction avec le même type de régions, ceci corroborerait l'hypothèse de l'existence d'une telle compétition entre les deux protéines. Ce mécanisme participerait à la précision du processus de maturation 3' par prévention d'une reconnaissance de sites poly(A) cryptiques par Rna15p (Bucheli et al. 2007). La nouvelle structure de Rna15p (Leeper et al. 2010) conteste cet hypothèse étant donné que Rna15p interagit, en présence de Hrp1p, avec un fragment d'ARN riche en A. D'autres études seront nécessaires pour établir une théorie valable.

Le recrutement précoce des facteurs de polyadénylation est à présent une certitude, mais il reste à déterminer si la réaction de clivage précède la terminaison de la transcription ou si ce moment diffère d'un gène à l'autre (Perales and Bentley 2009). Il a été observé que le site de clivage n'est pas suffisant pour détacher l'ARN synthétisé du TEC. Chez les mammifères, ce mécanisme requière la participation du CTD (Custodio et al. 2007) et la fin de l'épissage (Rigo and Martinson

2009). Chez la levure, le complexe **THO** (Hpr1p, Tho2p, Thp1p et Mft1p), l'hélicase **Sub2p**, le facteur d'export **Mex67p** et la phosphatase **Glc7p** sont nécessaires pour la libération de l'ARNm (Gilbert and Guthrie 2004) (Qu et al. 2009) (Rougemaille et al. 2008).

D'autre part, les signaux de poly(A) sont associés à une réduction de nucléosomes sur l'élément de séquence PAS et à un enrichissement en aval de ce site (Spies et al. 2009). La fonction de cette chromatine caractéristique n'est pas bien définie mais elle pourrait provoquer une augmentation du taux **d'élongation de la transcription** et influencer, de cette façon, le processus de maturation 3'.

Le modèle « **double pause** » (« dual-pausing ») décrit par Glover-Cutter (Glover-Cutter et al. 2008), explique la coordination entre la transcription et la maturation des pré-ARNms formés à partir des gènes humains. Pendant la transcription, l'ARN pol II effectue une pause en 5' et une autre en 3' (fig. 26). Selon le modèle présenté, les sites « pause » de l'ARN pol II déterminent le moment et l'endroit où les éléments nécessaires pour la formation de la coiffe et pour la polyadénylation sont incorporés. Ainsi, des facteurs participant à la maturation 5' sont recrutés sur le site de **pause 5'**. Quand la pol II effectue la **pause en 3'** (Boireau et al. 2007) (Darzacq et al. 2007), à approximativement 1kb en aval du site poly(A), le CTD de l'ARN pol II se trouve fortement phosphorylé en **Ser<sub>2</sub>** et les niveaux des facteurs impliqués dans la polyadénylation en association avec le TEC est maximal (Glover-Cutter et al. 2008).



**Figure 26.** Modèle de double pause qui propose une maturation 5' et 3' co-transcriptionnelles des pré-ARNms (Glover-Cutter et al. 2008). Les pauses de l'ARN pol II en 3' et 5' marquent le moment et l'endroit d'assemblage des complexes participant dans la polyadénylation et la formation de la coiffe.

De plus, la machinerie qui intervient dans la maturation 3' permet d'effectuer le couplage entre la **terminaison de la transcription** et la reconnaissance du signal de poly(A) (Rosonina et al. 2006). Deux modèles principaux ont été proposés pour expliquer la dissociation d'un TEC stable. Le **modèle « allostérique »** (Logan et al. 1987) suggère l'existence d'un changement de conformation dans le TEC, dépendant du site poly(A), que le rend moins processif. Le **modèle « torpédo »** (Connelly and Manley 1988) part de l'hypothèse que la coupure de l'ARNm transcrit permettrait l'accès à des exonucléases 5'-3'. Ces enzymes viendraient dégrader le fragment d'ARN clivé et déstabiliser le TEC. Les deux modèles partagent l'implication des facteurs de la maturation

3', comme Pcf11p et des exonucléases Xrn2/Rat1, avant la dissociation du TEC (Rosonina et al. 2006). En 2008, un **modèle combinant les deux précédents** a été proposé (West et al. 2008). Ces études montrent que la reconnaissance du site poly(A) s'effectue rapidement, qu'elle est essentielle pour rendre la pol II compétente pour la terminaison, et que la dégradation exonucléolytique 5'-3' du pré-ARNm clivé est antérieure au détachement de l'ARN pol II. D'autres recherches plus récentes ont démontré que le signal poly(A) a un rôle dans l'apparition de la pause transcriptionnelle, mais aussi dans la dégradation de transcrits d'ARNms non clivés (Kazerouninia et al. 2009). Ceci signifierait qu'une compétition signal poly(A)-dépendante peut exister entre les processus de maturation et de dégradation.

Une nouvelle connexion entre la terminaison de la transcription et la polyadénylation est mise en évidence grâce à l'interaction entre **CstF-64/Rna15p** et **PC4/Sub1p**. PC4/Sub1p fonctionne comme un facteur d'antiterminaison de la transcription grâce à son interaction avec CstF-64/Rna15p. Quand PC4 est phosphorylée, elle interagit avec le complexe de clivage/polyadénylation via CstF-64 et prévient la terminaison prématurée de la transcription. La déphosphorylation de ce facteur, provoque la libération du domaine C-terminal de CstF-64 et la terminaison. De cette façon, clivage et polyadénylation peuvent avoir lieu (Calvo and Manley 2001).

En conclusion, la maturation 3' est un processus qui a lieu co-transcriptionnellement. Des facteurs intervenant dans le clivage du pré-ARNm et dans la formation de la queue poly(A) sont présents dès l'initiation de la transcription. La connexion entre les deux processus est maintenue jusqu'à la terminaison de la transcription.

### c.) Connexion entre la modification post-transcriptionnelle en 3' et l'épissage

L'épissage consiste en l'excision des introns et la liaison des exons par le **splicéosome**. Ce complexe est composé des particules ribonucléoprotéiques (**snRNPs**) appelées **U1, U2, U3, U4, U5** et **U6** et d'autres protéines (Hastings and Krainer 2001).

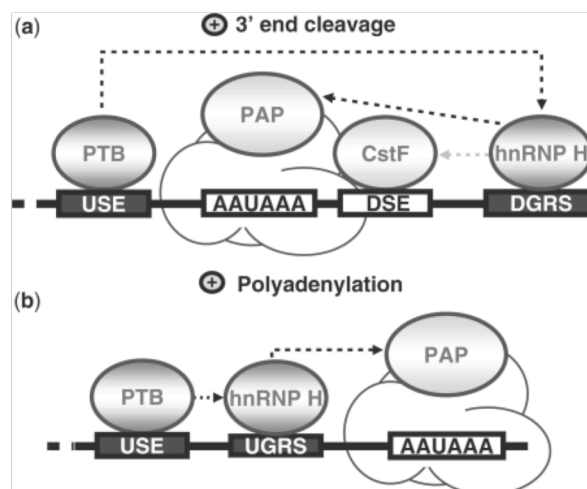
Plusieurs travaux ont démontré l'existence d'une connexion entre l'épissage et la formation de la queue poly(A). Des mutations sur les éléments *cis* impliqués dans l'épissage ont des effets sur les deux processus (Cooke and Alwine 1996) (Vagner et al. 2000). Le résultat est réciproque quand les mutations se trouvent sur des éléments *cis* de la maturation 3' (Niwa and Berget 1991) (Cooke et al. 1999). De plus, des expériences effectuées *in vivo* ont confirmé que la polyadénylation joue un rôle dans la sélection du site d'épissage (Nesic and Maquat 1994) (Tsai et al. 1998).

Plus précisément, certaines études ont montré l'importance du rôle de **U1 snRNP** dans le lien entre ces processus (*voir aussi la section A.1.f., exemple sur IgM*). La protéine **U1A** de U1 snRNP stimule la polyadénylation via son interaction avec **CPSF-160** (Lutz et al. 1996). D'un autre côté, U1A intervient aussi dans l'inhibition de la polyadénylation de son propre pré-ARNm grâce à son interaction avec **PAP** (Gunderson et al. 1994). La protéine **U1 70** de U1 snRNP est également

capable d'interagir avec **PAP** pour inhiber la polyadénylation (Gunderson et al. 1998). En outre, un nouveau lien entre le processus d'épissage et la maturation 3' est établie à travers l'interaction de U1 snRNP avec **CF Im** (Awasthi and Alwine 2003).

La découverte de l'implication d'autres snRNPs dans la connexion entre la maturation 3' et l'épissage est plus récente. Il a été décrit que des protéines de **U2 snRNP** sont associés au complexe **CPSF**. Ces interactions serviraient au couplage des deux processus et à la stabilisation des machineries impliquées (Kyburz et al. 2006).

De plus, un régulateur de l'épissage appelé **PTB** (polypyrimidine tract-binding protein), joue un rôle dans la régulation de la maturation 3'. PTB peut interagir avec des éléments de séquence auxiliaires associés au site poly(A). Il a été observé que l'association de PTB aux séquences **USEs** (upstream sequence elements) promeut, *in vitro*, le clivage et la polyadénylation (Danckwardt et al. 2007). Néanmoins, quand PTB interagit avec des séquences **DSEs**, elle diminue l'efficacité de la polyadénylation. Cet effet est peut être dû à une compétition avec **CstF-64** pour ces séquences (Castelo-Branco et al. 2004). L'intervention de PTB dans le processus favorise aussi le recrutement du facteur d'épissage **hnRNP H** vers des séquences riches en G associées au site de polyadénylation (Millevoi et al. 2009). L'interaction hnRNP H/ARN stimule la réaction de polyadénylation et aide à stabiliser l'union de CstF à l'ARN (Veraldi et al. 2001) (Bagga et al. 1998). Dans le modèle proposé (fig. 27), au moment du **clivage**, hnRNP H serait localisé en aval du site PAS, en interaction avec DSE auxiliaires (DGRS sur la figure). A cet instant, hnRNP H stabilise l'assemblage de la machinerie de la maturation 3' grâce à son interaction avec CstF et PAP. Pendant la réaction de **polyadénylation**, PTB stabilise l'union hnRNP H/ARN en amont du PAS. En ce moment, son interaction avec PAP pourrait servir à augmenter la processivité de l'enzyme (Millevoi et al. 2009).



**Figure 27.** Adaptée de (Millevoi et al. 2009). a) Pendant la réaction de clivage, PTB facilite l'union de hnRNP H sur un site DGRS (downstream G-rich sequence element). Le recrutement de PAP et CstF est favorisé. b) Pendant la polyadénylation, hnRNP H est fixé sur un site UGRS (upstream G-rich sequence element) en amont du PAS. PAP devient plus processive. Les interactions sont indiquées par des flèches.

D'autres facteurs de l'épissage lient directement des protéines impliquées dans la maturation, par exemple CPSF-100/U2 snRNP (Kyburz et al. 2006), ou interviennent dans la régulation de la maturation 3' à travers son association avec les séquences **USE auxiliaires**, par exemple **U2AF35** ou **U2AF65** (Danckwardt et al. 2007). Ainsi, des multiples interactions peuvent connecter les deux processus.

**d.) Connexion entre la formation de la queue poly(A) et l'export des ARNms vers le cytoplasme**

Les ARNms matures doivent être transportés vers le cytoplasme pour leur traduction. Les ARNms sont exportés sous forme de complexes **ribonucléoprotéiques messagers (RNPms)** qui s'assemblent de façon co-transcriptionnelle. Dans un premier temps, toutes les étapes de la maturation des ARNms doivent être complétées et les complexes impliqués dans ces étapes doivent être libérés de l'ARNm pour que l'export ait lieu.

Des études réalisées ces dernières années ont montré comment des modifications dans l'assemblage et l'export des RNPms affectent négativement la formation de la queue poly(A). Ces altérations peuvent rendre difficile la transcription sur le site poly(A), diminuer la synthèse de la queue, induire une rapide dégradation de l'extrémité 3', promouvoir l'accumulation de transcrits non traités et causer une hyperadénylation ou la rétention des facteurs impliqués dans la polyadénylation. Dans cette section nous donnerons seulement quelques exemples.

Il a été montré que le clivage et la polyadénylation sont nécessaires, mais ne sont pas suffisants, pour l'export. L'existence d'un lien entre le transport des ARNms vers le cytoplasme et la polyadénylation est mise en évidence lorsqu'on s'aperçoit que **l'addition de la queue poly(A) est requise pour un export efficace** (Brodsky and Silver 2000). Des mutations sur les sous-unités Rna14p, Rna15p et Pcf11p de CF IA, sur PAP ou des délétions des éléments en *cis* de la maturation 3', provoquent dans la plupart des cas seulement un défaut dans la formation de la queue poly(A). Cependant, dans d'autres occasions ces mutations bloquent en plus le transport des ARNms correctement polyadénylés vers le cytoplasme (Brodsky and Silver 2000) (Hammell et al. 2002). Ce fait démontre la participation des facteurs de la polyadénylation à l'export des ARNms.

La relation entre la formation de RNPs, la polyadénylation et l'export apporte une méthode additionnelle de régulation (Villa et al. 2008). Il a été montré que les défauts sur la croissance et sur la diminution des ARNms portant une queue poly(A), occasionnés par la mutation *pap1-1*, sont supprimés quand le gène *RRP6* est délété. **Rrp6p** est une des sous-unités nucléaires de l'exosome. Ces résultats suggèrent l'existence d'un mécanisme qui permet de cibler les ARNms insuffisamment polyadénylés avant leur export dans le cytoplasme (Burkard and Butler 2000). La participation de Rrp6 et d'autres sous-unités de l'exosome comme **Rrp41p** dans ce processus, a été également démontrée à partir des souches de *S. cerevisiae* portant les mutations *rna14p-1* ou

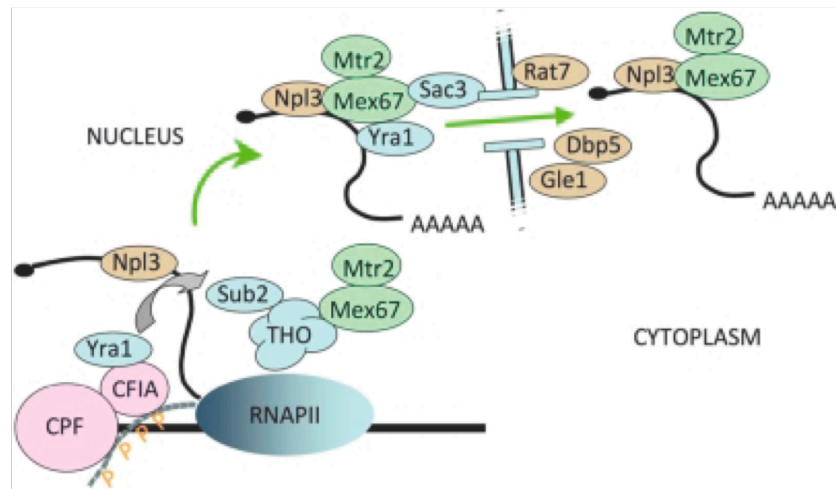
*rna15p-2*. Ces mutations affectent les réactions de clivage, de polyadénylation et la terminaison de la transcription (Birse et al. 1998) (Minvielle-Sebastia and Keller 1999). Dans ce contexte, la délétion de Rrp6p et de Rrp41p, permet là aussi la stabilisation des ARNms aberrants (Libri et al. 2002) (Torchet et al. 2002). Une compétition cinétique entre la maturation 3' et la dégradation pourrait avoir lieu dans ce type de régulation (Villa et al. 2008).

La relation entre les processus de maturation 3' et d'export a été par la suite validée grâce à l'existence des interactions des facteurs qui y participent. Ainsi, **Xpo1p**, impliqué dans l'export des ARNms, interagit avec plusieurs facteurs de la maturation 3' comme **Pab1p** (Hammell et al. 2002).

D'un autre côté, **Yra1p**, un facteur d'export des ARNms (Stutz et al. 2000), interagit avec CF IA à travers la sous-unité **Pcf11p**. Le site de Pcf11p impliqué dans l'interaction avec Yra1p coïncide avec celui utilisé pour l'association avec Clp1p. Dans le modèle proposé (Johnson et al. 2009) (Saguez and Jensen 2009), Yra1p serait recruté par Pcf11p dans un premier temps et ensuite transféré à Sub2p. En support de cette théorie, il a été montré que Yra1p utilise la même région pour l'interaction avec ces deux partenaires, Pcf11p et Sub2p.

En effet, pour l'export, il est nécessaire que le **THO** (Hpr1p, Tho2p, Thp1p et Mft1p) s'associe avec **Sub2p** et **Yra1p** pour former le complexe **TREX** (transcription and export). Par interaction avec la sous-unité **Hpr1p**, THO recrute aussi **Mex67p**. Cette dernière protéine forme un hétérodimère avec **Mtr2p** et les deux effectuent le transport des RNPms matures à travers les pores nucléaires. Des altérations du complexe **THO/Sub2p** provoquent une dégradation 3'→5' rapide des transcrits via la dégradation spécifique de **Fip1p**, le régulateur de Pap1p. Ce fait met en évidence une nouvelle interaction entre les deux processus (Saguez et al. 2008).

L'influence des deux processus est **bidirectionnelle**. Les facteurs de l'assemblage de RNPms jouent un rôle dans la libération des complexes de la maturation 3' quand la polyadénylation est terminée (Qu et al. 2009) (*fig. 28*). Des essais d'immunoprécipitation d'ARN permettent de connaître la quantité de **Rna15p** associée à l'ARN poly(A)<sup>+</sup>. A partir de ces études, il a été montré que les mutations de certains facteurs d'export nucléaires comme **Mex67p** inhibent la libération de Rna15p de la machinerie.



**Figure 28.** Adaptée de (Qu et al. 2009). Modèle d'export des RNPms du noyau vers le cytoplasme. Le complexe THO aide au recrutement de **Sub2p** et **Mex67**. Les facteurs impliqués dans la maturation 3', CPF et CF IA, interagissent avec le CTD phosphorylé de l'ARN pol II. **Yra1p** est recruté par CF IA, et probablement, quand la polyadénylation a lieu, une réorganisation des facteurs se produit et Yra1p est transférée pour former un complexe avec **Sub2p**. Yra1p peut ainsi interagir avec **Mex67p/Mtr2p**. L'ARNm mature traverse enfin le pore nucléaire à l'aide d'autres facteurs d'export comme **Sac3**. Sur la figure, des mutations sur les facteurs colorés en bleu et vert provoquent la rétention de Rna15p sur l'ARNm. L'alteration des éléments colorés en marron n'ont pas d'effet sur cette sous-unité de CF IA.

En outre, l'activité phosphatase de **Glc7p** semble participer aussi à l'export des ARNms (Gilbert and Guthrie 2004). **Npl3p**, une protéine requise pour l'export des ARNms, est sensible à une phosphorylation en Ser par une kinase cytoplasmique, **Sky1p** (Siebel et al. 1999). Cette phosphorylation est essentielle pour qu'elle puisse accomplir sa fonction. Il a été montré que Glc7p promeut la déphosphorylation de Npl3p dans le noyau (Gilbert et al. 2001). Le recrutement de **Mex67p** est aussi dépendent de Glc7p, probablement grâce à la déphosphorylation de Npl3p (Gilbert and Guthrie 2004). Ainsi, un cycle de phosphorylation cytoplasmique - déphosphorylation nucléaire (par Glc7p et probablement par d'autres enzymes) jouerait un rôle dans le processus d'export des ARNms matures. De plus, Rna15p entre en compétition avec **Npl3p** (nucleolar protein 3) pour l'association au même site du pré-ARNm (voir Chapitre I.A.2.b) (Bucheli et al. 2007). Cette interaction corrobore l'implication de Npl3p dans le processus de polyadénylation et vice-versa.

e.) **Connexion entre la formation de la queue poly(A) et la traduction**

L'ARNm mature est transporté vers le cytoplasme. L'association de PABP avec la queue poly(A) promeut la formation des liaisons qui stimulent l'initiation de la traduction (Jacobson and Favreau 1983) (Jacobson and Peltz 1996). La formation d'une riboncléoprotéine pseudocirculaire dépend des interactions entre **PABP** et **eIF4G** et entre ce dernier et **eIF4E** (un facteur qui reconnaît la coiffe). Cet assemblage est nécessaire pour recruter la sous-unité ribosomale **40S** (Tarun and Sachs 1996) (Wells et al. 1998).

Le domaine C-terminal de PABP peut aussi s'associer à la région N-terminale du facteur de terminaison de la traduction **eRF3** (Hoshino et al. 1999) (Cosson et al. 2002). Cette interaction induit le « recyclage » du ribosome qui peut se réinstaller sur l'extrémité 3' pour recommencer la traduction du même ARNm (Uchida et al. 2002). L'interaction PABP/eRF3 est aussi impliquée dans la terminaison de la traduction (Cosson et al. 2002). De plus, la formation de ce complexe contribue à la diminution de la quantité de monomères de PABP sur la queue poly(A) facilitant l'accès des enzymes qui la dégradent (Hoshino et al. 1999). De ce fait, un lien entre la terminaison de la transcription et la dégradation de la queue d'adénines a pu être établi.

Finalement, il a été montré que la **dégradation de l'ARNm** peut être initié par trois événements : un clivage endonucléolytique, l'élimination de la coiffe, et le raccourcissement de la queue poly(A) (Jacobson and Peltz 1996). Chez la levure, l'analyse de ce processus montre qu'il existe d'abord une réaction nucléolytique qui réduit la queue de poly(A) en 10-15 nucléotides. Ensuite, la coiffe est éliminée par le complexe **Dcp1p-Dcp2p** et le transcrit est alors digéré par des exoribonucléases 5'→3' comme **Xrn1p**, et/ou par l'exosome en direction 3'→5' (Mangus et al. 2003) (Decker and Parker 2002).

En **conclusion**, le processus de maturation 3' est intimement connecté avec toutes les étapes de l'expression génique. Certains des facteurs qui interviennent accompagnent le pré-ARNm dès sa transcription jusqu'au son export vers le cytoplasme. L'interaction des facteurs de la polyadénylation avec ceux qui participent dans les autres processus permet une régulation générale de l'expression génique et assure la terminaison correcte de la transcription et la traduction des ARNm matures dans le cytoplasme.



## B. Objectifs

Chez les eucaryotes, les pré-ARNms sont synthétisés par l'ARN pol II et doivent subir des modifications post-transcriptionnelles avant leur transport dans le cytoplasme où ils seront traduits. Le sujet de cette thèse est focalisé sur l'étude de protéines impliquées dans une de ces modifications, la **maturation 3'**. Ce processus inclut généralement une étape de clivage qui est suivie de la formation d'une queue d'adénines (poly(A)). Il a été démontré que la queue poly(A) est essentielle pour la stabilité, la translocation vers le cytoplasme et la traduction de l'ARNm mature.

Le principe général des événements de la maturation 3' est connu, cependant, beaucoup de détails de ce processus et de la connexion entre celui-ci et d'autres étapes de l'expression génique nous échappent encore de nos jours. Un grand nombre de protéines intervenant dans ce processus ont été identifiées. Certaines d'entre elles participent directement dans la réaction de clivage ou dans l'addition d'adénines et d'autres ont un rôle régulateur ou stabilisateur. Nous nous sommes intéressés plus précisément aux complexes homologues **CstF** ou « Cleavage stimulation Factor » des mammifères et **CF IA** ou « Cleavage Factor IA » des levures. Le premier est indispensable pour le clivage du pré-ARNm tandis que CF IA participe aux deux étapes de la maturation 3' chez la levure. Le but de cette thèse a été d'approfondir la connaissance de l'organisation des sous-unités de ces deux complexes pour mieux comprendre leur participation dans le processus. Malgré les avancés des dernières années dans le domaine, plusieurs questions en rapport avec CstF et CF IA restent pour l'instant sans réponse et ont motivé ces travaux :

1. Quelle est l'architecture moléculaire de chaque protéine dans les deux complexes ?
2. Comment les différentes sous-unités interagissent entre elles ?
3. Comment les protéines de CstF et de CF IA permettent la connexion entre le processus de maturation 3' et d'autres étapes de l'expression génique ?
4. Quelles différences/similitudes peut-on trouver entre le complexe chez les mammifères et chez les levures ?

Mes travaux ont consisté principalement en l'analyse des sous-unités de CstF et de CF IA isolées ou en complexe. Pour cela, nous avons utilisé principalement une **approche structurale**, qui combinée avec des techniques de biologie moléculaire et de biochimie, constitue un outil puissant pour l'étude des protéines. De cette façon, des étapes de clonage, expression et purification des protéines, ont été effectuées. Quand ce fut possible, ces expériences ont été suivies des études structurales par cristallographie, dans le laboratoire, ou par RMN (Résonance Magnétique Nucléaire) en collaboration avec C. Mackereth (dans l'IECB).

**CstF** est composé de trois sous-unités, **CstF-77**, **CstF-64** et **CstF-50**. La structure de la première, à l'exception de la région située sur l'extrémité C-terminale, a été résolue au laboratoire et dans un autre groupe de recherche, (Legrand et al. 2007) (Bai et al. 2007) respectivement. A partir de ces modèles structuraux, une organisation dimérique a été attribuée à cette sous-unité de

CstF. Nous disposons également des modèles structuraux des domaines RRM (RNA Recognition Motif) et C-terminal de CstF-64 (Perez Canadillas and Varani 2003) (Qu et al. 2007). Cependant, aucune information sur l'architecture de CstF-50 n'a été révélée et la forme d'assemblage des trois protéines est aussi inconnue. La première partie du Chapitre II est dédiée aux travaux effectués pour le complexe CstF. L'étude structurale d'une région de la protéine CstF-50 et les résultats obtenus sur l'interaction des deux autres sous-unités, CstF-77 et CstF-64, seront exposés dans cette section.

D'autre part, le complexe CF IA de levures est composé de quatre protéines, **Rna14p** et **Rna15p**, les orthologues de CstF-77 et CstF-64 respectivement, **Pcf11p** et **Clp1p**. La structure de cette dernière en complexe avec une région de Pcf11p a été résolue (Noble et al. 2007). De plus, l'architecture moléculaire de la zone de Pcf11p impliquée dans l'interaction avec le CTD de l'ARN pol II est connue (Meinhart and Cramer 2004). Dans le Chapitre III, j'exposerai donc les études structurales réalisées sur l'interaction Rna14p/Rna15p et les résultats obtenus à partir de la reconstitution du complexe CF IA *in vitro*.

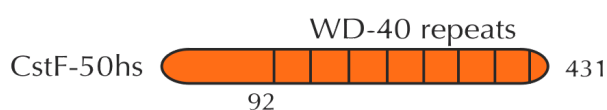
## Chapitre II : Résultats obtenus pour le complexe CstF

### A. Etude structurale du domaine N-terminal de la sous-unité CstF-50

CstF-50 participe au processus de maturation 3' des pré-ARNms en étant essentielle pour la réaction de clivage. La connaissance de l'architecture de cette protéine peut être cruciale pour la compréhension du rôle du complexe CstF dans le processus de **maturation 3'** et dans la connexion de cette étape avec la **transcription** et avec la **réponse au dommage à l'ADN**.

La protéine CstF-50 est constituée d'au moins deux domaines (*fig. 29*). Comme il a été exposé dans le premier chapitre de cette thèse, un de ces domaines est composé d'une série de sept répétitions du motif **WD-40**. Cette région de CstF-50 est responsable de l'interaction avec **CstF-77** et avec d'autres protéines comme **BARD1** (Takagaki and Manley 2000) (Kleiman and Manley 1999). Dans la protéine humaine, ce domaine commence probablement à l'acide aminé 95 et il s'étend jusqu'au dernière résidu. Ainsi, cette région correspond à plus des trois quarts de la sous-unité CstF-50. L'étude de la région **WD-40** aiderait à comprendre l'organisation générale du complexe CstF à travers l'interaction avec son partenaire CstF-77. De plus, la résolution de sa structure contribuerait à l'interprétation des modifications existantes au niveau de ce complexe en réponse au dommage de l'ADN (Kleiman and Manley 2001).

L'autre domaine protéique inclut la zone **N-terminale** de la protéine. A partir de la résolution de la structure du dimère CstF-77, l'hypothèse émise (Bai et al. 2007) propose l'organisation de CstF en **hexamère**. Selon ce modèle, deux copies de chaque sous-unité seraient présentes au moins un moment pendant le processus de maturation 3'. En accord avec cette théorie, des expériences de « far western » ont suggéré la capacité de la zone N-terminale de la protéine CstF-50 à interagir avec elle-même (Takagaki and Manley 2000). Cependant, il n'y a aucune autre évidence sur l'existence d'un tel type d'assemblage pour cette protéine. D'autre part, il a été montré que l'activation du processus de maturation 3' du pré-ARNm requière un contact entre le CTD et la sous-unité CstF-50 (McCracken et al. 1997). Il a été observé que la région N-terminale de CstF-50 est impliquée dans l'association avec le **CTD de l'ARN pol II** (Fong and Bentley 2001). L'interaction **CstF-50/CTD** participerait au recrutement du complexe CstF à la machinerie de maturation 3'. De cette façon, des progrès sur ces deux problématiques, l'organisation de CstF et la connexion maturation 3'-transcription, peuvent être effectuées *via* l'étude de la région **N-terminale** de la protéine CstF-50.



**Figure 29.** Schéma représentant les domaines de CstF-50 avec une région N-terminale suivie de sept répétitions du motif WD-40.

## 1. Etude de la région CstF-50 (1-92)\* de *Drosophila melanogaster*

### a.) Identification du domaine pour la réalisation des travaux

Pour les premiers travaux réalisés au laboratoire nous avons utilisé la sous-unité CstF-50 entière, incluant les deux domaines cités. Cependant, les tests d'expression effectués pour la protéine originaire de différentes espèces (*Homo sapiens*, *Drosophila melanogaster*, *Encephalitozoon cuniculi*) n'ont jamais donné des résultats exploitables en raison de l'insolubilité de la protéine. Nous avons donc choisi de travailler avec des versions tronquées de CstF-50 qui intégraient l'un ou l'autre domaine. Le clonage des différentes régions de la sous-unité CstF-50 a été effectué à partir des **analyses d'homologie des séquences** et des **études bioinformatiques** qui donnent des informations sur la **structure secondaire** des protéines (fig. 30). Les programmes utilisés pour ces analyses ont été principalement :

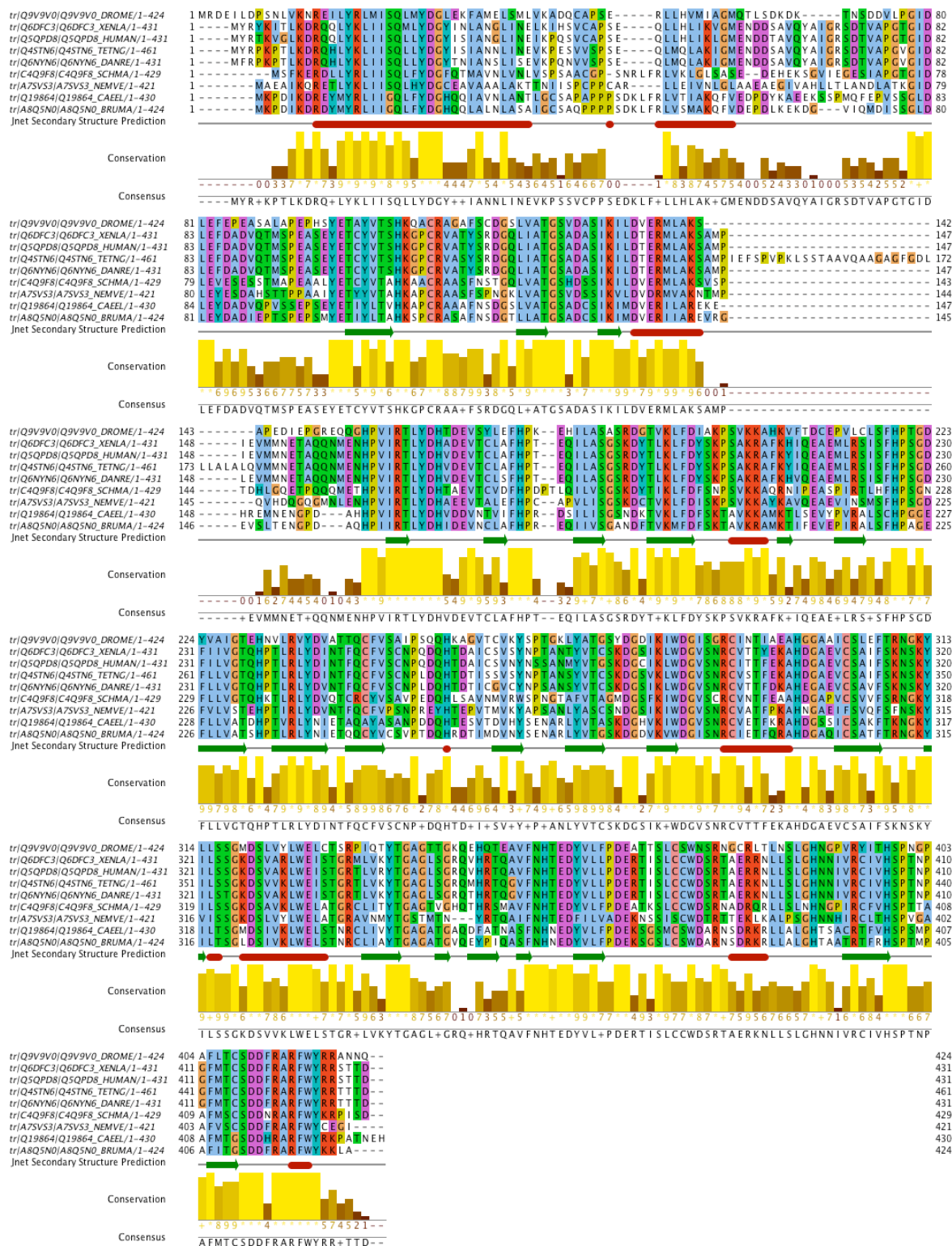
- BLAST (Altschul et al. 1997): Blastp (utilisable à partir du serveur ExPASy) a été employé pour la recherche de séquences similaires disponibles dans les bases de données protéiques.
- CLUSTALW (Thompson et al. 1994): est un programme d'alignement multiple de séquences.
- Jalview (Waterhouse et al. 2009) : est un éditeur d'alignement multiple de séquences. Ce programme a été employé pour la visualisation des alignements.
- Jnet Secondary Structure Prediction : permet d'effectuer la prédiction de la structure secondaire d'une protéine à partir de sa séquence. La méthode est basée sur la composition d'acides aminés et sur la similarité entre la séquence d'étude et d'autres séquences de structure secondaire connue. Ce programme est accessible sur Jalview. Les résultats ont été corroborés à l'aide d'autres programmes comme GOR, Porter ou Psipred.

Grâce à ces outils, nous pouvons orienter la stratégie pour la réalisation du clonage des différentes régions de la protéine. Par exemple, nous essayons d'éviter de couper au milieu d'une zone très conservée ou dans un endroit impliqué dans la formation d'une hélice. Dans un premier temps, nous avons cloné séparément les régions N-terminales et les domaines WD-40 des protéines CstF-50 humaine et de drosophile : CstF-50hs (1-95)<sup>#</sup>, CstF-50dm (1-92), CstF-50hs (95-431), CstF-50hs (92-424). La plupart des travaux de cette thèse ont été focalisés sur la partie **N-terminale** de la protéine. En effet ce domaine, structuralement inconnu, est impliqué et dans la formation possible d'un **dimère** et dans l'interaction avec le **CTD**, ce qui la rend essentielle dans la connexion entre la maturation 3' et la transcription.

---

\* Les numéros entre parenthèses correspondent au premier et dernier acide aminé inclus dans le fragment de la protéine d'étude.

<sup>#</sup> Pour faciliter l'écriture, les abréviations « hs » et « dm » sont utilisées pour faire référence aux espèces *Homo sapiens* et *Drosophila melanogaster* respectivement.



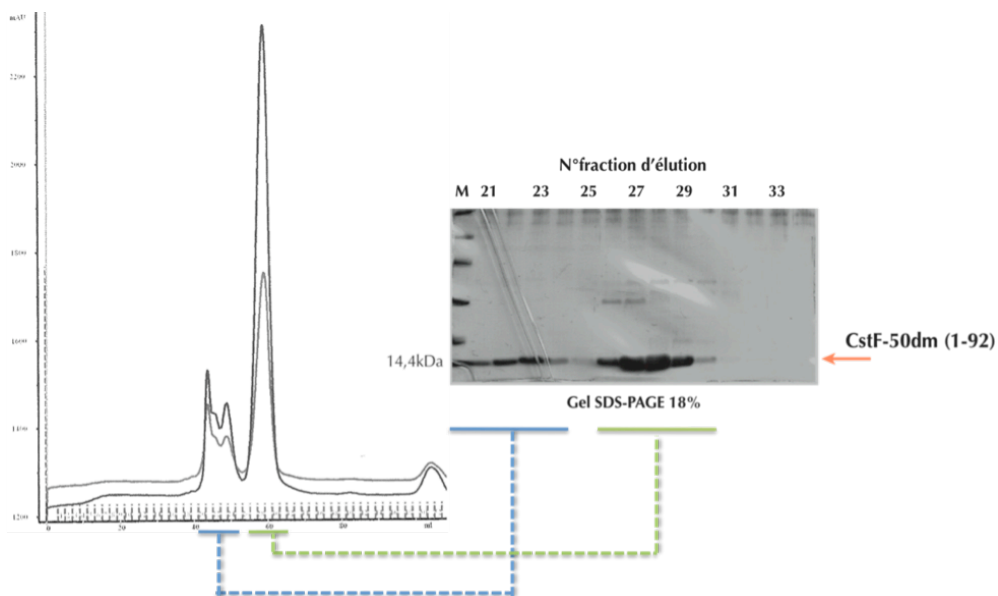
**Figure 30.** Alignement de séquences de CstF-50. Un schéma de la prédiction de la structure secondaire est situé en dessous des séquences. Les flèches vertes représentent les feuilletts bêta et les cylindres rouges les hélices alpha. La conservation des séquences est exprimée par un diagramme où la couleur jaune et la hauteur des barres indiquent une conservation plus importante. Dans l'alignement : DROME: *Drosophila melanogaster*, XENLA: *Xenopus laevis*, HUMAN: *Homo sapiens*, TETNG: *Tetraodon nigroviridis*, DANRE: *Danio rerio*, SCHM: *Schistosoma mansoni*, NEMVE: *Nematostella vectensis*, CAEEL: *Caenorhabditis elegans*, BRUMA: *Brugia malayi*.

**b.) Expression et purification de CstF-50 (1-92) de drosophile**

Pour le démarrage des études de ce domaine de la protéine CstF, les régions correspondant aux résidus 1 à 94 de la protéine humaine et 1 à 92 de la protéine de drosophile ont été clonées. Pour la réalisation du clonage, l'ADN correspondant à ces zones de CstF-50hs et de CstF-50dm est inséré dans un vecteur *pETMCN-EA-vH* (voir Chapitre VI : Matériels et Méthodes) qui contient une étiquette composée de 6 histidines (« tag H<sub>6</sub>») clivable par la protéase *TEV*. A partir des plasmides obtenus, différents tests d'expression ont été réalisés afin de déterminer la souche d'*E. coli* pour laquelle le domaine protéique est le mieux surexprimé. Ces résultats ont permis de sélectionner la souche **BL21 (DE3)** pour la poursuite des travaux.

Comme il a été montré dans la section antérieure, le domaine N-terminal de CstF-50 est très conservé entre les différentes espèces. Le choix de l'espèce utilisée pour les étapes suivantes a été réalisé en fonction de la solubilité et du niveau d'expression de la protéine correspondante. Ses deux critères ont déterminé la sélection de la protéine CstF-50 de *Drosophila melanogaster*.

La purification de **CstF-50dm (1-92)** s'est déroulée en trois étapes chromatographiques. D'abord, la protéine chimérique est purifiée grâce à la rétention de la protéine de fusion sur une résine de Co<sup>2+</sup> (voir section de Matériels et Méthodes). Dans la deuxième étape, j'ai utilisé une colonne d'échange ionique commerciale HiTrap<sup>TM</sup> S Fast Flow (Amersham Biosciences). Finalement, après la coupure du « tag » par la *TEV* et la concentration de la protéine, une chromatographie d'exclusion moléculaire (HiLoad 16/60 Superdex<sup>TM</sup> 75 pg) a permis l'obtention d'une protéine suffisamment pure pour la réalisation des études de cristallisation (fig. 31).



**Figure 31.** Dernière étape de purification de la protéine **CstF-50dm (1-92)**. 10µl de chaque fraction d'élution ont été chargées sur un gel SDS-PAGE. Sur le profil d'élution deux pics ont été différenciés. Pour les expériences de cristallisation, nous avons sélectionné seulement le deuxième pic afin d'avoir un échantillon homogène. La présence de contaminants pourrait expliquer l'existence de ces deux pics.

### c.) Etude structurale de CstF-50 dm (1 - 92)

Après une analyse sur gel SDS-PAGE, les fractions correspondantes à la protéine étudiée ont été groupées. La protéine a été concentrée jusqu'à **15-25mg/ml**. De nombreux essais de cristallisation, avec des kits commerciaux disponibles au laboratoire, criblant plusieurs centaines de conditions de cristallisation ont été réalisés. Cependant ce criblage a seulement abouti à la formation de petits cristaux non reproductibles (*fig. 32*).

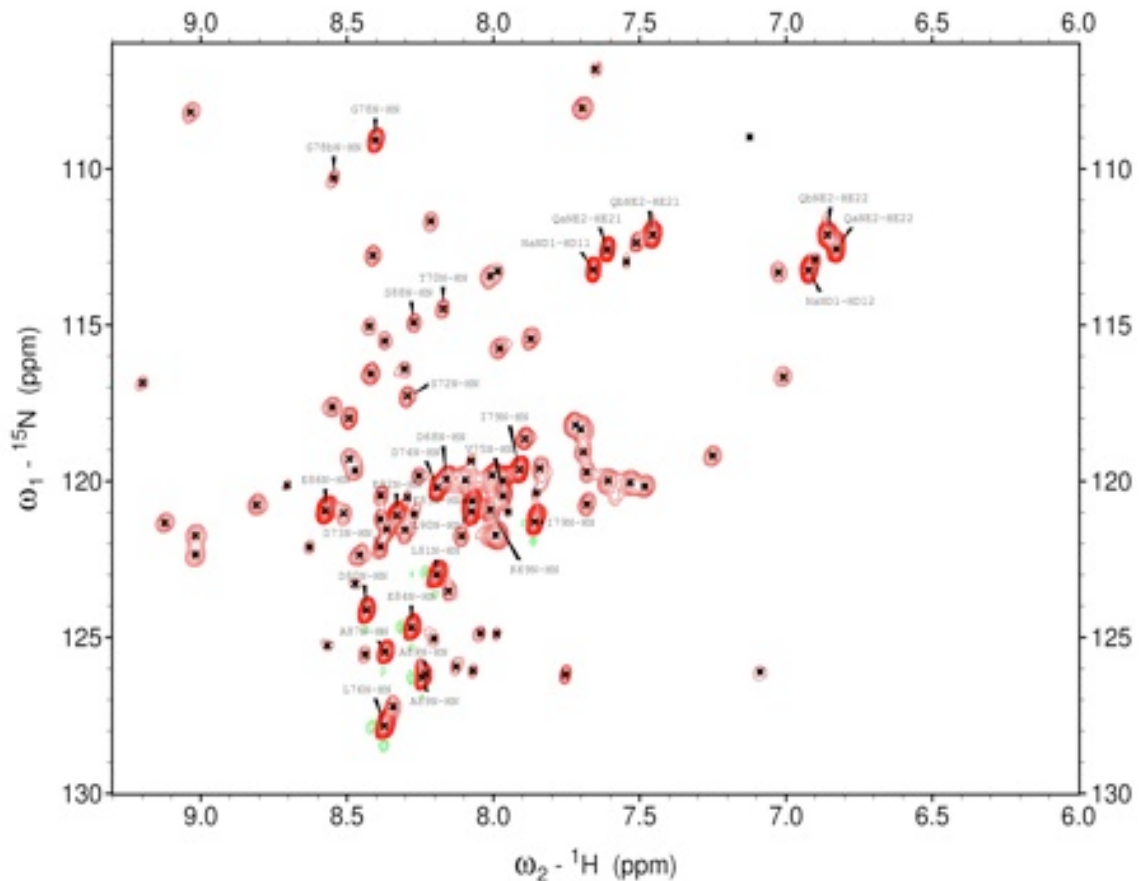


*Figure 32. Photographie des cristaux obtenus pour CstF-50 dm (1-92).*

La cristallisation n'ayant pas donné de résultats exploitables et puisque la taille de la protéine est faible, approximativement 10kDa, nous avons envisagé de résoudre sa structure par **RMN** (Résonance Magnétique Nucléaire). Ainsi, une collaboration avec le groupe « NMR spectroscopy of protein-nucleic acid complex » dirigé par le **Dr. Cameron Mackereth** à l'IECB, a été établie. La RMN offre l'avantage d'étudier la protéine directement en solution et apporte des informations sur le repliement général de la protéine. La résolution de la structure est également possible par cette méthode pour une protéine de cette taille.

Les expériences de RMN ont nécessité la réalisation de nouvelles purifications de la protéine marquée avec  $^{15}\text{N}$  et/ou  $^{13}\text{C}$  (*voir section de Matériels et Méthodes*). À partir de CstF-50 dm (1-92) pure et marquée, plusieurs expériences ont été effectuées. Après analyse des résultats nous avons pu constater que **la partie C-terminale du fragment étudié était très flexible et non-repliée** (*fig. 33*). En particulier, les valeurs des déplacements chimiques pour les noyaux de la chaîne principale ainsi que l'analyse des paramètres de relaxation révèlent que les résidus 65 à 92 sont en dehors du domaine replié. Ainsi, on peut imaginer qu'il existe une région flexible, située entre le domaine N-terminal et les répétitions du motif WD-40, qui permettrait, éventuellement, la jonction entre les deux domaines. De cette façon, la zone N-terminale et la C-terminale disposent d'une certaine indépendance dans l'ensemble de la protéine. La flexibilité de cette région peut être responsable d'un effet négatif sur la cristallisation dans les conditions testées.





**Figure 33.** Caractérisation de la région de CstF-50dm du résidu 1 au 92. Spectre  $^1\text{H},^{15}\text{N}$ -HSQC obtenu pour CstF-50dm (1-92). Les pics correspondant aux résidus spécifiques de la zone C-terminale de CstF-50dm (1-92) ont été notés avec le type et le numéro de résidu. Nous avons observé que la plupart de ces pics étaient regroupés dans la région indiquant une absence de structure secondaire (valeurs  $^1\text{H}$  entre 8 et 8,5 ppm).

## 2. Etude structurale de CstF-50 (1-65) de drosophile

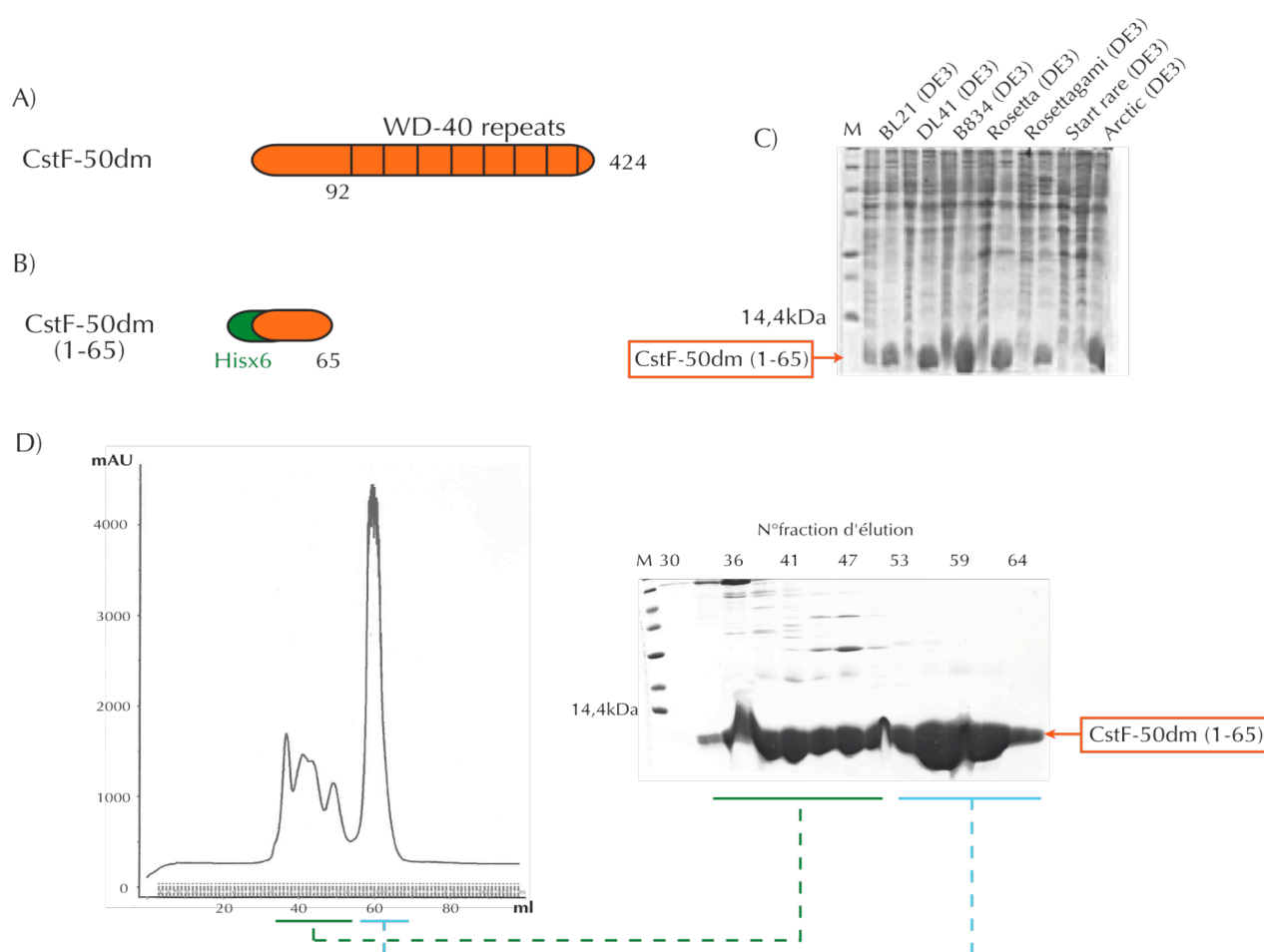
### a.) Clonage, expression et purification de la protéine CstF-50dm (1-65)

A partir des résultats obtenus par RMN, et de l'impossibilité de produire des cristaux à partir de CstF-50dm (1-92), nous avons décidé de recommencer toutes les étapes à partir d'un fragment de la protéine plus court. L'insert d'ADN souhaité a été acquis par PCR à partir d'un nouvel oligonucléotide (voir *Matériels et Méthodes*). L'ADN amplifié a été digéré avec les enzymes *NdeI* et *BamHI* pour son insertion dans le même vecteur contenant l'étiquette  $\text{H}_6$ , *pETMCN-EA-vH*.

Le clonage et les tests d'expression ont été effectués pour la protéine humaine et pour celle de drosophile, cependant, la région clonée pour la première (CstF-50hs (1-63)) exprimait une protéine insoluble. Pour cette raison j'ai décidé de continuer à travailler avec la protéine de drosophile, maintenant appelée **CstF-50dm (1-65)**, dont les tests montraient une importante surexpression dans différentes souches bactériennes (fig. 34).



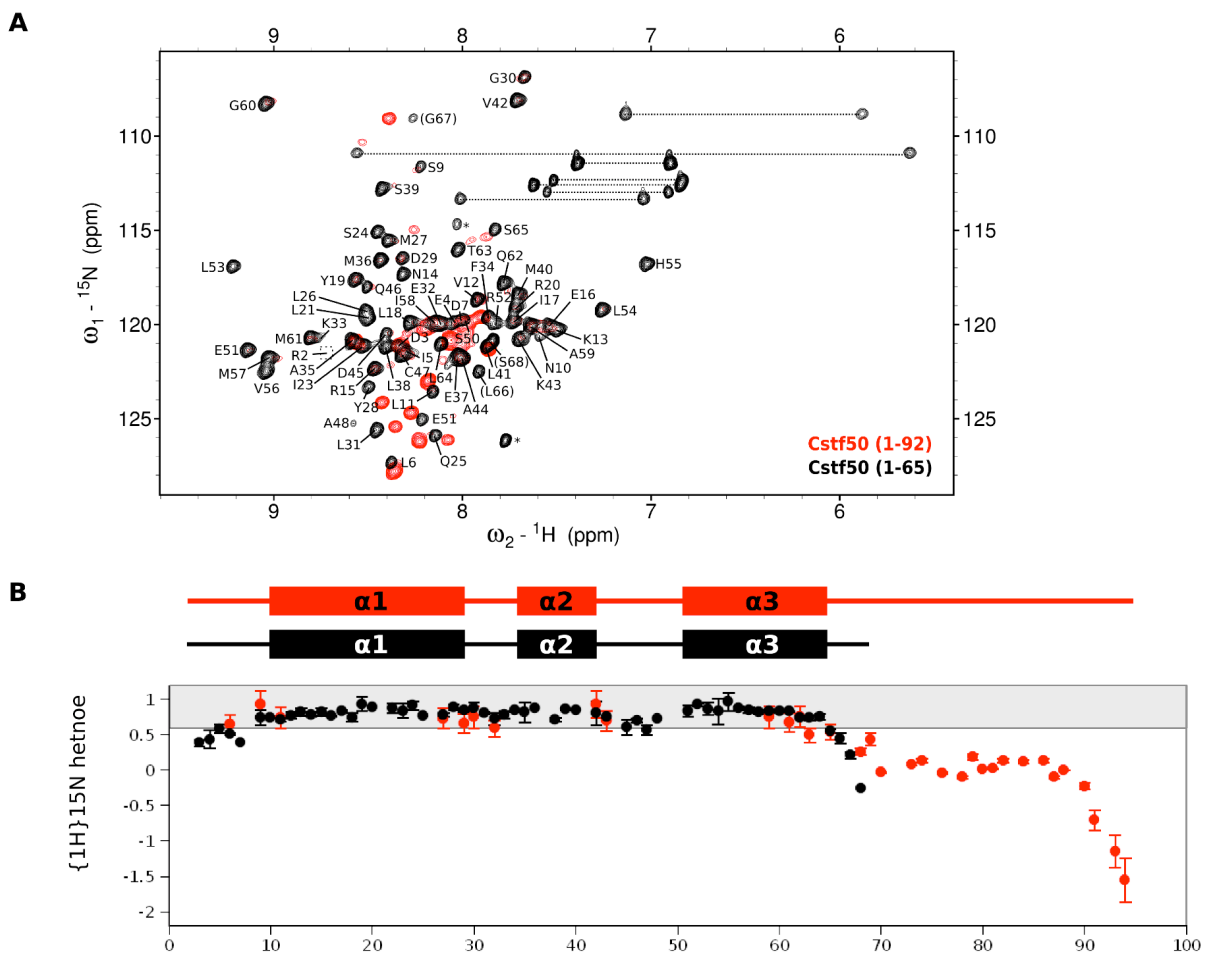
La production de la protéine à grande échelle a été de nouveau réalisée à partir de la souche **BL21(DE3)** et de 2 litres de culture dans un milieu LB (*Lysogeny broth* ou *Luria-Bertani*). Dans ce cas, la purification a pu être exécutée en deux étapes, une chromatographie **d'affinité**, grâce à l'étiquette de six histidines, et une autre **d'exclusion moléculaire**. Un gel montrant le niveau de pureté est présenté dans la *figure 34*. Sur les fractions obtenues à partir des premiers pics (approximativement de la fraction 30 à la 50), des bandes ne correspondant pas à la protéine d'intérêt sont visibles sur le gel. Il est possible que la présence de ces contaminants modifie le volume d'éluion de CstF-50dm (1-65). Les deux espèces protéiques obtenues dans les deux pics d'éluion différents ont été traitées séparément mais nous n'avons trouvé aucune différence au niveau de la cristallisation.



**Figure 34.** Premières étapes dans la résolution de la structure de **CstF-50dm (1-65)**. **A)** Schéma représentant les domaines de la protéine CstF-50dm entière. **B)** Fragment de la protéine qui a été cloné avec un « tag » de 6 His marqué en vert. **C)** Gel SDS-PAGE à 18% d'acrylamide obtenu à partir du test d'expression de ce fragment de la protéine. La flèche orange indique un poids moléculaire d'approximativement 7,5kDa. **D)** Dernière étape de purification de la protéine CstF-50dm (1-65). A gauche, le profil d'éluion obtenu à partir d'une colonne d'exclusion moléculaire. A droite, le gel SDS-PAGE 18% où 10µl des différentes fractions d'éluion ont été chargés.

**b.) Comparaison des constructions CstF-50dm (1-92) et CstF-50dm (1-65)**

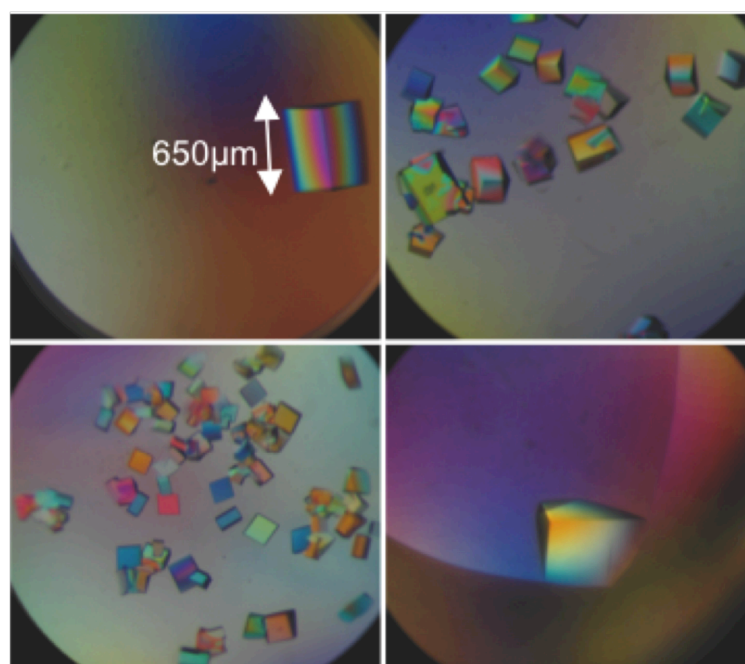
Nous avons purifié les deux versions tronquées de la protéine CstF-50dm, CstF-50dm (1-92) et CstF-50dm (1-65), marquées à l'azote 15 ( $^{15}\text{N}$ ). L'analyse du déplacement chimique du squelette peptidique et des expériences de relaxation de la construction plus longue suggèrent que les résidus placés immédiatement après le 65 ne se trouvent pas dans la région repliée de la protéine. La comparaison de ces résultats avec les spectres obtenus pour **CstF-50dm (1-65)**, a mis en évidence la présence commune d'un domaine correspondant aux résidus du 1 au 65 (fig. 35). L'élimination de la région C-terminale n'affecte donc pas le repliement du reste de la protéine. De plus, les résultats obtenus pour CstF-50dm (1-65) indiquent l'absence des régions de flexibilité importante. De cette façon, les essais de cristallisation ont pu à nouveau être envisagés avec la nouvelle construction de CstF-50dm.



**Figure 35.** Caractérisation de la région N-terminale de CstF-50dm. **A)** Superposition des spectres  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC obtenus pour **CstF-50dm (1-92)**, en rouge, et pour **CstF-50dm (1-65)**, coloré en noir. Les pics correspondant à cette dernière ont été notés avec le type et le numéro de résidu. Les résidus communs aux deux constructions présentent des pics avec une position similaire. On observe que les résidus spécifiques de la zone C-terminale de CstF-50dm (1-92) sont regroupés dans la région indicative de peptides désordonnés (valeurs  $^1\text{H}$  entre 8 et 8,5 ppm). **B)** Analyses de relaxation hétéronucléaire à partir de la protéine CstF-50dm (1-92) marquée  $^{15}\text{N}$ . Les valeurs de relaxation révèlent une flexibilité importante dans la région C-terminale (valeurs en dessous de 0,6). Une prédiction de la structure secondaire des chaînes principales est montrée au-dessus.

### c.) Cristallisation de CstF-50dm (1-65)

Après purification, la protéine est concentrée à **15-20 mg/ml**. Pour trouver les conditions de cristallisation de la protéine j'ai d'abord réalisé des tests à partir des kits commerciaux disponibles au laboratoire sur des boîtes de 96 puits. Plusieurs conditions dans les kits MbClass Suite II (Qiagen), Classics Suite (Qiagen), Crystal Screen I et II (Hampton) ont permis la formation de cristaux. L'optimisation a été effectuée sur des boîtes de 24 puits à l'aide des techniques de **goutte assise** et de **micro-ensemencement** (*détaillées dans la section Matériels et Méthodes*). Cette stratégie a permis l'obtention de cristaux exploitables de grande taille (300-600 nm) et d'une croissance rapide de seulement quelques heures (*fig. 36*).



**Figure 36.** Photographies des cristaux de CstF-50dm (1-65). Les images ont été prises en présence d'un filtre capable de polariser la lumière.

Finalement, la condition optimale de cristallisation contenait : **100mM HEPES à pH 7.5, 3% de PEG<sub>400</sub> et 1.0-1.6M de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>**. De plus, l'ajout de 10 à 20% de **glycérol** dans cette condition ne perturbe pas la cristallisation et permet, avec les molécules de PEG<sub>400</sub>, d'acquies des cristaux dans des conditions **cryoprotectrices** améliorant la collecte de données.

La région N-terminale de CstF-50 ne présente aucun homologue de séquence dans la PDB (Protein Data Bank). Ceci a empêché la résolution du problème des phases par la méthode du remplacement moléculaire. Le remplacement du soufre des Méthionines de la protéine par du **Sélénium** a donc été utilisé pour la résolution de la structure par la méthode du **SAD** (Single-wavelength Anomalous Diffraction). Ce remplacement a été possible grâce à l'introduction de **Se-Méthionine (Se-Met)** dans le milieu de culture ce qui provoque la substitution des Met par Se-Met dans la séquence protéique (*voir Matériel et Méthodes*). Les étapes de purification n'ont pas eu besoin d'être modifiées par rapport à celles utilisées pour la protéine native (sans Se). La

crystallisation de la protéine sélénée s'est achevée relativement facilement étant donné qu'elle cristallisait dans les mêmes conditions que la protéine native.

La rapidité de cristallisation et l'obtention de cristaux à partir de la protéine marquée ont été déterminantes dans le choix de la technique de **crystallographie**, plutôt que de RMN, pour la résolution de la structure du domaine N-terminal de CstF-50.

#### **d.) Résolution de la structure de CstF-50dm (1-65)**

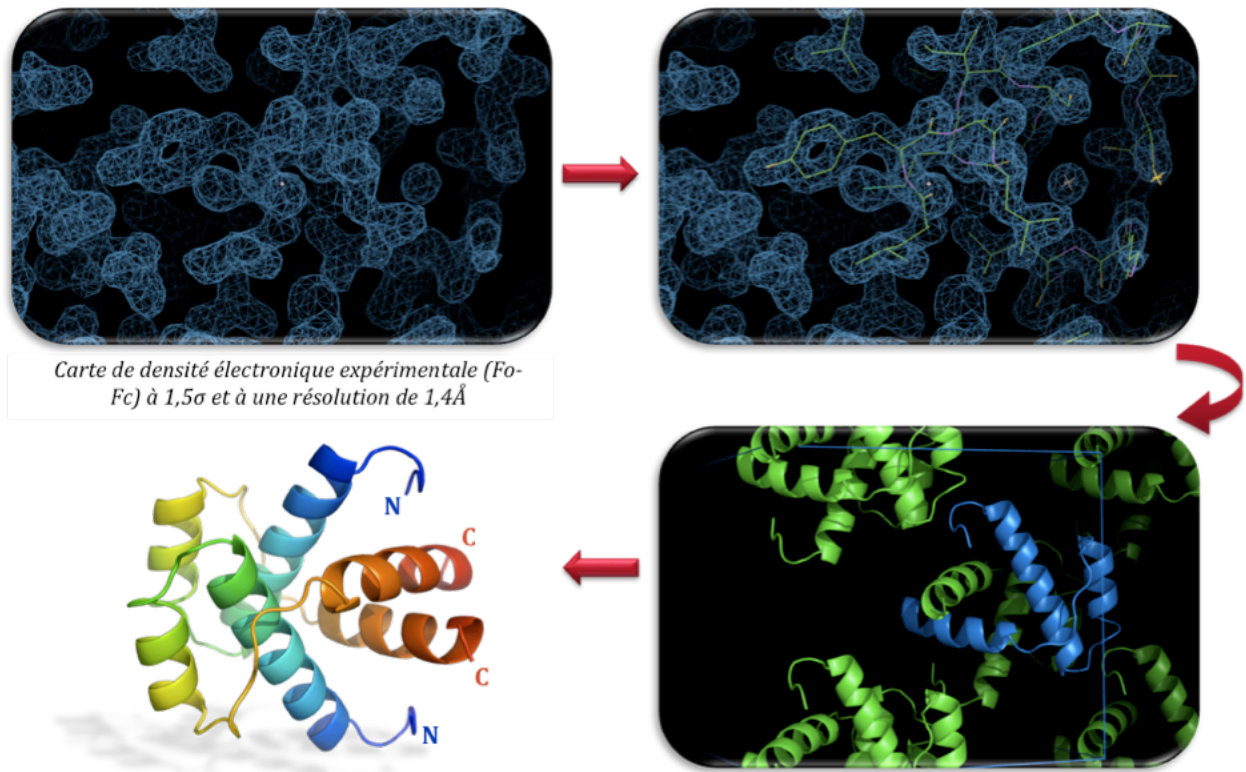
Les données de diffraction obtenues à partir des cristaux ont permis la détermination de la structure de ce domaine de CstF-50 à une résolution de **1,4 Å**. En effet, le fragment de 65 acides aminés contient dans la séquence 7 méthionines ; la substitution de ces résidus par des méthionines sélénées a été nécessaire pour l'obtention d'un modèle structural. Ainsi, nous avons pu utiliser la technique du **SAD** (Single-wavelength Anomalous Dispersion) pour résoudre le « problème des phases » (*plus en détail dans la section Matériels et Méthodes*). La protéine cristallise dans la condition mentionnée ci-dessus, sous sa forme dimérique et dans le groupe d'espace **c222<sub>1</sub>**. Les statistiques du traitement des données et de l'affinement de la structure sont présentées sur le *tableau 4*.

Le traitement des données a permis l'obtention d'une carte de densité électronique où l'on identifie clairement les chaînes latérales de la plupart des acides aminés (*fig. 37*). Après plusieurs étapes d'affinement nous avons obtenu le modèle final du domaine N-terminal de CstF-50dm. La protéine est entièrement  **$\alpha$ -hélicoïdale**. Les hélices 1 et 2 sont parallèles et la troisième est située de façon perpendiculaire aux autres.

Sur la *figure 37*, en bas à droite, il est montré un détail de la maille élémentaire comportant différentes unités asymétriques. Malgré une disposition avec une molécule par unité asymétrique, l'étude de l'arrangement de la protéine révèle la présence d'un **homodimère** qui repose sur l'axe de symétrie cristallographique. Cette observation serait en accord avec les études réalisées par « far-western blot » qui démontraient l'existence d'une association de CstF-50 avec elle-même (Takagaki and Manley 2000).

<i>Paramètres de la collecte de données</i>		
<b>Ligne de lumière</b>	SAD (ID14-4 ESRF)	Native (PROXIMA1-SOLEIL)
<b>Groupe d'espace</b>	C222 <sub>1</sub>	C222 <sub>1</sub>
<b>Paramètres de maille</b>		
<i>a b c (Å)</i>	45.90 51.90 57.40	46.20 52.25 57.85
<i>α β γ (°)</i>	90.0 90.0 90.0	90.0 90.0 90.0
<b>Longueur d'onde (Å)</b>	0.9765	0.9184
<b>Limites de résolution</b>	26.0 – 1.9	30.0 – 1.4
<b>Réflexions (totales/uniques)</b>	36406/10081	49453/16519
<b>R<sub>sym</sub> (%)</b>	2.00 (3.50)	3.90 (27.60)
<b>I/σ</b>	48.49 (27.78)	21.56 (5.13)
<b>Complétude (%)</b>	96.5 (91.60)	99.6 (99.3)
<b>Redondance</b>	3.61 (2.90)	3.52 (3.47)
<i>Phasage SAD</i>		
<b>Sites de Se trouvés/espérés</b>	4/7	
<b>Figure de mérite</b>	0.34	
<i>Paramètres d'affinement</i>		
<b>Rwork (%)</b>	-	18.17 (20.2)
<b>Rfree (%)</b>	-	21.74 (22.0)
<b>Facteur B moyen (Å<sup>2</sup>)</b>	-	16.75
<b>Nombre d'atomes</b>		
Protéine		543
Na <sup>+</sup>	-	4
PEG		7
H <sub>2</sub> O		41
<b>R.m.s.</b>		
Longueur de liaison (Å)	-	0.038
Angles (°)		2.988
<b>Solvant contenu (%)</b>	-	49
<b>Ramachandran (%)</b>	-	100

*Tableau 4. Statistiques cristallographiques de CstF-50dm (1-65). Les chiffres entre parenthèses correspondent à la tranche de plus haute résolution.*



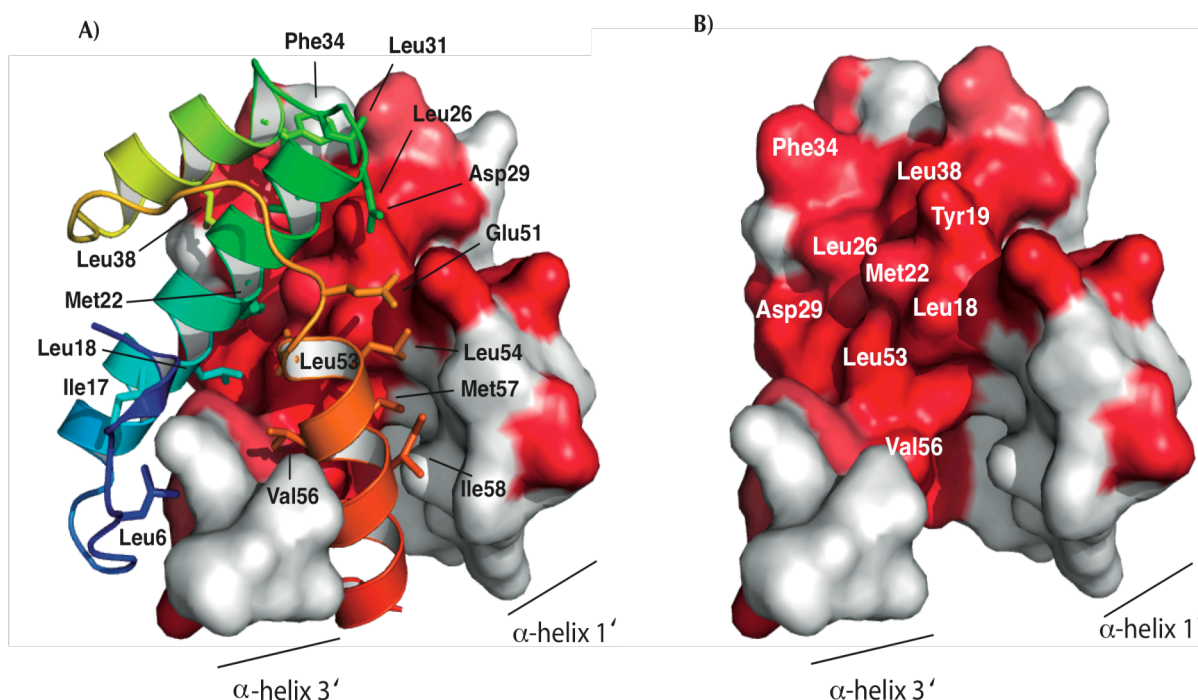
**Figure 37.** Les images montrent le chemin vers l'obtention du modèle structural de *CstF-50dm* (1-65) à partir d'une carte de densité électronique expérimentale. Les acides aminés doivent être placés dans cette densité. La haute résolution des données permet l'appréciation des contours des chaînes latérales et rend plus facile et plus exact le placement des résidus. En bas, à droite, l'image d'une maille élémentaire avec les différentes unités asymétriques et un des monomères coloré en bleu. A gauche, le modèle final de la structure coloré de bleu à rouge, de la région N-terminale à la C-terminale.

**e.) Le domaine N-terminal de CstF-50 forme un homodimère conservé**

Dans le dimère de *CstF-50dm* (1-65), la surface de contact entre les deux monomères est importante. Certains résidus forment une interface hydrophobe qui permet l'interaction entre les deux monomères. Les résidus Leu6, Leu18, Tyr19, Met22 et Leu26 de l'hélice  $\alpha_1$  et Leu31, Phe34, Leu38 et Leu41 de l'hélice  $\alpha_2$  sont impliqués dans la dimérisation. En outre, des contacts additionnels sont générés par les acides aminés Glu51, Leu53, Leu54, Val56, Met57, Ile58 et Met61. De cette façon, les trois hélices participent à la formation du dimère (fig. 38).

De plus, nous avons observé que la surface d'interaction entre les deux dimères est très conservée. Sur la *figure 38B*, la surface d'un des monomères est colorée avec un gradient de blanc à rouge des régions les moins conservées à celles que le sont le plus. Nous avons constaté que la surface impliquée dans la dimérisation est fortement conservée et donc, que les résidus dans le cœur hydrophobe de l'interface du dimère sont invariants entre les espèces.



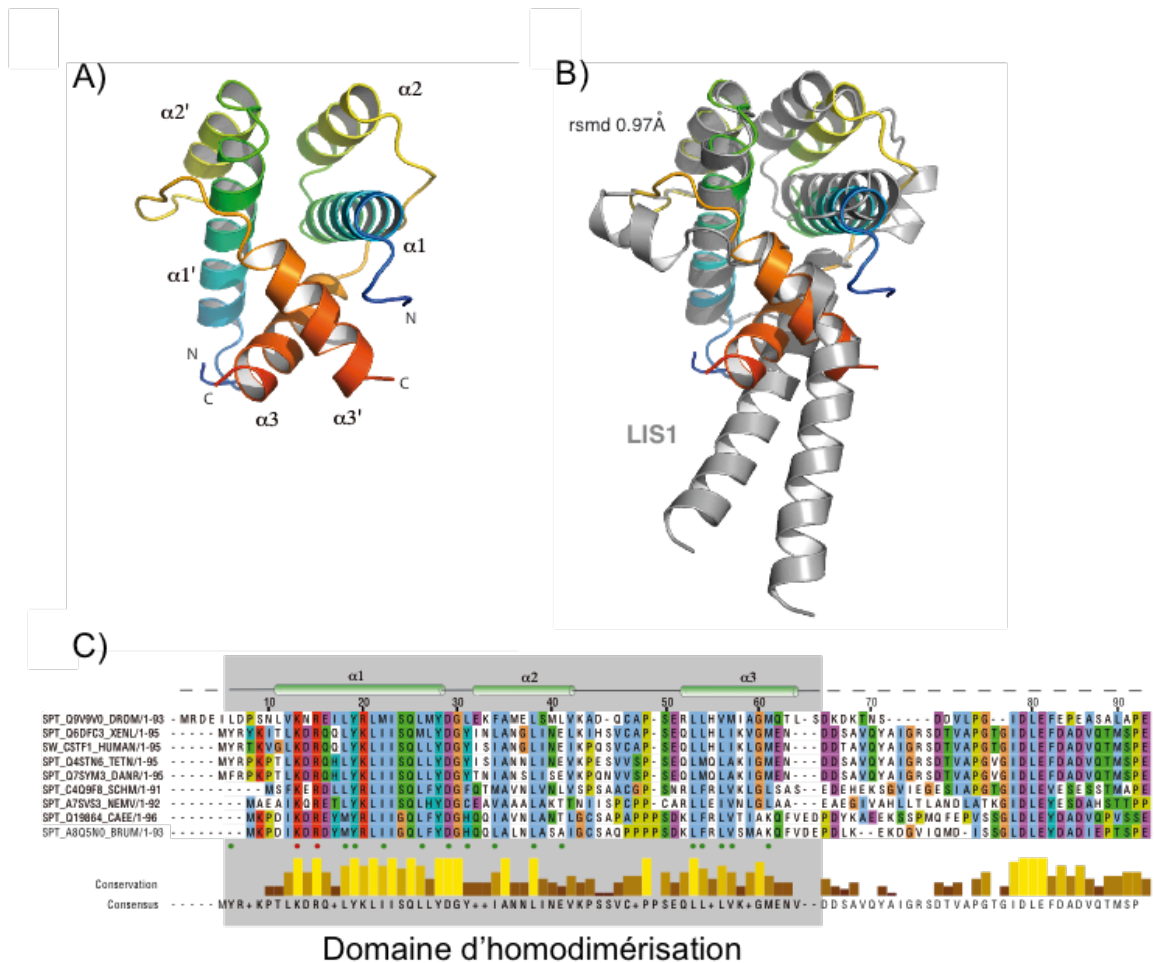


**Figure 38.** **A)** Un des monomères de CstF-50dm (1-65) est dessiné en mode « cartoon », coloré de bleu à rouge de la partie N-terminale à la partie C-terminale. L'autre monomère est représenté en surface. La coloration correspond à la conservation de séquence (de blanc à rouge du résidu moins conservé au plus conservé). Les acides aminés impliqués dans l'homodimérisation sont marqués dans le monomère représenté en « cartoon ». **B)** Un des monomères a été omis et sur la surface du deuxième on observe la conservation de la région d'interaction. Pour la réalisation de la figure nous avons utilisé le programme ESPript (Gouet et al. 2003) pour la coloration en fonction de la conservation associé au programme Pymol pour la représentation graphique.

D'autre part, des études de comparaison des structures en 3D réalisées à l'aide du serveur DALI (Holm and Rosenstrom 2010) ont montré l'existence d'un motif dont l'architecture ressemble à celle de CstF-50dm (1-65). Il s'agit du motif **LISH** (LIS1 homology motif) qui doit son nom à la protéine **LIS1** (lissencephaly protein 1) (Emes and Ponting 2001). La **structure** cristallographique du domaine **N-terminal de LIS1** a été résolue (Kim et al. 2004) (code PDB : 1UUJ). Cette région inclut le motif LISH (du résidu 1 au 39) suivi d'un fragment « coiled-coil » ou superhélice (de l'acide aminé 39 jusqu'au 86). Les deux domaines sont impliqués dans la dimérisation thermodynamiquement stable de LIS1 mais la présence du « coiled-coil » n'est pas critique (Tai et al. 2002) (Kim et al. 2004). Néanmoins, le « coiled-coil » est labile et peut adopter différentes conformations ce qui pourrait signifier l'existence d'une fonction structurale encore inconnue (Mateja et al. 2006).

La structure de la région N-terminale de LIS1 et celle obtenue pour CstF-50dm (1-65) au laboratoire ont été superposées. Les deux modèles structuraux présentent un rmsd (root mean square deviation) de **0.97 Å** (fig. 39). Sur la figure 39B on observe que les hélices  $\alpha_1$ ,  $\alpha_1'$ ,  $\alpha_2$  et  $\alpha_2'$  de

NtCstF-50dm sont parfaitement superposées avec le motif LISH de Lis1. Ceci démontre l'homologie entre les deux régions de ces protéines.



**Figure 39.** *A)* Représentation en « cartoon » du dimère CstF-50dm (1-65). Dans la figure, chaque monomère est coloré de bleu à rouge de la partie N-terminale à la C-terminale. *B)* Superposition des domaines N-terminaux de CstF-50dm (1-65) et de son homologue structural LIS1 (en gris). L'orientation de la première figure a été maintenue. *C)* Alignement de séquences des orthologues de la région N-terminale de CstF-50. Une représentation de la structure secondaire est marquée au-dessus. Les points rouges et verts situés en dessous de l'alignement correspondent aux résidus impliqués dans la dimérisation et aux résidus exposés au solvant, respectivement. La conservation de la séquence est représentée par un histogramme. Les espèces utilisées pour l'alignement sont similaires à celles employées pour l'alignement de la protéine entière (fig. 30).

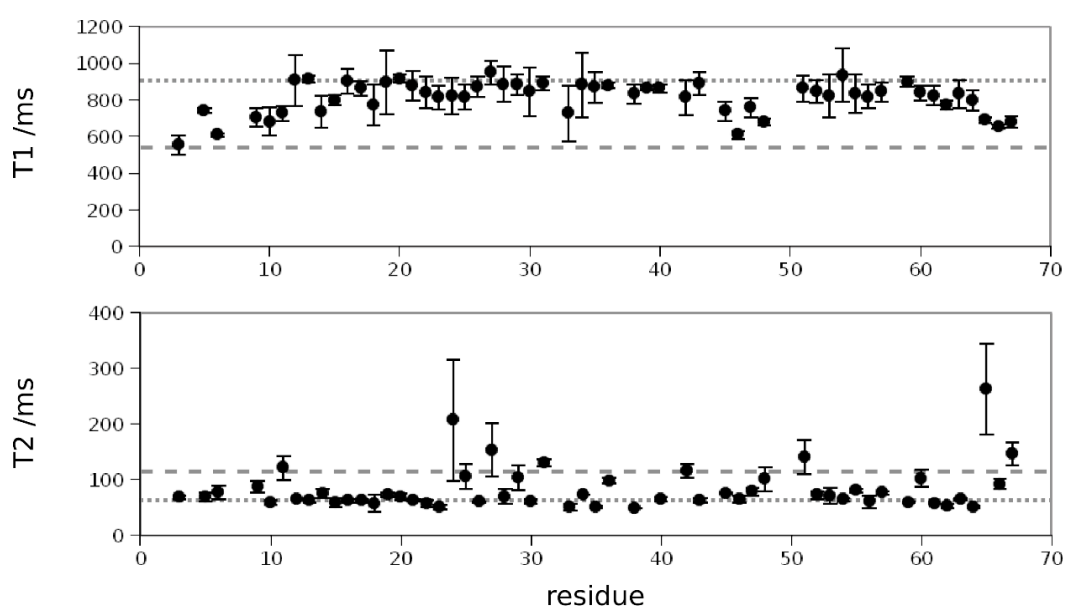
Le motif LISH est généralement composé de 33 acides aminés et est caractérisé par la présence de résidus hydrophobes (Emes and Ponting 2001). Il a été identifié dans au moins 100 protéines eucaryotes. La présence d'un motif conservé implique souvent des caractéristiques fonctionnelles communes. Le motif LISH est connu pour sa capacité à former des dimères, (Emes and Ponting 2001) son principal rôle biologique étant celui de faciliter l'oligomérisation (Emes and Ponting



2001) (Kim et al. 2004). Cependant, d'autres fonctions lui ont été attribuées. Ainsi, il a été montré que LISH peut affecter la demi-vie d'une protéine et influencer sa localisation cellulaire (Gerlitz et al. 2005).

LISH est retrouvé, par exemple, dans une des sous-unités du TFIIB (un facteur d'initiation de la transcription), **TAFII-100**. Ce motif va permettre l'interaction de cette sous-unité avec le reste du complexe (Dubrovskaya et al. 1996). De plus, de nombreux facteurs contenant un motif LISH comptent aussi un domaine **WD-40** (Emes and Ponting 2001). LIS1, TAFII-100, et potentiellement **CstF-50**, feraient partie de ce groupe de protéines.

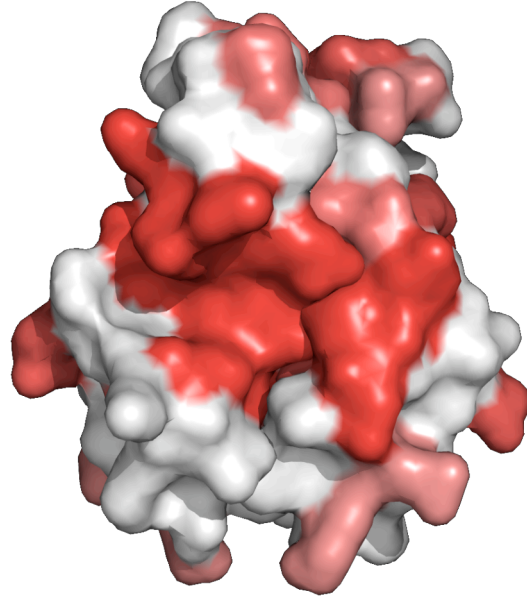
Nous pouvons ainsi conclure que le domaine N-terminal de CstF-50dm est organisé en un **homodimère**. Les résidus impliqués sont **conservés** à travers les espèces. Nous avons montré que les résidus **1-65** sont les responsables de la dimérisation et ils sont suffisants pour la formation du dimère. De plus, l'homologie de structure avec un motif généralement associé à la formation de dimères est en accord avec nos conclusions. Nous avons voulu confirmer ces résultats et analyser le comportement de ce domaine de CstF-50 en solution. Pour cela, nous avons procédé à l'étude de la même construction de CstF-50dm par spectroscopie **RMN**. Des expériences de relaxation effectuées à partir de la protéine marquée  $^{15}\text{N}$  indiquent que la taille de la protéine est de approximativement **16kDa** et non 8kDa comme on devrait l'observer pour le monomère (*fig. 40*). De plus, le même domaine contenant une mutation dans l'interface de dimérisation, **Y19A**, a été analysé par RMN et une perte de la dimérisation a pu être observée. Ces résultats indiquent que CstF-50 (1-65) est organisé en dimère en solution comme dans le cristal.



**Figure 40.** Expérience de relaxation  $^{15}\text{N}$ -amide à 298K et 700 MHz. Les valeurs T1 et T2 sont cohérentes avec celles prédites pour l'homodimère (ligne pointillée fine) et non pour le monomère (trait discontinu).

**f.) Conservation de la surface du dimère de CstF-50dm (1-65).**

L'alignement de séquences réalisé par ESPript (Gouet et al. 2003) a été utilisé pour la représentation des résidus conservés sur la surface de la protéine. Sur la *figure 38*, il a été montré l'existence d'une forte conservation des résidus situés à la surface d'interaction des deux monomères. D'autre part, une région très conservée est révélée lorsque nous effectuons le même type de représentation de conservation de surface sur la totalité du dimère (*fig. 41*). La zone conservée inclut la surface formée à partir de l'interaction entre les deux monomères. Par ailleurs, la rotation de 90° du modèle structural sur le même axe dévoile l'existence d'un type de surface similaire de l'autre côté du dimère. Ces régions comprennent les acides aminés Val12, Lys13, Arg15, Tyr19, Arg20, Tyr28, Asp29, Gly30 et Val42 de chaque monomère et forment une cavité qui pourrait accueillir d'autres éléments.

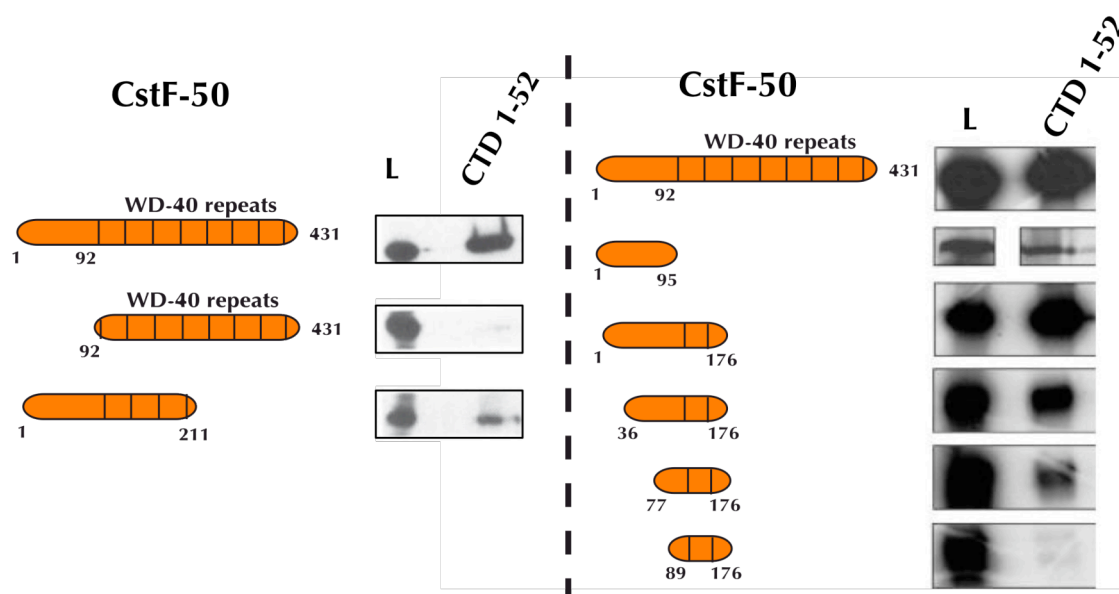


**Figure 41.** Représentation de la conservation des séquences sur la surface du dimère de la protéine CstF-50dm (1-65).

En effet, des résidus conservés et exposés à la surface peuvent être des bons candidats pour intervenir dans l'interaction avec d'autres facteurs protéiques. Il a été observé que la région N-terminale de CstF-50 était capable d'interagir avec le domaine carboxy terminal (CTD) d'une des sous-unités de l'ARN pol II, Rpb1 (Fong and Bentley 2001) (McCracken et al. 1997). De cette façon, notre première hypothèse propose que cette zone conservée, située à la surface du dimère formé par la région N-terminale de CstF-50dm, pourrait être impliquée dans l'interaction avec le CTD de l'ARN pol II.

**g.) Etudes de l'interaction de CstF-50 avec le CTD de l'ARN pol II**

Deux travaux différents, effectués dans le même groupe de recherche, ont mis en évidence l'interaction de la protéine CstF-50 humaine avec un GST-CTD immobilisé (Fong and Bentley 2001) (McCracken et al. 1997). Les expériences ont été réalisées à partir des fragments de la protéine CstF-50 partiellement purifiés ou transcrits *in vitro* par la technique du TNT (transcription-traduction *in vitro*). Des expériences de « pull-down » suivis d'immunoprécipitation avec des anticorps ont montré l'existence d'une interaction entre les deux facteurs. Il a été observé que la région N-terminale de CstF-50, du premier résidu au résidu 93, était suffisante pour l'interaction (*fig 42*).



D'après McCracken et al. 1997

D'après Fong & Bentley, 2001

**Figure 42.** Etudes précédentes réalisées pour l'identification des domaines de CstF-50 impliqués dans l'interaction avec le CTD. Différents fragments de la sous-unité CstF-50 obtenus par traduction 'in vitro' sont incubés avec GST-CTD. A gauche les résultats obtenus par D. Bentley et ses collaborateurs en 1997 et à droite ceux obtenus par N. Fong et D. Bentley en 2001. Les différents fragments de CstF-50 utilisés pour chaque expérience sont indiqués. L : « load », 10% de l'extrait total a été chargé sur le gel ; CTD 1-52 : représente 50% des éluats obtenus à haute concentration de sel après incubation GST-CTD/CstF-50.

D'un autre côté, il est connu que CstF-50hs lie de façon préférentielle les répétitions 27 à 52 des heptapeptides  $Y_1S_2P_3T_4S_5P_6S_7$  où la plupart des Ser situées à la septième position de l'heptapeptide ( $S_7$ ) sont remplacées par des Lys ( $K_7$ ) (Fong and Bentley 2001). Par ailleurs, les protéines de la maturation 3' qui s'associent au CTD, comme Pcf11p, le font avec une préférence pour des heptapeptides phosphorylés en  $S_2$ ,  $pS_2$  (Buratowski 2003) (Meinhart et al. 2005).

A partir de ces informations, nous avons envisagé la réalisation de différentes expériences qui détermineraient quels résidus de la région N-terminale de CstF-50 participent directement dans l'interaction avec le CTD de l'ARN pol II. Malgré les efforts et beaucoup de temps investi, aucune évidence claire de cette interaction n'a pu être révélée. Les expériences réalisées sont présentées dans les sections suivantes.

Pour les premières études d'interaction nous avons fait synthétiser des peptides dont les caractéristiques sont en accord avec les données trouvées dans la littérature. Une liste des peptides synthétiques utilisés est présentée ci-dessous (tableau 5) :

Nom dans cette thèse	Séquence du Peptide Synthétisé	Fournisseur
Btn-CTD	Biotine- Y <sub>1</sub> S <sub>2</sub> P <sub>3</sub> T <sub>4</sub> S <sub>5</sub> P <sub>6</sub> S <sub>7</sub> Y <sub>1</sub> ' S <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' S <sub>5</sub> ' P <sub>6</sub> ' S <sub>7</sub> '	Eurogentec
Btn-CTDpS <sub>2</sub>	Biotine- Y <sub>1</sub> pS <sub>2</sub> P <sub>3</sub> T <sub>4</sub> S <sub>5</sub> P <sub>6</sub> S <sub>7</sub> Y <sub>1</sub> ' pS <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' S <sub>5</sub> ' P <sub>6</sub> ' S <sub>7</sub> '	Eurogentec
Btn-CTDpS <sub>5</sub>	Biotine- Y <sub>1</sub> S <sub>2</sub> P <sub>3</sub> T <sub>4</sub> pS <sub>5</sub> P <sub>6</sub> S <sub>7</sub> Y <sub>1</sub> ' S <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' pS <sub>5</sub> ' P <sub>6</sub> ' S <sub>7</sub> '	Eurogentec
CTDpS <sub>2</sub>	P <sub>6</sub> S <sub>7</sub> Y <sub>1</sub> ' pS <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' S <sub>5</sub> ' P <sub>6</sub> ' S <sub>7</sub>	Eurogentec
Btn-CTD-K	Biotine- S <sub>5</sub> P <sub>6</sub> K <sub>7</sub> Y <sub>1</sub> ' S <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' S <sub>5</sub> ' P <sub>6</sub> ' K <sub>7</sub> ' Y <sub>1</sub> '' S <sub>2</sub> '' P <sub>3</sub> '' T <sub>4</sub> '' S <sub>5</sub> '' P <sub>6</sub> '' K <sub>7</sub> ''	Sigma
Btn-CTD-KpS <sub>2</sub>	Biotine- S <sub>5</sub> P <sub>6</sub> K <sub>7</sub> Y <sub>1</sub> ' S <sub>2</sub> ' P <sub>3</sub> ' T <sub>4</sub> ' S <sub>5</sub> ' P <sub>6</sub> ' K <sub>7</sub> ' Y <sub>1</sub> '' pS <sub>2</sub> '' P <sub>3</sub> '' T <sub>4</sub> '' S <sub>5</sub> '' P <sub>6</sub> '' K <sub>7</sub> ''	Sigma

**Tableau 5.** Peptides synthétiques utilisés pour les études d'interaction entre CstF-50 et le CTD.

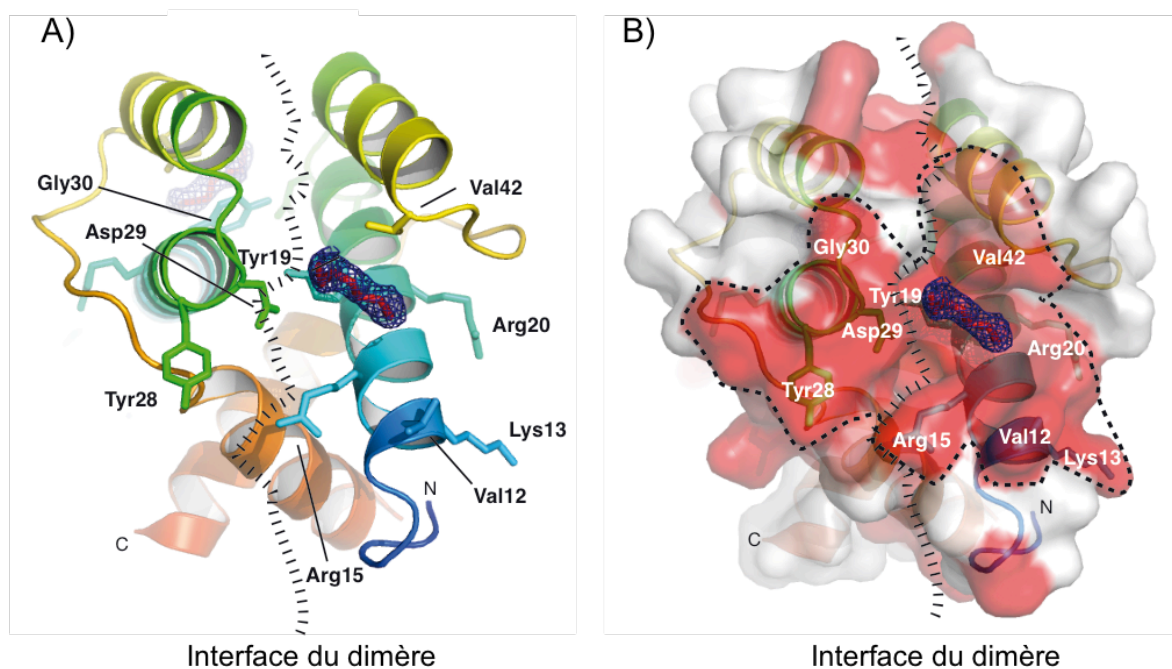
*g.1) Co-cristallisation et trempage*

Pour les essais de **co-cristallisation**, la protéine a été pré-incubée avec le peptide **CTDpS<sub>2</sub>** avant les essais de cristallisation. Selon notre hypothèse, si l'interaction a lieu entre CstF-50dm (1-65) et le CTD, elle doit impliquer le dimère formé par NtCstF-50 et deux molécules du peptide CTD pS<sub>2</sub>, étant donné que deux sites potentiels d'interaction sont situés des deux côtés du dimère.

De cette façon, nous avons combiné les deux substrats, peptide et protéine, dans une stœchiométrie de 1 mole de peptide pour 1 mole de protéine (1 :1). Après une nuit d'incubation à 4°C, nous avons réalisé les essais de cristallisation utilisant les mêmes conditions que pour la cristallisation de la protéine native. La concentration finale de l'échantillon était de 14mg/ml.

L'autre méthode utilisée pour la cristallisation de la protéine avec le peptide CTDpS<sub>2</sub> est le **trempage** (ou « soaking »). Cette technique implique la réalisation de boîtes de cristallisation de la protéine seule et l'addition postérieure du peptide dans chaque goutte contenant des cristaux. Ainsi, nous avons d'abord obtenu des cristaux de la protéine par la méthode expliquée dans la *section 2.1.b*. Deux jours plus tard, 0,5µl de peptide, à une concentration de 9,5µM, ont été ajoutés dans chaque goutte contenant des cristaux afin que le peptide diffuse dans les cristaux déjà formés.

A partir de ces deux stratégies, nous avons obtenu le même type de cristaux que pour la protéine seule. L'analyse des données de diffraction obtenues au synchrotron à partir de ces cristaux indique l'absence du CTD. Cependant, dans la zone pensée pour l'interaction, nous avons observé la présence d'une densité électronique non attribuée à la protéine. Après l'étude des différentes possibilités nous avons conclu que cette densité correspondrait à une molécule de **PEG**, un des composants de la solution de cristallisation (*voir aussi sur le tableau 4*). De ce fait, on ne peut exclure que cette molécule de PEG entre en compétition avec le CTD pour l'interaction avec NtCstF-50dm (*fig. 43*).



**Figure 43.** Disposition d'une molécule de PEG dans la cavité conservée formée à la surface du dimère CstF-50dm (1-65). La densité électronique observée (ici Fobs-Fcalc, contourée à  $1.5\sigma$ ) correspond à une molécule de PEG présente dans la condition de cristallisation. A) Représentation en « cartoon » de la protéine où l'on observe les molécules de PEG dans les deux cotés du dimère. B) Représentation de la surface de la protéine orientée de la même manière. Les résidus impliqués sont indiqués.

g.2) Etudes de l'interaction de CstF-50dm (1-65) avec le CTD de l'ARN pol II par la technique de Résonance Magnétique Nucléaire (RMN)

La technique de **RMN** est utilisée pour la résolution de structures des protéines de petite taille en solution, mais elle est aussi un outil important pour l'étude d'interactions **protéine-ligand**. Pour ces expériences, un spectre à 2 dimensions HSQC référence de la protéine seule est enregistré et représente une empreinte RMN de cette protéine. La fixation d'un ligand sur une protéine modifie l'environnement des atomes de la protéine et donc le déplacement chimique de ces atomes. L'interaction protéine-ligand provoque donc une modification de l'empreinte de la protéine.

Pour les expériences d'interaction avec le CTD, nous avons utilisé le peptide **CTDpS<sub>2</sub>**. Le titrage de  $^{15}\text{N}$  CstF-50dm (1-65) avec ce peptide a été réalisé à partir de la protéine à une concentration constante de  $880\mu\text{M}$ . A cette quantité fixe de protéine nous avons ajouté du peptide en solution pour obtenir une gamme d'équivalents molaires peptide : protéine de  $100\mu\text{M}$  à  $1\text{mM}$ .

Nous avons observé que les spectres de RMN en présence du peptide ne variaient que très légèrement par rapport au spectre référence, même dans les cas où la concentration du peptide était très importante. Nous ne pouvons pas conclure si les changements observés sont provoqués par une interaction spécifique ou aspécifique de deux composants. Si une interaction spécifique a lieu, la constante d'affinité est de l'ordre du mM ( $K_d > 1\text{mM}$ ).

g.3) Etudes d'interaction de CstF-50dm (1-65) et le CTD de l'ARN pol II par SPR (Surface Plasmon Resonance)

La **résonance plasmonique de surface** est un phénomène physique utilisé comme méthode de mesure de la liaison d'un ligand sur un récepteur fixé à la surface d'une couche métallique. Un système de détection SPR mesure la variation de l'indice de réfraction au voisinage de l'interface lorsque le ligand se fixe aux récepteurs.

Les études d'interaction se sont déroulées à l'IECB sur un système **BIAcore 3000™** (GE Healthcare) en collaboration avec le **Dr. Carmelo Di Primo** au sein de l'équipe « Aptamers and artificial small regulatory RNA » (U869) dirigé par Jean-Jaques Toulmé. Dans les premiers essais, le récepteur correspondait aux peptides du CTD biotinylés, phosphorylés ou non phosphorylés (**Btn-CTD**, **Btn-CTDpS<sub>2</sub>**, **Btn-CTDpS<sub>5</sub>**). La surface du métal a été fonctionnalisée avec de la streptavidine pour permettre la liaison des peptides biotinylés. Les peptides ont été immobilisés sur deux pistes différentes d'une même puce et une solution contenant la protéine a été injectée. Malgré la variation de la concentration protéique et du tampon de solution, aucun signe d'interaction entre les deux substrats a été observé.

Ainsi, de nouveaux essais ont été réalisés en immobilisant cette fois CstF-50dm (1-65) sur une puce préalablement préparée. Des concentrations croissantes (de 30µM à 2mM) des différents peptides CTD (Btn-CTD, Btn-CTDpS<sub>2</sub>, Btn-CTDpS<sub>5</sub>) ont été injectés. Après différents essais, nous n'avons pas obtenu par cette méthode de résultats nous permettant de conclure sur l'existence d'une interaction entre la protéine CstF-50dm et le CTD.

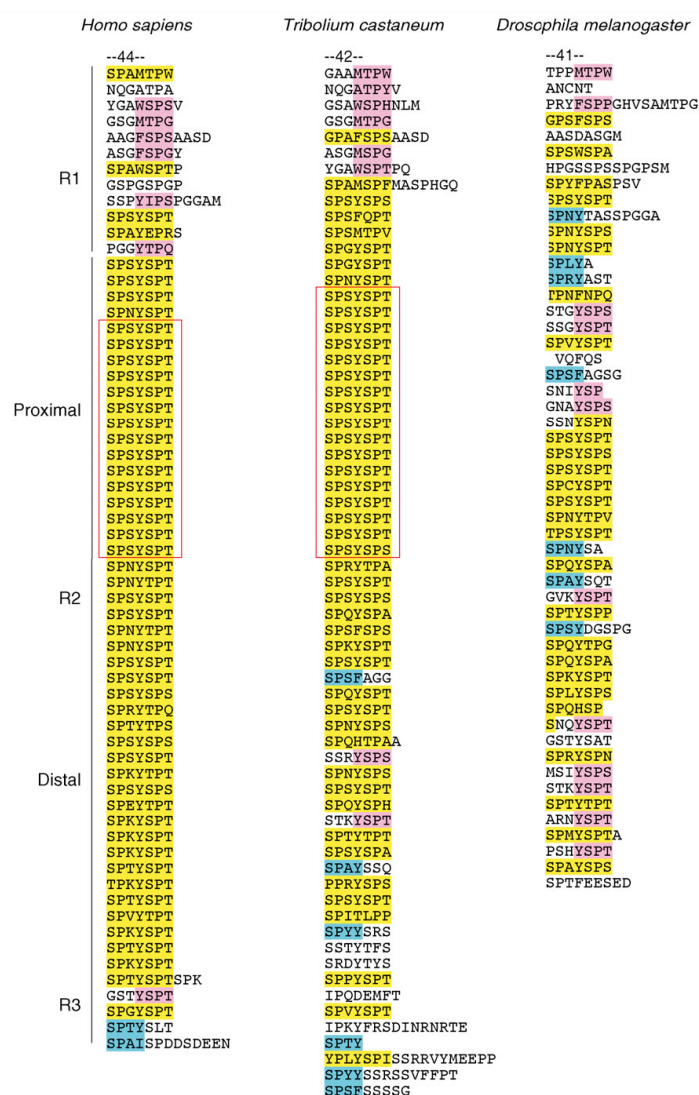
g.4) Etudes d'interaction de CstF-50 de drosophile et le CTD de l'ARN pol II par des expériences de « pull-down »

Pour ces expériences, les constructions de la protéine CstF-50dm (1-65) et (1-92) ont été incubées pendant la nuit à 4°C avec les peptides **Btn-CTD-K** ou **Btn-CTD-KpS<sub>2</sub>**. La substitution de la Ser en position 7 par une Lys est basée sur les observations effectuées par D. Bentley et ses collaborateurs (Fong and Bentley 2001). Comme contrôle positif nous nous sommes servis de la région de la protéine **Pcf11p** recombinante qui intervient dans l'interaction avec le CTD (CID, CTD Interaction Domain) (Barilla et al. 2001) (Meinhart and Cramer 2004). L'utilisation des billes de streptavidine commerciales (*Dynabeads MyOne Streptavidin T1, Invitrogen*) permet la récupération des peptides biotinylés et des protéines potentiellement associées à ces peptides. Les billes ont été ajoutées au mélange CTD-CstF50dm et l'ensemble a été incubé pendant 2 heures à 4°C. Après trois lavages avec un tampon contenant 250mM de NaCl, les billes sont chargées directement sur un gel SDS-PAGE. Après plusieurs essais avec des variations du tampon de lavage, du temps d'incubation ou du nombre de lavages, le CID de Pcf11p et la construction CstF-50dm (1-92) ont montré une interaction non spécifique aux billes. De plus, l'interaction entre les deux peptides CTD (phosphorylé et non phosphorylé) et CstF-50dm (1-65) n'a pas pu être mise en évidence.



g.5) Vers une nouvelle stratégie : utilisation des protéines humaines pour les nouvelles expériences

Dans la composition du CTD de la majorité des espèces étudiées chez les mammifères, et même dans d'autres organismes très éloignés comme *Tribolium castaneum* (le petit ver de farine), la plupart des peptides répétés sont similaires à l'exception de l'acide aminé de la septième position qui est plus variable. Néanmoins, il a été observé que le CTD de *Drosophila melanogaster* contient une quantité plus faible d'heptapeptides avec la même séquence (Chapman et al. 2008) (fig. 44). De cette façon, établir les bases de l'interaction à partir des protéines de drosophile peut s'avérer compliqué. Il est possible que le choix des séquences peptidiques employées pour les expériences présentées précédemment ne soit pas adéquat.



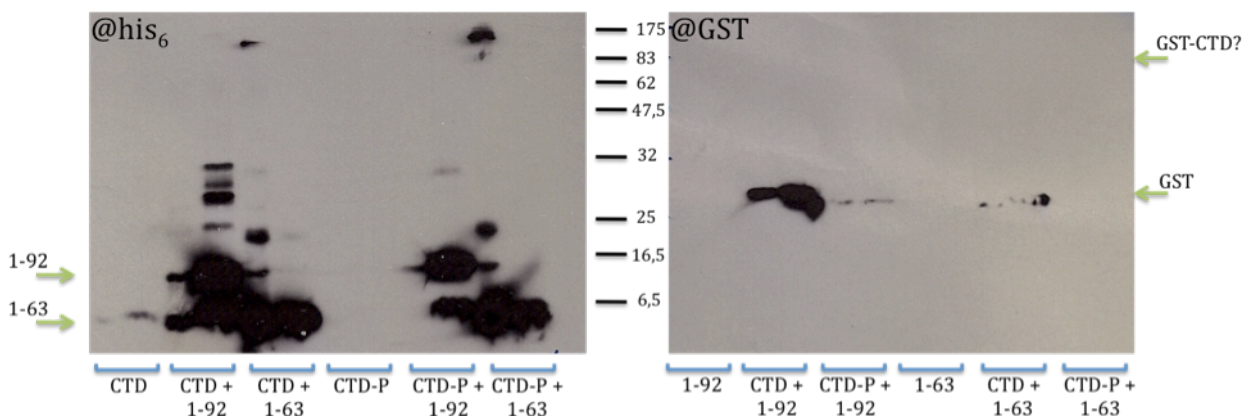
**Figure 44.** Adaptée de (Chapman et al. 2008). Comparaison du CTD de différents organismes. Le CTD de la sous-unité Rbp1 de l'ARN pol II humaine et de celui de *Tribolium castaneum* contiennent un nombre important de séquences canoniques de l'heptapeptide YSPSSPT (boîtes rouges), contrairement à *D. melanogaster* où le nombre de peptides avec une séquence similaire est plus faible.

De cette manière, une nouvelle stratégie pour l'étude de l'interaction CstF-50/CTD a été envisagée. Des clonages pour l'obtention des constructions nécessaires de la protéine CstF-50hs ont été effectués. Le plasmide contenant l'ADN codant pour une protéine GST-CTD humaine a été fourni par K. Ryan, du « Department of Chemistry and Biochemistry, City University of New York ». La suite des expériences décrites (g.6 à g.9) repose sur l'utilisation des protéines humaines.

*g.6) Etudes d'interaction de CstF-50hs, purifiée, et le CTD de l'ARN pol II par « pull-down »*

Des essais de phosphorylation ont d'abord été réalisés sur la protéine GST-CTDhs grâce à l'utilisation de kinases commerciales capables de phosphoryler le CTD. **Cdk1-cyclin B** catalyse de façon spécifique la phosphorylation en Ser<sub>2</sub> et Ser<sub>5</sub>, et **Cdk7** uniquement en Ser<sub>5</sub>. L'anticorps monoclonal **H14**, qui reconnaît la phosphorylation en Ser<sub>5</sub>, ainsi que l'anticorps monoclonal **H5**, qui reconnaît pSer<sub>2</sub>, ont été utilisés pour vérifier les réactions. Les deux anticorps reconnaissent le CTD phosphorylé sur n'importe quelle Serine, ce qui indique que les anticorps ne sont pas très spécifiques ou que les enzymes n'étaient pas restrictives d'un site de phosphorylation. Cependant, le CTD non phosphorylé n'était pas reconnu, donc pour les expériences suivantes nous avons utilisés les deux, CTD et CTD-P.

Pour ces expériences, les constructions de la protéine humaine de l'acide aminé 1 au 95 et du résidu 1 au 63, His<sub>6</sub>CstF-50hs (1-92) et His<sub>6</sub>CstF-50hs (1-63), sont purifiées et incubées avec le GST-CTDhs, toute la nuit à 4°C. Ensuite, 40µl de résine Co<sup>2+</sup> ou de résine Glutathione Sepharose sont ajoutés et l'ensemble est incubé pendant 2 heures à 4°C. Après trois lavages, 10µl de résine sont chargés sur un gel SDS-PAGE puis un Western blot est réalisé. Pour le « pull-down » effectué avec la résine Cobalt nous avons utilisé un anticorps anti-GST, et pour celui avec la résine Glutathione Sepharose, un anti-HisTag. Les résultats obtenus étaient décevants, GST-CTD n'était pas présent, ou en quantités très faibles, ou il n'était pas reconnu par l'anticorps anti-GST. Nous n'avons pas pu conclure l'existence d'une interaction à partir de cette expérience (fig. 45).



**Figure 45.** Résultats de l'expérience de « pull-down » de CstF-50hs et GST-CTD. La révélation a été effectuée sur des films après « western blot ». Sur la figure, 1-92 correspond à la protéine CstF-50hs (1-92) et 1-63 à CstF-50hs (1-63). CTD est la protéine GST-CTDhs et CTD-P correspond à la même protéine phosphorylée.



g.7) Etudes d'interaction de CstF-50hs, obtenue à partir des extraits totaux, et le CTD de l'ARN pol II par « pull-down »

Les expériences de pull-down précédemment décrites impliquent la purification des différents fragments de la protéine CstF-50hs. Ces purifications se sont avérées compliquées, certains des fragments étant insolubles. Pour les nouvelles expériences de « pull-down » nous avons donc décidé de partir **d'extraits totaux** obtenus après la surexpression des protéines ou de fragments de protéines chez *E. coli*. Plusieurs types d'expériences ont été effectués en partant de deux stratégies différentes :

i) Les protéines GST-CTDhs et GST (contrôle) sont d'abord immobilisées sur une **résine GST** (Gluthatione Sepharose™ 4 Fast Flow, *Amersham Biosciences*). Ensuite, la résine liée aux protéines est incubée, pendant 4-16 heures à 4°C, avec les extraits totaux des bactéries où CstF-50hs, différentes régions de CstF-50hs ou CstF-50hs contenant différentes mutations ponctuelles ont été surexprimés. Après trois lavages avec un tampon PBS, dont nous avons varié la concentration saline selon l'expérience, les échantillons (10µl de résine) sont chargés sur un gel SDS-PAGE. Finalement, un « Western Blot » est effectué utilisant l'anticorps anti-His<sub>6</sub> pour révéler les différentes régions de CstF-50hs et l'anticorps anti-GST pour contrôler la présence de GST et de GST-CTDhs.

ii) Des extraits totaux de la protéine CstF-50hs, des fragments de CstF-50hs ou des mutants ponctuels de celle-ci, sont d'abord incubés avec une **résine d'affinité pour l'étiquette de 6 His**. Les résines Ni<sup>2+</sup> ou Co<sup>2+</sup> (HIS-Select® Nickel Affinity Gel ou HIS-Select® Cobalt Affinity Gel, *Sigma*), et des billes magnétiques Co<sup>2+</sup> (Dynabeads® His-Tag Isolation & Pulldown, *Invitrogen*) ont été utilisées. La résine employée était ensuite incubée avec des extraits solubles contenant les protéines GST ou GST-CTDhs pendant 4-16 heures à 4°C. Après trois lavages avec un tampon PBS, les échantillons (10µl de résine) sont chargés sur un gel SDS-PAGE. Ensuite, un « Western Blot » est effectué utilisant l'anticorps anti-GST pour révéler GST et GST-CTDhs et avec l'anticorps anti-His<sub>6</sub> pour contrôler la présence de CstF-50hs.

En général, les tampons utilisés pour les lavages ont été les suivants :

- Tampon A : PBS 1x, 0,5% NP40, 250mM NaCl
- Tampon B : PBS 1,5x, MgAcétate 1mM, NP-40 0,1%, Imidazole 20mM, Glycerol 10%
- Tampon C : PBS 1x, 0,5% NP40, 250mM NaCl, 20mM glutathion réduit (tampon d'élution, dans ces cas, les protéines sont d'abord détachées de la résine GST et ensuite chargées sur le gel).

Les protéines tronquées de CstF-50hs employées sont : CstF-50hs (1-63), CstF-50hs (1-92), CstF-50hs (64-431) et CstF-50hs (92-431). Par ailleurs, cinq mutants de CstF-50hs ont été créés pour les expériences. Parmi les cinq, quatre contiennent une seule mutation ponctuelle (Y15A, Y24G, D25G ou G26N) et le cinquième porte quatre mutations sur la région située entre le domaine N-terminal et le WD-40 (D82S, E84S, D86S, D88S).

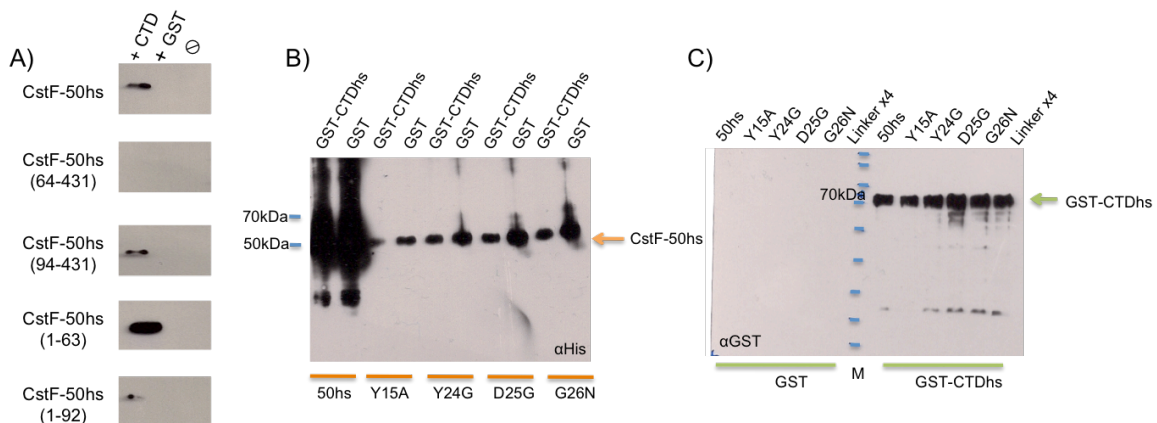
Nous avons récapitulé sur le *tableau 6* les expériences effectuées par cette méthode :

Résultats

	GSTCTDhs	GST	Anti-His	Anti-GST	50hs	Délétions	Mutations	Tampon A	Tampon B	Tampon C	Problèmes constatés
Résine GST	✓	✓	✓	✓	✓	x	✓	✓	x	✓	Contrôle GST +50hs positif → Précipitation 50hs (fig. 46B)
	✓	✓	✓	✓	✓	✓	x	✓	x	x	Précipitation/expression 50hs (fig. 46A)
	✓	✓	✓	✓	✓	✓	x	✓	x	✓	Précipitation/expression 50hs
Résine Co <sup>2+</sup>	✓	✓	✓	✓	✓	x	✓	✓	✓	x	Contrôle GST-CTDhs + résine positif (fig. 46C)
Résine Ni <sup>2+</sup>	✓	✓	✓	✓	✓	x	✓	✓	✓	x	Contrôle GST-CTDhs + résine positif
Billes Co <sup>2+</sup>	✓	✓	✓	✓	✓	x	✓	✓	✓	x	?*

**Tableau 6.** Expériences « pull down » couplées au « western blot » pour l'interaction entre des extraits totaux de CstF-50hs, ou des versions de cette protéine modifiée, et le CTD de l'ARN pol II. \*Dans ce cas-là, les expériences sont en cours au laboratoire, le contrôle GST-CTDhs + billes étant négatif nous pouvons procéder à la réalisation des expériences d'interaction.

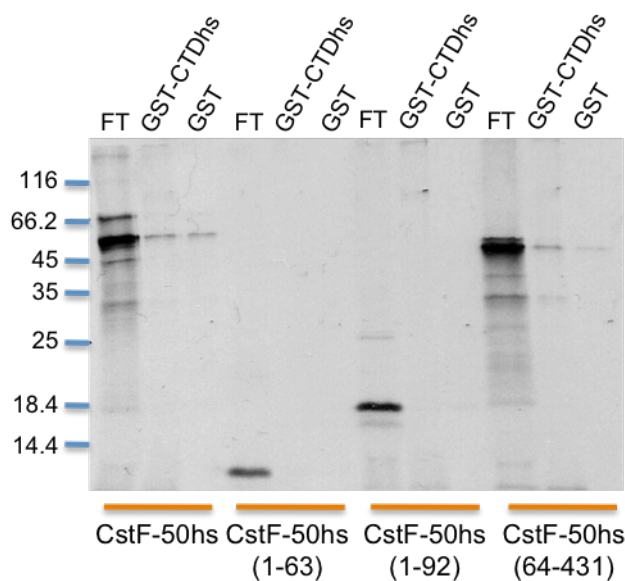
Les résultats obtenus à partir des expériences de « pull down » avec une résine GST (fig. 46A) ont montré l'existence d'une interaction pour certaines constructions de la protéine CstF-50 avec le CTD, mais ces résultats n'étaient pas reproductibles. Probablement les variations d'expression des protéines et/ou leur insolubilité ont influencé les différences des effets observés à partir de la même expérience. D'autres essais (fig. 46B) laissent imaginer l'existence d'une interaction CstF-50hs/GST ou d'une possible précipitation de CstF-50 avec la résine. La deuxième hypothèse a été confirmée par des résultats ultérieurs. Pour le « pull down » avec une résine Co<sup>2+</sup> (fig. 46C), les résultats, qui semblaient positifs pour l'interaction du CTD avec toutes les constructions de CstF-50, se sont révélés inexacts après la découverte de l'association de GST-CTDhs avec la résine et non avec CstF-50hs. Donc, les résultats obtenus ne nous permettent pas de conclure sur l'interaction étudiée.



**Figure 46.** A) « Pull down » avec une résine GST. La révélation a été effectuée avec un anticorps anti-His. B) « Pull down » avec une résine GST. Révélation avec un anticorps anti-His. C) « Pull down » avec une résine Co<sup>2+</sup>. La révélation a été réalisée avec un anticorps anti-GST.

*g.8) Etudes d'interaction de CstF-50hs, obtenue à partir de la technique de Transcription-Traduction 'in vitro' (TNT), et le CTD de l'ARN pol II par « pull-down »*

Dans les travaux effectués par Bentley et ses collaborateurs (Fong and Bentley 2001), la sous-unité CstF-50 de rat et des fragments de cette protéine ont été produits à partir d'expériences de traduction 'in vitro'. Ils ont incubé différentes constructions de CstF-50 avec un GST-CTD de souris immobilisé sur une résine glutathione sepharose. Nous avons essayé de reproduire ces expériences afin d'infirmer ou de confirmer l'implication de CstF-50hs (1-65) dans l'interaction avec le CTD. Pour cela j'ai réalisé la traduction 'in vitro' des différentes constructions de CstF-50hs grâce à l'utilisation d'un kit TNT (Quick coupled Transcription/Translation Systems, Promega). Ce kit commercial permet la traduction de protéines et leur marquage avec  $^{35}\text{S}$  par introduction de Méthionines marquées avec cet isotope (voir aussi Chapitre VI : Matériels et Méthodes). Après la vérification de la production des différents fragments protéiques par cette méthode, nous avons procédé à la réalisation des expériences de « pull-down ». Pour cela, les extraits contenant les protéines marquées sont incubés avec la résine GST préalablement liée à GST-CTDhs ou à GST. Après trois lavages avec un tampon à 500mM NaCl, la résine est chargée sur un gel SDS-PAGE qui est ensuite séché. L'exposition est effectuée sur la nuit avec un film spécial pour les essais de radioactivité (Kodak BioMax MR Film). J'ai réalisé l'expérience plusieurs fois, dans différentes conditions, cependant nous n'avons pas observé les mêmes résultats que ceux obtenus dans les travaux réalisés par D. Bentley. Sur la *figure 42*, on peut observer l'existence d'une interaction entre CstF-50 (1-95) et le CTD, même si elle est faible par rapport à celle obtenue pour la protéine entière. Contrairement à ces travaux, aucune interaction n'a été observée entre CstF-50hs (1-92) et le CTD. De la même manière, nous n'avons pas pu montrer l'existence d'une association entre CstF-50hs (1-63) et le CTD. Pour CstF-50hs (64-431) et la protéine entière, une interaction était visible mais de façon beaucoup moins évidente que dans leurs résultats.

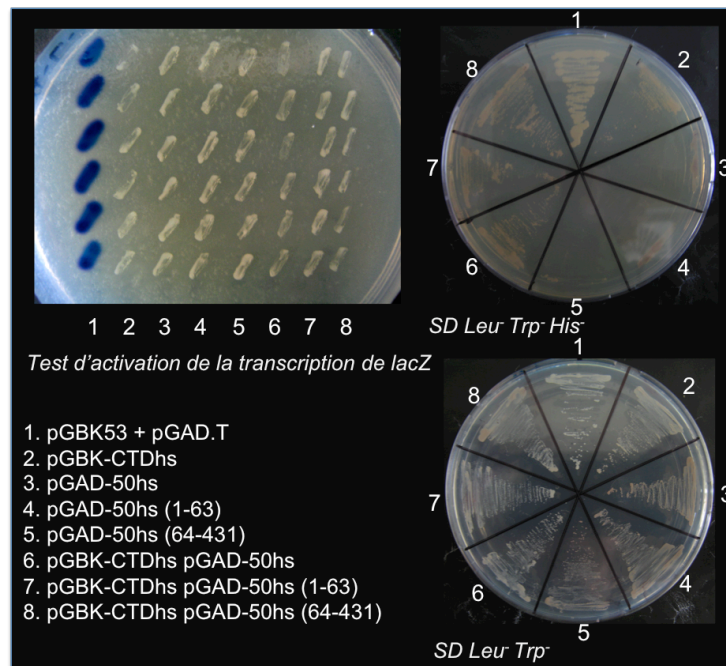


**Figure 47.** Expérience d'interaction GST-CTDhs et GST avec CstF-50hs par « pull down ». CstF-50hs, et des versions tronquées de celle-ci, ont été synthétisées par la technique du TNT. CstF-50hs et CstF-50hs (64-431) interagissent avec les deux, GST et GST-CTD. Le problème de cette non-spécificité a pu être contourné en augmentant la concentration saline du tampon de lavage (à 500mM de NaCl). Cependant, aucune interaction n'est visible pour les deux autres constructions de CstF-50 avec le CTD.

*g.9) Double hybride*

La technique de **double hybride** est une technique de biologie moléculaire permettant de détecter une interaction physique, directe ou indirecte, entre deux protéines. Nous pouvons supposer que l'interaction entre le domaine carboxy-terminal de la sous-unité Rbp1 de l'ARN pol II et CstF-50hs n'est pas très stable, la même région du CTD interagissant avec un grand nombre d'autres protéines. De plus, cette interaction peut avoir une vitesse de dissociation-association élevée. La technique de double hybride est appliquée à l'étude de ce type d'interactions. L'association Pcf11p/CTD a été mise en évidence par cette méthode (Sadowski et al. 2003), ainsi, nous avons opté pour l'utilisation de cette technique pour notre problématique.

Pour cette étude, j'ai cloné CstF-50hs et différents fragments de cette protéine dans un vecteur **pGAD-T7** qui contient le domaine activateur (« activating domain ou AD ») et le CTDhs dans un vecteur **pGBK-T7** qui contient le domaine de liaison (« binding domain », BD) (*plus détaillé dans la section de Matériels et Méthodes*). Les plasmides ont été transformés en deux souches différentes de levure, AH109 pour pGAD et Y187 pour pGBK. Nous avons étalé les levures transformées avec nos plasmides sur un milieu SD en absence de leucine (Leu<sup>-</sup>) pour la sélection de pGAD ou de tryptophane (Trp<sup>-</sup>) pour sélectionner les levures contenant le plasmide pGBK. Après croisement des deux souches, les cellules sont étalées, cette fois-ci dans un milieu Leu<sup>-</sup> Trp<sup>-</sup>. Après 48h d'incubation, le **test d'activation de la transcription du gène reporteur lacZ**, n'a donné aucun signe d'interaction entre CstF-50hs et le CTD. Nous avons réalisé également le test de transcription du **marqueur d'auxotrophie HIS3** mais nous n'avons pas obtenu de résultats permettant de conclure sur l'existence d'une interaction (*fig. 48*).



**Figure 48.** Résultats obtenus pour l'interaction CstF50hs/CTD à partir des expériences de double hybride. A gauche, les résultats du test d'activation de la transcription de lacZ et à droite de HIS1.

Les numéros correspondent à la liste de plasmides représentée sur la gauche dans tous les cas.

### 3. Conclusions et perspectives

Nous avons résolu la **structure du domaine N-terminal de la protéine CstF-50**. Des études préliminaires ont permis de sélectionner la région et l'espèce adéquate pour l'étude de l'architecture de cette région par cristallographie. Finalement, la construction CstF-50dm (1-65) a permis la formation de cristaux qui diffractaient jusqu'à 1,4Å de résolution. A partir des cristaux de la protéine « sélénée » et à l'aide de la méthode du SAD, nous avons déterminé la structure de ce domaine. Ces résultats ont été objet d'un article qui a été soumis récemment (Moreno-Morcillo et al., RNA, 2010).

La protéine a cristallisé sous sa forme **dimérique** dans le groupe d'espace C222<sub>1</sub>. Cette dimérisation a été également observée à partir de la protéine en solution grâce à l'utilisation de la technique de RMN. D'autre part, des études de comparaison des structures ont montré l'existence d'un motif dont l'architecture ressemble à celle de CstF-50dm (1-65). Il s'agit du motif **LISH** (LIS1 homology motif), connu pour son implication dans l'oligomérisation des protéines. L'homologie des structures de CstF-50dm (1-65) et du motif LISH a été corroborée par la superposition des deux modèles. L'homologie avec ce motif d'oligomérisation est en accord avec le modèle dimérique obtenu pour la CstF-50 au laboratoire.

Trois zones dans la région N-terminale de CstF-50 sont particulièrement **conservées**. La première est impliquée dans la formation du **dimère**. Nous suggérons que cette propriété est donc présente dans d'autres espèces et qu'elle a été maintenue à travers l'évolution. La deuxième et la troisième région conservées sont identiques et elles sont situées sur les côtés opposés de la surface. Ces régions exposées sont générées par la formation du dimère et constituent deux cavités qui pourraient être impliquées dans **l'interaction avec des partenaires**. Des travaux précédents ont décrit la région N-terminale de CstF-50 comme responsable de l'interaction avec le **CTD de l'ARN pol II**. Nos efforts ont été ainsi focalisés sur l'approfondissement de la description du rôle de la partie N-terminale de CstF-50 dans cette interaction.

Des expériences de co-cristallisation, de RMN, de SPR, de double hybride et des diverses variantes de la technique du « pull down » **n'ont pas apporté des évidences claires de l'interaction entre le CTD et CstF-50**. Pour certaines expériences, le choix des séquences des CTD synthétiques, le choix de l'espèce de laquelle proviennent les protéines et l'instabilité ou la sous-expression des protéines mises à l'étude ont pu être responsables de l'absence de résultats clairs. D'autres expériences basées sur l'utilisation des billes magnétiques de Co<sup>2+</sup> sont en cours au laboratoire. Des nouveaux essais en RMN ou SPR à partir de la protéine humaine pourraient aussi être envisagés, en espérant trouver une différence avec les études précédentes.

D'autre part, les travaux de D. Bentley et ses collaborateurs ont montré qu'une interaction existe entre le CTD et une région de la protéine CstF-50 commençant sur le résidu 77 (CstF-50 (77-176)). Cependant, la délétion des acides aminés du 77 au 89 empêche cette association (Fong and Bentley

2001). Ceci suggère que la zone sur laquelle nous nous sommes focalisés peut être impliquée dans l'interaction mais elle n'est pas indispensable. Cette étude suggère également que la zone située entre la région N-terminale et le WD-40 serait nécessaire pour l'interaction. Par des expériences de RMN nous avons observé que cette région de la protéine est désordonnée. Probablement, dans le contexte de la protéine entière, ces résultats diffèrent. Nonobstant, il est possible que ce « linker » soit flexible pour faciliter la mobilité des deux régions de la protéine et permettre au même temps l'interaction avec le CTD. Ce « linker » en étant indispensable à l'interaction, la construction des résidus 1 à 65 serait alors trop courte pour révéler l'interaction.

Si la cavité conservée formée par le dimère n'est pas impliquée dans l'interaction avec le CTD, il serait intéressant de pouvoir découvrir la fonction qui lui est attribuée. La recherche d'autres partenaires différents au CTD pourrait être abordée depuis différentes perspectives. Par exemple, des études en protéomique par double hybride ou par la méthode TAP (Tandem Affinity Purification) pourrait être envisagées.

Finalement, l'étude du domaine WD-40 de CstF-50 peut être envisagé. Cette région de la protéine n'est pas soluble mais ce problème pourrait être contourné en utilisant une de ces stratégies :

- l'addition d'une étiquette, différente à l'étiquette H<sub>6</sub>, sur l'extrémité N-terminale ou C-terminal du domaine,
- l'optimisation du tampon utilisé au moment de la lyse et de la purification,
- la formation d'un complexe entre ce domaine et un de ses partenaires. L'interaction de ce domaine avec des protéines comme **CstF-77** ou **BARD1** pourrait également être envisagée. La résolution des structures des complexes, accompagnée des études d'affinité, peut participer à la compréhension de la connexion du processus de maturation avec le système de réparation de l'ADN. En effet, CstF-50 utilise la même région, son domaine WD-40, pour l'association avec ces deux protéines. Il est possible qu'une compétition entre ces deux partenaires s'établisse au moment de cet endommagement.

## B. Interactions dans le complexe CstF

Le complexe CstF est composé de trois sous-unités appelées CstF-77, CstF-64 et CstF-50 en fonction de leur poids moléculaire. CstF-77 a un rôle structural essentiel puisqu'elle interagit avec les deux autres protéines et maintient le complexe formé. Le domaine HAT de CstF-77 est suivi d'une **région dite riche en Pro**. Des études réalisées par « Far Western » (Takagaki and Manley 2000) suggèrent que cette zone de la protéine serait impliquée dans l'interaction avec le domaine « **hinge** » de CstF-64 et avec le domaine **WD-40 de CstF-50**. L'interaction entre CstF-77 et CstF-64 a été également démontrée pour leurs homologues chez la levure, Rna14p et Rna15p (Noble et al. 2004) (Legrand et al. 2007).

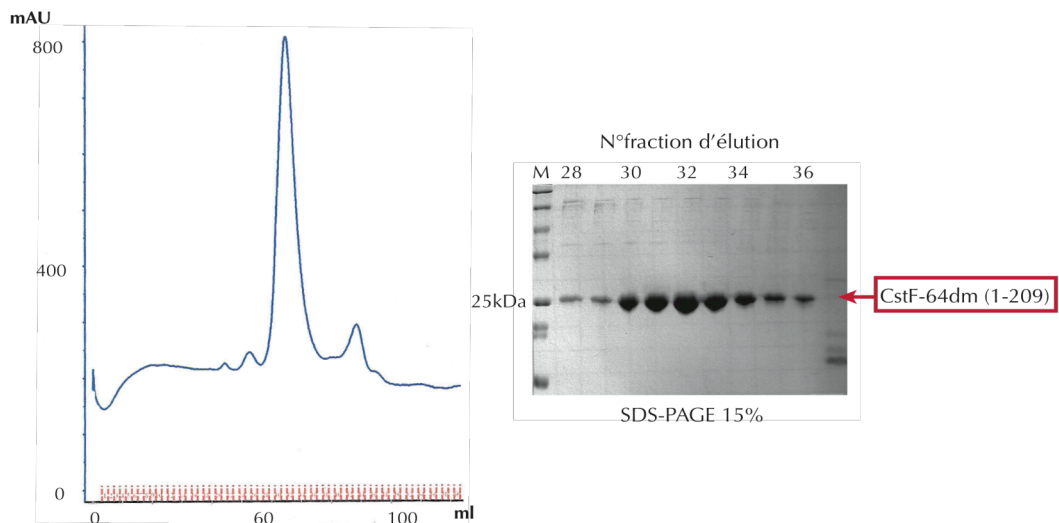
L'objectif de nos travaux est de résoudre la structure du complexe CstF, à partir des protéines ou des domaines, isolés ou en interaction avec leurs partenaires. Certaines de ces structures sont déjà connues. En effet, la structure de **CstF-77**, pour la protéine de la superfamille des muridés (Bai et al. 2007) et pour la protéine provenant de l'espèce *Encephalitozoon cuniculi* (Legrand et al. 2007), a été résolue. L'architecture moléculaire du domaine **RRM (RNA Recognition Motif)** et de la région **C-terminale de CstF-64** est également connue (Perez Canadillas and Varani 2003; Qu et al. 2007). Enfin, dans la dernière section de ce chapitre, une description de l'architecture de la région **N-terminale de CstF-50** a été présentée. Néanmoins, les détails sur l'interaction entre les trois sous-unités ainsi que la structure de certains domaines indispensables pour l'accomplissement de la fonction du complexe n'ont pas été dévoilés.

### *1. Etude des domaines impliqués dans l'interaction entre CstF-77 et CstF-64*

#### **a.) Travaux préliminaires : étude du domaine « hinge » de CstF-64**

Des études réalisées précédemment au laboratoire à partir de la **protéine CstF-64 entière** (humaine et de drosophile) n'ont pas pu aboutir à la résolution de la structure. A mon arrivée au sein du laboratoire, j'ai débuté mes travaux à partir de la protéine **CstF-64 de *E. cuniculi*** (CstF-64ec). La purification de cette sous-unité s'est révélée difficile. Après plusieurs essais, il a été constaté qu'une diminution de la concentration de NaCl, de 150mM à 25mM, était nécessaire pour éviter la précipitation de la protéine. Finalement, malgré l'obtention d'une protéine pure à une concentration de 4 mg/ml, les tests de cristallisation de la protéine CstF-64ec n'ont pas abouti à la formation de cristaux exploitables.

Nous nous sommes focalisés sur l'étude du **domaine « hinge » de CstF-64**. La région N-terminale (du résidu 1 au 209) contenant le RRM et le « hinge » de la protéine CstF-64 de *D. melanogaster*, a été clonée dans un vecteur qui porte une étiquette de 6 His (H<sub>6</sub>) dans sa séquence. La protéine recombinante a été exprimée chez *E. coli* (BL21 (DE3)) et ensuite purifiée en trois étapes, une chromatographie d'affinité, suivi d'une échangeuse d'ions et finalement une chromatographie d'exclusion stérique (fig. 49). L'étiquette H<sub>6</sub> a été clivée après la deuxième étape de purification à l'aide de la *TEV*. Des essais de cristallisation ont été réalisés à partir de la protéine concentrée à **12-16 mg/ml**. Cependant, malgré l'utilisation d'un grand nombre de kits commerciaux de cristallisation, aucune condition n'a permis la formation des cristaux de CstF-64dm (1-209).

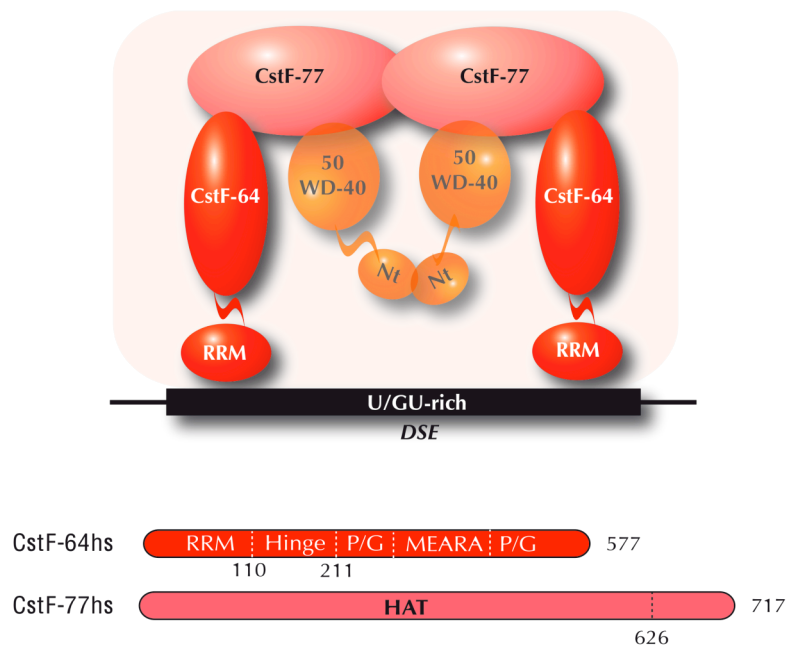


**Figure 49.** Purification de **CstF-64dm (1-209)**. A gauche, le profil d'élution obtenu après une colonne de filtration su gel. A droite, un gel dénaturant qui montre le niveau de pureté de la protéine.

A partir de ces résultats, la stratégie de travail a consisté en la reconstitution de **complexes**. Ces complexes incluent essentiellement le domaine « hinge » de CstF-64 auquel d'autres régions, de la protéine CstF-64 elle-même ou de CstF-77, sont ajoutées. La cartographie des interactions des sous-unités de CstF et la purification des complexes, en plus de participer à la compréhension de l'assemblage de CstF, peuvent contribuer à l'augmentation des probabilités de cristallisation.

Les résultats présentés ci-dessous décrivent l'interaction entre CstF-77 et CstF-64 (fig. 50). D'autres expériences concernant la liaison de CstF-50 avec CstF-77 ont été réalisées mais nous n'avons pas pu apporter de nouvelles informations pour cette association.





**Figure 50.** Schéma montrant les protéines du complexe CstF et leurs interactions.  
 Cette section des résultats porte sur l'étude des protéines CstF-64 et CstF-77.

### **b.) Identification des régions impliquées dans l'interaction entre CstF-77 et CstF-64**

Parallèlement aux travaux réalisés pour CstF, nous avons obtenu des informations sur l'interaction entre les homologues de CstF-77 et CstF-64 chez la levure, **Rna14p** et **Rna15p** respectivement. Ces résultats ont été décisifs pour l'étude de l'association des sous-unités humaines **CstF-77hs** et **CstF-64hs** et seront présentés ultérieurement (*Chapitre III*). A partir des expériences de « pull-down » avec différentes régions de Rna14p et de Rna15p, nous avons pu déterminer les zones minimales de chaque protéine participant à l'interaction. Ces expériences ont montré que la zone C-terminale de **Rna14p (626-677)** et la région correspondant au « hinge » de **Rna15p (127-232)** sont suffisantes pour l'association des deux sous-unités. Les expériences réalisées en « far western » (Takagaki and Manley 2000) et les résultats obtenus pour les protéines de levures, suggèrent que la région « hinge » de CstF-64 est impliquée dans l'interaction avec CstF-77. Cette région est en effet fortement conservée entre les mammifères et les levures. Nonobstant, la conservation de la zone C-terminale de CstF-77 est faible et la zone d'interaction est moins bien définie. Notre premier objectif a été focalisé sur la détermination des résidus intervenant dans l'interaction CstF-64/CstF-77.

Pour la formation des complexes hétérodimériques nous utilisons au laboratoire deux techniques différentes. La première est basée sur le mélange soit de protéines purifiées, soit des extraits totaux précédant l'étape de rupture cellulaire. L'autre méthode est la **co-expression** des protéines. Pour ceci, nous pouvons nous servir de deux plasmides différents contenant les séquences d'ADN qui

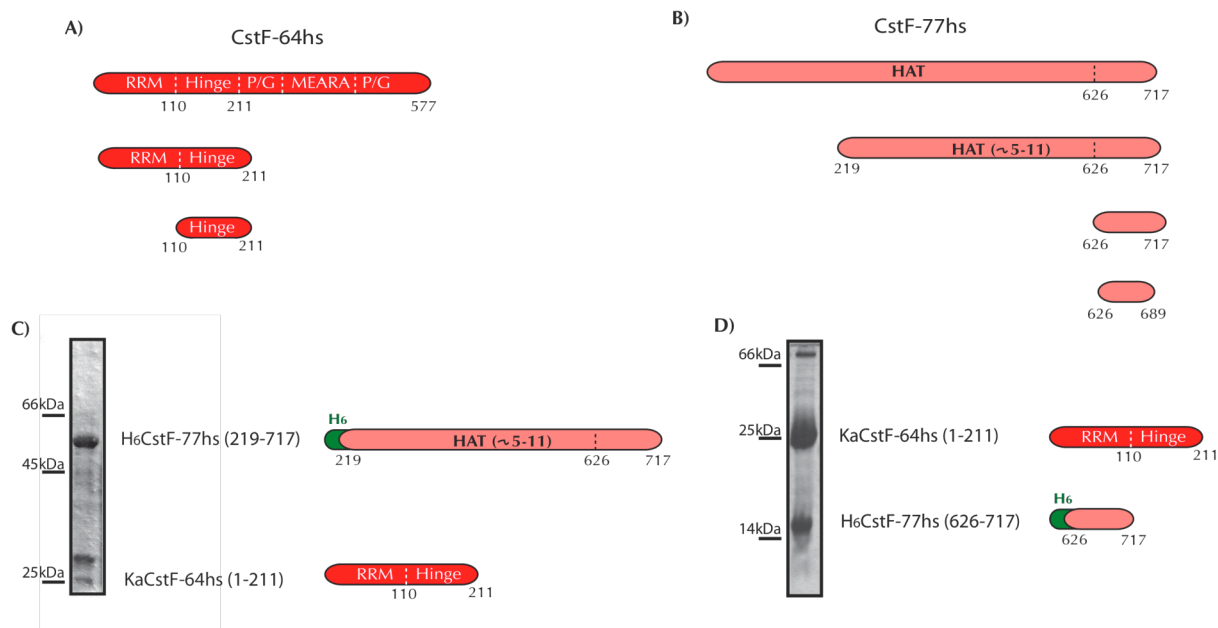
codent pour chacune des protéines. L'autre stratégie de la co-expression consiste en l'utilisation d'un seul plasmide où nous intégrons en tandem, et sous contrôle du même promoteur, les deux séquences nécessaires pour l'expression des deux protéines d'intérêt (*voir dans Matériels et Méthodes la section correspondant aux vecteurs bicistroniques*). La co-expression présente des avantages par rapport à l'autre méthode sur différents points:

- la solubilité : le repliement des protéines intégrées dans un complexe peut augmenter leur solubilité et leur stabilité par rapport à la protéine isolée.
- la stœchiométrie du complexe : les deux protéines s'assemblent sous leur forme naturelle dans la cellule et le problème de stœchiométrie peut être ainsi contourné.
- la facilité d'exécution et le temps requis : la co-transformation de deux plasmides, ou la transformation d'un plasmide bicistronique, n'implique aucune manipulation supplémentaire par rapport à une transformation de routine. De plus, le complexe est formé chez *E. coli* et nous pouvons le traiter de la même façon qu'une protéine seule. L'autre méthode peut nécessiter que les deux protéines soient purifiées séparément ce qui implique un investissement de temps plus important.

De cette façon, au laboratoire, nous essayons d'atteindre la formation d'un complexe par co-expression et en cas d'échec la méthode de « co-purification » est employée.

Pour les protéines CstF-64 et CstF-77, des nouvelles constructions ont été créées pour la réalisation des tests de co-expression. Le but est de préciser les régions des deux sous-unités impliquées dans la formation du complexe. Pour CstF-64hs nous avons cloné dans différents vecteurs la région qui contient le motif RRM et le « hinge », **CstF-64hs (1-211)**, et une autre région où le RRM est absent, **CstF-64hs (110-211)**. Un des plasmides créés pour CstF-77hs permet l'expression de la zone C-terminale de la protéine, de l'acide aminé 626 jusqu'au dernier résidu, **CstF-77hs (626-717)**. Dans un autre plasmide de la sous-unité CstF-77hs les derniers résidus moins conservés ont été éliminés, **CstF-77hs (626-689)**. Finalement, nous avons construit également un plasmide qui contient certaines répétitions du motif HAT, **CstF-77hs (219-717)** (*fig. 51*).

A partir de ces plasmides, j'ai effectué des tests pour étudier la surexpression des protéines seules ou en complexe. Les constructions CstF-64hs (110-211) et CstF-77hs (626-689) n'ont pas permis la surexpression des protéines correspondantes seules ou en interaction avec leur partenaire. La co-transformation de **H<sub>6</sub>CstF-77hs (219-717)** ou **H<sub>6</sub>CstF-77hs (626-717)** avec **KaCstF-64hs (1-211)** dans différentes souches d'*E. coli* a permis la surexpression des complexes (*fig. 51*). Les étapes de purification et de cristallisation ont été poursuivies pour le complexe **H<sub>6</sub>CstF-77hs (626-717)/ KaCstF-64hs (1-211)**.



**Figure 51.** Tests de co-expression des domaines d'interaction de CstF-64 et CstF-77. A) et B) Constructions protéiques créées pour CstF-64 et pour CstF-77, respectivement. C) Résultats obtenus pour les constructions  $H_6$ CstF-77hs (219-717) et KaCstF-64hs (1-211). D) Test de co-expression des constructions  $H_6$ CstF-77(626-717) et KaCstF-64hs (1-211) sur un gel SDS-PAGE 18%. Dans les deux cas, 10 $\mu$ l de la résine  $Co^{2+}$  contenant les protéines d'intérêt sont directement chargés sur le gel.

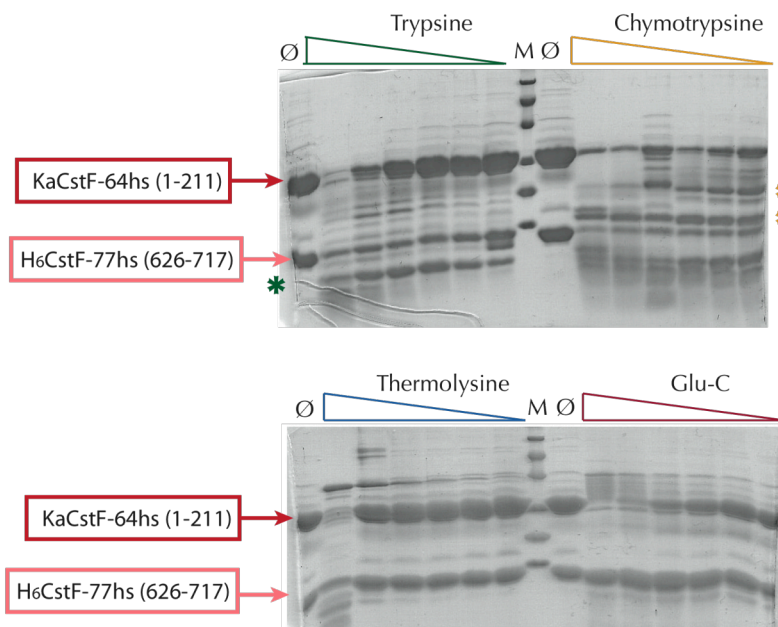
### c.) Analyse du complexe CstF-77hs (626-717) / CstF-64hs (1-211)

Le complexe formé entre les deux sous-unités a été exprimé dans la souche bactérienne **BL21 (DE3)** et ensuite purifié en trois étapes. La réalisation d'une étape de chromatographie d'affinité sur une résine de cobalt a été possible grâce à l'étiquette de 6 His située sur la région N-terminale de la construction de CstF-77hs. Cette étape a été suivie d'une chromatographie d'échange cationique et d'une colonne d'exclusion stérique. Le « tag »  $H_6$  a été éliminé avant la dernière étape de purification par digestion avec la protéase *TEV*.

Le rendement de la purification était convenable pour les essais de cristallisation étant donné qu'à partir d'un litre de cultures nous avons obtenu jusqu'à **8 mg** de protéine. Des gouttes de cristallisation ont été réalisées à partir du complexe concentré de **15 à 40 mg/ml**. Malgré l'utilisation de différents kits commerciaux, nous n'avons pas trouvé une condition permettant la **cristallisation** du complexe. En effet, la présence de précipité dans les gouttes pour les conditions essayées est très rare malgré la concentration importante des protéines. Ceci indiquerait que le complexe est extrêmement soluble et trouver des conditions induisant sa nucléation peut s'avérer compliqué.

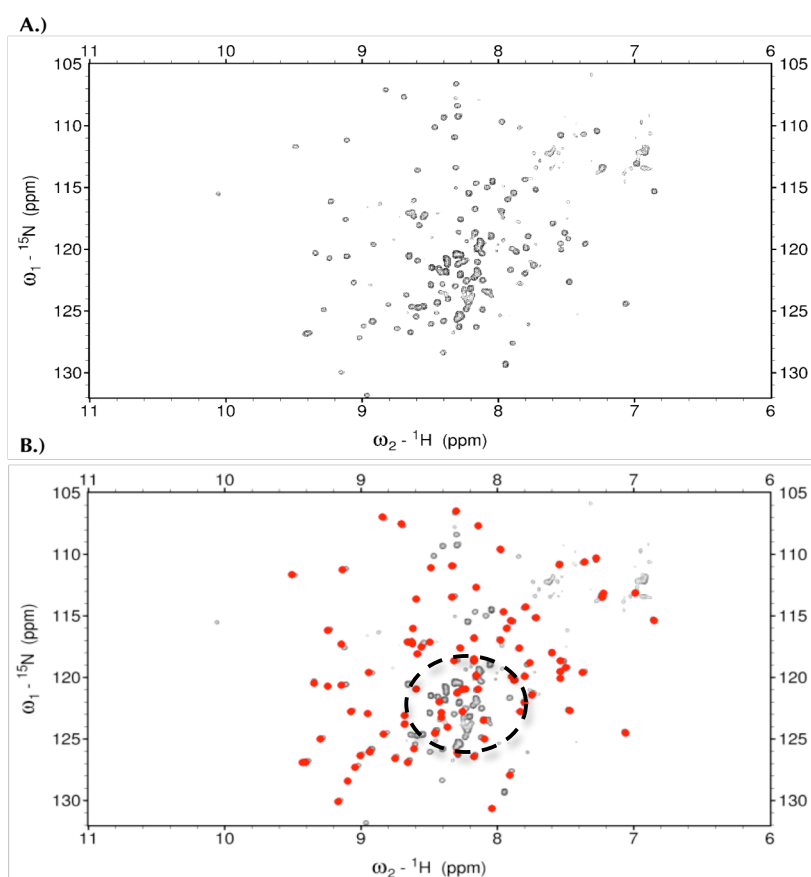
Afin de comprendre l'organisation et la stabilité du complexe protéique et de trouver les conditions qui pourraient permettre la cristallisation, nous avons effectué des expériences de **protéolyse ménagée** et de **RMN**. Pour les premières, nous disposons au laboratoire de 4 protéases différentes, **Trypsine**, **Chymotrypsine**, **Thermolysine** et **Glu-C**. A une concentration adéquate d'enzyme et de protéine, ce type d'expériences peut donner un indice des domaines repliés présents dans une protéine et des régions impliquées dans une interaction. La digestion enzymatique aura lieu dans les endroits plus accessibles, correspondant souvent aux régions inter-domaine, et les zones protégées par une interaction ou par le propre repliement de la protéine seront plus difficilement digérées.

Un gradient de concentration de protéase, réalisé par des dilutions en série (de 333µg/ml à 13µg/ml), a été ajouté à une quantité fixe du complexe H<sub>6</sub>CstF-77hs (626-717)/ KaCstF-64hs (1-211) (10µl à 0,8mg/ml) (fig. 52). Seulement la digestion par la trypsine montre une région de CstF-77hs (626-717) facilement digérable qui n'a pas été identifiée. La digestion chymotrypsique a créé deux bandes pour CstF-64hs (1-211) probablement dues à la présence des deux domaines, le motif RRM et l'« hinge ». Cependant, la construction contenant seulement la région « hinge » de CstF-64hs avait déjà été clonée (CstF-64hs (110-211)) mais l'expression de la protéine seule ou en complexe avec CstF-77hs est très faible. Ces expériences ont montré que le complexe était relativement stable avec peu de zones flexibles. Nous avons donc poursuivi les essais de cristallisation avec ces constructions.



**Figure 52.** Analyse des digestions par protéases sur des gels SDS-PAGE colorés au Bleu de Coomassie. Les triangles montrent la quantité décroissante d'enzyme. Les bandes correspondant aux deux protéines d'étude sont indiquées. Les astérisques vert et jaunes indiquent les bandes obtenues à partir des digestions par trypsine et chymotrypsine, respectivement.

L'expérience de **RMN** (**R**ésonance **M**agnétique **N**ucléaire) a requis le marquage des protéines à l'azote 15 et au carbone 13 (*voir Chapitre VI*). Une des approches de RMN pour la résolution de structures est l'expérience **TROSY** (**T**ransverse **R**elaxation-**O**ptimized **S**pectroscopy). Cette technique permet l'étude de macromolécules biologiques de taille plus importante. 500 $\mu$ l du complexe H<sub>6</sub>CstF-77hs (626-717)/ KaCstF-64hs (1-211), à une concentration de 3mg/ml, ont été analysés sur un spectromètre à 700MHz disponible à l'IECB (expériences réalisées en collaboration avec le **Dr. Mackereth**). Le tampon final était composé de 20mM NaHPO<sub>4</sub> à pH 7.5 et 50mM de NaCl. Les expériences ont été réalisées à une température de 298K. Les résultats préliminaires montrent que, dans l'ensemble, le complexe est replié. Le spectre obtenu à partir de cette expérience a été comparé à ceux acquis pour le RRM de la protéine CstF-64 isolée (Perez Canadillas and Varani 2003) (*fig. 53*). Nous avons remarqué que la plupart de pics obtenus pour le complexe coïncident avec ceux obtenus pour le domaine RRM seul. Parmi les pics restants, certains (ceux entourés d'un cercle en pointillé sur la *figure 53B*) indiquent un comportement non ordonné des résidus impliqués. Ceci signifie que probablement une partie du domaine « hinge » ou de la région C-terminale de CstF-77 est moins bien repliée. Si ces résidus correspondent à CstF-77, ces résultats seraient cohérents avec ceux obtenus par protéolyse dirigée. D'autres expériences devraient confirmer ces résultats et identifier les régions qui se trouvent mal repliées.



**Figure 53.** A.) Spectre de RMN  ${}^1\text{H}, {}^{15}\text{N}$ -TROSY obtenu pour le complexe CstF-64hs (1-211)/CstF77hs (626-717). B.) Superposition du spectre A.) avec celui correspondant au motif RRM de CstF-64 en rouge (code : 5774 de BMRB- **B**iological **M**agnetic **R**esonance **D**ata **B**ank).

**d.) Le complexe CstF-77hs (626-717) / CstF-64hs (1-211) interagit avec une séquence d'ARN riche en G/U**

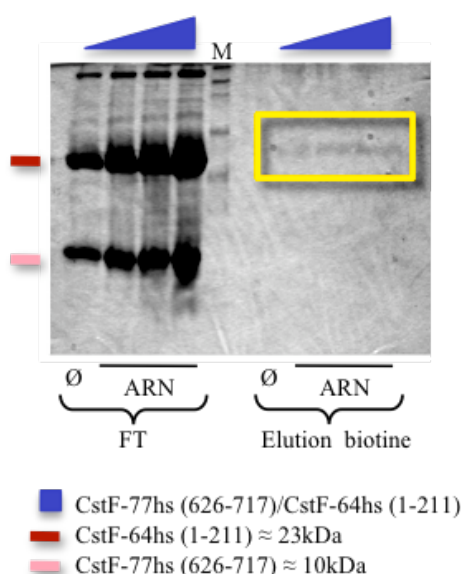
Il a été montré que CstF-64hs reconnaît une région conservée de l'ARNm riche en G/U. L'interaction avec l'ARN a lieu à travers un motif **RRM** situé dans la région N-terminale de la protéine qui comprend les résidus du 1 au 110 (Takagaki and Manley 1997). Le modèle structural de cette région de CstF-64 en complexe avec l'ARN a été obtenu à partir des expériences de RMN (Perez Canadillas and Varani 2003). Le domaine présente le repliement typique du motif de reconnaissance de l'ARN (RRM) plus deux hélices, une située à l'extrémité N-terminale et l'autre en C-terminale. Cette dernière hélice pourrait avoir un rôle régulateur dans l'interaction avec l'ARN. La liaison avec l'ARN s'effectue à travers les feuillets bêta du motif RRM.

Des études d'affinité, basées sur des observations en RMN, entre la sous-unité CstF-64 et différentes séquences d'ARN ont montré l'existence d'une interaction plus importante quand deux uraciles en tandem, **UU**, sont présents dans la séquence d'ARN (Perez Canadillas and Varani 2003). A partir de ces expériences, il a été observé que des séquences contenant exclusivement des répétitions GU avaient une affinité de l'ordre de  $\mu\text{M}$ . Contrairement, le RRM de CstF-64 possède une affinité de  $\sim 100\text{nM}$  pour les séquences incluant deux U en tandem (Takagaki and Manley 1997) (Perez Canadillas and Varani 2003).

Les détails sur l'interaction entre CstF-64 et l'ARN sont connus mais, dans les expériences réalisées précédemment seulement le RRM de la sous-unité CstF-64 était présente (Perez Canadillas and Varani 2003) (Deka et al. 2005). L'addition de la région de CstF-77hs interagissant avec la protéine CstF-64hs pourrait modifier la spécificité de l'association. De cette façon, notre premier objectif a été de démontrer l'existence d'une interaction entre un ARN riche en G/U (avec UU dans sa séquence) et le complexe CstF-77hs (626-717) / CstF-64hs (1-211).

Nous avons synthétisé un oligonucléotide riche en G/U imitant les caractéristiques des séquences utilisées dans des travaux précédents de SELEX (Systematic Evolution of Ligands by EXponential enrichment) (Takagaki and Manley 1997) ou de RMN (Perez Canadillas and Varani 2003) (Deka et al. 2005). Ainsi, pour nos études,  $0,5\mu\text{g}$  d'un ARN dont la séquence est **5' (GU)<sub>5</sub>(UG)<sub>5</sub> 3' -Biotine** a été mis en contact avec différentes quantités (de 0 à  $0,2\text{mg}$ ) du complexe CstF-77hs (626-717) / CstF-64hs (1-211) partiellement purifié, pendant 16h à  $4^\circ\text{C}$ . L'ARN biotinylé a d'abord été déprotégé et purifié par chromatographie d'échange ionique (HiTrap<sup>TM</sup> Q) (voir *Chapitre VI Matériels et Méthodes*). Des expériences de « pull-down » avec des billes de **streptavidine** commerciales (Dynabeads MyOne Streptavidin T1, *Invitrogen*) ont été réalisées. Pour ceci, le mélange ARN/protéine a été incubé avec des billes de streptavidine 2h à  $4^\circ\text{C}$ . Ensuite, trois lavages ont été effectués avec un tampon composé de 20mM HEPES pH 7.5, 75mM NaCl, 2mM  $\text{Mg}(\text{CH}_3\text{COO})_2$ , 0,1% NP-40, 10% Glycerol, 1mM DTT. L'élution a été réalisée à l'aide d'un tampon contenant du détergent SDS qui permet la séparation de la streptavidine et la biotine. Ces expériences ont mis en évidence l'association entre l'ARN et le complexe protéique (*fig. 54*).

L'absence d'interaction avec un ARN poly(A) contrôle a démontré la spécificité de cette interaction (résultats non présentés sur la figure).



**Figure 54.** Interaction entre l'ARN et le complexe CstF77hs (626-717) / CstF-64hs (1-211). Ø représente le contrôle en absence d'ARN. A gauche du marqueur 10 $\mu$ l du surnageant obtenu au premier lavage ont été chargés. A droite du marqueur, résultats de l'éluion après trois lavages. Seulement une faible quantité de CstF-64hs est observé mais l'existence d'une interaction est présumée.

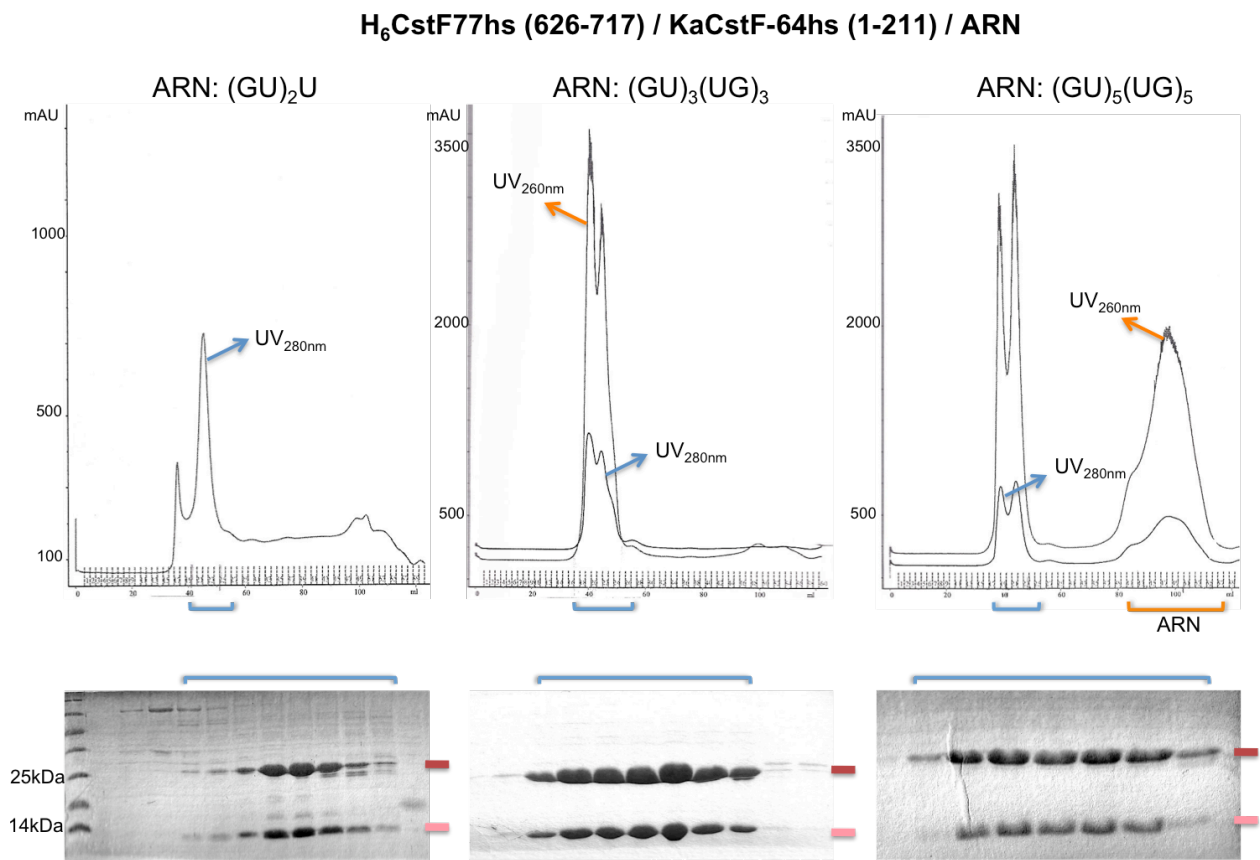
**e.) Essais de cristallisation du complexe CstF-77hs (626-717) / CstF-64hs (1-211) / ARN**

L'objectif est de résoudre la structure du complexe CstF-77hs (626-717) / CstF-64hs (1-211) avec ou sans l'ARN. Des essais de cristallisation réalisés à partir du complexe protéique seul ont montré la difficulté à obtenir des cristaux exploitables. L'addition de l'oligonucléotide pourrait stimuler la cristallisation du complexe.

A partir de la littérature (Auweter et al. 2006) (Maris et al. 2005) (Lunde et al. 2007), il est connu qu'un domaine RRM reconnaît approximativement 4-5 nucléotides d'une séquence d'acide nucléique (2 nucléotides au minimum et 8 au maximum). Des ARNs de différentes tailles susceptibles d'interagir avec l'hétérodimère CstF-77hs (626-717) / CstF-64hs (1-211) ont été synthétisés :

- **GU-20** : 5' GUGUGUGUGUUGUGUGUGUG 3'
- **GU-12** : 5' GUGUGUUGUGUG 3'
- **GU-5** : 5' GUGUU 3'

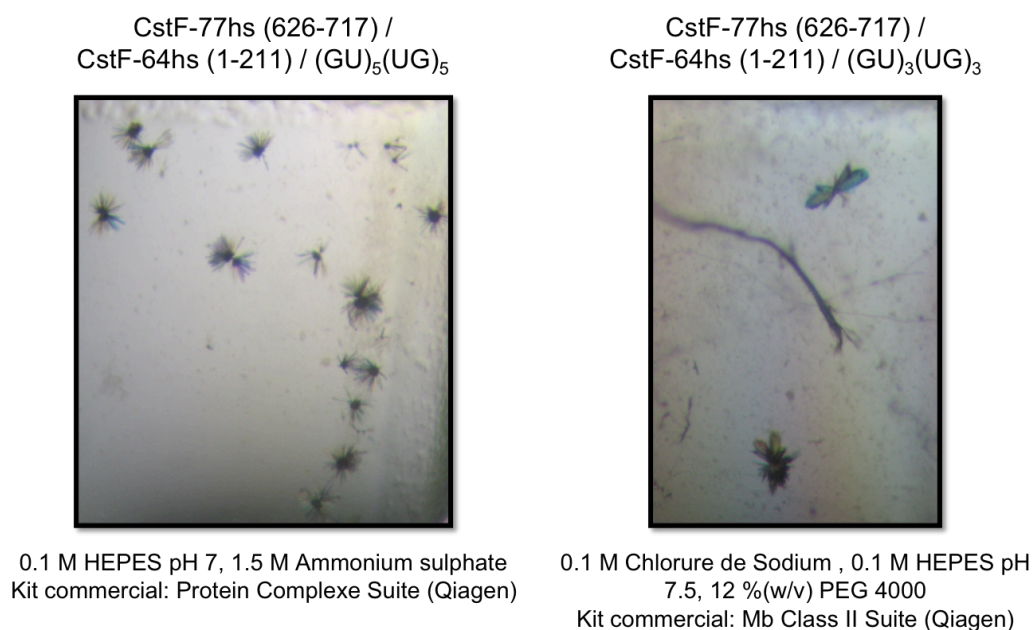
Dans le protocole général, l'ARN purifié (voir Chapitre VI : Matériels et Méthodes) est ajouté au complexe protéique pur dans un rapport équimolaire ou en léger excès (1,2 mol d'ARN pour 1 mol de complexe protéique). Après 2-4 heures d'incubation à 4°C, ce mélange est à nouveau purifié sur une colonne de filtration sur gel. De cette façon, un éventuel pic correspondant à l'ARN seul en excès peut être observé sur le profil d'élution. Seules les fractions correspondant à l'association **CstF-77hs (626-717) / CstF-64hs (1-211) / ARN** ont été réunies et concentrées pour la réalisation des tests de cristallisation (fig. 55).



**Figure 55.** Dernière étape de purification pour le complexe H<sub>6</sub> CstF-77hs (626-717) / Ka CstF-64hs (1-211) et les différentes séquences d'ARN. Les profils obtenus pour chaque type de complexe à partir d'une chromatographie d'exclusion stérique sont représentés sur la partie supérieure de la figure et les gels (18% SDS-PAGE) correspondant à ceux-ci sont situés en dessous. Le volume d'élution se trouve à environ 40ml sur une colonne HiLoad 16/60 Superdex<sup>TM</sup> 75 pg. La bande de couleur rouge indique la protéine CstF-64hs (1-211) sur le gel et celle en couleur saumon montre le niveau de migration de CstF-77hs (626-717).

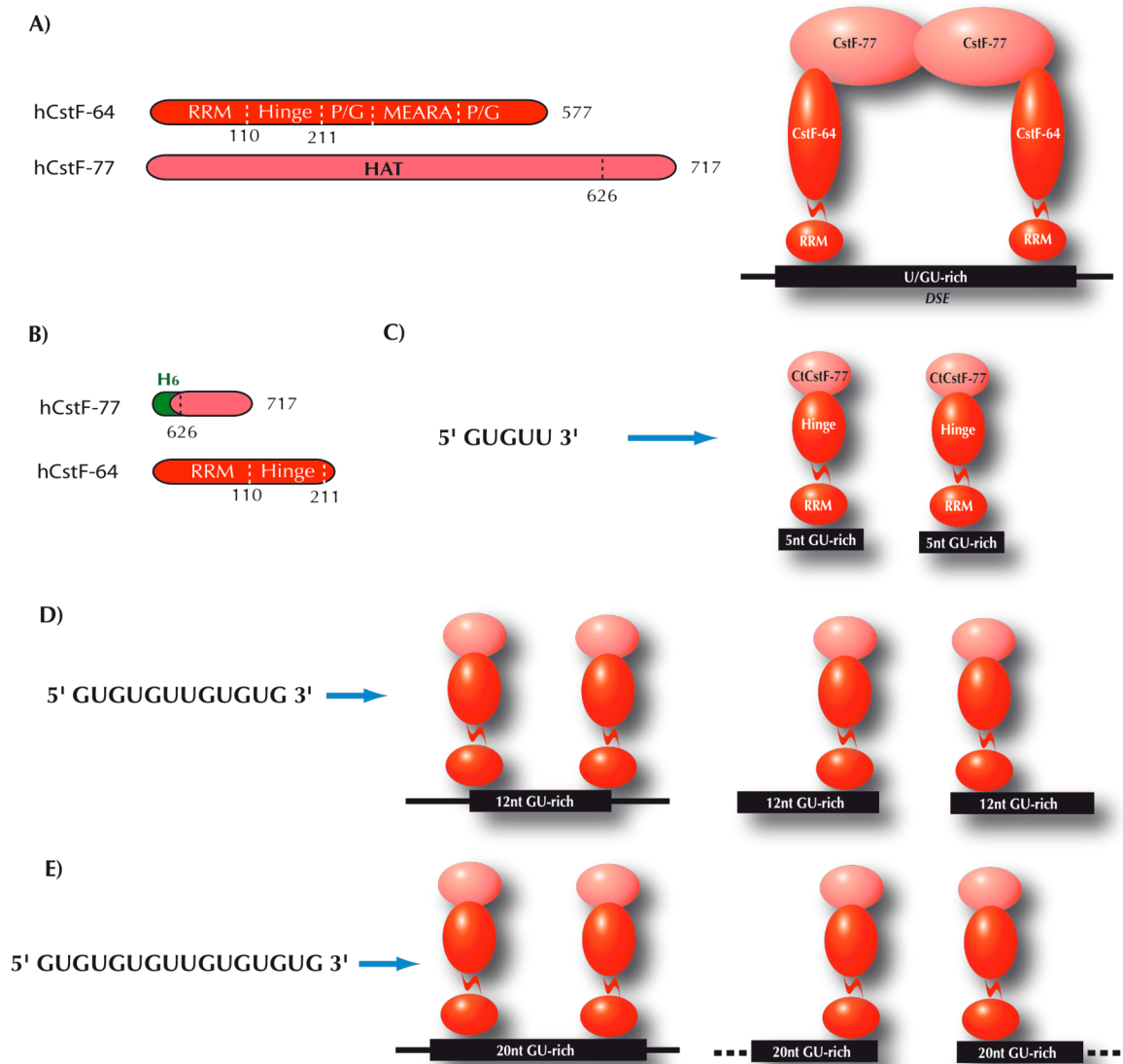


Nous avons utilisé une large gamme de conditions pour la réalisation des essais de cristallisation à partir du complexe concentré à **10-25 mg/ml**. Des cristaux en forme « d'oursin », comme ceux montrés sur la *figure 56*, ont été obtenus dans des conditions de cristallisation différentes mais les tentatives d'optimisation pour modifier leur forme, leur taille ou leur diffraction n'ont pas donné des résultats favorables.



**Figure 56.** Exemples des cristaux obtenus à partir du complexe CstF-77hs (626-717) / CstF-64hs (1-211) / ARN.

La difficulté à cristalliser le complexe CstF-77hs (626-717) / CstF-64hs (1-211) / ARN pourrait dériver d'un problème au niveau de **l'homogénéité de l'échantillon**. A partir des études précédentes réalisées sur l'interaction entre l'ARN et CstF-64 et selon les données connues sur le motif RRM, nous pouvons présumer que l'ARN de 5 nucléotides interagit avec un seul domaine RRM de la sous-unité CstF-64. Cependant, pour les autres oligonucléotides synthétisés, il est probable qu'une seule molécule d'ARN soit reconnue par deux de ces domaines. Ceci peut être important au niveau de la **stœchiométrie** du complexe. En effet, CstF-77 forme un dimère dans des conditions natives. Cependant, des travaux réalisés pour son homologue chez la levure dans notre laboratoire (*voir Chapitre III*) et des expériences de « Far Western » effectuées dans d'autres équipes de recherche (Takagaki and Manley 2000), suggèrent que la région de CstF-77 employée, du résidu 626 au 717, n'est pas suffisante pour la formation d'un dimère. Ainsi, différentes possibilités existent pour l'interaction du complexe protéique avec les différents ARNs synthétiques (*fig. 57*). Les épaulements observés au niveau des pics d'éluion obtenus à la dernière étape de purification pour les complexes formés avec (GU)<sub>3</sub>(UG)<sub>3</sub> et avec (GU)<sub>5</sub>(UG)<sub>5</sub>, supporteraient cette théorie (*fig. 57*).



**Figure 57.** *A) Schéma représentatif des domaines des protéines CstF-77 et CstF-64 et modèle de l'interaction des deux sous-unités entières avec l'ARNm. B) Régions de CstF-64hs et de CstF-77hs employées pour nos études. C) Modèle proposé pour l'interaction avec un ARN de 5 nucléotides. D) et E) Deux possibilités pour l'interaction entre le complexe protéique et un ARN de 12 et 20 nucléotides, respectivement.*

## 2. Conclusions et perspectives

A partir de nos études, les régions minimales participant à l'interaction des sous-unités CstF-77hs et CstF-64hs ont été précisées. Nous avons mis en évidence que pour **CstF-77hs** la zone nécessaire pour l'association correspond à moins de 100 résidus situés dans sa région C-terminale. La structure générale de CstF-77 a été résolue au laboratoire (Legrand et al. 2007) et dans un autre groupe de recherche (Bai et al. 2007). Cependant, dans aucun des cas, cette région C-terminale n'a pu être déterminée. L'absence d'une densité électronique pour la région C-terminale dans les données de diffraction obtenus pour la protéine CstF-77 de *E. cuniculi*, a été sûrement provoquée par une mobilité importante de cette zone de la protéine.

La région « hinge » de **CstF-64hs**, responsable de l'association de la protéine au complexe CstF (Hatton et al. 2000; Takagaki and Manley 2000) (Hockert et al. 2010), n'a pas pu être surexprimée chez la bactérie. Nous avons inclus le motif RRM pour effectuer l'interaction avec CstF-77. Le complexe **CstF-77hs (626-717) / CstF-64hs (1-211)** a été purifié et des essais de cristallisation ont été réalisés.

Des expériences de « far western » (Takagaki and Manley 2000) ont montré que l'extrémité C-terminale de CstF-77 n'est pas essentielle pour l'interaction avec CstF-64 et qu'une région riche en Pro, comprenant les acides aminés du 580 au 626, est nécessaire. Nos études, ne coïncidant pas avec ces résultats, prouvent que la région de CstF-77hs indispensable pour l'interaction commence au résidu 626 et se termine au dernier acide aminé.

L'interaction de CstF-64 avec l'ARN a été décrite auparavant (Takagaki and Manley 1997) (Perez Canadillas and Varani 2003) (Deka et al. 2005). Nos résultats ont mis en évidence l'existence d'une association entre le complexe CstF-77hs (626-717) / CstF-64hs (1-211) et une séquence **d'ARN riche en GU** et incluant UU dans sa séquence. La présence de « l'hinge » de CstF-64 et de la région C-terminale de CstF-77 pourrait modifier les constantes d'affinité déterminées pour le motif RRM isolé. Des études sur l'affinité entre ces deux partenaires, complexe protéique et ARN, peuvent être envisagés au laboratoire par des méthodes biophysiques de RMN (résonance magnétique nucléaire), de SPR (résonance plasmonique de surface) ou d'ITC (calorimétrie à titrage isotherme).

Dans le cadre de cette thèse, des essais de **cristallisation** du complexe protéique avec l'ARN ont été effectués mais nous n'avons pas obtenu de cristaux exploitables pour la diffraction par rayons X. La caractérisation biophysique du complexe peut aider à déterminer la stabilité et le comportement des deux protéines. Ainsi, d'autres expériences de RMN sont en cours au laboratoire. Par ailleurs, l'utilisation des techniques comme la **DLS** (diffusion dynamique de la lumière ou Dynamic Light Scattering), l'**ultracentrifugation analytique** ou le **dichroïsme circulaire**, pourraient également être envisagée. L'application de ces méthodes au complexe CstF-77hs (626-717) / CstF-64hs (1-211) avec ou sans ARN, permettraient l'obtention des informations sur sa conformation/repliement,

sur l'homogénéité des échantillons, sur les changements structuraux induits par la l'addition de l'ARN ou sur la stœchiométrie du complexe protéique avec différentes séquences d'ARN. Sur la base de ces résultats, la poursuite des essais de cristallisation peut être envisagée à partir de quatre stratégies différentes :

- l'optimisation des conditions de cristallisation initiales pour l'obtention de cristaux exploitables,
- l'utilisation d'ARNs qui diffèrent en taille et/ou en séquence à ceux utilisés dans nos études,
- l'emploi d'autres kits commerciaux pour trouver des nouvelles conditions,
- la modification de la concentration protéique ou de la solution de départ pour la réalisation des gouttes.

En outre, l'utilisation des nouvelles constructions des deux sous-unités peut être envisagée. En effet, la co-expression de **CstF-77hs (219-717)** et **CstF-64hs (1-211)** permet la formation du complexe chez la bactérie. L'utilisation d'une construction plus longue de la protéine CstF-77hs peut modifier fortement le comportement du complexe au niveau de sa stabilité et de sa capacité à former des cristaux. La production à grande échelle, la purification et la cristallisation de ce nouveau complexe pourrait aboutir à la résolution de la structure de CstF-77hs (219-717) / CstF-64hs (1-211), avec ou sans l'ARN. Pour continuer l'étude dans cette voie, la stœchiométrie du complexe devra être prise en compte. La nouvelle construction de CstF-77hs est prédite comme formant un dimère ce qui pourrait changer les propriétés du complexe et son interaction avec l'ARN puisque deux motifs RRM seraient présents. La stœchiométrie du complexe peut être étudiée par des techniques comme la RMN, ou la spectrométrie de masse.

Dans le *Chapitre I: Introduction*, l'importance de CstF-64 comme **régulateur de la polyadénylation tissu-spécifique** a été décrite. Un aspect intéressant que nous pourrions développer serait l'étude des variants de CstF-64, en déterminant comment les modifications engendrées dans la protéine peuvent affecter son interaction avec CstF-77. Le variant de CstF-64 nommé  **$\tau$ CstF** contient une substitution Pro→Ser dans le domaine RRM, 9 répétitions de la séquence MEARA à la place de 12 et deux insertions qui ne se trouvent pas sur CstF-64. Cette version de CstF-64 a été associée à certains problèmes neurologiques et d'infertilité masculine (Dass et al. 2002) (Dass et al. 2007).  **$\tau$ CstF-64** varie considérablement de la protéine CstF-64 ce qui pourrait influencer son interaction avec CstF-77.  **$\beta$ CstF-64** est un des autres variants de CstF-64 identifiés (Shankarling et al. 2009). Dans cette isoforme de CstF-64, la région riche en Pro/Gly est interrompue par l'insertion d'un autre domaine. Cette région n'est pas directement impliquée dans la formation du complexe CstF-77/CstF-64 mais l'insertion des résidus peut induire un changement de conformation qui affecterait l'interaction.

D'autre part, l'implication de la région « hinged » de CstF-64 dans l'interaction avec la **symplekine** a été démontrée dans la littérature (Takagaki and Manley 2000). La symplekine est une protéine qui participe au clivage et à la polyadénylation des ARNm (Hatton et al. 2000). De ce fait,

il est possible qu'il y ait une compétition entre les deux protéines, symplekine et CstF-77, pour le domaine « hinge » de CstF-64 (Takagaki and Manley 2000). Il serait intéressant de connaître les détails de l'interaction CstF-64 / symplekine et ainsi déterminer la signification de cette éventuelle compétition.

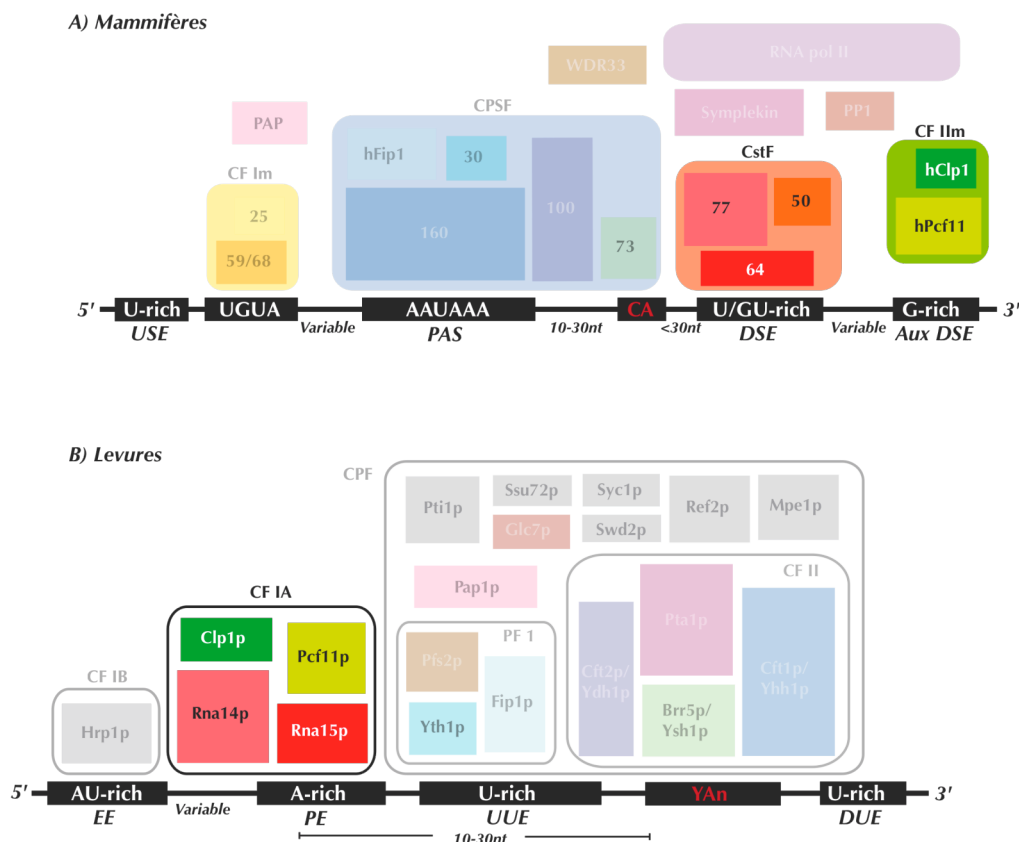
Finalement, une des perspectives pour la continuation de ces travaux serait l'étude des interactions inter-domaines existant entre les trois protéines de **CstF**. La région de **CstF-50** chargée de la maintenir associée au complexe, le domaine WD-40, n'est pas soluble. Le comportement de ce domaine rend difficile la reconstitution du complexe. Cependant, l'utilisation des tampons permettant la stabilisation de la protéine, l'addition des domaines de CstF-77 ne participant pas directement à l'interaction ou des combinaisons de CstF-50/CstF-77 avec la sous-unité CstF-64 pourraient être envisagées.

## Chapitre III : Résultats obtenus pour le complexe CF IA

### A. Etat de l'art : comparaison des complexes CstF et CF IA

Le complexe CF IA de levures est composé de quatre protéines : **Rna15p**, **Rna14p** (Minvielle-Sebastia et al. 1994), **Pcf11p** (Amrani et al. 1997) et **Clp1p** (Minvielle-Sebastia et al. 1997). Les deux premières correspondent aux homologues des protéines humaines CstF-64 et CstF-77, respectivement. Pcf11p équivaut à la région N-terminale de la protéine humaine hPcf11 (de Vries et al. 2000) et Clp1p est l'homologue de hClp1.

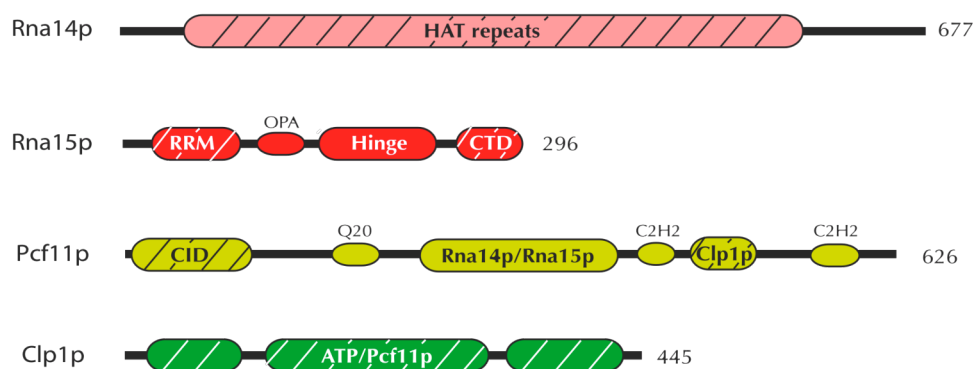
Chez les mammifères, CstF-64 et CstF-77 font partie du complexe **CstF** avec CstF-50 (protéine dont l'homologue chez la levure n'a pas été identifié). L'interaction existante entre hPcf11 et hClp1 est conservée chez la levure mais ces deux protéines font partie, chez les mammifères, d'un complexe isolé appelé **CF IIm** (fig. 58). De plus, aucune interaction n'a été détectée entre ce complexe et CstF. Une autre caractéristique importante du complexe de levure est que, à la différence de leurs homologues humains, les quatre sous-unités de CF IA participent aux deux étapes de la maturation 3' des ARNm (Minvielle-Sebastia et al. 1994) (Minvielle-Sebastia et al. 1997) (Amrani et al. 1997) (Gross and Moore 2001).



**Figure 58.** Schéma des facteurs impliqués dans la maturation 3' des ARNm. Les protéines qui ont fait sujet de cette thèse font partie de complexes différents chez les mammifères (CstF et CF IIm) et chez la levure (CF IA).

Les protéines intégrant les deux facteurs, CstF et CF IA, diffèrent. Cependant, malgré les dissimilarités dans l'organisation des complexes, les **séquences** des quatre protéines de CF IA (Rna14p, Rna15p, Pcf11p et Clp1p) et les **interactions** entre elles ont été conservées à travers l'évolution. De plus, une nouvelle caractéristique commune aux deux groupes a été mise en évidence. Aussi bien Rna14p que son homologue CstF-77 forment un homodimère (Legrand et al. 2007) (Bai et al. 2007) (Noble et al. 2004) déterminant, de cette manière, la **stœchiométrie** des complexes CF IA et CstF, respectivement. L'étude structurale de CF IA de levure peut, de cette façon, participer à la compréhension du mécanisme de polyadénylation chez les mammifères.

Comme décrit dans le *Chapitre I*, certaines des **interactions** entre les différents partenaires de CF IA sont connues. L'association entre Rna14p et Rna15p a été mise en évidence grâce à des expériences de co-immunoprécipitation (Kessler et al. 1996) et plus tard elle a été corroborée dans d'autres groupes de recherche (Kessler et al. 1997) (Gross and Moore 2001) (Noble et al. 2004). Les détails moléculaires sur l'interaction Clp1p/Pcf11p ont été obtenus à partir de la résolution de la structure de Clp1p en complexe avec l'ATP et une région minimale de Pcf11p indispensable pour la liaison (Noble et al. 2007). De plus, des expériences de « pull-down » réalisées à partir de la protéine recombinante GST-Pcf11p, ont montré l'existence d'une interaction entre cette sous-unité de CF IA et les trois autres protéines du complexe (Gross and Moore 2001) (*fig. 59*). Cependant, les régions impliquées dans l'association des différentes sous-unités, les détails moléculaires des interactions et l'assemblage général du complexe doivent être explorés pour comprendre la fonction de chaque protéine et le mécanisme général du processus de clivage et de polyadénylation dans lequel CF IA participe.



**Figure 59.** Schéma des domaines protéiques présents dans les sous-unités du complexe CF IA.

Les traits diagonaux représentent les zones dont la structure est connue.

## **B. Analyse structurale et fonctionnelle du complexe CF IA**

### ***1. Etude structurale des domaines d'interaction des sous-unités Rna15p et Rna14p***

#### **a.) Etudes précédentes réalisées sur l'interaction entre Rna14p et Rna15p**

La participation de Rna15p et Rna14p dans le processus de maturation 3' des pré-ARNms et l'hypothèse de la formation d'un complexe entre les deux protéines a été mise en évidence par L. Minvielle-Sebastia en 1994 (Minvielle-Sebastia et al. 1994). Cette interaction a été démontrée plus tard par co-immunoprécipitation (Kessler et al. 1996) (Gross and Moore 2001), par des expériences de « pull down » (Legrand et al. 2007) et par des études protéomiques (Nedeia et al. 2003) (Gavin et al. 2002).

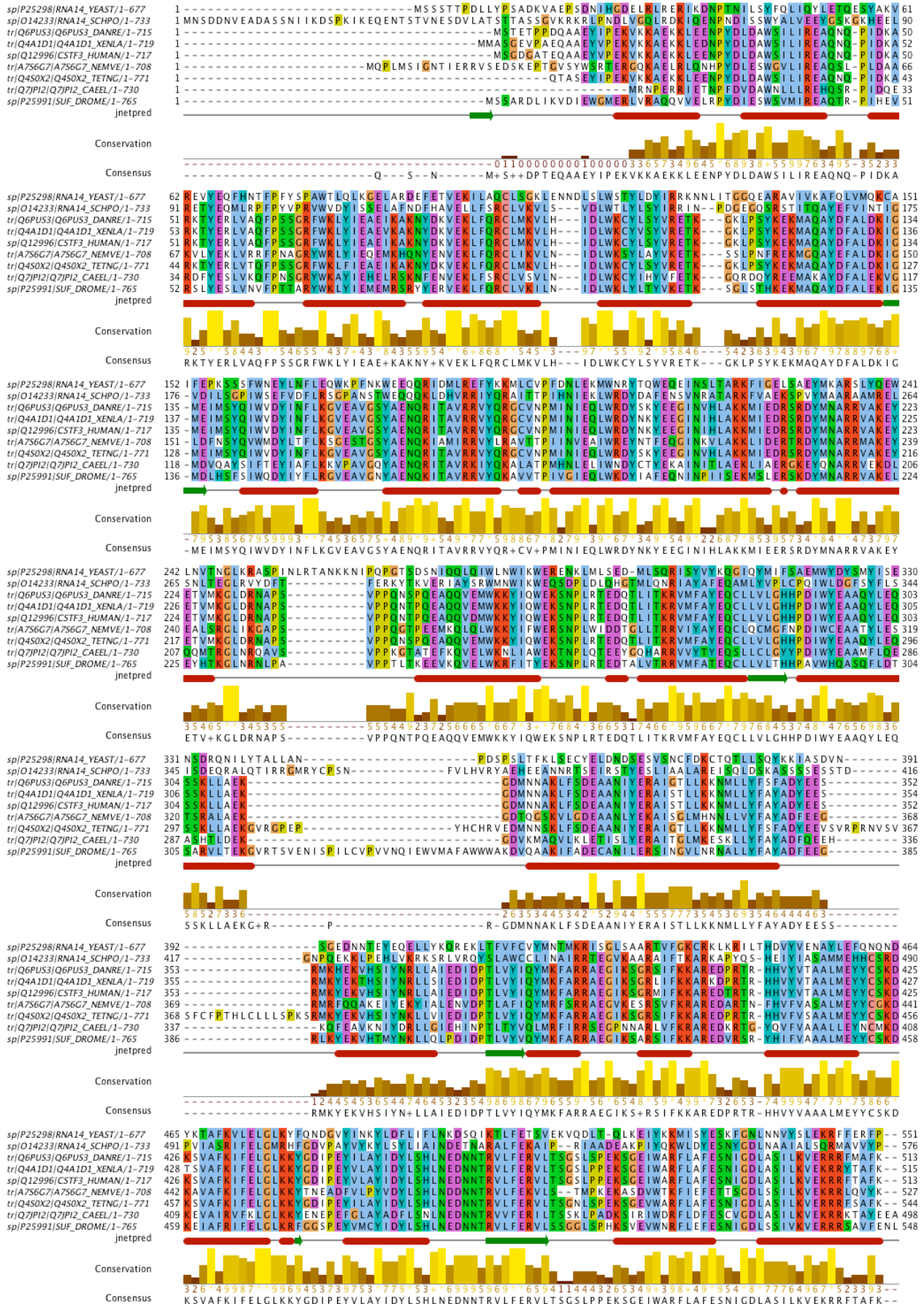
Dans notre laboratoire l'interaction des deux sous-unités a été étudiée à partir de la **co-expression** chez *E. coli* des protéines recombinantes. Le projet a été démarré avant mon arrivée au laboratoire par **Lynn Gregory** (postdoctorante au laboratoire, 2005-2007). La co-expression des deux protéines entières, la purification du complexe et des essais de cristallisation ont été réalisés. Des essais de co-purification des protéines exprimées séparément ont aussi été effectués. Cependant, l'absence de cristaux exploitables nous a mené vers une nouvelle stratégie basée sur la détermination des régions impliquées dans la formation de l'hétérodimère.

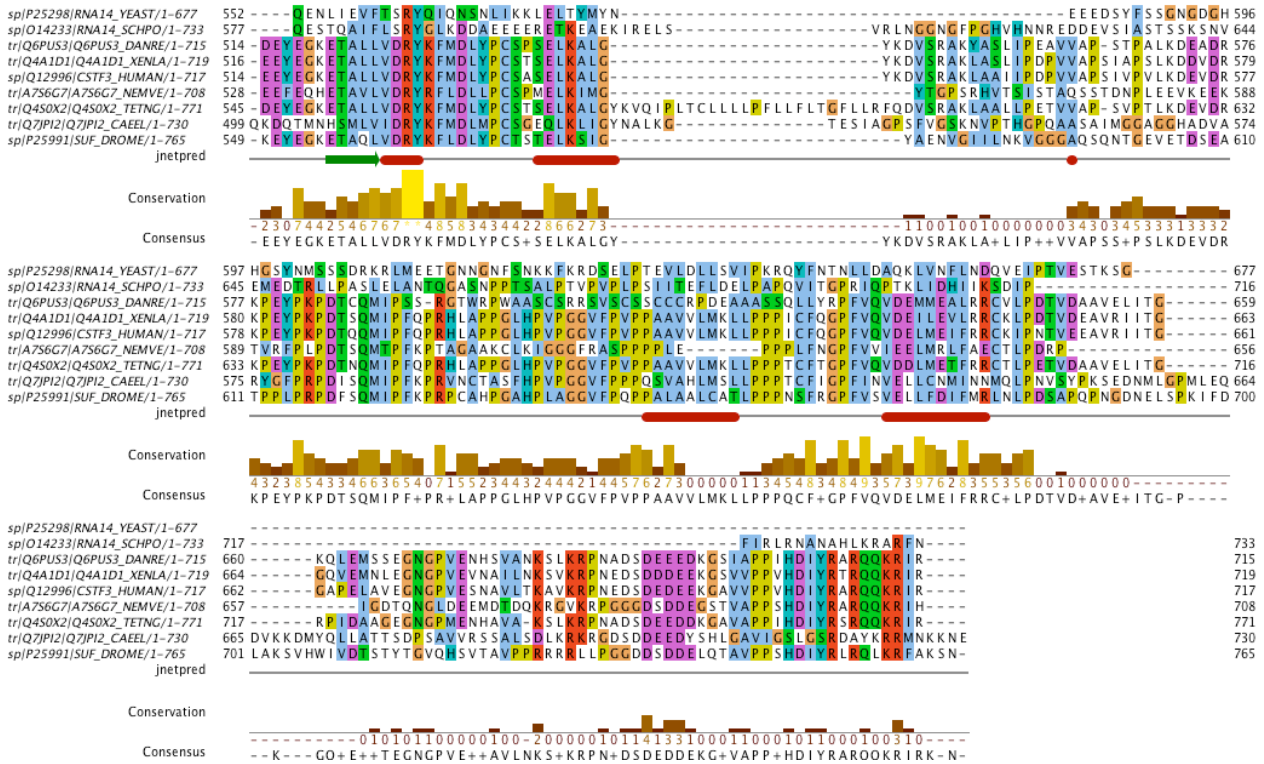
#### **b.) Identification des régions de Rna14p et de Rna15p impliquées dans l'interaction**

Le but est de trouver les constructions permettant la résolution de la structure et de révéler ainsi les bases moléculaires de l'association Rna14p-Rna15p.

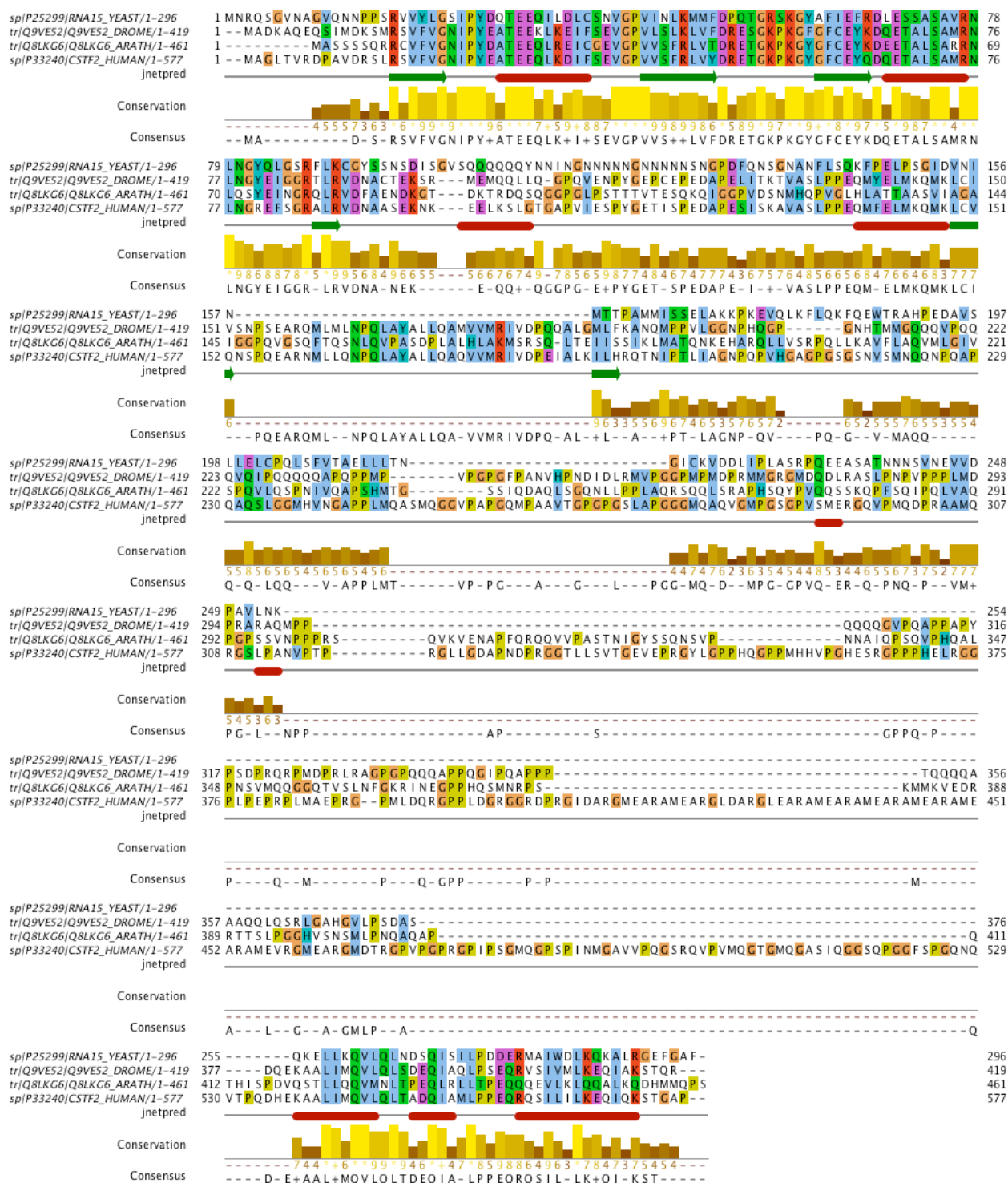
Pour **cartographier** l'interaction entre ces deux sous-unités, différentes régions de chacune des protéines ont été clonées dans divers vecteurs. L'utilisation de deux vecteurs portant de la résistance à des antibiotiques distincts et l'emploi de vecteurs bicistroniques ont participé à la **co-expression** des deux protéines (*voir Chapitre VI : Matériels et Méthodes*). Le choix des régions protéiques mises à l'étude a été établi à partir des observations des séquences, de la conservation des résidus, de la prédiction de la structure secondaire des deux sous-unités et des données obtenues à partir de la littérature (*figs. 60, 61 et 62A*). L'implication de la région C-terminale de Rna14p dans l'interaction avec Rna15p a été mise en évidence dans notre groupe (Legrand et al. 2007). Concernant Rna15p, la région « hinge » de son homologue CstF-64 est responsable de l'interaction avec CstF-77 (Takagaki and Manley 2000) (Hatton et al. 2000). Nous pourrions anticiper que, de la même façon, ce type de domaine interviendra dans la formation du complexe Rna15p/Rna14p.







**Figure 60.** Alignement de séquences de la protéine Rna14p. La conservation des résidus est représentée par un histogramme où les barres plus hautes et la couleur jaune indiquent une conservation plus importante. Une représentation de la prédiction de la structure secondaire est placée en dessous de l'alignement. Les flèches vertes montrent les feuillets bêta potentiels et les cylindres rouges les hélices alpha. YEAST : *Saccharomyces cerevisiae* ; SCHPO : *Schizosaccharomyces pombe* ; DANRE : *Danio rerio* ; XENLA : *Xenopus laevis* ; HUMAN : *Homo sapiens* ; NEMVE : *Nematostella vectensis* ; TETNG : *Tetraodon nigroviridis* CAEEL : *Caenorhabditis elegans* ; DROME : *Drosophila melanogaster*.



**Figure 61.** Alignement de séquences de la protéine Rna15p. La conservation des résidus est représentée par un histogramme où les barres plus hautes et la couleur jaune indiquent une conservation plus importante. Une représentation de la prédiction de la structure secondaire est placée en dessous de l'alignement. Les flèches vertes montrent les feuilletés bêta potentiels et les cylindres rouges les hélices alpha. YEAST : *Saccharomyces cerevisiae* ; DROME : *Drosophila melanogaster* ; ARATH : *Arabidopsis thaliana* ; HUMAN : *Homo sapiens*. Les programmes utilisés pour les alignements et pour les prédictions des structures secondaires sont : Blastp, ClustalW, Jalview et Jpred (Altschul et al. 1997) (Thompson et al. 1994) (Waterhouse et al. 2009) (Cole et al. 2008).

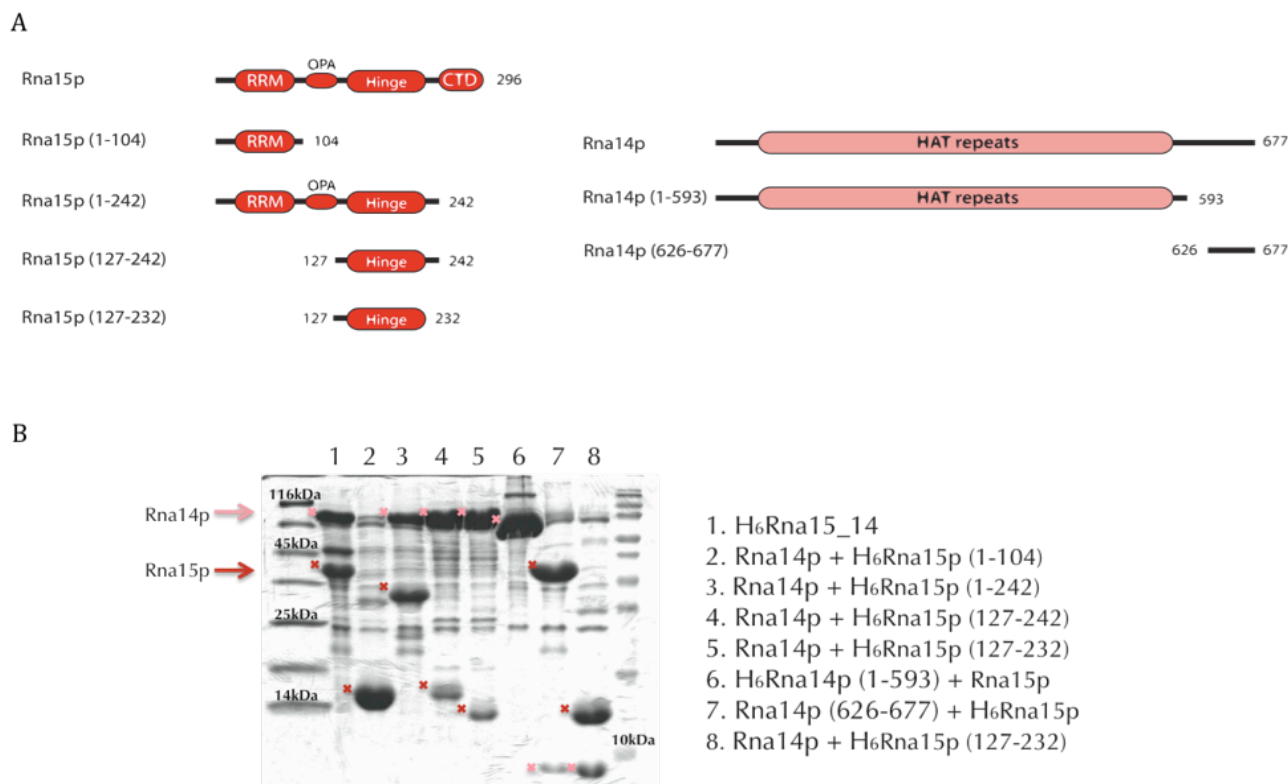
Nous avons observé que les versions entières de Rna14p et de Rna15p peuvent être co-exprimées chez *E. coli* à partir d'un vecteur bicistronique qui contient l'ADN codant pour les deux protéines en tandem (*fig. 62B, puits 1*). La co-transformation de deux vecteurs contenant l'information pour les deux protéines est également efficace (résultats non présentés sur le gel). Les informations obtenues à partir de ces résultats sont :

- la région de Rna15p qui comprend les acides aminés du 1 au 104, correspondant au motif RRM (Minvielle-Sebastia et al. 1991), ne participe pas à l'interaction (*fig. 62B, puits 2*),
- l'addition de résidus au motif RRM de Rna15p dans la partie C-terminale, pour former la version de la protéine Rna15p (1-242), permet l'association avec Rna14p (*fig. 62B, puits 3*),
- la zone minimale d'interaction pour Rna15p contient les résidus **127 à 232** (*fig. 62B, puits 4 et 5*),
- la région C-terminale de Rna14p, **Rna14p (626-677)**, est nécessaire et suffisante pour l'association avec Rna15p. Ainsi, Rna14p (1-593) n'est pas capable d'interagir avec Rna15p (*fig. 62B, puits 6 et 7*).

De cette façon, nous avons pu conclure que la partie C-terminale de Rna14p, incluant les acides aminés du **626 au 677**, et les résidus correspondant au « hinge » de Rna15p, du **127 au 232**, étaient suffisants pour l'interaction. Des expériences de RMN réalisées sur le complexe H<sub>6</sub>Rna15p (127-242)/ StrRna14p (626-677) ont contribué à l'identification de la région minimale de Rna15p participant dans l'association avec Rna14p. En effet, les résultats obtenus ont montré la présence des zones très flexibles et non repliées situées sur les extrémités de Rna15p (127-242) (*voir section B.1.c.4 de ce chapitre*).

Dans la *figure 62*, seuls les principaux tests effectués pour cartographier l'interaction Rna14p-Rna15p ont été représentés. D'autres constructions protéiques ont été réalisées pour l'identification des régions minimales impliquées dans l'interaction. Dans les cas où un complexe a été formé, nous avons poursuivi son étude par sa purification et par des essais de cristallisation. Les résultats obtenus sont présentés ci-dessous.



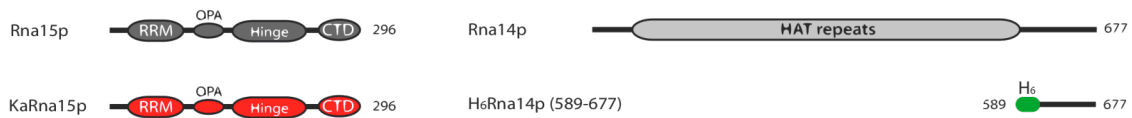


**Figure 62.** Tests de co-expression des sous-unités Rna14p et Rna15p. A.) Régions des sous-unités Rna14p et Rna15p employées pour les tests de co-expression. B.) A gauche, sur le gel 18% SDS-PAGE, les bandes correspondants aux fragments de Rna14p sont identifiées par une croix couleur saumon et pour Rna15p par une croix rouge. Deux marqueurs de protéines, chargés dans le premier et le dernier puits, ont été utilisés pour mieux identifier les différentes régions protéiques. Le numéro 1 correspond à l'expression à partir d'un vecteur bicistronique qui inclue les deux sous-unités entières. A droite, les différentes constructions utilisées pour l'expérience. H<sub>6</sub>: l'ADN d'intérêt est cloné dans un vecteur qui confère résistance à l'ampicilline et contient une étiquette de 6 His ; Dans 2, 3, 4 et 5, le vecteur employé pour Rna14p porte la résistance au chloramphénicol ; Dans le numéro 6, Rna15p a été cloné dans un vecteur résistant à la kanamycine ; Dans 7 et 8, le vecteur utilisé pour le clonage de Rna14p porte la résistance à la streptomycine.

**c.) Etudes préliminaires : purification et essais de cristallisation des complexes formés entre différentes régions de Rna14p et Rna15p**

Nous prétendons décrire les bases moléculaires de l'interaction Rna14p/Rna15p par la résolution de sa structure. La cristallisation du complexe s'est avérée difficile malgré les nombreux essais à partir de différentes régions des deux protéines. Dans cette section, les combinaisons de Rna15p et Rna14p faisant l'objet de purification et de tests de cristallisation sont présentées. Pour les protéines entières, ces essais avaient été effectués par L. Gregory mais des cristaux du complexe n'ont jamais été obtenus.

c.1) Purification et essais de cristallisation du complexe formé entre Rna14p (589-677) et Rna15p

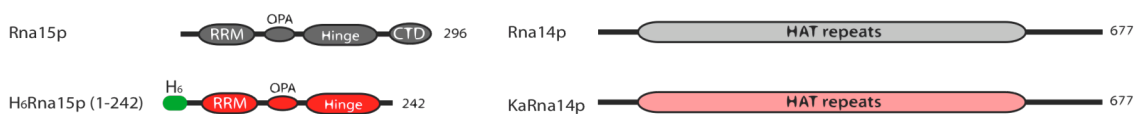


La co-expression de H<sub>6</sub>Rna14p (589-677) et KaRna15p\* a été réalisée à partir de la souche **Rosetta (DE3)**. Après trois étapes de purification (chromatographie d'affinité, échange ionique et filtration sur gel) (*fig. 63*), le complexe a été concentré. Le rendement de la production du complexe a été d'approximativement 1mg de protéine par litre de culture bactérienne dans un milieu LB.

Les premiers essais de cristallisation ont été réalisés à partir du complexe H<sub>6</sub>Rna14p (589-677)/ KaRna15p seul, mais pour les expériences suivantes un oligonucléotide d'ARN, dont la séquence est **UGCGUAUUCUCC**, a été ajouté. Cette longueur d'ARN permet potentiellement d'interagir avec 2 motifs RRM. Le rapport pour la formation du complexe a été de 1,2 moles d'ARN pour 2 moles de protéine. En effet, pour ces essais, l'hypothèse adoptée reposait sur la dimérisation de la protéine Rna14p. Dans ce cas, deux motifs RRM de Rna15p seraient également présents et une seule mole de cet ARN aurait pu interagir avec 2 moles du complexe protéique (en prenant comme masse du complexe la somme des masses des deux protéines). Cependant, nous savons maintenant que la zone C-terminale de Rna14p n'est pas suffisante pour l'oligomérisation de la protéine. Nous aurions donc dû utiliser un rapport de 1 mole d'ARN pour 1 mole de complexe.

Dans ces travaux réalisés à partir du complexe H<sub>6</sub>Rna14p (589-677)/ KaRna15p, malgré l'utilisation de différents screens commerciaux et la présence/absence d'ARN, nous n'avons pas pu obtenir des cristaux.

c.2) Purification du complexe Rna14p / Rna15p (1-242)

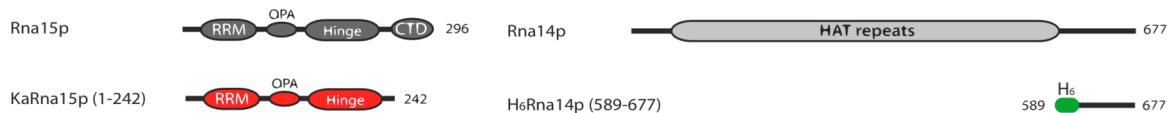


Pour le complexe formé entre H<sub>6</sub>Rna15 (1-242) et KaRna14p la purification a été réalisée en trois étapes différentes de celles suivies pour le complexe précédent. Nous avons commencé par une chromatographie d'affinité qui a été suivie d'une colonne héparine (HiTrap™ Heparin HP) et finalement une d'exclusion stérique (*fig. 63*). La stabilité de ce complexe était plus faible que pour le complexe H<sub>6</sub>Rna14p (589-677)/ KaRna15p et une partie de H<sub>6</sub>Rna15 (1-242) s'est dissociée du complexe après la dernière étape de purification. La concentration du complexe H<sub>6</sub>Rna15 (1-242)/

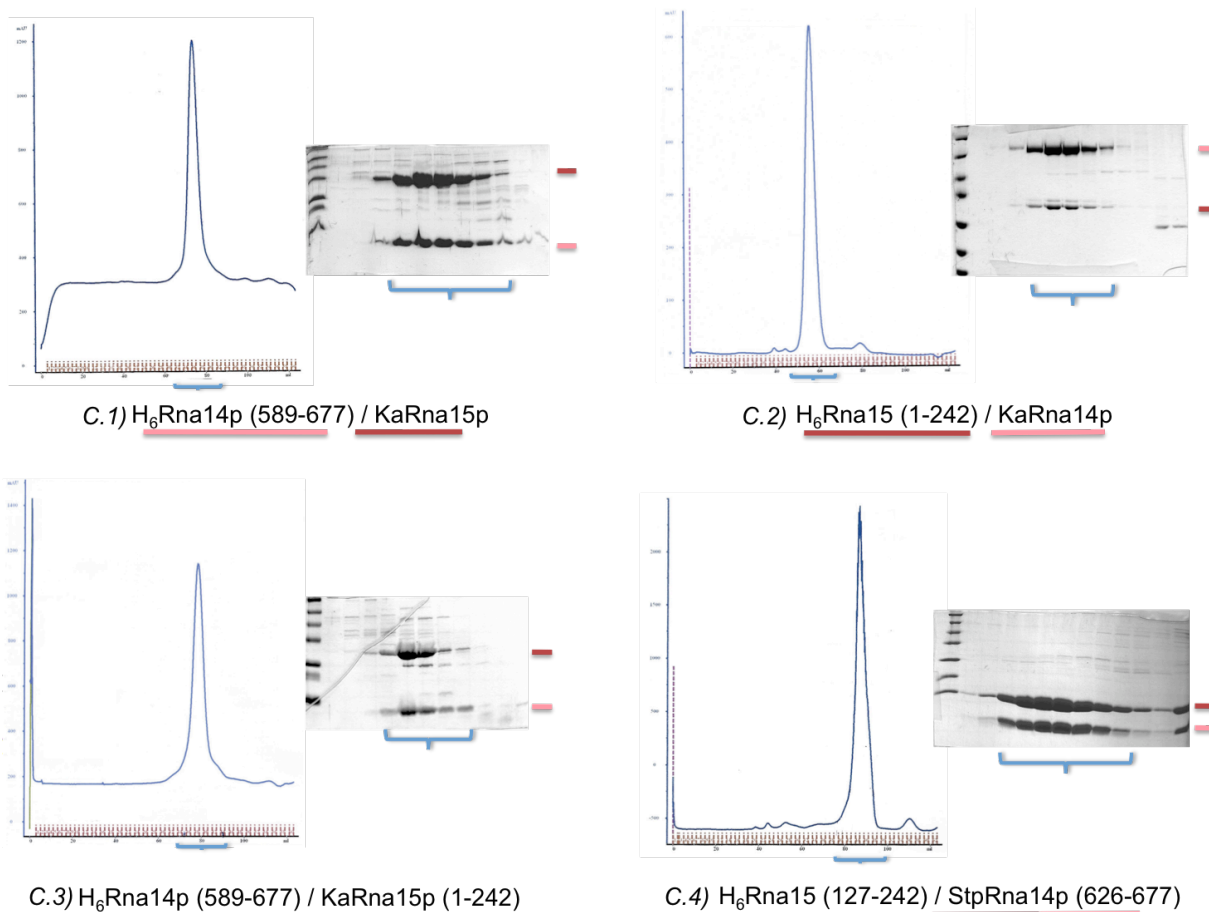
\* L'ADN codant pour Rna15p est intégré dans un vecteur résistant à la kanamycine.

KaRna14p, effectuée à l'aide d'un tube Amicon<sup>®</sup>, a induit la précipitation des protéines et nous n'avons pas pu procéder aux étapes de cristallisation.

*c.3) Purification et essais de cristallisation du complexe Rna14p (589-677) / Rna15p (1-242)*

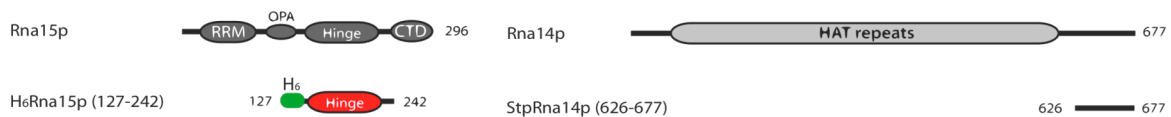


Ce complexe a été également purifié à partir d'une chromatographie d'affinité, une colonne d'échange ionique et une d'exclusion stérique (fig. 63). Une meilleure expression a été atteinte avec l'étiquette de H<sub>6</sub> située sur l'extrémité N-terminale de Rna14p. La formation du complexe avec l'ARN a été effectuée avec une stœchiométrie de 2 moles de protéine pour 1 mole d'ARN. Les tests de cristallisation réalisés à partir du complexe Rna14p (589-677) / Rna15p (1-242) en absence/présence d'ARN n'ont pas dévoilé des conditions permettant la formation de cristaux.



**Figure 63.** Dernière étape de purification pour chaque complexe étudié. Rna15p est représentée par la couleur rouge et Rna14p par la couleur saumon.

*c.4) Purification et essais de cristallisation du complexe Rna14p (626-677) / Rna15p (127-242)*



Jusqu'à présent, le motif RRM de Rna15p avait été inclus dans le complexe. Pour les expériences suivantes ce motif a été éliminé et de cette façon, la possibilité de la formation d'un complexe avec l'ARN a aussi été écartée. Ainsi, la nouvelle version de Rna15p contient seulement le motif « hinge », du résidu 127 au 242 et l'étiquette H<sub>6</sub> sur l'extrémité N-terminale. La séquence de la région C-terminale de Rna14p employée a été aussi raccourcie par rapport à celle utilisée antérieurement. Les acides aminés 589 à 625, qui ne sont pas conservés entre les espèces, ont été retirés.

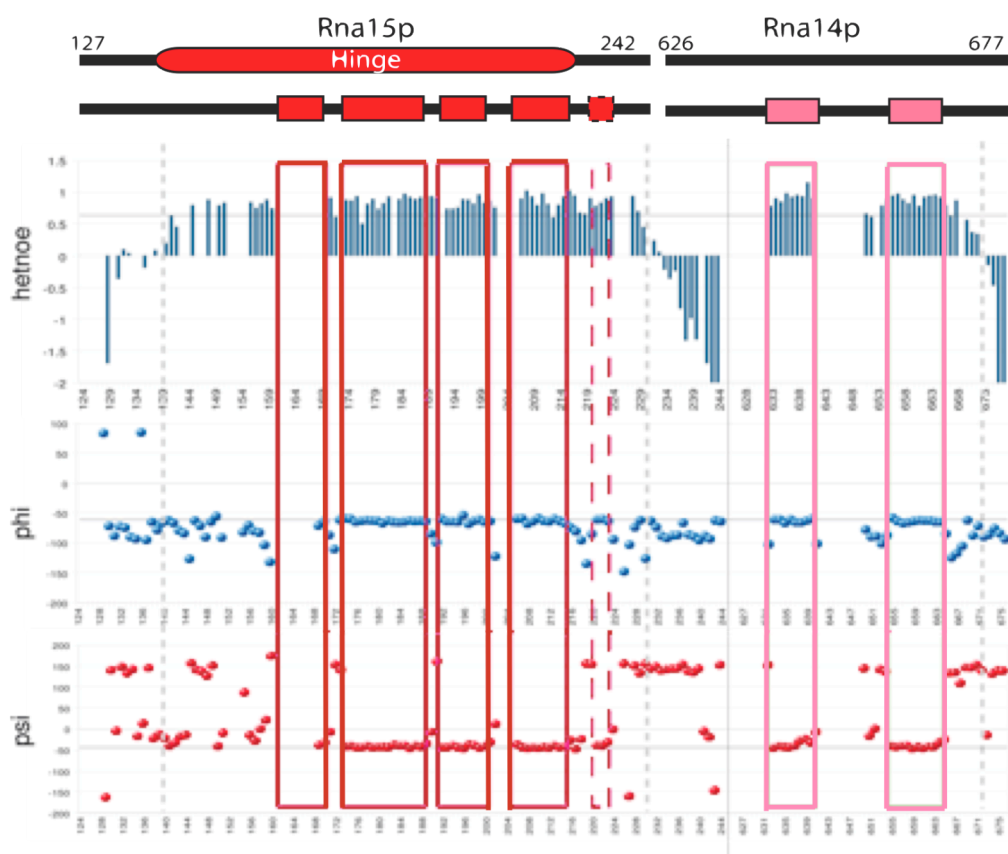
Pour l'obtention de ce complexe pur, nous avons eu besoin de seulement deux étapes de purification, une chromatographie d'affinité suivie d'une d'exclusion stérique (*fig. 63*). Des essais de cristallisation ont été réalisés à partir du complexe protéique à une concentration de **25mg/ml**. De nouveau, les conditions essayées n'ont pas permis la formation de cristaux capables de diffracter les rayons X.

Après les nombreux échecs et afin d'étudier l'organisation générale du complexe, nous avons effectué des expériences de **RMN** à partir de H<sub>6</sub>Rna15 (127-242) / StpRna14p (626-677)\*, la taille des protéines le permettant. La technique de RMN nous a servi dans d'autres occasions (ex. CstF-50) à trouver les constructions protéiques pouvant susciter leur cristallisation. Pour ces expériences, le marquage du complexe avec <sup>15</sup>N et <sup>13</sup>C a été nécessaire (*voir Chapitre VI*). Ces travaux ont été réalisés en collaboration avec le **Dr. C. Mackereth**. L'interprétation des premiers résultats obtenus est représentée sur la *figure 64*. Ces expériences ont montré que certaines régions étaient repliées et d'autres plus flexibles. Ces dernières pourraient empêcher éventuellement la cristallisation.

Nous avons pu conclure que, pour Rna15p, les deux extrémités sont très flexibles, spécialement sur la région C-terminale. La partie C-terminale correspondant à la fin de la protéine Rna14p était aussi mal repliée. A partir de ces résultats, des nouvelles constructions susceptibles d'augmenter les probabilités de cristallisation ont été clonées. De cette façon, deux nouveaux plasmides ont été créés permettant l'expression d'autres régions de Rna15p : **Rna15p (127-232)**, **Rna15p (138-232)**. Pour la deuxième construction, l'expression chez *E. coli* n'était pas optimale et la poursuite des expériences a été réalisée à partir de Rna15p (127-232). La cristallisation du complexe, malgré la modification du nombre de résidus intégrant la protéine Rna15p, n'a pas été possible. La structure de cette région de Rna15p en complexe avec Rna14p (626-677) a été résolue par RMN et sera présentée ultérieurement.

\* L'ADN codant pour Rna14p est intégré dans un vecteur résistant à la streptomycine.

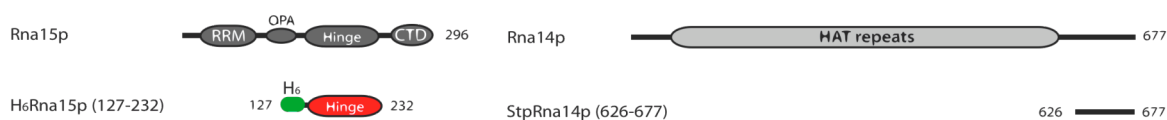




**Figure 64.** Schéma représentative des résultats de RMN obtenus à partir du complexe  $H_6Rna15$  (127-242) /  $StpRna14p$  (626-677). Dans la partie supérieure du diagramme (hetnoe), les barres situées dans le côté positif correspondent aux résidus plus ordonnés et celles qui se trouvent sur le côté négatif indiquent une flexibilité plus importante. Les deux autres diagrammes donnent aussi une idée du repliement du complexe à partir des angles Phi et Psi des protéines.

#### d.) Structure des domaines d'interaction des protéines Rna14p et Rna15p

##### d.1) Production, purification et marquage du complexe pour les expériences de RMN

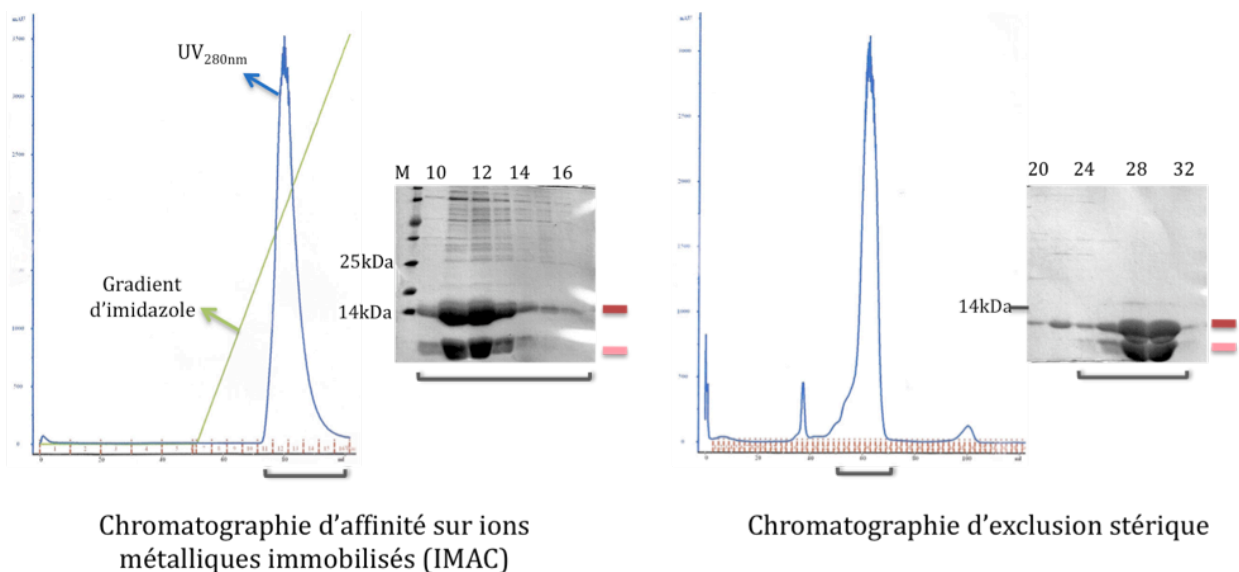


La co-expression des plasmides  **$H_6Rna15p$  (127-232)** et  **$StpRna14p$  (626-677)**, permet la formation d'un complexe stable qui est surexprimé chez *E. coli*. La transformation indépendante des deux constructions,  $H_6Rna15p$  (127-232) et  $H_6Rna14p$  (626-677), a impliqué des niveaux plus faibles d'expression et de stabilité des protéines. Des spectres  $^1H^{15}N$ -HSQC de RMN réalisés à partir des peptides séparés, montrent une dispersion des pics moins importante. Ceci indique un repliement moins complet quand les protéines sont exprimées séparément. L'addition de quantités

équimolaires de H<sub>6</sub>Rna15p (127-232) et H<sub>6</sub>Rna14p (626-677) augmentait la qualité du spectre, mais les résultats les plus favorables ont été observés à partir de la co-expression des deux (*voir annexe I pour consulter les spectres*). Dans notre cas, l'inconvénient de la co-expression est que les deux protéines ne peuvent pas être marquées séparément et ceci rend plus difficile la résolution de leur structure par RMN. Néanmoins, cette difficulté est compensée par une quantité de complexe soluble plus importante obtenu à partir de la co-expression, que ce soit dans des conditions natives ou dans le cas de marquage isotopique.

Pour la production du complexe, les plasmides H<sub>6</sub>Rna15p (127-232) et StpRna14 (626-677) ont été co-transformés dans la souche **BL21 (DE3)**. Des colonies bactériennes contenant les deux plasmides, donc capables de pousser sur un milieu LB-agar avec ampicilline et streptomycine, sont transférées dans un volume de 15ml de LB liquide. Cette pré-culture nous permet, le lendemain, d'inoculer 2 litres de milieu minimum contenant de l'azote 15 (<sup>15</sup>N) pour le **marquage** des protéines (*voir Chapitre VI : Matériels et Méthodes*). Pour les premières expériences, seulement ce marquage a été nécessaire, cependant, la résolution de la structure a nécessité plusieurs purifications du complexe marqué à l'azote 15 (<sup>15</sup>N), au carbone 13 (<sup>13</sup>C) et/ou au deutérium (<sup>2</sup>H).

Les premières purifications ont été effectuées en deux étapes, une chromatographie d'affinité au Co<sup>2+</sup> et une d'exclusion stérique sur une colonne HiLoad 16/60 Superdex™ 75 pg (*fig. 65*). L'étiquette H<sub>6</sub> a été clivée avec la protéase TEV pendant 16h, à 22°C. Par la suite, la purification sur une colonne de filtration sur gel a été remplacée par une deuxième colonne d'affinité sur une résine Ni<sup>2+</sup>. Cette nouvelle étape permet la rétention de l'étiquette de 6 His et de la protéine non clivée. Le complexe a été concentré et stocké dans un tampon final adéquat pour les expériences de RMN et pour la stabilité de la protéine : 20 mM sodium phosphate à pH 6.5, 50 mM NaCl et 2 mM DTT.

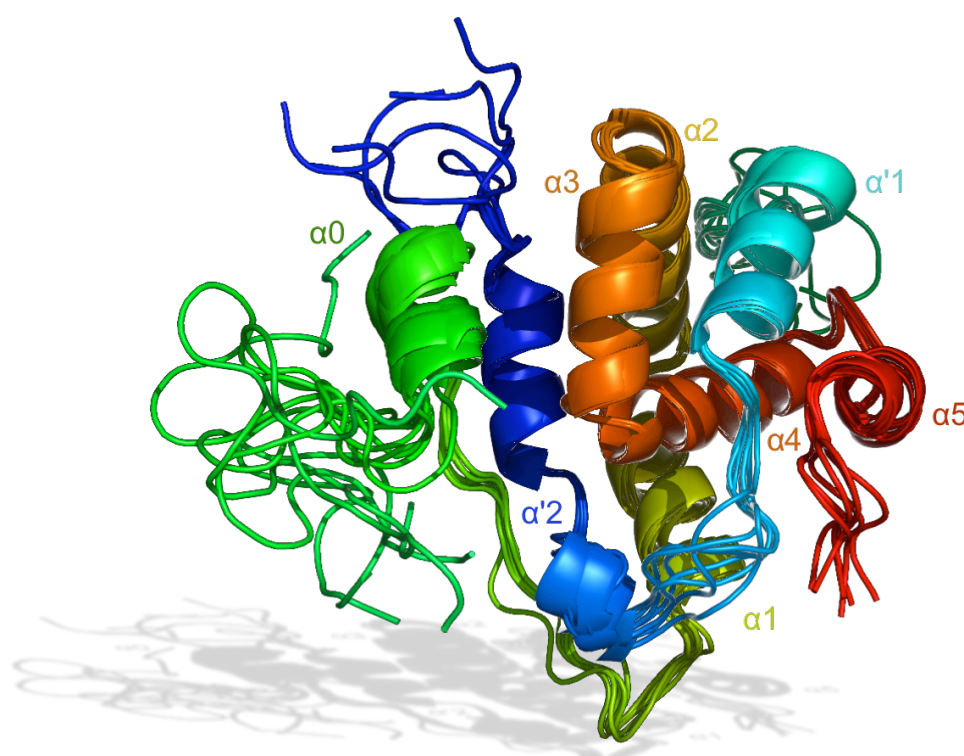


**Figure 65.** Etapes de purification du complexe Rna15p (127-232) / Rna14p (626-677). L'étiquette H<sub>6</sub> a été éliminée entre les deux étapes. Les bandes rouges et saumon indiquent les niveaux de migration pour la construction de Rna15p et de Rna14p, respectivement.

*d.2) Résolution de la structure du complexe Rna15p (127-232) et Rna14p (626-677)*

• **Structure du complexe H<sub>6</sub>Rna15p (127-232) / StpRna14p (626-677)**

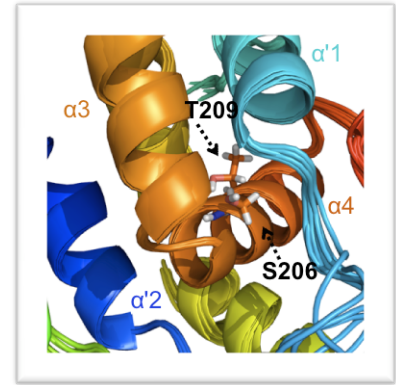
La structure du complexe a été résolue par notre collaborateur **C. Mackereth** par spectroscopie RMN (le spectre <sup>1</sup>H<sup>15</sup>N-HSQC avec l'attribution des résonances appartenant à chaque acide aminé a été représenté sur l'annexe 2). Le domaine de Rna15p employé pour l'interaction est composé de six hélices, α0-α5. Les hélices α'1 et α'2 de la région C-terminale de Rna14p forment une enveloppe autour du « hinge » de Rna15p. Les deux protéines, étant fortement associées, forment **une seule entité protéique** correspondant à l'hétérodimère minimal (*fig. 66*).



**Figure 66.** Modèle structural du complexe Rna15p (127-232) / Rna14p (626-677) résolu par RMN (superposition de 7 modèles). Les deux protéines sont colorées dans un gradient du bleu au rouge commençant par la région N-terminale de Rna14p jusqu'aux résidus situés sur l'extrémité C-terminale du « hinge » de Rna15p. Pour simplifier l'interprétation, nous avons respecté le code de couleur indiquant le nom de chaque hélice, α pour Rna15p et α' pour Rna14p.

Un élément de la structure qu'il faut souligner est la présence de plusieurs **résidus polaires** placés dans le cœur structural du complexe. Ce phénomène a été observé auparavant et dans ces cas-là, un rôle stabilisateur a été attribué à ce type de résidus. La **Ser<sub>206</sub>** et la **Thr<sub>209</sub>** de la région « hinge » de Rna15p ne se trouvent pas exposées au solvant (*fig. 67*). La **Ser<sub>206</sub>** est un acide aminé conservé chez certaines espèces de levures mais elle est substituée par une Ala chez les métazoaires.

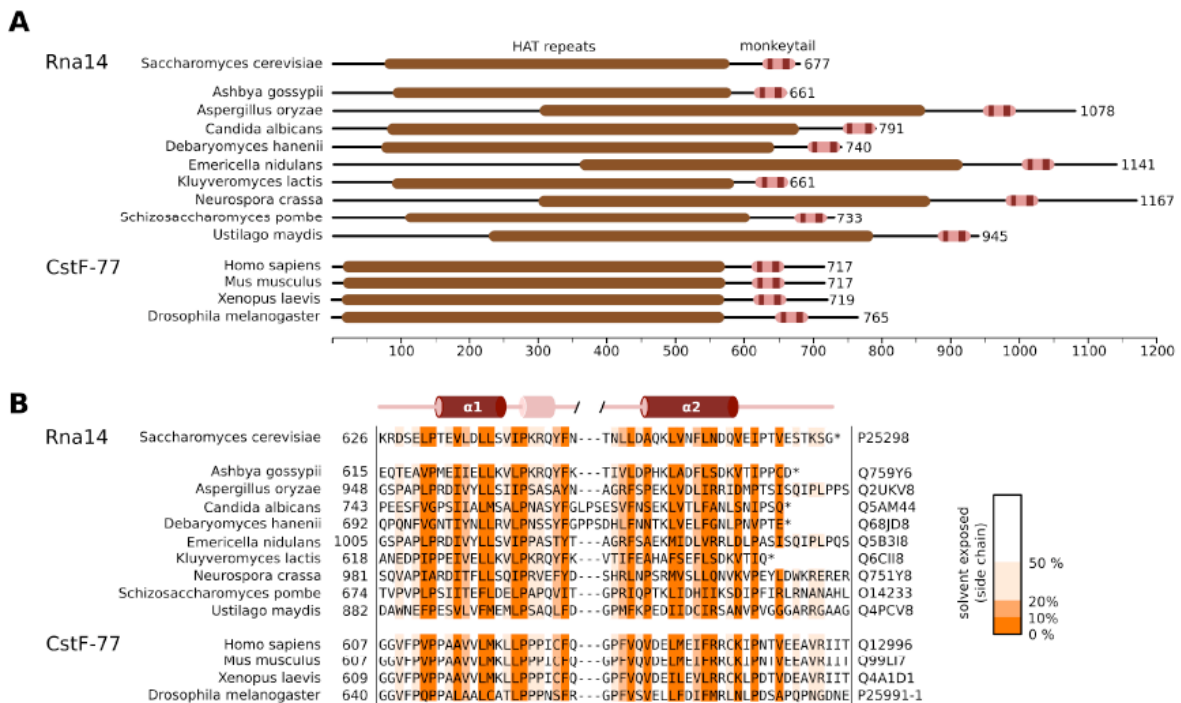
La **Thr<sub>209</sub>** est seulement présente chez *S. cerevisiae* et elle est substituée dans d'autres espèces par des résidus hydrophobes. Des expériences de RMN réalisées au laboratoire ont mis en évidence la formation d'un complexe Rna14p/Rna15p stable malgré la mutation de ces deux acides aminés en Ala. Nous avons observé que la mutation T209A affecte légèrement la stabilité thermique du complexe contrairement à la mutation S206A qui favorise cette stabilité. La conservation de la Ser<sub>206</sub> peut être due à une augmentation de la solubilité de Rna15p dans la cellule avant la formation du complexe ou à une augmentation de la spécificité dans la reconnaissance entre les deux protéines.



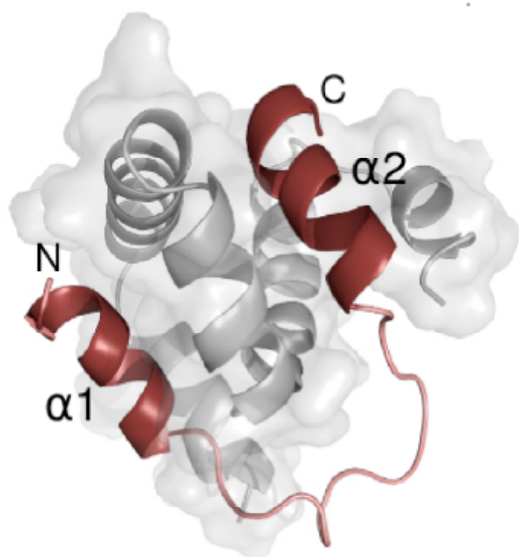
**Figure 67.** Détail de la structure de l'hétérodimère Rna14p/Rna15p où l'on observe les résidus Ser<sub>206</sub> et Thr<sub>209</sub>.

• **Description du domaine « monkeytail » de Rna14p**

Des études sur l'homologie des séquences de la région C-terminale de Rna14p (du résidu 626 au 677) et la protéine CstF-77, montrent une conservation limitée de cette zone entre les espèces incluses dans l'alignement (Legrand et al. 2007). Cependant, à partir d'une combinaison de la prédiction structurale de cette région et de l'identification de résidus hydrophobes, cette zone peut être localisée dans les homologues de Rna14p/CstF-77 chez les levures, les champignons ou les métazoaires (fig. 68).



**Figure 68. A)** Schéma des domaines présents dans Rna14 et CstF-77. Le domaine « monkeytail » a été retrouvé dans toutes les espèces à partir de l'identification des résidus hydrophobes et de la prédiction structurale des protéines. **B)** Détail du domaine « monkeytail ». La structure secondaire de la région « monkeytail » est indiquée en dessus des séquences. Les niveaux d'exposition au solvant pour chaque résidu sont représentés par un gradient orange, les résidus les plus exposés étant colorés en blanc (figure réalisée par C. Mackereth).



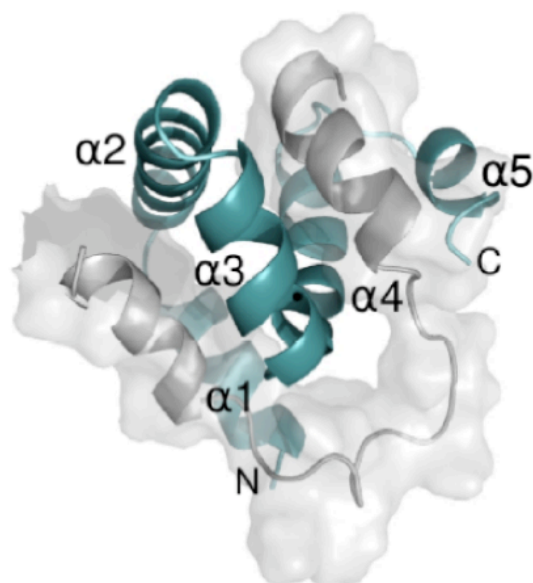
**Figure 69.** Domaine « monkeytail » de Rna14p. Le domaine « hinge » (de l'hélice  $\alpha 1$  à la  $\alpha 5$ ) est représenté en gris.

Dans le modèle structural obtenu pour le complexe, nous pouvons apprécier comment la « queue » C-terminale de Rna14p enveloppe le domaine « hinge » de Rna15p. Pour cette raison, cette région particulière a reçu le nom de domaine « **monkeytail** » ou « queue de singe » (fig. 69). Malgré des distances variables du domaine « monkeytail » par rapport aux répétitions HAT ou par rapport au dernier résidu de la région C-terminale, ce domaine est caractéristique des protéines Rna14p/CstF-77.

- **Description du domaine « hinge » de Rna15p**

Nos travaux ont donné pour la première fois un indice sur l'organisation structurale du domaine « hinge » (fig. 70). Ces dernières années, le rôle du domaine « hinge » de CstF-64 dans l'interaction avec CstF-77 et la symplektine ainsi que sa participation dans le transport de CstF-64 dans le noyau, ont été montrés (Hatton et al. 2000) (Takagaki and Manley 2000) (Hockert et al. 2010). L'alignement de séquences de CstF-64/Rna15 a confirmé la présence de ce domaine dans toutes les espèces étudiées (fig. 71). Ainsi, son implication dans la formation d'un hétérodimère CstF-64/Rna15 : CstF-77/Rna14 peut être, en principe, généralisée.

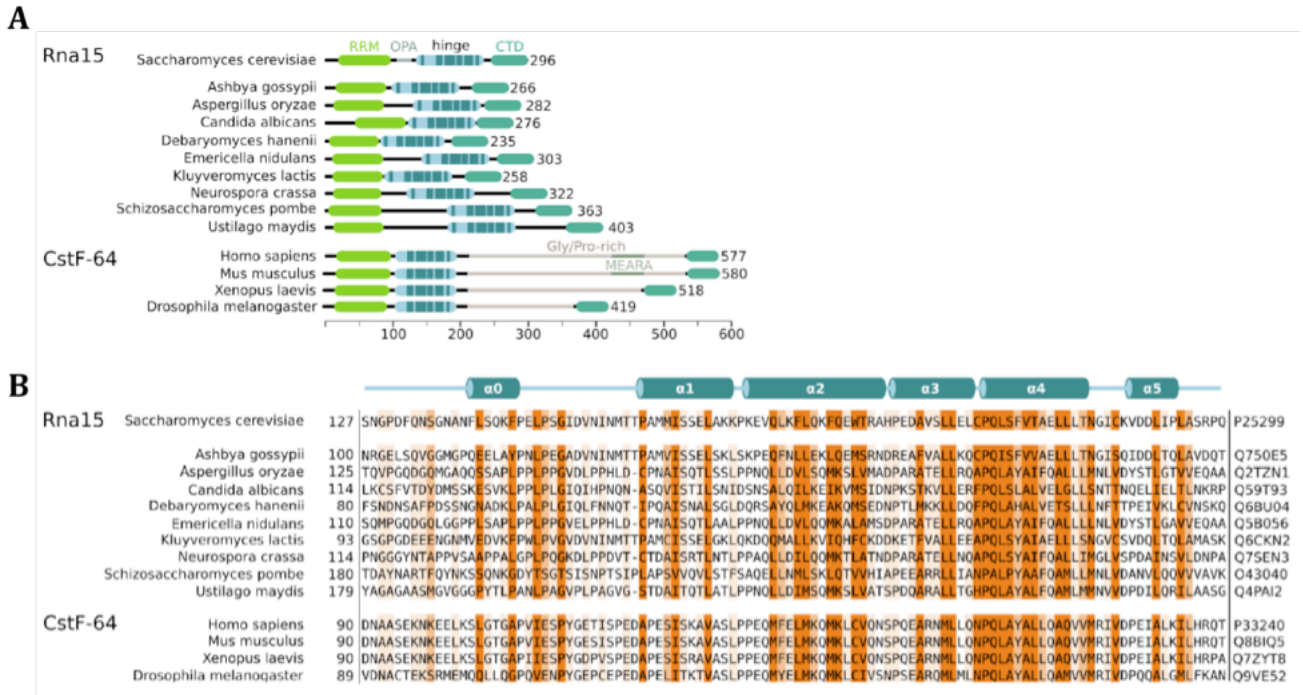
Nous avons observé que le domaine « hinge » est moins conservé autour de l'hélice  $\alpha 0$  correspondant aux résidus situés sur la zone N-terminale (approximativement les acides aminés du 127 au 145). Les expériences de RMN présentées antérieurement sur le complexe H<sub>6</sub>Rna15 (127-242) / StpRna14p (626-677), ont montré que cette région de la protéine était très flexible et n'était pas repliée (fig. 64). De plus, elle présente un degré important de résidus manifestement exposés au solvant (fig. 71). Pour mieux délimiter la région « hinge », le Dr. C. Mackereth a analysé par spectroscopie RMN la dénaturation de résidus induite par un **gradient d'urée**. Ces expériences



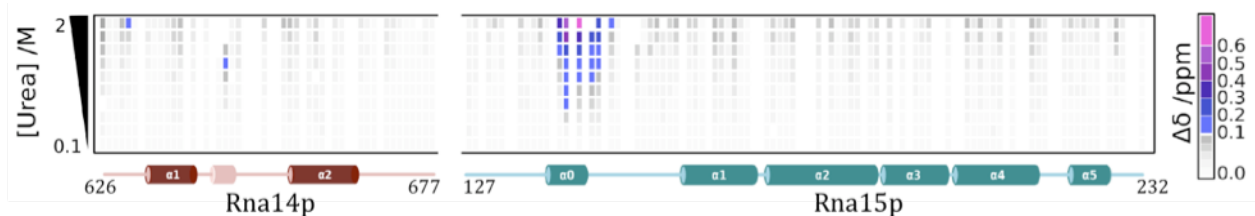
**Figure 70.** Domaine « hinge » de Rna15p (de l'hélice  $\alpha 1$  à la  $\alpha 5$ ). Le domaine « monkeytail » est représenté en gris.



permettent de localiser les zones moins stables d'une protéine. Les résultats montrent qu'à une basse concentration d'urée, seulement la zone correspondant à l'hélice  $\alpha 0$  est perturbée (fig. 72) indiquant sa plus faible stabilité. De cette façon, le « hinge » serait vraisemblablement délimité par les résidus 145 et 228. Cependant, les tests effectués à partir de la construction **H<sub>0</sub>Rna15p (138-232)** n'ont pas permis la surexpression de cette région chez la bactérie. En conséquence, la région initiale du « hinge », incluant approximativement les résidus 127 à 145, est importante pour l'expression du complexe, mais pourrait être dispensable pour l'interaction avec Rna14p.



**Figure 71. A)** Schéma des domaines présents dans Rna15 et CstF-64. Le domaine « hinge » est retrouvé par homologie de séquence dans tous les homologues analysés. **B)** Détail du domaine « hinge ». La structure secondaire est indiquée en dessus des séquences. Les niveaux d'exposition au solvant sont indiqués pour chaque résidu par un gradient orange, les résidus les plus exposés étant colorés en blanc et ceux qui le sont le moins en orange (figure réalisée par C. Mackereth).



**Figure 72.** Analyses par spectroscopie RMN du dépliement des résidus induit par un gradient d'urée. Chaque colonne du diagramme correspond à un acide aminé. La couleur des rectangles montre les variations de déplacement chimique ( $\Delta\delta$ ) de chaque résidu à une concentration d'urée déterminée. Ce déplacement est plus important pour les résidus colorés en rose et moins pour ceux qui le sont en blanc. Il est observé qu'à basse concentration d'urée, la zone correspondant à l'hélice  $\alpha 0$  est perturbée.

- **La structure de H<sub>6</sub>Rna15p (127-232) / StpRna14p (626-677) est conservé dans le complexe formé pour les protéines entières**

La comparaison des spectres <sup>1</sup>H<sup>15</sup>N-TROSY obtenus pour l'hétérodimère minimal et ceux acquis pour la protéine Rna15 entière en complexe avec StpRna14p (626-677), montre une ressemblance significative. Ceci suggère que le même type d'organisation « monkeytail » / « hinge » est présent dans le contexte d'un complexe plus grand.

Le spectre du complexe H<sub>6</sub>Rna15p / StpRna14p (626-677) a été également comparé à celui obtenu pour le domaine RRM de Rna15p isolé (H<sub>6</sub>Rna15p (1-104)). Les déplacements chimiques des résidus correspondant au RRM sont les mêmes dans le cas du RRM seul ou dans le cas du complexe avec Rna15p entière. Ceci signifie que la structure du domaine RRM ne varie pas dans le contexte du complexe Rna15p/Rna14p.

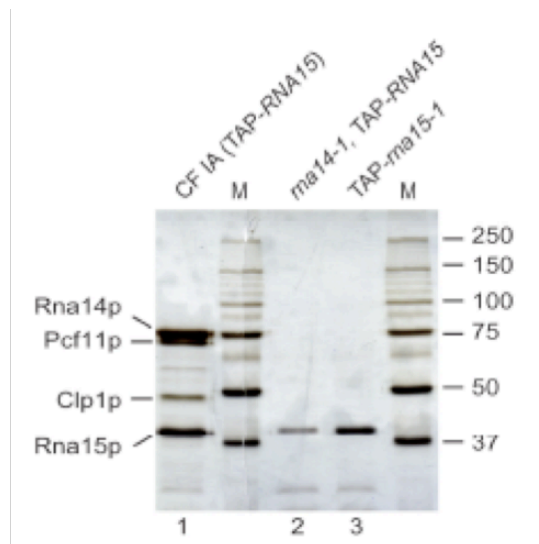
Le Dr. Mackereth a réalisé le même type d'expériences pour le complexe formé par les deux protéines entières. Ce complexe est supposé former un tétramère avec deux copies de chaque protéine. La taille importante du complexe rend l'interprétation des spectres difficiles, de nombreux pics étant superposés. Étonnement, des pics correspondant au motif RRM de Rna15p ont été facilement repérés, indiquant sûrement sa mobilité indépendante par rapport au complexe entier. De plus, d'autres pics ont été observés et, grâce au déplacement chimique particulier des tryptophanes, certains de ces pics ont pu être attribués dans la séquence de Rna15p à W187 et W282. Nous pouvons donc imaginer que ces autres pics correspondent à la région C-terminale de Rna15p. Ainsi, cette région, dont la structure est connue pour son homologue CstF-64 (Qu et al. 2007), aurait aussi une mobilité indépendante du reste du complexe (*voir annexe 3 pour consulter les spectres*).

**e.) L'association « hinge » / « monkeytail » est nécessaire pour l'assemblage de Rna15p au complexe CFIA**

Des mutants de Rna14p et de Rna15p présentant un phénotype thermosensible ont été identifiés (Bloch et al. 1978). Ces mutations ont un effet sur la maturation 3' des ARNm. Elles provoquent l'inhibition du clivage (Minvielle-Sebastia et al. 1994), la réduction de la stabilité de l'ARNm et de la longueur de la queue poly(A) (Minvielle-Sebastia et al. 1991) et des désordres dans la terminaison de la transcription (Torchet et al. 2002).

Les allèles mutants les plus utilisés et les mieux caractérisés, *rna14-1* et *rna15-1*, contiennent des altérations localisées dans les régions nécessaires pour la formation de l'hétérodimère. La mutation *rna14-1*, engendre une erreur dans le cadre de lecture qui génère la substitution des 16 derniers résidus de la protéine (LNDQVEIPTVESTKG) en seulement 2 résidus (FK) (Rouillard et al. 2000). Cette modification affecte l'hélice  $\alpha$ 2 du domaine « monkeytail ». D'un autre côté, la protéine transcrite à partir de *rna15-1*, contient une mutation simple, où l'acide aminé Leu 214 est substitué par une Pro (**L214P**) (Qu et al. 2007). Cette mutation est située dans la région C-terminale du domaine « hinge », au niveau de l'hélice  $\alpha$ 4.

A partir de la structure, nous pouvons présumer que ces mutations déstabilisent l'hétérodimère minimal. Sur la base de ces résultats, nous suggérons que le même effet est produit pour les protéines entières. Si ceci est exact, les mutations mentionnées devraient provoquer la dissociation de Rna15p du sous-complexe Rna14p/Rna15p et probablement du complexe CF IA. Dans le laboratoire, L. Minvielle-Sebastia a effectué des expériences à partir du complexe CF IA en utilisant la technique du TAP (voir Matériels et Méthodes et la section B.2. de ce chapitre : Reconstitution du complexe CF IA). Cette méthode permet la récupération et la purification du complexe à partir des cellules de *S. cerevisiae*. Des souches contenant TAP-Rna15p/*rna14-1* et TAP-*rna15-1* ont été élaborées. Ces cellules sont thermosensibles, leur croissance étant déficiente à 37°C et normale à 24°C. La purification du complexe quand une des deux sous-unités est mutée, a occasionné l'obtention de Rna15p (ou *rna15-1*) isolée (fig. 73). Ces résultats indiquent que la sous-unité Rna15p est incapable de s'incorporer à CF IA lorsque une mutation est présente dans les régions d'interaction avec Rna14p.

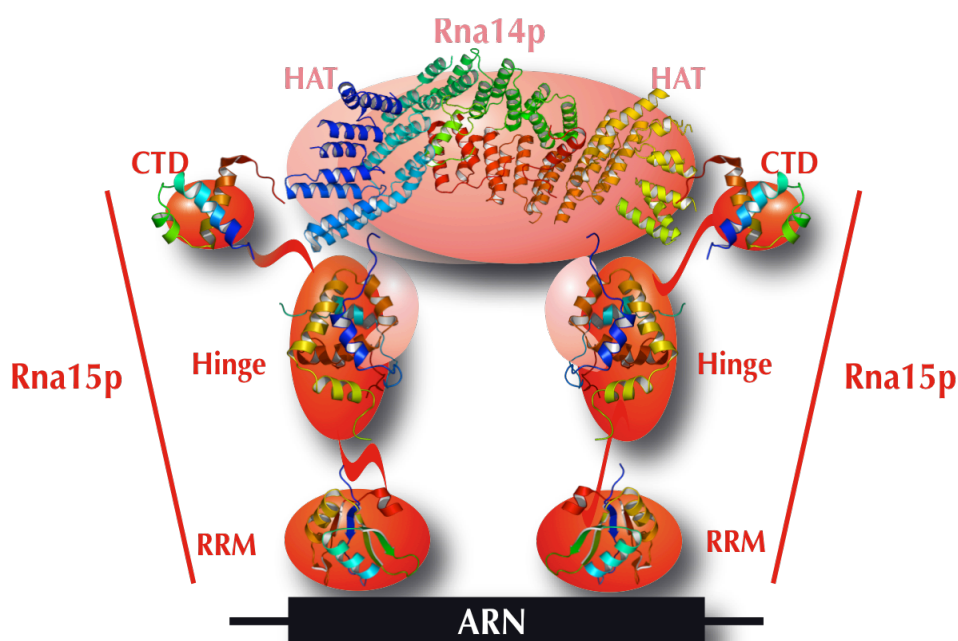


**Figure 73.** Effet des mutations *rna14-1* et *rna15-1* sur l'assemblage du complexe CF IA.  
Le puits 1 correspond au complexe CF IA intact. Dans les puits 2 et 3 seulement la protéine Rna15p est récupérée après purification.



### f.) Conclusions et perspectives pour l'interaction Rna15p/Rna14p

L'association entre les protéines Rna14p/CstF-77 et Rna15p/CstF-64 a été montrée auparavant, mais les détails moléculaires sur la formation du complexe restaient inconnus. Les travaux réalisés au cours de cette thèse, ont permis d'**identifier les domaines** de Rna15p et de Rna14p essentiels pour l'interaction. Une combinaison des techniques de biologie moléculaire, biochimie et spectroscopie RMN a mis en évidence la participation des régions comprenant le domaine « hinge » de Rna15p, du résidu 127 au 232, ainsi que la région C-terminale de Rna14p, de l'acide aminé 626 au 677, dans la formation de l'hétérodimère. Un modèle d'organisation de ces deux sous-unités de CF IA est proposé sur la *figure 74*.



**Figure 74.** Modèle de l'organisation de l'hétérodimère Rna14p/Rna15p. Pour les domaines dont la structure est connue pour Rna14p, Rna15p ou leurs homologues CstF-77 et CstF-64, celle-ci a été incluse dans le dessin. (Pancevac et al. 2010), (Legrand et al. 2007), (Qu et al. 2007), (Moreno-Morcillo et al., *Molecular Cell*, en préparation).

La détermination de la **structure du complexe** a été effectuée à partir des expériences de RMN. Cette structure montre l'organisation du complexe formant une entité unique où les deux protéines sont fortement associées. La région de Rna14p participant dans l'interaction a reçu le nom de domaine « **monkeytail** » en raison de sa manière d'envelopper la zone « hinge » de Rna15p. Le domaine « **hinge** » a été, comme supposé, celui impliqué dans l'interaction. Nous avons montré l'organisation de cette région en 5 hélices, plus une sixième dans la zone N-terminale chargée de stabiliser le complexe et permettant son expression chez *E. coli*. La présence de ces deux domaines chez leurs homologues métazoaires, CstF-77 et CstF-64, suggère la conservation de ce type

d'interaction à travers les espèces. De plus, la ressemblance entre les spectres RMN obtenus pour le complexe Rna15p (127-232)/ Rna14p (626-677) et ceux obtenus pour le complexe CstF-64 (1-211)/ CstF-77 (626-717) est un indice d'un repliement similaire pour les deux complexes (*pour consulter les spectres voir annexe 4*).

Les détails moléculaires de l'interaction Rna14p/Rna15p, indiquent que les protéines transcrites à partir des allèles *rna14-1* et *rna15-1* déstabilisent la formation de l'hétérodimère. A partir des expériences de TAP-tagging, nous avons mis en évidence la perte d'association de Rna15p au complexe CF IA lorsque une de ces mutations est présente. La poursuite de ces travaux devrait éclaircir différents aspects. L'absence d'autres sous-unités à partir du TAP-*rna15-1* indique que, soit la protéine mutée n'interagit pas avec Pcf11p/Clp1p, soit la disruption de l'hétérodimère Rna15p/Rna14p implique la rupture de l'interaction avec le sous-complexe Pcf11p/Clp1p.

D'un autre côté, nous avons montré que la mutation *rna14p-1* inhibe l'association avec Rna15p parce qu'elle empêche la co-purification des deux protéines. Nonobstant, nous ignorons si Rna14p mutée est capable d'établir un contact avec le reste des composants de CF IA. La préparation de souches de levure contenant un TAP-Rna14p ou un TAP-*rna14-1* apporterait ces informations. En outre, la réalisation d'autres expériences '*in vitro*' serait nécessaire pour approfondir sur l'implication de *rna15-1* et/ou *rna14-1* dans la perte d'interaction avec Pcf11p. De plus, nous voudrions examiner à nouveau les effets de ces mutations dans la maturation 3' des ARNms afin d'interpréter les résultats depuis cette nouvelle perspective.

Par ailleurs, il a été montré que Rna14p/CstF-77 est un **dimère** en solution (Bai et al. 2007) (Legrand et al. 2007) (Noble et al. 2004). L'analyse de l'interaction de ce dimère avec le domaine « hinge » de Rna15p peut aider à la compréhension de l'organisation générale du complexe. En ajoutant le domaine **RRM** au complexe Rna15p (127-232) / Rna14p (626-677), il serait possible d'étudier l'implication de cette interaction dans l'association avec l'ARN au niveau de sa spécificité et/ou de son affinité.

Des nouvelles perspectives pour ces travaux ont été suscitées quand l'existence d'une possible compétition entre la **symplekine** et **CstF-77** pour l'interaction avec le domaine « hinge » de CstF-64 a été décrite (Hatton et al. 2000). Chez la levure, la co-purification de Rna15p avec l'homologue de la symplekine, **Pta1p**, a été mise en évidence (Gavin et al. 2002). Ces résultats laissent imaginer l'existence d'une interaction entre Rna15p et Pta1p apportant un nouvel élément commun aux mammifères et levures. La co-expression de Pta1p et Rna15p (127-232), ou d'une autre région parmi les nombreuses constructions réalisées, peut être envisagée à partir des outils disponibles au laboratoire. De plus, lors d'une éventuelle résolution de la structure du complexe entre les deux protéines, la connaissance de l'architecture du « hinge » faciliterait l'interprétation des données.

---

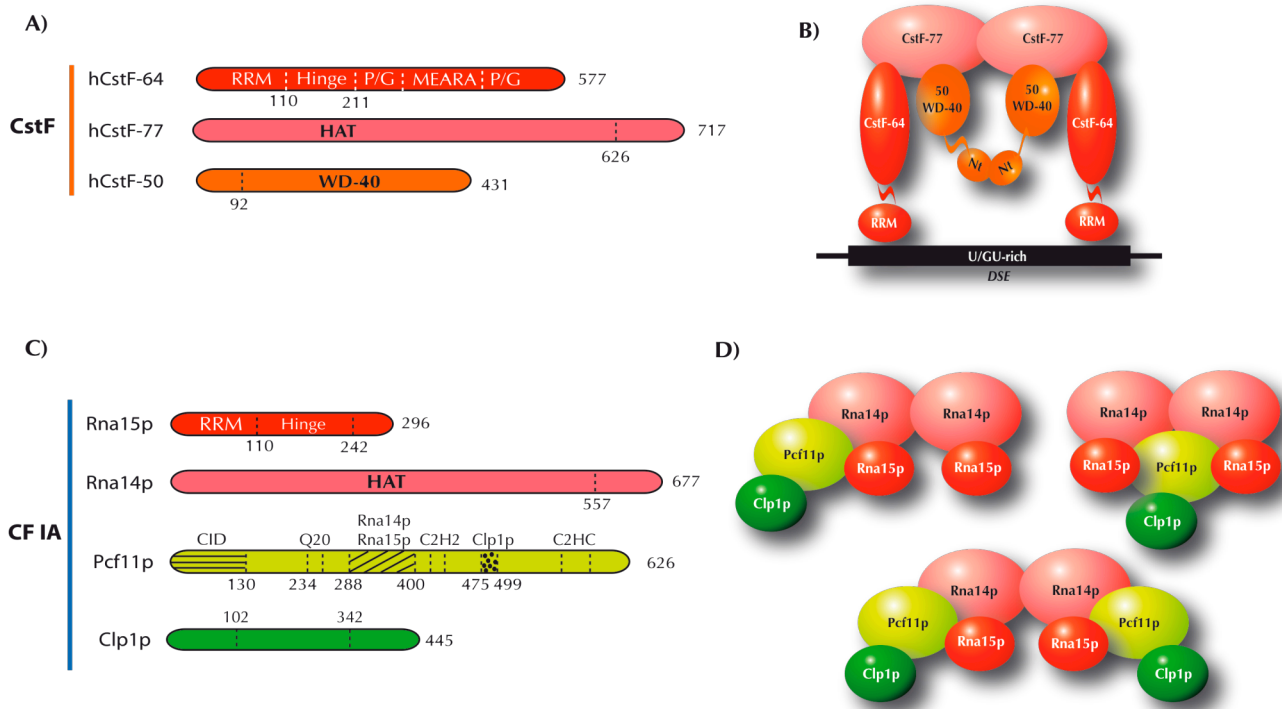
En outre, le domaine « hinge » est aussi requis pour l'internalisation de la protéine dans le noyau, suggérant que son interaction avec CstF-77/Rna14p est importante pour la **localisation nucléaire** (Hockert et al. 2010). En effet, CstF-77 tout comme son homologue chez la levure, possèdent un domaine NLS (Nuclear Localization Signal). Des expériences de localisation cellulaire à partir des protéines Rna15p et/ou Rna14p tronquées ou contenant des mutations ponctuelles dans les régions d'interaction, pourraient être effectuées. De cette façon nous pourrions étudier la participation de ces régions dans ce phénomène.

D'autres travaux peuvent être envisagés à partir des résultats obtenus. Il a été montré que la protéine **Pti1p** fait partie du complexe CPF associé au **snoRNA** (Nedea et al. 2003). Cette protéine n'est pas indispensable pour la formation de la queue poly(A), mais elle intervient dans la sélection du site de clivage et dans la longueur de la queue (Skaar and Greenleaf 2002). Pti1p partage avec Rna15p/CstF-64 le domaine « hinge », le RRM et sa région C-terminale. Son interaction avec Rna14p et Pta1p a aussi été démontrée la rapprochant des caractéristiques de Rna15p. Il a été proposé que cette protéine puisse former un complexe **CF IA alternatif** où Rna15p est substituée par Pti1p (Skaar and Greenleaf 2002). L'étude de l'association du domaine « hinge » de Pti1p ou de la protéine entière avec Rna14p donnerait les bases de cette interaction potentielle. Une preuve appuyant l'hypothèse décrite serait apportée si la région de Rna14p impliquée coïncidait avec celle utilisée pour la liaison avec Rna15p. Dans ce cas, d'autres tests doivent être réalisés pour étudier les effets produits par la formation d'un complexe CF IA contenant Pti1p.

## 2. Reconstitution du complexe CF IA

Nous commençons à comprendre maintenant l'organisation du complexe **CstF**. Un modèle d'assemblage a été proposé dans le *Chapitre II*. En effet, à partir des résultats obtenus dans notre laboratoire pour CstF-77 et CstF-50, nous pouvons affirmer que deux copies de chacune de ces protéines sont présentes dans la formation du complexe (*fig. 75 A et B*).

Cependant, l'organisation générale de son complexe homologue chez la levure, **CF IA**, reste inconnue. La formation d'un **hétérotétramère** compatible avec la présence de 2 copies de **Rna14p** et de **Rna15p** a été mise en évidence par microscopie électronique et ultracentrifugation analytique (Noble et al. 2004). D'un autre côté, les résultats montrés dans la section précédente ont donné les bases moléculaires de l'interaction entre ces deux sous-unités. L'interaction entre **Pcf11p** et **Clp1p** a été démontrée et la structure de Clp1p en complexe avec une région de Pcf11p a été résolue (Noble et al. 2007). Les quatre protéines co-purifient (Minvielle-Sebastia et al. 1997) mais, comment s'assemblent-elles pour former le complexe CF IA ? Si Rna14p et Rna15p forment un hétérotétramère, est-ce que Pcf11p et Clp1p sont aussi présentes en deux copies dans le complexe ? De quelle manière les deux sous-complexes, Rna14p-Rna15p et Pcf11p-Clp1p, interagissent-ils ? (*fig. 75 C et D*).



**Figure 75.** CstF versus CF IA. A) Protéines du complexe CstF. B) Modèle proposé pour l'organisation de CstF. C) Sous-unités faisant partie du facteur CF IA. D) Trois modèles possibles de l'assemblage de CF IA chez la levure.

Pour répondre à ces questions et mieux comprendre le mécanisme de maturation 3', nous avons adopté deux stratégies différentes. D'une part, le complexe CF IA natif peut être purifié à partir d'extraits de levure pour étudier ensuite sa structure et sa fonction. D'autre part, la reconstitution du complexe peut être effectuée en partant des protéines recombinantes. Les objectifs des deux méthodes sont :

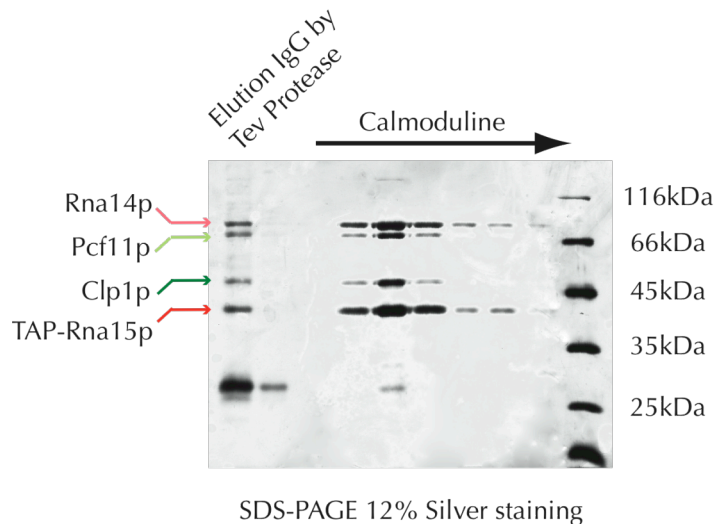
- l'identification des régions impliquées dans les interactions entre les différentes protéines,
- l'étude de l'organisation générale du complexe par microscopie électronique ou par cristallographie de rayons X.

#### **a.) Etudes préliminaires : purification du complexe natif**

Rna15p, Rna14p, Pcf11p et Clp1p co-purifient dans le complexe CF IA (Minvielle-Sebastia et al. 1997). Les expériences permettant cette co-purification ont été développées par notre ancien collaborateur et actuel membre du laboratoire, le Dr. Minvielle-Sebastia (directeur de recherche) à l'aide du Dr. Haddad. Ensuite, dans le laboratoire, les travaux réalisés par **L. Minvielle-Sebastia** et **F. Maurice** (postdoctorante) ont permis d'établir un protocole optimisé pour l'obtention efficace du complexe pur à partir de *S. cerevisiae*.

Le complexe CF IA a été produit à partir de la levure par la méthode du « **TAP** » (« Tandem Affinity Purification ») (Puig et al. 2001). La méthode du TAP permet une récupération fiable et efficace des protéines présentes en faibles quantités dans la cellule en conditions natives. Cette méthode de purification implique la fusion d'une étiquette TAP à la protéine d'intérêt, **Rna15p** dans notre cas, et l'introduction de cette construction dans l'organisme. Le « TAP tag », ou étiquette TAP, contient le domaine d'interaction avec l'IgG de la protéine A de *Staphylococcus aureus* et un peptide d'union à la calmoduline. Le « TAP tag » peut être clivé par la protéase TEV. Ainsi, le complexe CF IA peut être isolé en seulement deux étapes de purification grâce à l'utilisation de l'étiquette TAP fusionnée à la protéine Rna15p et à l'association du reste des sous-unités à celle-ci (*voir aussi Matériels et Méthodes*) (fig. 76).

Le complexe CF IA natif obtenu par cette méthode était pur et des expériences de microscopie électronique ont été envisagées. Cependant, la difficulté pour atteindre les quantités de protéine nécessaires pour la cristallisation du complexe nous a incité à adopter une nouvelle stratégie, la production du complexe chez *E. coli* utilisant des protéines recombinantes.



**Figure 76.** Purification par la méthode du TAP de CF IA à partir de 12 litres de culture de levure. Le gel est coloré au nitrate d'argent. Nous pouvons observer l'éluion avec la TEV et les différentes fractions d'éluion obtenues après la deuxième étape de purification. La pureté du complexe peut être remarquée sur le gel. Ces travaux ont été réalisés par F. Maurice et L. Minvielle-Sebastia.

#### **b.) Reconstitution du complexe CF IA : tests d'expression**

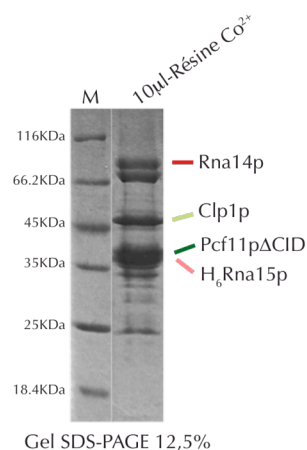
Le but est d'obtenir le complexe CF IA en quantité suffisante pour la réalisation des essais de cristallisation et éventuellement de microscopie électronique. Pour cela, nous pouvons produire le complexe à partir des quatre protéines recombinantes. Des nombreux tests ont été effectués au laboratoire afin de trouver les conditions optimales pour la **co-expression** des quatre sous-unités de CF IA. Pour chaque test, seulement une des protéines, Rna15p, Rna14p, Pcf11p ou Clp1p, contient une étiquette de 6 His. Différents plasmides contenant l'ADN codant pour ces protéines sont introduits dans *E. coli* et les bactéries co-transformées sont sélectionnées par une combinaison adéquate d'antibiotiques dépendante des vecteurs utilisés. Par la technique de « pull-down » et à l'aide d'une résine de  $\text{Co}^{2+}$ , la protéine contenant le « tag » d'histidines est récupérée et ses partenaires co-éluent avec elle. Après plusieurs lavages, la plupart des contaminants sont éliminés et l'on peut examiner, sur un SDS-PAGE coloré au bleu de Coomassie, si le complexe a été formé.

La production du complexe '*in vitro*' à partir des **protéines recombinantes** a été décisive pour l'obtention des quantités nécessaires pour la suite des expériences, mais elle confère également un autre avantage. L'utilisation de cette méthode permet en effet de varier plus facilement les composants de CF IA. Ainsi, nous pouvons imaginer la formation d'un complexe qui ne contient pas le motif RRM de Rna15p ou, comme il a été le cas, un complexe où le CID, le domaine d'interaction avec le CTD de l'ARN pol II, de Pcf11p est absent. De ce fait, nous pourrions comparer un CF IA où toutes les protéines sont complètes, avec un autre dans lequel l'un des domaines ou l'une des sous-unités est absent. De cette manière, l'utilisation de protéines

recombinantes donne la possibilité de trouver la combinaison des sous-unités et des domaines permettant une surexpression plus importante. La variation des composants de CF IA peut consister également à l'élaboration des mutations ponctuelles. Ces mutants peuvent être employés pour l'étude des effets des résidus mutés dans la formation des interactions entre les différentes sous-unités.

Des nombreux clones ont été créés pour la réalisation des essais. L'étiquette  $H_6$  a été placée finalement sur la protéine Rna15p, la sous-unité capable de récupérer le complexe CF IA complet 'in vivo'. L'utilisation de **plasmides bicistroniques** contenant les gènes codants pour deux protéines a été décisive. Pour la reconstitution finale du complexe 'in vitro', nous avons utilisé un vecteur où l'ADN codant pour la protéine **Rna15p** avec l'étiquette mentionnée en position N-terminale, était suivi de l'ADN codant pour **Rna14p**. Un vecteur résistant à la kanamycine a été employé pour l'expression de **Clp1p**. La surexpression la plus favorable a été atteinte avec la sous-unité **Pcf11p** sans son domaine CID, c'est-à-dire, avec une version protéique comprenant les résidus 261 à 626. Cette région de Pcf11p a été intégrée dans un vecteur résistant à la streptomycine. En résumé, pour la reconstitution de CF IA, il a été nécessaire de co-transformer la souche BL21 (DE3) avec les plasmides  **$H_6$ Rna15p\_Rna14p (Amp<sup>R</sup>)**, **Ka-Clp1p** et **Stp-Pcf11p $\Delta$ CID**.

Sur la *figure 77*, un gel SDS-PAGE montre les résultats obtenus à partir du test d'expression. Nous avons ainsi démontré que la reconstitution du complexe est possible et la production de ce complexe à grande échelle, pour la réalisation des tests de cristallisation et des expériences en microscopie électronique, a pu être envisagée.



**Figure 77.** Test d'expression de CF IA $\Delta$ CID.

*$H_6$ Rna15p et Pcf11p $\Delta$ CID ont des tailles proches. Sur les gels ultérieurs nous avons pu apprécier deux bandes bien différenciées.*

### c.) Pcf11p interagit avec toutes les sous-unités du complexe CF IA ?

Les détails des interactions entre les différentes protéines ont été également l'objet de nos études. Les associations Rna14p-Rna15p et Pcf11p-Clp1p sont bien connues. Par contre, l'assemblage entre ces deux sous-complexes n'est pas clair.

En 1997, le groupe de F. Lacroute (Amrani et al. 1997) a obtenu, par des expériences de **double hybride**, des résultats positifs pour l'interaction entre Pcf11p et Rna14p et entre Pcf11p et Rna15p. Ils ont aussi défini la région de Pcf11p de l'acide aminé **288 au 400** comme étant suffisante pour ces interactions. Cependant, ils ont proposé deux hypothèses à partir de ces résultats, soit Pcf11p interagit directement avec les deux sous-unités, soit elle s'associe seulement à une des deux et le signal positif obtenu pour l'autre provient de l'étroite relation existant entre Rna14p et Rna15p.

D'un autre côté, dans la littérature, des expériences de « **pull-down** » analysées par « western blot » à partir de la protéine recombinante **GST-Pcf11p**, indiquent qu'une interaction existe entre cette sous-unité et les trois autres protéines du complexe (Gross and Moore 2001).

Dans le laboratoire, nous avons mis en évidence une interaction directe entre Rna15p et Pcf11p (ou Pcf11p $\Delta$ CID) et la formation d'un complexe '*in vitro*' contenant ces deux protéines et Clp1p. Selon nos résultats, Rna15p ne nécessite pas la présence de Rna14p pour interagir avec Pcf11p. Des expériences de « pull down » ont été effectuées pour localiser la région de Rna15p nécessaire pour l'interaction avec Pcf11p. Les résultats préliminaires obtenus indiquent que cette région est située dans la partie C-terminale de la protéine Rna15p, du résidu 242 au 296. Ces expériences corroboreraient les résultats obtenus par C. Moore et ses collaborateurs (Qu et al. 2007), qui démontrent l'existence d'une interaction entre MBP (myelin basic protein)-Pcf11p et la région C-terminale de Rna15p.

Une interaction directe entre Rna14p et Pcf11p a été montrée par des expériences de « pull-down » couplées à des « western blot », à partir de la protéine recombinante GST-Pcf11p (Gross and Moore 2001). Dans le laboratoire, la démonstration de cette association a été moins évidente que pour Pcf11p/Rna15p. En effet, une faible bande au niveau de la taille de Pcf11p apparaît sur le gel coloré au Bleu Coomassie lorsque nous effectuons un « pull-down » avec la protéine Rna14p marquée avec 6 His. Cependant, d'autres expériences, comme la réalisation d'un western blot avec des anticorps spécifiques, devraient être effectuées pour confirmer la correspondance de cette bande avec la protéine Pcf11p.

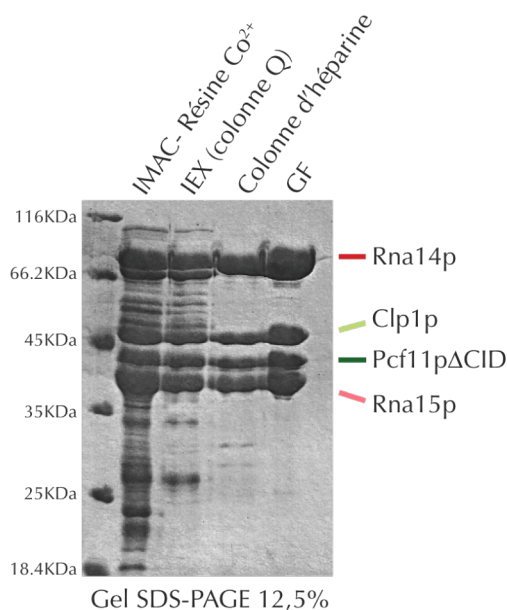
#### **d.) Purification du complexe CF IA à partir des protéines recombinantes**

Quatre étapes de purification sont nécessaires pour l'obtention d'un complexe suffisamment pur pour la réalisation d'essais de cristallisation et/ou de microscopie électronique. Pour la purification nous sommes partis de 6 litres de culture de la souche BL21 (DE3), dans un milieu LB.

La première étape consiste en une **chromatographie d'affinité**, réalisable grâce à une étiquette H<sub>6</sub> située à l'extrémité N-terminale de Rna15p. Après une colonne **d'échange cationique** et une colonne **héparine** (HiTrap™ Q Sepharose HP et HiTrap™ Heparin HP, respectivement) le complexe est concentré. A cette étape une bande est observée en dessous de celle correspondant à Rna14p. Cette protéine paraissant interagir avec le complexe, la bande a été analysée par spectrométrie de masse et elle correspond à la protéine bifonctionnelle résistante à la polymyxine, une protéine naturelle de la bactérie *E. coli*. Finalement, une étape chromatographique **d'exclusion moléculaire** permet l'obtention d'un complexe pur à plus de 95% (aucune autre bande correspondant à un contaminant n'est observée sur un gel SDS-PA) (*fig. 78*) (*voir aussi la figure 95 du Chapitre VI : Matériels et Méthodes*). Un élément à remarquer dans la purification de CF IA est la nécessité d'une concentration saline de **220mM** de NaCl pour maintenir la stabilité du complexe. La solution finale contient le complexe protéique avec l'étiquette de 6 His dans un



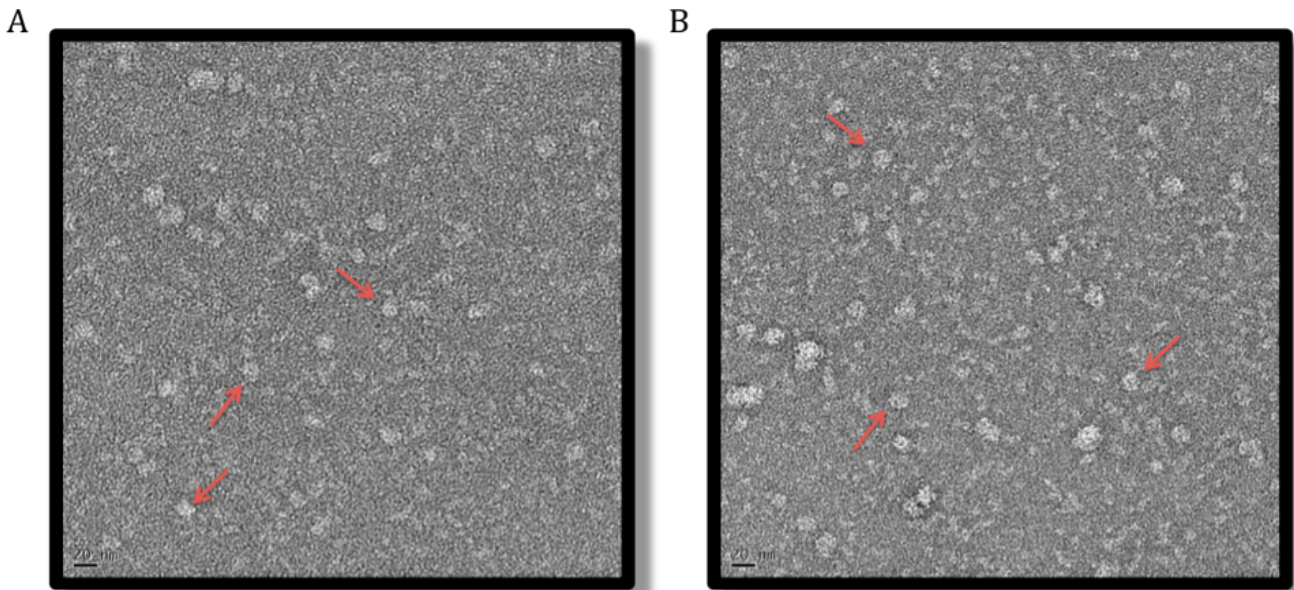
tampon comportant : 50mM de Tris HCl pH 7.5, 220mM de NaCl et, seulement dans certains cas, 1mM de DTT.



**Figure 78.** Purification de CF IAΔCID. 10μl de solution obtenue à partir de chaque étape de purification ont été chargés sur un gel 12,5% SDS-PAGE. IMAC : « immobilized metal affinity chromatography ». IEX : « Ion exchange chromatography ». GF : « gel filtration ».

Le rendement maximal du complexe obtenu a été de 4mg à partir de 6 litres de culture. Ceci nous a permis d'effectuer des **essais de cristallisation** à des concentrations situées entre 3mg/ml et 10mg/ml. Malgré l'utilisation de plusieurs kits commerciaux, aucun cristal correspondant au complexe n'a été formé. Cependant, l'aspect des gouttes de cristallisation nous permet de rester optimistes et la poursuite de ces travaux est en cours au laboratoire.

D'un autre côté, des expériences de **microscopie électronique** sont en cours en collaboration avec **Patrick Bron**, du Centre de Biochimie Structurale (Montpellier). Les premières images obtenues sont présentées sur la *figure 79*. D'autres tests pour trouver les conditions idéales de concentration protéique, de pureté et de tampon sont en train d'être effectués. De nouvelles préparations seront envoyées à nos collaborateurs prochainement en espérant qu'elles soient convenables pour la résolution de la structure du complexe par cette technique.



**Figure 79.** Images de microscopie électronique obtenues pour le complexe CF IA ( $\Delta$ CID). A) Image acquise par Patrick Bron (CBS, Montpellier). B) Image obtenue par F. Maurice (cours EMBO). Les flèches indiquent les formes correspondant probablement au complexe CF IA.

**e.) Conclusion et perspectives pour la reconstitution ‘in vitro’ du complexe CF IA**

À la différence de CstF, le complexe CF IA est composé de 4 sous-unités, les homologues de CstF-77 et CstF-64, **Rna14p** et **Rna15p**, respectivement, et deux protéines supplémentaires, **Pcf11p** et **Clp1p** dont leurs homologues chez les mammifères, hPcf11 et hClp1, sont associées dans le complexe CF IIm.

Les techniques de microscopie électronique et d’ultracentrifugation analytique ont mis en évidence la formation d’un sous-complexe compatible avec la présence de 2 copies de Rna14p et de Rna15p (Noble et al. 2004). La structure de leurs domaines d’interaction a été résolue dans le laboratoire et présentée dans la section précédente. L’architecture du RRM de Rna15p a été définie récemment en complexe avec l’ARN (Pancevac et al. 2010) (Leeper et al. 2010). La structure du domaine CID de Pcf11p avec deux heptapeptides du CTD et l’architecture de Clp1p, avec une région de Pcf11p et une molécule d’ATP, sont également connues (Noble et al. 2005) (Noble et al. 2007). A partir des homologues, des informations structurales sur le domaine HAT de Rna14p et le domaine C-terminal de Rna15p sont aussi disponibles (Bai et al. 2007) (Legrand et al. 2007) (Qu et al. 2007). Tous ces travaux nous permettent de commencer à imaginer l’assemblage de CF IA, cependant, aucune information sur la structure générale du complexe n’a été dévoilée. Nous espérons que l’aboutissement de ces travaux révélera les bases de cet assemblage et que l’interprétation de la structure aidera à la compréhension de son rôle dans le processus de maturation 3’.

Ces travaux nous ont permis d'effectuer la reconstitution du complexe CF IA où seulement le domaine CID de Pcf11p est absent. Des essais de cristallisation de ce complexe en présence/absence d'ARN sont en cours au laboratoire et une collaboration a été établie pour la résolution de sa structure par microscopie électronique. Des expériences réalisées récemment en DLS (Dynamic Light Scattering) montrent une différence de conformation quand l'ARN est ajouté au complexe. De plus, ces essais ont permis de corroborer la nécessité d'un tampon riche en sel (220mM de NaCl) pour maintenir le complexe stable. La récente publication de la structure des domaines RRM de Hrp1p et de Rna15p en complexe avec l'ARN (Leeper et al. 2010), ouvre de nouvelles voies pour nos essais. Des nouvelles séquences d'ARN, plus en accord avec celles utilisées dans ces travaux, pourraient être envisagées.

D'autre part, des nouvelles reconstructions du complexe peuvent être réalisées grâce à l'incorporation d'autres protéines. La formation d'un complexe CF I entier, où l'on ajoute la sous-unité Hrp1p au complexe précédent, a déjà été considérée dans notre groupe. Des clonages et des tests d'expression permettant la reconstitution de l'ensemble CF IA / CF IB sont en cours.

Au même temps, nous pouvons imaginer aussi la formation du complexe CF IA sans une des sous-unités ou des domaines. L'étude comparative d'un CF IA entier avec un CF IA dans lequel l'une des protéines/régions protéiques serait absente, peut nous aider à décrire les changements de conformation qui ont lieu dans le complexe.

Ces travaux, peuvent non seulement conduire à la compréhension du mode d'organisation de CF IA, mais ils ouvrent aussi un éventail de possibilités pour la continuation du projet. Les plasmides capables de surexprimer CF IA ont été déterminés. Le protocole à suivre pour sa purification a aussi été défini. Des adaptations de cette méthode devraient être suffisantes pour l'obtention de variantes du complexe.

Finalement, les études réalisées sur la fonction de Clp1p au laboratoire apporteront aussi de la lumière à la compréhension de la formation du complexe CF IA. Son interaction avec l'ATP fait de cette protéine une kinase potentielle mais son activité reste inconnue. Il est possible que l'ATP aie un rôle dans l'interaction avec Pcf11p, dans l'assemblage de ces deux protéines avec le sous-complexe Rna14p/Rna15p ou dans l'intégration de CF IA dans la machinerie de la maturation 3', mais nous devons attendre les résultats des nouvelles expériences pour pouvoir établir une théorie valable.



## Chapitre IV : Discussion

A partir des résultats obtenus dans le cadre de cette thèse et des travaux effectués dans le domaine, la problématique peut être abordée depuis deux perspectives différentes : un aspect structural et un aspect fonctionnel. Depuis plusieurs années les études pour découvrir les mystères qui englobent le processus de maturation 3' ont ciblé l'identification des protéines participant au processus. La fonction de ces protéines est intimement liée à leur structure. Cette approche a été adoptée pour l'interprétation des résultats obtenus dans cette thèse et pour un examen plus global de la thématique.

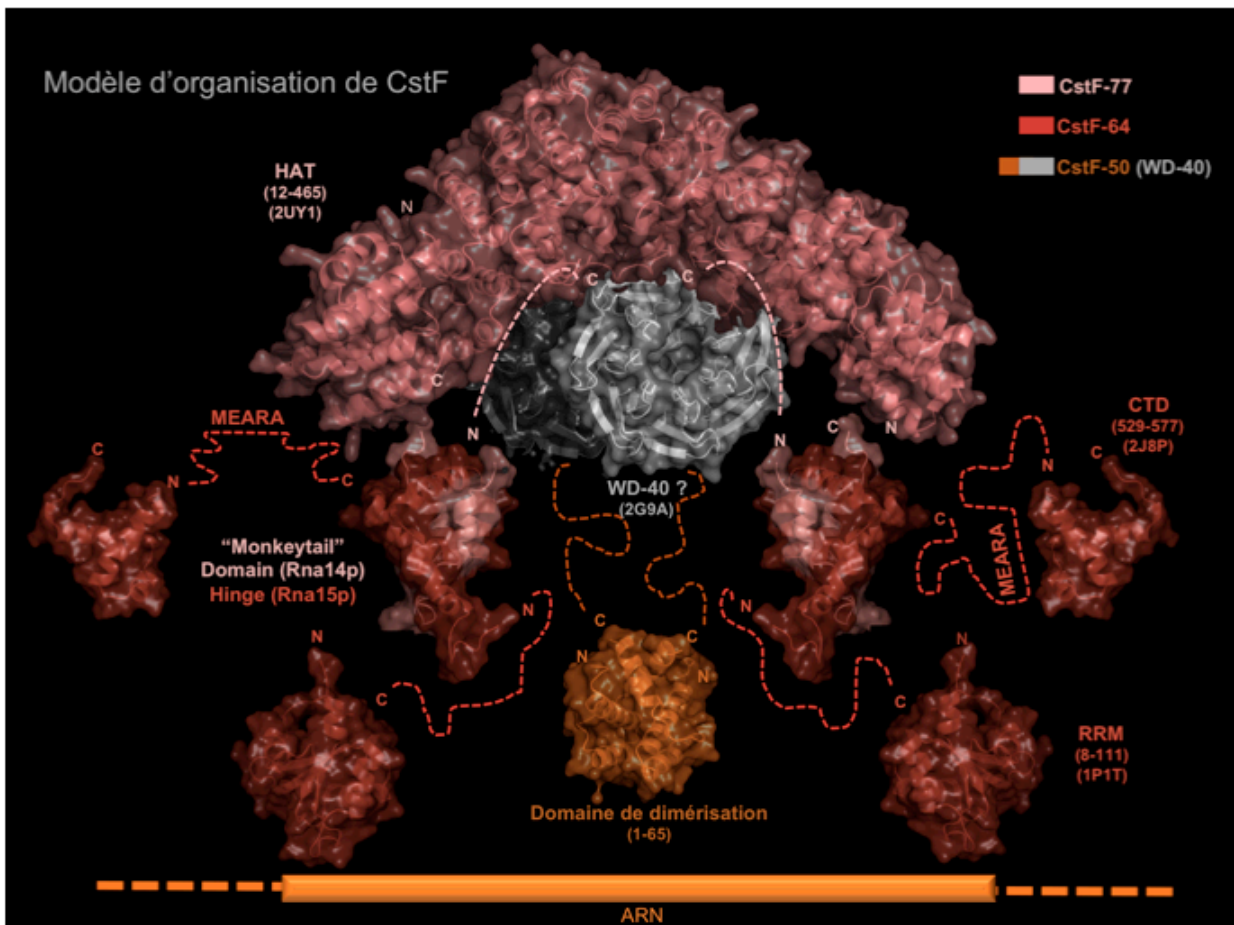
### A. Modèles structuraux des complexes CstF et CF IA

La plupart des structures des protéines impliquées dans la maturation 3' ont été obtenues durant la dernière décennie. Les sous-unités formant les complexes CstF et CF IA n'ont pas été une exception et plusieurs domaines comptent un modèle structural. Nous avons rassemblé toutes les informations pour proposer des **schémas de l'organisation de CstF et de CF IA** (*fig. 80 et fig. 82, respectivement*). Mes travaux ont contribué directement à ces modèles grâce à la résolution de la structure du domaine N-terminal de CstF-50 et des domaines d'association de Rna14p et Rna15p.

Le premier détail que nous pouvons remarquer sur les figures des modèles structuraux est la présence de **deux copies** de toutes les sous-unités intégrant le facteur CstF et d'au moins les protéines Rna14p et Rna15p de CF IA. La présence d'une seule copie des sous-unités Pcf11p et de Clp1p doit être confirmée mais des résultats préliminaires supportent cette hypothèse.

La capacité de dimérisation de **CstF-77** a été suggérée à partir des expériences de « far western » (Takagaki and Manley 2000) et ensuite validée par la résolution de sa structure cristallographique (Bai et al. 2007) (Legrand et al. 2007). Des images obtenues par microscopie électronique du complexe **Rna14p/Rna15p** ont montré la formation de particules « bacilliformes » présentant un angle de 110° (Noble et al. 2004), le même type d'angle que celui trouvé pour le coude formé par CstF-77. De plus, des expériences d'ultracentrifugation analytique ont mis en évidence la capacité du complexe Rna14p/Rna15p à former un hétérotétramère (Noble et al. 2004). Ces données, additionnées à la conservation des séquences au niveau des répétitions HAT, nous permettent de présumer que l'organisation dimérique de CstF-77 est similaire à celle de Rna14p chez la levure (*figs. 80 et 82*).

La résolution de la structure du **domaine N-terminal de CstF-50** au sein du laboratoire corrobore le modèle stœchiométrique initialement proposé (Bai et al. 2007) (*fig. 80*). Nous avons observé que le domaine N-terminal de CstF-50dm est organisé en un **homodimère** et que cette architecture est **conservée** à travers les espèces. Les résidus 1 à 65 sont responsables et suffisants pour la formation du dimère.



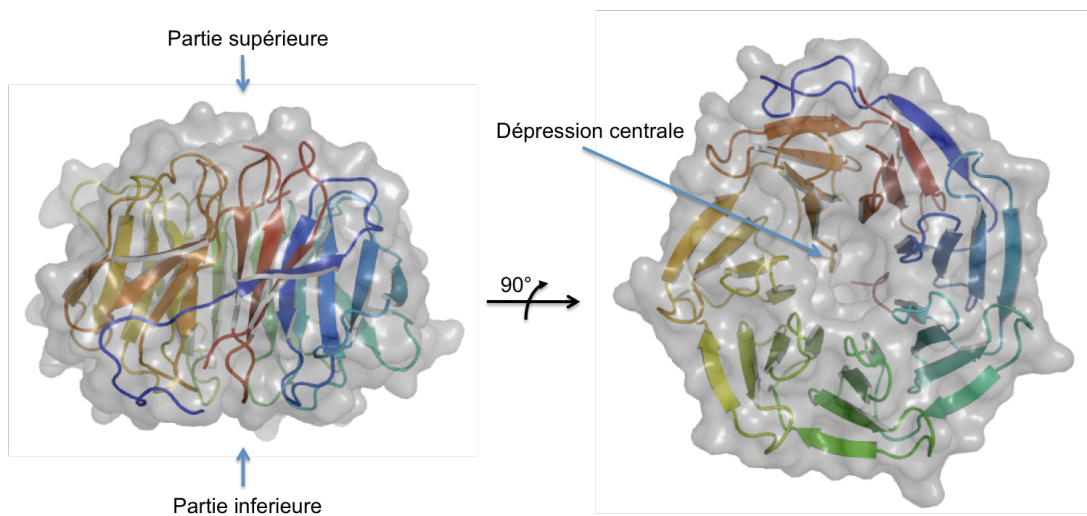
**Figure 80.** Modèle d'organisation du complexe CstF. Les structures des domaines connus sont représentées en style « cartoon » et en surface. La structure du domaine WD-40 de CstF-50 (en gris) n'a pas été résolue. Nous nous sommes servis du modèle structural d'un motif similaire appartenant à la protéine WDR5 afin de nous donner une idée plus réaliste de l'ensemble. Les pointillés indiquent des régions dont la structure n'est pas connue. Pour les protéines dont nous ne disposons pas de structure, celle des homologues leur a été attribuée. Les résidus et les codes PDB, si ils sont disponibles, ont été indiqués pour chaque domaine. Les figures ont été réalisées avec « PyMOL ».

La dimérisation de la région N-terminale de CstF-50 implique la présence de deux domaines **WD-40** dans le complexe CstF. La structure de ce domaine de la protéine CstF-50 est inconnue\*. Il a été montré que ce type de domaines présente une architecture en hélice- $\beta$ , ou «  $\beta$ -propeller » avec une dépression centrale. Le domaine WD-40 est connu pour sa capacité à interagir avec des protéines, des peptides ou des acides nucléiques. Ces interactions peuvent avoir lieu avec la partie supérieure de la dépression centrale, avec la partie inférieure ou avec les résidus qui font partie de la circonférence (Stirnimann et al. 2010) (*fig. 81*). Des expériences de « far western » ont montré que cette région de CstF-50 interagit avec CstF-77. Selon ces travaux, toutes les délétions des différentes parties ou de la totalité du domaine WD-40 empêchent l'association (Takagaki and

\* Dans la *figure 81*, une structure d'un domaine WD-40, n'appartenant pas à CstF-50, a été mise à sa place pour mieux imaginer l'ensemble du complexe.



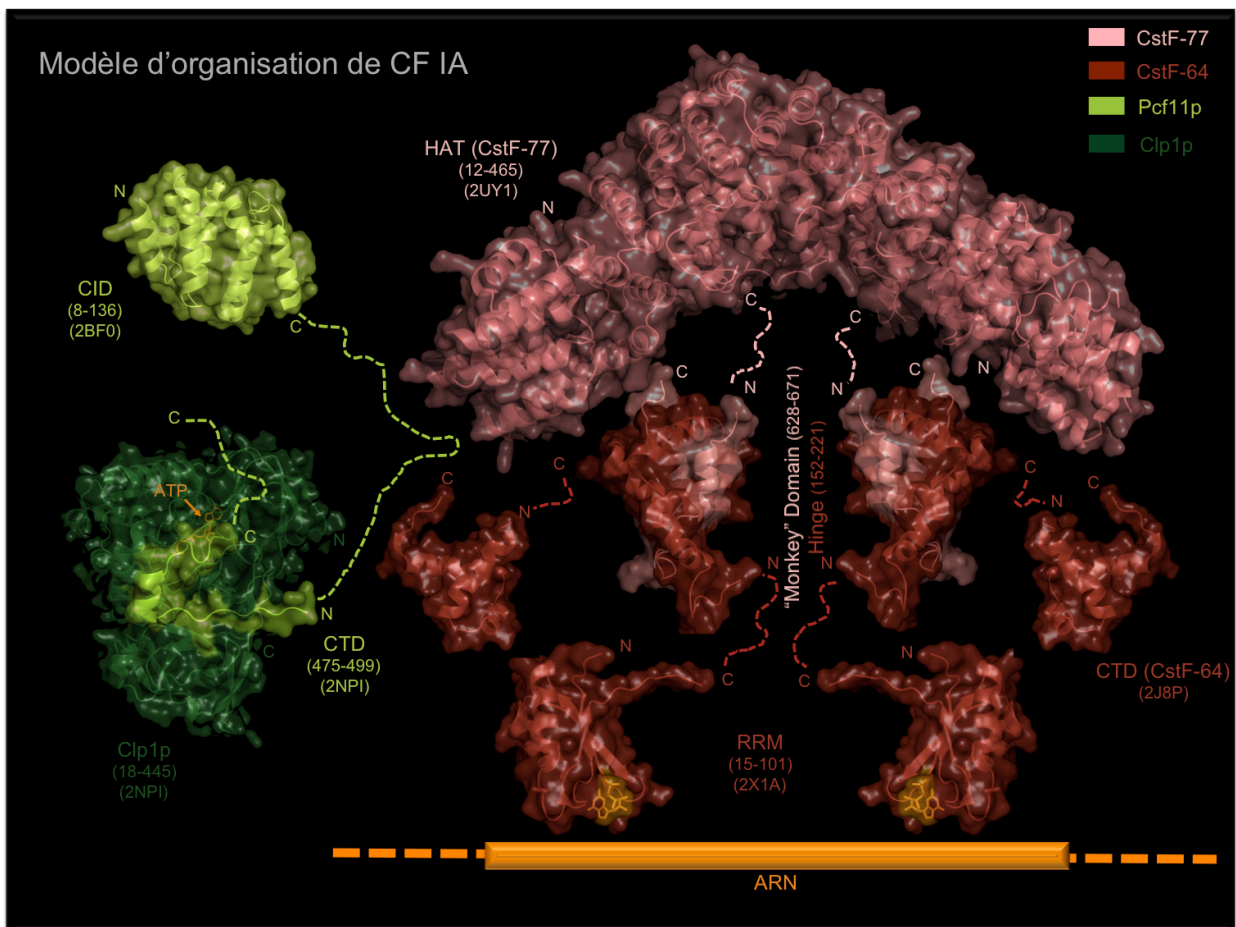
Manley 2000). Ceci suggère que chacune des délétions déstabiliserait l'architecture du domaine entier ou que CstF-50, comme la plupart des protéines de cette famille, nécessiterait la dépression formée au centre de sa structure pour l'interaction (Russell et al. 1998) (Stirnemann et al. 2010). La zone de CstF-77 impliquée est située dans sa région C-terminale. Pour que l'interaction ait lieu, au moins les deux derniers motifs HAT et la région riche en Pro doivent être présents (Takagaki and Manley 2000). Au même temps, des analyses biochimiques ont montré que le domaine HAT-C de CstF-77 est impliqué dans l'interaction avec CPSF-160 (Murthy and Manley 1995) (Bai et al. 2007). La structure de cette sous-unité de CPSF n'a pas été résolue, mais à partir de sa séquence on prévoit qu'elle contient environ 18 domaines «  $\beta$ -propeller » (16 chez son homologue de levures, Yhh1p) (Neuwald and Poleksic 2000). Des expériences de double hybride ont identifié la région de CPSF-160 incluant les résidus 400 à 973 comme celle responsable de l'association avec CstF-77 (Bai et al. 2007). Cette zone comprend la région d'interaction avec l'ARN (467-741, 5 motifs «  $\beta$ -propeller ») (Dichtl et al. 2002) et au moins 2 répétitions en plus du motif «  $\beta$ -propeller ». CstF-77 aurait donc une forte capacité à interagir avec les domaines  $\beta$ -propeller. D'autre part, CstF-77 étant un dimère, il ne serait pas surprenant que deux domaines WD-40 de CstF-50, provenant de chaque monomère, et deux des domaines «  $\beta$ -propeller » de CPSF-160 interagissent à la fois avec CstF-77.



**Figure 81.** Modèle structural typique du domaine WD-40 (ici appartenant à la protéine WDR5, code PDB : 2G29). La partie supérieure et la partie inférieure du domaine sont indiquées sur la figure à gauche. La dépression centrale est mieux aperçue lorsqu'on fait une rotation de 90° à partir de la figure située à gauche.

La protéine **CstF-64**, ne formant pas un dimère, serait également présente en deux copies. Ceci signifie que deux **RRMs** interviennent dans la reconnaissance de l'ARNm (fig. 80). Puisque Rna14p forme un dimère, cette hypothèse peut être élargie au système des levures pour la protéine Rna15p. Les structures du domaine RRM et de la région C-terminale de **CstF-64** ont été résolues (Perez Canadillas and Varani 2003; Qu et al. 2007). De plus, les travaux effectués dans cette thèse ont permis de décrire l'architecture de la région intermédiaire, le domaine « **hinge** », pour son homologue chez *S. cerevisiae*, la protéine Rna15p.

Le modèle structural de l'interaction **Rna14p et Rna15p**, obtenu à partir de nos expériences de RMN, a participé à la compréhension de l'organisation du complexe CF IA. Contrairement aux expériences réalisées pour leurs homologues humains par « far western » (Takagaki and Manley 2000), nous avons montré que la région minimale d'interaction de Rna14p comprenait les acides aminés 626 à 677. Cette région contient un domaine que nous avons appelé « **monkeytail** » et maintient une étroite association avec le domaine « **hinge** » de Rna15p. Nous n'avons pas réussi à obtenir des cristaux de l'hétérodimère CstF-77/CstF-64, néanmoins, les résidus impliqués dans l'interaction ont été identifiés et un complexe contenant ces régions, CstF-77 (626-717) / CstF-64 (1-211), a été purifié. A partir des études bioinformatiques pour la localisation des résidus hydrophobes, le domaine « monkeytail » a été repéré dans la séquence de CstF-77 des métazoaires, plus précisément, dans la région indispensable pour l'interaction avec CstF-64. Le domaine « hinge » est également présent dans la protéine CstF-64. Nous proposons que la stratégie d'association de ces deux sous-unités dans le facteur CstF soit du même type que celle observée pour leurs homologues chez la levure. La ressemblance des spectres TROSY (RMN) obtenus au laboratoire pour les sous-complexes de levure et humain seraient en appui à cette théorie.



**Figure 82.** Modèle d'organisation du complexe CF IA. Les structures des domaines connus sont représentées en style « cartoon » et en surface. Les pointillés indiquent des régions dont la structure n'est pas connue. Nous avons placé la structure des homologues quand celle de la protéine de levure correspondante n'était pas connue. Les résidus et les codes PDB, si ils sont disponibles, ont été indiqués pour chaque domaine.

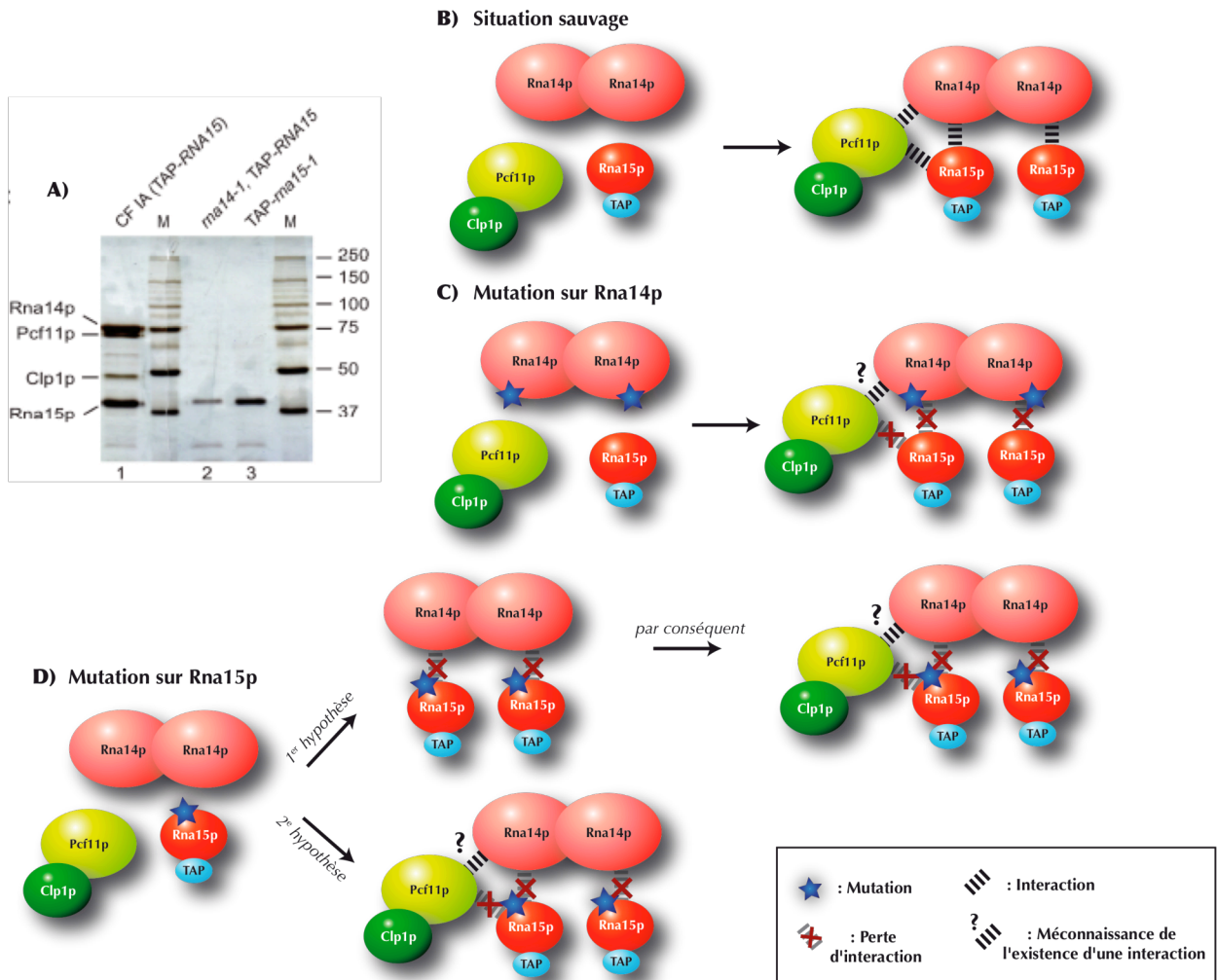


Les régions participant aux interactions dans le complexe CstF ont été globalement déterminées. Dans CF IA, l'organisation des sous-complexes Rna15p/Rna14p et Pcf11p/Clp1p a été définie, mais l'assemblage des quatre protéines n'a pas pu être caractérisé. Dans le modèle proposé pour CF IA (fig. 82), la région située entre le CID et le domaine C-terminal de Pcf11p (CTD), de structure inconnue, interagit avec la région C-terminale de Rna15p et avec une zone non-identifiée de Rna14p. Des données précédentes (Gross and Moore 2001) (Qu et al. 2007) et nos expériences de « pull down » ont permis de cartographier les régions d'interaction de Rna15p et Pcf11p. Pour Rna14p les informations sont plus confuses et des nouveaux tests doivent être effectués. En outre, dans l'hypothèse de la présence d'une seule copie de Pcf11p, il serait intéressant d'étudier si Pcf11p est capable d'interagir avec deux sous-unités de Rna15p et de Rna14p au même temps à travers des motifs qui seraient répétés dans sa séquence.

Par ailleurs, à partir de l'analyse de la structure des domaines d'interaction de Rna14p et Rna15p, nous avons constaté que les protéines transcrites à partir des allèles *rna14-1* et *rna15-1* déstabilisent la formation de l'hétérodimère. De plus, ces deux mutations impliquent la perte d'association de Rna15p au complexe CF IA (fig. 83A). Pourquoi ces mutations impliquent la séparation de Rna15p et Pcf11p ? Lorsque la mutation se trouve sur le domaine « hinge » de Rna15p (fig. 83D), non seulement la protéine n'interagit plus avec Rna14p mais l'association avec Pcf11p est aussi brisée. Ceci signifierait que la mutation **L214P** de Rna15p est capable d'inhiber l'association avec les deux protéines. La perte d'association entre Rna15p et Pcf11p peut être indirecte ou directe, ainsi, nous proposons deux hypothèses différentes :

- Perte d'interaction indirecte : la rupture de l'interaction entre Rna15p et Pcf11p peut être générée par un changement de conformation induit par l'absence d'association avec Rna14p. Nonobstant, nous avons observé que la région C-terminale de Rna15p peut interagir avec Pcf11p 'in vitro' sans requérir la présence de Rna14p. Il est possible que la formation d'un hétérodimère Rna14p/Rna15p soit indispensable pour l'assemblage de Rna15p dans un système 'in vivo', même si 'in vitro' nous sommes capables de reconstruire le complexe Rna15p/Pcf11p/Clp1p.
- Perte d'interaction directe : la deuxième hypothèse serait que le changement de conformation se produise dans la région de Rna15p nécessaire pour l'interaction avec Pcf11p. Le résidu muté est situé dans la partie C-terminale du domaine « hinge », très proche de l'extrémité C-terminale de la protéine. La mutation L214P pourrait engendrer la modification de l'architecture de la région C-terminale de Rna15p et empêcher son association avec Pcf11p.

La délétion de la région C-terminale de la protéine transcrite par l'allèle *rna14-1* inhibe l'association avec Rna15p (fig. 83C). L'effet est similaire au précédent, TAP-Rna15p étant incapable de d'interagir avec le reste des protéines du complexe. Ceci appuie notre hypothèse sur le changement de conformation de Rna15p quand l'hétérodimère avec Rna14p n'est pas formé. Cependant, nous ignorons si Rna14p mutée est capable d'établir des contacts avec le reste des composants. La création d'une souche où Rna14p contient le « tag-TAP » pourrait être envisagée.



**Figure 83.** Effet des mutations *rna15-1* et *rna14-1* dans la formation du complexe CF IA. La protéine Rna15p contient une étiquette TAP (TAP-tag). A) Gel SDS-PAGE obtenu à partir des expériences de TAP (voir Chapitre III). B) Le complexe CF IA peut être purifié à partir du TAP-tag de Rna15p. C) et D) Quand une des deux mutations est présente, seulement Rna15p est récupérée. En C) la mutation est située sur la région C-terminale de Rna14p et en D) sur le « hinge » de Rna15p. Le sous-complexe Pcf11p/Clp1p pourrait interagir avec les deux copies de Rna14p et de Rna15p.

## B. Interprétation et analyse des modèles structuraux : vers un modèle fonctionnel

La structure et la fonction d'une protéine sont intimement reliées. Les modèles structuraux présentés sont un indice du comportement des protéines qui l'intègrent. Dans cette section de la discussion, trois aspects sont analysés. Le premier point est la question que pose l'arrangement dimérique de CstF et d'au moins deux des sous-unités de CF IA, « quelle est la raison d'une telle complexité ? » D'un autre côté, nous proposons un schéma de la participation des complexes CstF et CF IA dans le processus de maturation 3' obtenu après l'analyse de nos résultats et de ce qui était connu auparavant. Finalement, le troisième aspect est focalisé sur l'interprétation dynamique du système où CstF et CF IA sont inclus.

- **Les trois protéines de CstF et au moins Rna15p et Rna14p de CF IA, sont présentes en deux copies dans la cellule. Quelle est la nécessité d'une telle complexité ?**

Plusieurs hypothèses sur l'importance de cette stœchiométrie peuvent être proposées pour les différentes sous-unités des deux complexes.

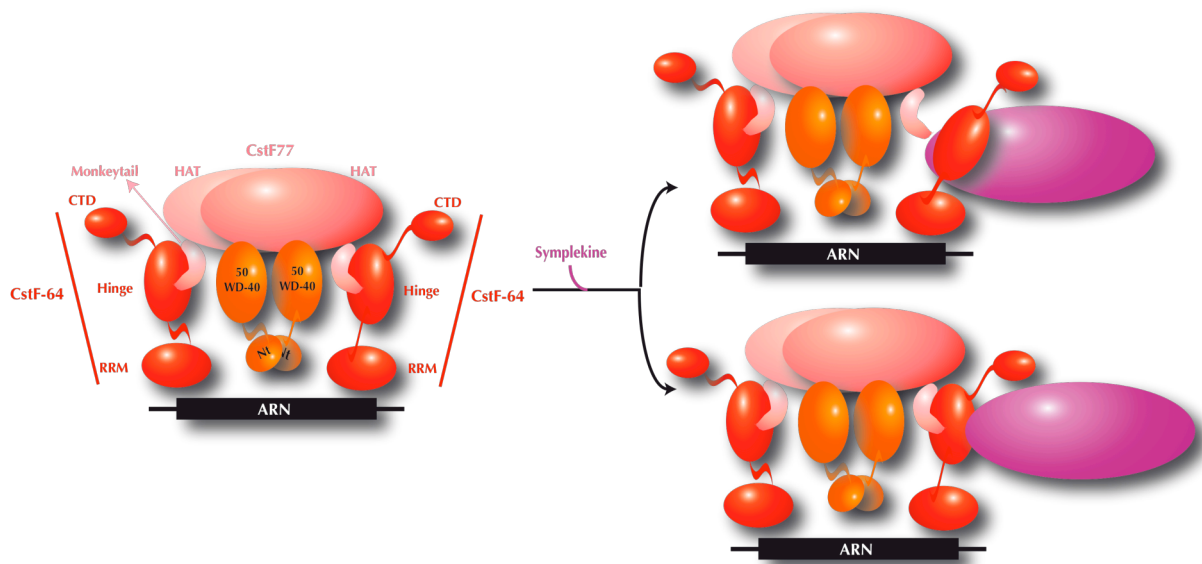
La dimérisation d'une protéine implique souvent une surface d'exposition plus importante. Un site spécifique d'interaction avec une protéine, un peptide ou un acide nucléique sera doublement exposé sur la surface d'une protéine dimérique. Ce fait favoriserait ses associations avec plusieurs composants, ou avec plusieurs domaines d'un même composant, à la fois. Dans les complexes étudiés, les protéines **CstF-77 et Rna14p** établissent la jonction avec l'autre grand complexe de la maturation 3', CPSF ou CPF respectivement. CstF-77 interagit avec les deux autres protéines du complexe et avec CPSF-160 (*voir aussi la section A de ce chapitre*), Fip1p et le CTD de l'ARN pol II. Rna14p s'associe au moins à Rna15p, Pcf11p, Pfs2p, Fip1p, Cft1p/Yhh1p, le CTD de l'ARN pol II et Hrp1p. De cette façon, le fait de posséder une plus grande surface notablement exposée aiderait à ce que ces interactions se produisent.

Nos résultats démontrent que le **domaine N-terminal de CstF-50** est suffisant pour l'oligomérisation. L'explication biologique de la dimérisation de CstF-50 n'est pas claire. Des résidus conservés forment deux zones exposées identiques de deux côtés de la surface du dimère. Des expériences de co-cristallisation, de RMN et de SPR et des analyses biochimiques ont été utilisées pour caractériser l'implication de cette région (CstF-50dm (1-65)) dans la reconnaissance du **CTD** de l'ARN pol II. Aucune de ces techniques n'a permis d'établir les bases de cette interaction. Des travaux précédents ont montré que la protéine tronquée CstF-50hs (77-176) est capable d'interagir avec le CTD (Fong and Bentley 2001). L'ensemble de ces données supportent une contribution du « linker », situé entre le domaine de dimérisation et le domaine WD-40 (approximativement des résidus 65 à 95), dans la reconnaissance du CTD de l'ARN pol II. Dans ce cas-là, les surfaces conservées formées par le dimère dans la région N-terminale pourrait être

employée pour l'interaction avec un partenaire non identifié. Ceci pourrait constituer la signification de sa organisation dimérique.

La dimérisation de la région N-terminale de CstF-50 signifie également que deux domaines **WD-40** sont présents dans le complexe CstF. Ce domaine possède un rôle essentiel dans la réponse cellulaire aux dommages de l'ADN interagissant avec **BARD1** (**B**RCA1 **A**ssociated **R**ING **D**omain **1**), un facteur de réplication/réparation de l'ADN (Kleiman and Manley 1999) (Mirkin et al. 2008). La présence des deux domaines WD-40 permet d'imaginer des stades intermédiaires de CstF en réponse au dommage de l'ADN. Par exemple, lorsque une réponse aux dommages de l'ADN est activée, CstF-50 pourrait être associée au complexe CstF *via* un de ses domaines WD-40 et à BARD1/PARN à travers l'autre. De cette manière, selon cette hypothèse, CstF-50 fonctionnerait comme un point de connexion entre l'ARN pol II, le complexe CstF et la machinerie de réparation de l'ADN s'associant aux trois simultanément.

La dimérisation de CstF-77 implique également la présence de deux sous-unités **CstF-64**. De cette façon, deux domaines « **hinge** » sont présents. Par des expériences de « far western », le groupe du Dr. J. Manley (Takagaki and Manley 2000), a mis en évidence la capacité de ce domaine de CstF-64 à interagir avec CstF-77 et la **Symplekine**. Un hétérotrimère CstF-64/CstF-77/Symplekine peut être formé, cependant, l'interaction CstF-64/CstF-77 s'affaiblit lorsque la Symplekine est ajoutée. Leurs conclusions se dirigent vers l'existence d'une compétition entre CstF-77 et la Symplekine pour le domaine « hinge ». En effet, un modèle de CstF hexamérique rend cette théorie plus vraisemblable. La formation initiale d'un complexe CstF avec deux copies de chaque protéine pourrait laisser la place à un complexe où l'un des deux domaines « hinge » est occupé par la Symplekine et non par CstF-77. L'autre possibilité provient de l'analyse de la structure de Rna14p/Rna15p qui montre que les deux protéines sont fortement associées. Il est possible que l'entité unique formée par l'hétérodimère CstF-77/CstF-64, interagisse directement avec la Symplekine, à travers des résidus présents sur la surface de CstF-64 non impliqués dans la reconnaissance de CstF-77 (*fig. 84*). De cette façon, les deux protéines peuvent interagir simultanément avec CstF-64 et une compétition n'aurait pas lieu. Sur la base de l'interaction intime observée pour la structure des domaines « monkeytail » et « hinge », nous proposons le deuxième modèle comme plus probable. L'association de la Symplekine au complexe CstF, pourrait avoir lieu au moment du **recrutement** étant donnée que cette protéine interagit également avec CPSF et ceci renforcerait les liaisons entre les deux grands complexes, CstF et CPSF. Le même principe pourrait être appliqué aux homologues de CstF-64 et de la Symplekine chez la levure, Rna15p et **Pta1p** respectivement.



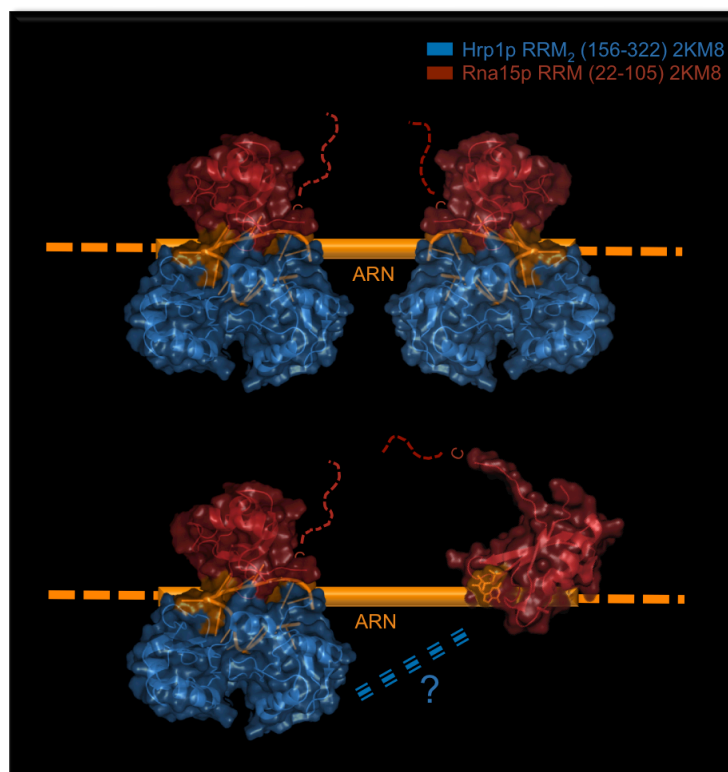
**Figure 84.** Deux hypothèses sur l'interaction CstF-64/Symplekine. La présence de la Symplekine dans les environnements de CstF pourrait provoquer, possiblement au moment du recrutement des composants de la maturation 3', la dissociation des domaines « hinge » et « monkeytail » ou plus probablement, son assemblage à la surface libre de CstF-64 liée à CstF-77.

Un des points clés dans la dimérisation de CstF/CF IA est la double présence du motif **RRM** de CstF-64 ou Rna15p. Différentes études ont démontré la reconnaissance de plus de 10 nucléotides riches en GU par le complexe CstF (Takagaki and Manley 1997) (Beyer et al. 1997). Toutefois, le RRM de CstF-64 isolé interagit avec 4-5 nucléotides (Perez Canadillas and Varani 2003). La présence des deux RRM peut ainsi être nécessaire pour la reconnaissance de la totalité du motif riche en GU du pré-ARNm. Ceci pourrait augmenter l'affinité ou la spécificité de l'interaction.

Concernant le complexe CF IA, la structure du motif RRM de **Rna15p** en complexe avec l'ARN a été récemment résolue (Pancevac et al. 2010). Ces travaux ont montré une préférence du motif isolé pour des séquences d'ARN riches en GU et ont établi les bases moléculaires de cette interaction. Néanmoins, une étude plus récente (Leeper et al. 2010) décrit l'architecture de Rna15p en complexe avec une séquence d'ARN et deux motifs RRM de la protéine **Hrp1p**. Ces travaux démontrent une spécificité de Rna15p pour des séquences riches en A en présence de Hrp1p avec qui, en plus, Rna15p établit des contacts. *In vivo*, ce complexe ternaire est stabilisé par **Rna14p**, capable de s'associer aux deux protéines (Kessler et al. 1996) (Gross and Moore 2001). Il a été montré que les mutations *rna14-1* et *rna15-1* affectent significativement l'interaction de CF IA avec l'ARN (Birise et al. 1998). En outre, des expériences de SPR et des essais de retard sur gel (Noble et al. 2004), indiquent que l'interaction de Rna15p avec le dimère Rna14p et l'incorporation à CF IA augmentent la spécificité et l'affinité d'interaction de cette sous-unité avec l'ARN (Noble et al. 2004) (Gross and Moore 2001). De cette façon, les interactions de Rna15p avec Hrp1p et avec Rna14p régulent la liaison de CF IA à l'ARN.

La faible spécificité de la sous-unité Rna15p isolée pour une séquence d'ARN pourrait être une adaptation pour la régulation de la machinerie. Il est possible que Rna15p interagisse avec une séquence d'ARN pas très spécifique et, qu'avec l'aide de Rna14p, le recrutement de Hrp1p ait lieu sur une séquence très spécifique permettant ainsi la fixation du complexe CF I sur une région précise de l'ARN. Rna15p interviendrait dans les premières étapes du recrutement de la machinerie pour cibler des régions d'ARN, mais ces régions ne seront clivées et polyadénylées que si Hrp1p intervient et le complexe est fixé.

D'autre part, la structure de CstF-77 et d'autres études, évoquent la dimérisation de Rna14p, néanmoins, ce point n'est pas considéré dans le modèle proposé par T. C. Leeper et ses collaborateurs (Leeper et al. 2010). Si Rna14p est un dimère et si deux RRM de Rna15p font partie du complexe CF IA, comment ces deux protéines participent-elles à l'interaction avec Hrp1p ? Est-il possible que deux copies de Hrp1p soient aussi présentes ?, ou est-ce qu'au contraire une seule copie de Hrp1p intervient dans l'interaction ? et, dans ce cas-là, est-ce que Hrp1p est capable d'établir une connexion avec les deux motifs RRM de Rna15p ou seulement avec un d'entre eux ? Un schéma de cette problématique est représenté sur la *figure 85*. L'adjonction du dimère Rna14p à l'ensemble Rna15p/Hrp1p/ARN permettrait de fixer les bases de la stœchiométrie du complexe CF I.



**Figure 85.** Modèle des motifs RRM présents dans CF I (code PDB : 2KM8). Dans notre modèle deux motifs RRM de Rna15p sont présents. La première hypothèse soutiendrait que la protéine Hrp1p est aussi présente en deux copies comme la partie supérieure de l'image le montre. Cependant, si seulement une protéine Hrp1p est présente il se peut qu'elle interagisse avec une seule sous-unité de Rna15p ou qu'elle ait deux sites d'union pour chaque RRM de Rna15p (partie inférieure de l'image).

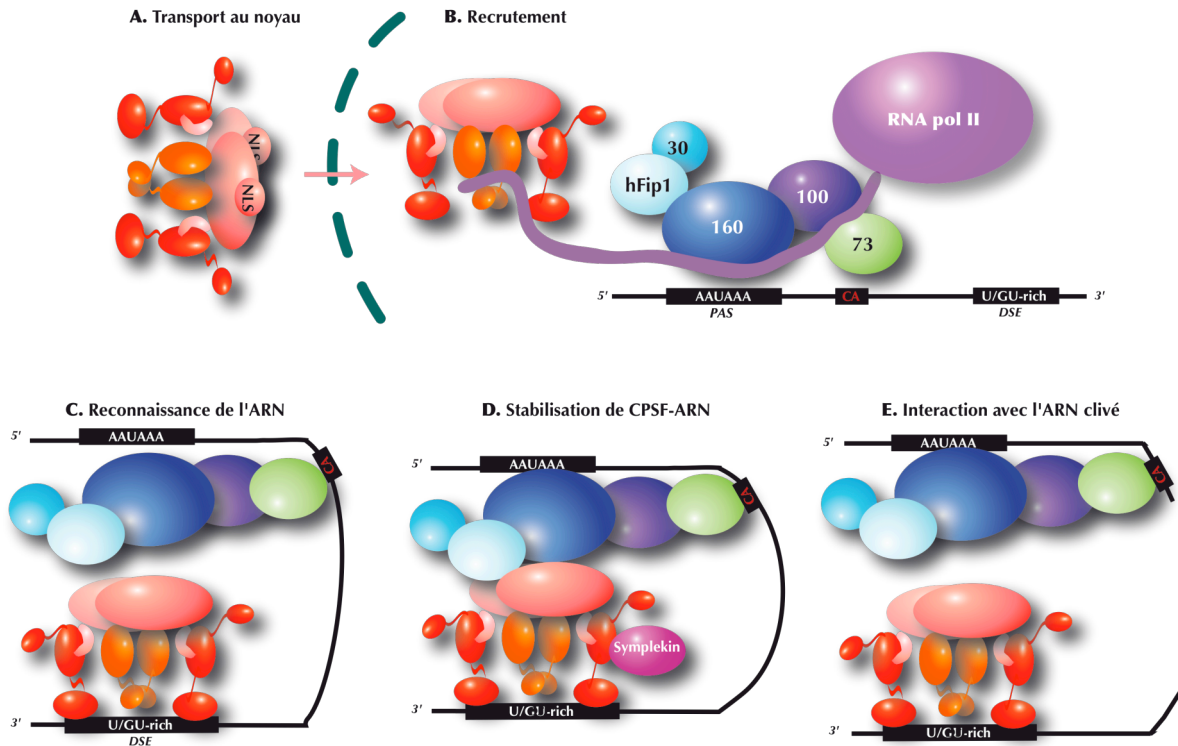
• **Vers un modèle fonctionnel de la participation de CstF et de CF IA dans le processus de maturation 3'**

Avant la réaction de clivage, les facteurs protéiques participant à la réaction doivent s'assembler et interagir avec le pré-ARNm de façon coopérative. Nous ne savons pas précisément si les interactions entre les complexes différents précèdent à l'assemblage avec l'ARN ou si les deux types de liaison ont lieu en même temps.

Chez les **mammifères**, l'assemblage du complexe **CstF** a vraisemblablement lieu avant son internalisation dans le noyau, étant donné que CstF-77 est la seule des trois protéines contenant une région **NLS** (**N**uclear **L**ocalization **S**ignal) (Hockert et al. 2010) (*fig. 86*). Il a été observé que le facteur **TFIIB**, qui participe à l'assemblage du PIC (**P**reinitiation **C**omplex) et au recrutement de la polymérase, interagit avec **CPSF-73** et **CstF-64** (Wu et al. 1999) (Knaus et al. 1996) (Wang et al. 2010). Ainsi, au moins deux protéines de la machinerie de polyadénylation sont présentes au moment de l'initiation de la transcription, ce qui suggère une implication des promoteurs dans le recrutement de facteurs de la maturation 3' (Perales and Bentley 2009). Par ailleurs, il existe différents mécanismes régulateurs qui empêchent le déroulement prématuré des réactions de clivage et polyadénylation. Par exemple, l'interaction entre CstF-64 et **PC4** fonctionne comme un facteur d'antiterminaison de la transcription. Quand PC4 est phosphorylée, elle interagit avec CstF-64 et prévient la terminaison de la transcription. La déphosphorylation de ce facteur, provoque la libération du domaine C-terminal de CstF-64 et la terminaison. De cette manière, clivage et polyadénylation peuvent avoir lieu (Calvo and Manley 2001).

Le type de phosphorylation du CTD déterminera le moment de ce recrutement (*Chapitre I.A.2.a*). Selon le modèle « **double pause** » (« dual-pausing ») décrit par Glover-Cutter (Glover-Cutter et al. 2008), quand la pol II effectue la **pause en 3'** (Boireau et al. 2007) (Darzacq et al. 2007), à approximativement 1kb en aval du site de poly(A), le CTD de l'ARN pol II se trouve fortement phosphorylé en **Ser<sub>2</sub>** et les niveaux des facteurs impliqués dans la polyadénylation en association avec le TEC (**T**ranscription **E**longation **C**omplex) sont maximaux (Glover-Cutter et al. 2008). A ce moment, le complexe CstF est recruté grâce à l'association de CstF-50 avec le CTD de l'ARN pol II. D'après nos résultats, cette interaction aurait probablement lieu à travers la région inter-domaine située entre la zone N-terminale et les motifs WD-40. Dans la réaction de clivage, **CPSF** reconnaît le signal poly(A) pour former un complexe protéine-ARN relativement instable (Gilmartin and Nevins 1989). A ce moment du processus, selon notre hypothèse, CstF-64 devrait être associée à CstF-77 pour que les deux motifs RRM soient capables de reconnaître plus efficacement la séquence d'ARN correcte. De façon parallèle ou quelques instants après, les composants de CstF interagissent avec CPSF afin de stabiliser l'interaction CPSF-ARN (Weiss et al. 1991). Il est possible que la dissociation de certaines sous-unités de CPSF ou de CstF ait lieu pendant le recrutement ou la stabilisation du complexe CPSF-CstF-ARN-ARN pol II. Une compétition des deux protéines pour une même région (d'une autre protéine de l'un des complexes)

peut survenir. La dimérisation de CstF pourrait permettre alors de doubler les possibilités d'interaction et ainsi maintenir plusieurs partenaires liés en même temps. Quand le clivage a été effectué toutes les interactions entre CstF et CPSF doivent être rompues et le CTD différemment phosphorylé. La dissociation de l'ARNm clivé constitue possiblement la dernière étape dans la participation de CstF (fig. 86).



**Figure 86.** Schéma représentant les étapes de la participation de CstF dans la réaction de clivage du processus de maturation 3'. A) Le complexe traverse la membrane nucléaire grâce au motif NLS présent dans la protéine CstF-77. B) CstF-50 interagit avec le CTD de l'ARN pol II permettant le recrutement de CstF. C) Les deux RRM de CstF-64 interagissent avec l'ARN. D) Cette étape peut avoir lieu de façon simultanée avec C. Les interactions de CstF avec CPSF stabilisent l'interaction de CPSF avec l'ARN. Les interactions des composants des deux facteurs peuvent impliquer des réarrangements dans chaque complexe. E) CstF est dissocié de la machinerie quand l'ARN a été clivé. Le moment de séparation de CstF de l'ARN est inconnu.

Chez la **levure**, dans un premier temps, au moins cinq protéines reconnaissent le pré-ARNm. Le processus général est commun à celui présenté pour les mammifères. Les complexes **CF I** et **CPF** lient des séquences spécifiques du pré-ARNm afin de déterminer le site de clivage. Quand la machinerie est fixée sur l'ARN, elle se stabilise *via* des interactions protéine-protéine et les réactions de clivage et polyadénylation peuvent ainsi avoir lieu (Zhao et al. 1999). Toutefois, la manière de procéder de CF IA est moins claire que celle observée pour CstF et plusieurs divergences entre les deux systèmes ont été constatées.



- 
- CF IA doit rester fixé sur l'ARN parce qu'il intervient dans les deux étapes de la maturation. De cette façon, ce complexe reconnaît une séquence d'ARN en aval du site de clivage et reste ainsi lié à la machinerie pour la réaction de polyadénylation.
  - Un homologue de CstF-50 n'a pas été identifié chez la levure et le recrutement du complexe par l'ARN pol II est principalement effectué à travers son interaction avec Pcf11p.
  - Chez la levure il a été observé que le CID de Pcf11p est responsable au même temps de l'interaction avec l'ARN et avec le CTD. L'existence d'une compétition pourrait jouer un rôle dans la libération des facteurs de la maturation 3' (Hollingworth et al. 2006). Chez les mammifères ce comportement du domaine CID n'a pas été observé.
  - Pcf11p et Clp1p font partie du complexe. Le rôle précis de Clp1p dans le processus n'a pas été déterminé, mais à partir des résultats obtenus au laboratoire par F. Maurice et L. Minvielle-Sebastia, il a été observé que l'interaction de cette protéine avec l'ATP paraît indispensable pour les réactions de clivage et polyadénylation et pour l'assemblage du complexe CF IA.
  - L'interaction avec CF IB (Hrp1p) pour former le complexe CF I n'existe pas chez les mammifères. Cette association possède un rôle essentiel chez la levure dans la détermination du site de clivage.
  - Finalement, l'existence d'un complexe alternatif de CF IA a été proposée. Dans ce nouveau complexe, Pti1p substituerait Rna15p (Skaar and Greenleaf 2002). **Pti1p** partage avec Rna15p et CstF-64 son domaine RRM, le domaine « hinge », et la région C-terminale, les trois zones étant fortement conservées dans les trois protéines. De plus, Pti1p est capable d'interagir avec Rna14p, avec Pcf11p et avec Pta1p, le rapprochant des caractéristiques de Rna15p. Cependant, d'autres études soutiennent l'hypothèse que Pti1p est un composant du complexe APT (associated with the Pta1p subunit of CPF) qui s'associe faiblement au complexe CF IA (Nedeia et al. 2003). APT est composé de 6 protéines (Pti1p, Swd2p, Glc7p, Ssu72, Ref2p et Syc1p) et interagit avec les composants de CPF pour former un complexe appelé « holo-CPF ». Ce nouveau complexe interviendrait principalement dans la maturation de snoRNAs (Small nucleolar RNAs) (Nedeia et al. 2003) (Dheur et al. 2003).

- **La machinerie de la maturation 3' des pré-ARNms constitue un système dynamique.**

La présentation d'un modèle structural d'un complexe protéique peut induire des biais et faire paraître un ensemble statique. Sur la base des informations extraites à partir de nos données, des travaux effectués au cours des dernières années et des hypothèses proposées, nous pouvons considérer l'existence d'un comportement dynamique du système de maturation 3'.

Les associations entre les différents éléments de la machinerie n'ont pas forcément lieu simultanément et ne sont pas maintenues intactes tout au long du processus. Par exemple, nous ne savons pas clairement si la dimérisation de CstF est un état transitoire ou si le complexe reste inamovible pendant tout le mécanisme de clivage et polyadénylation. De plus, nous avons présenté par exemple la possibilité d'une intervention de la **Symplekine** sur CstF au moment du recrutement. D'un autre côté, **Yra1p**, un facteur d'export des ARNms vers le cytoplasme (Stutz et al. 2000), interagit avec CF IA à travers la sous-unité **Pcf11p**. Le site de Pcf11p impliqué dans l'interaction avec Yra1p et avec Clp1p coïncident. Dans le modèle proposé (Johnson et al. 2009) (Saguez and Jensen 2009), Yra1p serait recruté par Pcf11p dans un premier temps et ensuite transféré à Sub2p. Ceci voudrait dire que Pcf11p est isolé de Clp1p au moins au moment de l'export des ARNms et que le complexe CF IA ne reste pas intact.

De plus, la composition des facteurs varie en fonction du type d'ARN, du type cellulaire ou des circonstances externes.

Selon le **type d'ARN** les facteurs qui interviennent varient. Les différences dans la composition protéique du complexe CPF au cours de la maturation des snoRNAs a été évoquée précédemment. Dans le processus de maturation 3' des ARNms **d'histones répllication-dépendantes**, la réaction de clivage n'est pas suivie de la formation de la queue poly(A). Malgré cela, les cinq sous-unités du complexe **CPSF**, dont **CPSF-73** comme endonucléase, et la **Symplekine** sont impliquées dans la réaction, mais seulement deux des trois protéines de **CstF (64kDa et 77kDa)** interviennent (Dominski et al. 2005).

Par ailleurs, certaines protéines de la maturation 3' participent dans la régulation d'une **polyadénylation tissu-spécifique**. Par exemple, la régulation de l'expression de la **chaîne lourde de l'IgM** pendant la différenciation de lymphocytes B dépend des variations dans la concentration de la sous-unité CstF-64. Le mécanisme exact n'a pas été défini, mais selon le modèle proposé (Takagaki and Manley 1997) (Danckwardt et al. 2008), pendant la différenciation de cellules B, une augmentation de CstF-64 favoriserait l'expression de la « forme sécrétée » de la chaîne lourde de l'IgM par rapport à la « forme membranaire ». Il est possible que la plasticité dans la reconnaissance des séquences d'ARN par CstF-64 joue aussi un rôle dans ce phénomène.

En outre,  **$\tau$ CstF** s'exprime de manière importante dans les cellules mâles germinales et participe à la polyadénylation spécifique de ce type cellulaire et dans le processus de spermatogenèse (Wallace et al. 1999). Cette spécificité est probablement occasionnée par la reconnaissance d'une séquence d'ARNm caractéristique de ce type de tissu *via* son RRM modifié

(Monarez et al. 2007). La variante  **$\tau$ CstF-64** a été associée à certains problèmes neurologiques et d'infertilité masculine (Dass et al. 2002) (Dass et al. 2007). Le « hinge » restant intact dans cette version de CstF-64, l'interaction avec CstF-77 ne serait pas affectée. De cette façon, un complexe CstF différent serait composé de CstF-77/ $\tau$ CstF-64/CstF-50. De la même façon, d'autres variations de CstF-64 comme  **$\beta$ CstF-64**, une protéine abondante dans les cellules du système nerveux, participeraient à la diversité de l'assemblage de CstF.

En plus du type cellulaire, des changements dans la formation ou dans la concentration du complexe CstF se manifestent au cours du cycle cellulaire. Le taux d'expression de CstF-64 augmente cinq fois pendant le cycle cellulaire entre la phase  $G_0$  et S, de cette façon, il y aura une quantité plus importante du complexe CstF à cette étape (Martincic et al. 1998). L'augmentation du CstF entraîne une accumulation d'ARNms avec une queue poly(A) dans les cellules en sortant de la phase  $G_0$  (Getz et al. 1976). La réduction de CstF-64 peut mener à un **arrêt du cycle cellulaire** dans la phase  $G_0/G_1$  et provoquer **l'apoptose** (Takagaki and Manley 1998) (Danckwardt et al. 2008).

Les **circonstances externes** peuvent modifier l'organisation générale de la machinerie. Nous avons cité, par exemple, les effets suivis après l'endommagement de l'ADN où **BARD1** et **PARN** doivent interagir avec CstF-50. Il a été montré que la mutation BARD1-Gln564His, détectée dans des cas de tumeurs ovariennes, réduit son affinité pour le WD-40 de CstF-50 et la rend inactive dans l'inhibition de la polyadénylation (Irminger-Finger and Leung 2002) (Danckwardt et al. 2008). D'autre part, une variante de BARD1, contenant la délétion dénommée  **$\Delta$ RIN**, a été détectée dans le **cancer du sein**. BARD1 $\Delta$ RIN pourrait entrer en compétition avec BARD1 pour l'interaction avec CstF-50 ou alternativement, empêcher l'entrée de CstF-50 dans le noyau (Tsuzuki et al. 2006). Dans les deux cas, BARD1 $\Delta$ RIN inhibe la formation du complexe CstF/BARD1/BRCA1 et ainsi la réponse cellulaire aux endommagements de l'ADN. Ces données mettent en évidence l'importance de l'adaptation de la machinerie de la maturation 3' face aux circonstances imposées.

En conséquence, un modèle exact du fonctionnement n'a pas pu être établi et beaucoup des détails nous échappent encore aujourd'hui. Nous pouvons commencer à décrire un modèle général de fonctionnement et d'assemblage pour la machinerie de maturation 3'. Ceci est une étape importante vers la constitution d'un modèle précis et la compréhension du mécanisme. Néanmoins ce modèle devra être adapté au type d'ARN, aux spécificités de chaque tissu, aux circonstances externes et surtout à l'espèce à laquelle nous faisons référence.



## Chapitre V : Conclusions et perspectives

La maturation 3' des pré-ARNms est un processus fondamental qui nécessite la participation d'un nombre important de facteurs protéiques. En général, les pré-ARNms sont soumis à une étape de clivage sur leur extrémité 3' qui est suivie de la formation de la queue poly(A). Les facteurs qui interviennent sont impliqués dans la reconnaissance du pré-ARNm, dans les réactions de clivage et/ou polyadénylation ou dans la régulation du processus. Le complexe **CstF** (Cleavage stimulation Factor) des mammifères est indispensable pour la réaction de clivage tandis que son homologue chez les levures, **CF IA** (Cleavage/polyadenylation Factor IA), est nécessaire pour les deux étapes, le clivage et l'addition d'adénines. Mes travaux ont été focalisés sur la description structurale des sous-unités intégrées dans ces deux facteurs et sur la manière dont elles s'assemblent.

Dans ce manuscrit nous avons abordé premièrement les études effectuées pour le complexe CstF. Ce complexe est composé de trois sous-unités, CstF-77, CstF-64 et CstF-50 appelées ainsi en raison de leur masse moléculaire. Des travaux précédents ont suggéré l'organisation hexamérique de ce facteur, formé ainsi par deux copies de chaque protéine. Les études réalisées au cours de cette thèse ont apporté des évidences corroborant cette hypothèse. La capacité de CstF-50 pour former un dimère a été suggérée par des expériences de « far western » (Takagaki and Manley 2000). En effet, la **détermination structurale du domaine N-terminal de CstF-50** montre qu'il constitue un module d'homodimérisation qui est conservé à travers les espèces (Moreno-Morcillo et al., RNA, 2010, *en révision*). La résolution de sa structure a permis d'établir les bases moléculaires de l'interaction entre les deux monomères et de découvrir la présence d'un motif LISH connu principalement pour sa capacité à former des dimères. De plus, des expériences de RMN (Résonance Magnétique Nucléaire) ont démontré que la dimérisation était aussi une caractéristique de la protéine en solution. La signification de cette dimérisation reste inconnue. Néanmoins, nous avons observé que le dimère permettait la formation de deux surfaces exposées (une de chaque côté du dimère) riches en résidus conservés. Ces régions pourraient être impliquées dans l'interaction avec un partenaire encore inconnu et apporter ainsi une signification biologique à la dimérisation de cette protéine. En outre, la dimérisation de CstF-50 implique la présence de deux domaines WD-40, la région responsable de l'interaction avec CstF-77 et avec BARD1. L'existence de deux domaines WD-40 augmente les possibilités d'association simultanée avec plusieurs partenaires. En outre, la structure de CstF-77 montre que cette protéine est également organisée en dimère (Bai et al. 2007) (Legrand et al. 2007). Finalement, CstF-64, ne formant pas un dimère, serait également présent en deux copies, ceci impliquant la présence de deux domaines RRM. La double présence de ce motif est peut-être indispensable pour une reconnaissance plus précise des séquences du pré-ARNm.

Concernant le complexe CstF, nous avons aussi identifié les régions d'interaction de CstF-77 et CstF-64. Contrairement à ce qui a été suggéré auparavant (Takagaki and Manley 2000), la région

riche en Pro de CstF-77 n'intervient pas dans l'interaction et seulement une région C-terminale contenant les résidus du 626 au 717 est impliquée. Malgré les nombreux essais de cristallisation, l'architecture des domaines d'interaction de ces deux protéines n'a pas pu être déterminée. Cependant, nous avons résolu la structure des régions impliquées pour leurs homologues chez la levure, Rna14p et Rna15p. Ces résultats, et d'autres obtenus pour le complexe CF IA entier, ont été exposés, dans la deuxième partie de ce manuscrit.

CF IA est composé de 4 sous-unités, Rna14p et Rna15p, les homologues de CstF-77 et CstF-64 respectivement, et deux protéines supplémentaires, Pcf11p et Clp1p dont leurs homologues chez les mammifères, hPcf11 et hClp1, sont associées dans le complexe CF IIm. Les travaux réalisés au cours de cette thèse, ont permis d'identifier les régions minimales de **Rna15p** et de **Rna14p** essentiels pour l'interaction. Nous avons employé la technique de RMN pour la détermination de la structure du sous-complexe (collaboration avec le Dr. Mackereth) (Moreno-Morcillo et al., *Molecular Cell*, 2010, *en préparation*). Dans le complexe, les deux protéines sont fortement associées et forment une entité unique. Pour l'interaction, la région C-terminale de Rna14p, qui a reçu le nom de domaine « monkeytail », s'entrelace intimement avec la région « hinge » de Rna15p. Nous avons démontré que ce type d'interaction est maintenu dans le contexte du complexe entier. La présence de ces deux domaines chez leurs homologues métazoaires, CstF-77 et CstF-64, suggère en plus la conservation de cette organisation à travers les espèces. En outre, nous nous sommes intéressés à **l'assemblage général de CF IA**. Des études préliminaires qui contribuent à la résolution de la structure du complexe ont été effectuées. Nous avons déterminé les constructions protéiques nécessaires pour l'expression du complexe CF IA *'in vitro'* où seulement le domaine CID de Pcf11p est absent. Les conditions idéales pour sa purification ont été également définies et des essais de cristallisation de ce complexe en présence/absence d'ARN sont en cours au laboratoire. De plus, les premières images de microscopie électronique (effectuées par Patrick Bron du CBS, Montpellier) ont été obtenues à partir de nos échantillons et nous permettent de rester optimistes pour la future détermination structurale du complexe par cette méthode.

Cette thèse constitue en plus un exemple clair des avantages remportés par la combinaison de différentes techniques dans la résolution de structures protéiques. Spécialement, la cristallographie des rayons X associée aux expériences de RMN a été une approche essentielle pour la détermination structurale du domaine N-terminal de CstF-50. De façon similaire, l'application des techniques de biologie moléculaire et biochimiques permet l'adaptation aux problèmes rencontrés au cours de la cristallisation ou des expériences de RMN.

Pour poursuivre les travaux sur les complexes CstF et CF IA, différentes stratégies peuvent être adoptées. Tout d'abord, le rôle de la dimérisation du domaine N-terminal de CstF-50 reste indéfini. Par exemple, des expériences de protéomique en double hybride pourraient aider au dépistage de candidats protéiques interagissant avec cette région de CstF-50. Les dernières expériences pour comprendre l'association de cette protéine avec le domaine CTD de l'ARN pol II sont actuellement

en cours au laboratoire. Il serait intéressant d'étudier également le domaine WD-40 pour déchiffrer le mécanisme d'interaction avec CstF-77 et avec la protéine BARD1. Nous considérons que l'étude de ces interactions est indispensable pour la compréhension du développement de certains cas de tumeurs du sein et des ovaires.

En outre, la caractérisation biophysique du complexe CstF-77/CstF-64 peut aider à déterminer le comportement des deux protéines. D'autres expériences de RMN et l'utilisation des techniques comme la DLS (diffusion dynamique de la lumière ou Dynamic Light Scattering), l'ultracentrifugation analytique ou le dichroïsme circulaire pourraient être envisagées. D'autre part, la co-expression, purification et des essais de cristallisation du complexe incluant une nouvelle construction plus longue de la sous-unité CstF-77hs sont en cours. L'étude structurale du sous-complexe formé à partir des protéines humaines pourrait aider à déterminer comment les modifications engendrées dans la protéine  $\tau$ CstF-64 peuvent affecter son interaction avec CstF-77 et au développement des certains problèmes neurologiques et d'infertilité masculine.

À partir de la structure du complexe Rna14p (626-677)/Rna15p (127-232), nous voudrions approfondir nos connaissances sur les effets produits par des mutations situées sur les régions d'interaction de Rna14p et Rna15p. Des expériences utilisant la méthode du TAP ou des techniques de co-expression chez la bactérie ont été envisagées pour élucider cette problématique. En outre, la poursuite des expériences à partir du complexe CF IA entier devrait permettre la résolution de sa structure. L'addition ou la suppression de domaines protéiques ou des protéines complètes peut être examinée pour l'étude de la modification de l'architecture globale en fonction de ces changements.

De futurs travaux peuvent également se focaliser sur la définition des **rôles** de chaque protéine intégrée dans les complexes CstF et CF IA. Dans le laboratoire, l'étude du rôle de **Clp1p** dans la maturation 3' est en cours et révèle l'importance de cette protéine dans l'organisation générale du complexe.

De plus, dans la littérature, il a été montré que toutes les étapes de l'expression génique étaient fortement coordonnées. La connexion entre le processus de la maturation 3' avec la terminaison de la transcription ou avec l'export des ARNms a été mise en évidence, mais le mécanisme exact d'enchaînement des événements est inconnu. La caractérisation des **interactions** physiques entre facteurs impliqués dans des processus différents est un outil crucial pour la compréhension de la régulation de l'expression génique.

L'importance de la **régulation** du processus de maturation 3' se reflète dans le nombre maladies occasionnées par des erreurs dans la machinerie. Depuis une perspective médicale, il serait intéressant d'explorer l'utilisation des polymorphismes des composants du processus de maturation comme biomarqueurs. La compréhension au niveau moléculaire des pathologies provoquées par des anomalies spécifiques, peut être considérée comme une approche pour la correction de ces troubles.







## Capítulo V : Conclusiones y perspectivas

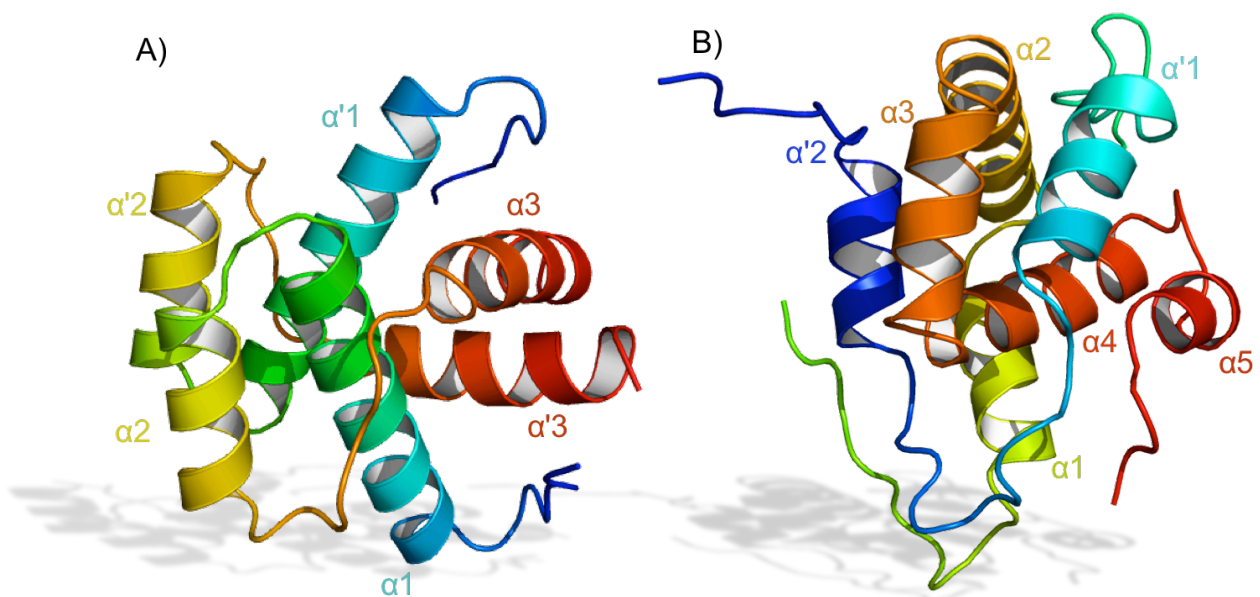
La maduración 3' del pre-ARNm es un proceso fundamental que precisa la participación de un número importante de factores proteicos. En general, los pre-ARNms están sujetos a una reacción de corte en el extremo 3', donde después será añadida una cola de adeninas (cola poliA). Los factores que intervienen en esta etapa de la expresión génica, pueden estar implicados en el reconocimiento del ARN, en las reacciones de corte y/o poliadenilación o en la regulación del proceso global. El complejo **CstF** (Cleavage stimulation Factor) de mamíferos es indispensable para la reacción de corte, mientras que su homólogo correspondiente en levaduras, **CF IA** (Cleavage/polyadenylation Factor IA), es necesario en ambas reacciones, en el corte y la poliadenilación. Mi Tesis se ha centrado en la descripción estructural de las diferentes subunidades que componen estos complejos y en cómo las proteínas que los integran interaccionan.

En esta Tesis se abordó en primer lugar el estudio del complejo **CstF**. Dicho complejo está formado por tres subunidades, CstF-77, CstF-64 y CstF-50, llamadas así en función de su peso molecular. La hipótesis de que este complejo presenta una organización hexamérica ha sido propuesta anteriormente por otros grupos de investigación (Bai et al. 2007). Si esta hipótesis es exacta, dos subunidades de cada proteína están presentes en el complejo, al menos en un punto del proceso de la maduración 3'. Los estudios realizados en el transcurso de esta Tesis han aportado una nueva prueba que permite corroborar esta teoría. La capacidad de CstF-50 para dimerizar fue en principio sugerida a partir de experimentos de "far western" (Takagaki and Manley 2000). Los resultados obtenidos mediante esta técnica han sido confirmados gracias a la **determinación estructural del dominio N-terminal de CstF-50** (Moreno-Morcillo et al., RNA 2010, *en revisión*) (*fig. 87A*). Hemos demostrado que esta región de la proteína constituye un módulo de homodimerización. La interacción entre ambos monómeros tiene lugar a través de residuos que se encuentran conservados entre las especies. La resolución de esta estructura nos ha permitido establecer la bases moleculares de la interacción entre los dos monómeros y descubrir la presencia de un motivo llamado LISH, conocido principalmente por su función en la dimerización de las proteínas en las que se encuentra. Por otro lado, por medio de la técnica de RMN (Resonancia Magnética Nuclear), hemos demostrado que la condición dimérica es también una característica de la proteína en solución. La importancia biológica de esta dimerización es por el momento desconocida. No obstante, la formación del dímero revela dos zonas idénticas, fuertemente conservadas y expuestas en lados opuestos de la superficie. Estas regiones podrían estar implicadas en la interacción con otra proteína aún desconocida, dando de esta forma un significado biológico a este tipo de organización. La dimerización de CstF-50 implica igualmente la presencia de dos dominios WD-40. Este dominio constituye la región responsable de la interacción con CstF-77 y con BARD1. La existencia de dos WD-40 aumentaría las posibilidades de una asociación con



varios elementos de manera simultánea. Por otra parte, la estructura de CstF-77 muestra que esta proteína forma también un dímero (Bai et al. 2007) (Legrand et al. 2007). CstF-64 no dimeriza, sin embargo la organización general del complejo CstF implica que dos copias de esta proteína están presentes y que, por tanto, dos dominios RRM's participan en el reconocimiento de la secuencia del pre-ARNm. La doble presencia de este motivo podría participar en la especificidad o en el incremento de la afinidad de la interacción CstF-ARN.

Con respecto al complejo CstF, nuestro trabajo también ha permitido la identificación de las regiones implicadas en la interacción entre CstF-77 y CstF-64. Contrariamente a lo establecido por estudios precedentes (Takagaki and Manley 2000), según nuestros análisis, la región rica en Pro de CstF-77 no interviene en la asociación sino que sólo la zona C-terminal de la proteína (del aminoácido 626 al 717) está implicada. A pesar de los numerosos tests de cristalización, la arquitectura de los dominios de interacción no ha podido ser determinada. Sin embargo, la estructura de las regiones implicadas ha sido resuelta en el laboratorio para sus proteínas homólogas en levaduras, Rna14p y Rna15p. Estos resultados, junto con otros obtenidos a partir del complejo CF IA completo, se muestran en la segunda parte de esta Tesis.



**Figura 87.** A) Estructura cristalográfica del dominio N-terminal de CstF-50. Cada monómero ha sido coloreado en un gradiente de azul a rojo correspondiendo a los residuos desde la región N-terminal a la C-terminal. Las hélices del primer monómero se denominan como  $\alpha 1$ ,  $\alpha 2$  y  $\alpha 3$  y las del segundo como  $\alpha'1$ ,  $\alpha'2$  y  $\alpha'3$ . B) Modelo estructural del subcomplejo Rna14p (626-677)/ Rna15p (127-232) en solución. Para simplificar, solamente uno de los modelos obtenidos por RMN ha sido utilizado para realizar la figura. De la misma manera que en A), el gradiente de colores comienza en la región N-terminal de Rna14p (hélices  $\alpha'1$  y  $\alpha'2$ ) en azul para finalizar en el extremo C-terminal de Rna15p ( $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 4$  y  $\alpha 5$ ) representado en rojo.

Las imágenes han sido realizadas con el programa PyMOL.



El complejo CF IA está compuesto por 4 subunidades, Rna14p y Rna15p, que corresponden a los homólogos de mamíferos CstF-77 y CstF-64 respectivamente, y dos proteínas suplementarias, Pcf11p y Clp1p. Los homólogos de estas dos últimas se encuentran en mamíferos formando parte de un complejo llamado CF IIm. Los trabajos llevados a cabo durante esta Tesis han permitido identificar las regiones mínimas indispensable para la asociación entre **Rna14p** y **Rna15p**. Nos hemos servido de la técnica de RMN para la resolución de la estructura formada por este subcomplejo (colaboración con el Dr. Mackereth) (*fig. 87B*). En el heterodímero, las dos proteínas se encuentran fuertemente asociadas formando una entidad única. Para la interacción, la región C-terminal de Rna14p, aquí bautizada como dominio “monkeytail”, mantiene una estrecha relación con el dominio “hinge” de Rna15p, de manera que las hélices de la dos proteínas quedan entrelazadas. Hemos demostrado además que este tipo de conformación es mantenida en el contexto de la versión integral del complejo Rna14p/Rna15p. La presencia de los dos dominios (“monkeytail” e “hinge”) en sus respectivos homólogos de mamíferos, CstF-77 et CstF-64, sugiere que la organización del subcomplejo está conservada entre las especies.

Por otro lado, nuestros estudios han sido dirigidos hacia el **ensamblaje general del complejo CF IA**. Se han realizado estudios preliminares para la resolución de su estructura. Entre estos estudios destaca la determinación de las construcciones proteicas necesarias para la expresión ‘*in vitro*’ de un complejo CF IA, donde solamente el dominio CID de Pcf11p ha sido eliminado. Por otro lado, hemos determinado las condiciones ideales para una purificación eficiente del complejo y los tests de cristalización en presencia/ausencia de ARN están siendo llevados a cabo actualmente en el laboratorio. Además, disponemos de las primeras imágenes de microscopía electrónica obtenidas a partir de nuestras muestras (colaboración con Patrick Bron du CBS, Montpellier). Nuevas purificaciones serán necesarias, pero estas primeras imágenes nos permiten ser optimistas en cuanto a la resolución de la estructura por este método.

Esta Tesis constituye además un ejemplo palpable de las ventajas adquiridas a partir de la combinación de diferentes técnicas para la resolución de estructuras proteicas. Especialmente, la cristalografía de rayos X asociada a experimentos de RMN ha sido determinante para la obtención del modelo estructural del dominio N-terminal de CstF-50. Del mismo modo, la aplicación de técnicas de biología molecular y bioquímicas, ofrece la posibilidad de adaptarse a los problemas que puedan aparecer durante la cristalización o los experimentos de RMN.

Para continuar el estudio de los complejo CstF y CF IA, podemos adoptar diferentes estrategias. En primer lugar, el papel de la dimerización de CstF-50 es aún desconocido. Estudios de proteómica en doble híbrido por ejemplo, podrían ayudar a la detección de posibles candidatos capaces de interactuar con la zona N-terminal de CstF-50. Actualmente, están siendo realizadas en el laboratorio las últimas pruebas destinadas a la comprensión de la asociación de CstF-50 con el CTD de la ARN pol II. Asimismo, sería interesante indagar en el estudio del dominio WD-40 para poder descifrar el mecanismo de la interacción que tiene lugar entre éste y CstF-77 o BARD1. Hay



que tener en cuenta que el análisis de este proceso puede ser indispensable para la comprensión de del desarrollo de ciertos tipos de cáncer de mama o de ovarios.

Por otra parte, la caracterización biofísica del complejo CstF-64/CstF-77 puede participar en la descripción del comportamiento de estas dos proteínas. Otras experiencias en RMN o el empleo de técnicas como la de DLS (difusión dinámica de la luz o Dynamic Light Scattering), la ultracentrifugación o el dicroísmo circular, podría ser útil para este fin. Por otro lado, la coexpresión, la purificación y los tests de cristalización del complejo formado a partir de una construcción más larga de la subunidad CstF-77 están siendo realizados en el laboratorio. El estudio estructural del subcomplejo formado por las proteínas humanas podría ayudar a definir cómo las modificaciones producidas en la proteína  $\tau$ CstF-64 afectan a la interacción con CstF-77 e implican el desarrollo de ciertos problemas neurológicos y de infertilidad masculina.

Con respecto a la estructura del complejo Rna14p (626-677)/ Rna15p (127-232), nos gustaría profundizar en los efectos producidos por las mutaciones situadas en las regiones de interacción entre Rna14p y Rna15p. Para aclarar la cuestión inducida por estas mutaciones en el ensamblaje del complejo CF IA, experimentos utilizando el método de TAP (Tandem Affinity Purification) o el método de coexpresión en *E. coli*, han sido previstos. Además, la continuación del proyecto que involucra el complejo CF IA podrá aportar la resolución de su estructura. En este complejo, la adición o la eliminación de dominios proteicos o de proteínas completas pueden ser tenidos en cuenta para analizar la modificación de la arquitectura global del factor CF IA en función de las variaciones efectuadas.

Los trabajos que se realicen en el futuro pueden igualmente estar enfocados en la descripción de las funciones de las proteínas que integran los dos complejos, CstF y CF IA. El estudio del papel de **Clp1p** en la maduración 3' está siendo examinado y por el momento, los primeros resultados indican que esta proteína participa en la asociación de Pcf11p al complejo, lo cual complica el mecanismo de ensamblaje de CF IA.

Además, según datos bibliográficos, todas las etapas de la expresión génica están fuertemente coordinadas. La conexión entre la poliadenilación y la transcripción o la exportación del ARNm ha sido demostrada. Sin embargo, el mecanismo exacto de la sucesión de estos acontecimientos es desconocido. La caracterización de las interacciones físicas que tienen lugar entre los factores implicados en los diferentes procesos, es una herramienta esencial para la comprensión de la regulación de la expresión génica. La importancia de la regulación del proceso de maduración 3' queda además reflejada por el número de enfermedades desarrolladas a partir de errores producidos en la maquinaria. Desde un punto de vista médico, sería interesante explorar la utilización de polimorfismos de los componentes del proceso de maduración como biomarcadores. La comprensión a nivel molecular de las patologías provocadas por anomalías específicas, puede ser considerada como una vía hacia la corrección de estos problemas.

## Chapitre VI : Matériels et Méthodes

### A. Protocoles de Biologie Moléculaire

#### 1. Le clonage

Pendant ma thèse, j'ai travaillé avec différentes protéines et différentes régions de certaines d'entre elles. L'expression de ces constructions a été possible grâce à l'insertion des fragments d'ADN codants pour les régions protéiques souhaitées dans des vecteurs adéquats. Ce processus est connu en Biologie Moléculaire comme **clonage**. La technique du clonage consiste à isoler un fragment d'ADN codant pour la protéine d'intérêt (ou un fragment de celle-ci) et à le multiplier à l'identique en l'insérant dans une molécule d'ADN porteuse, appelée **vecteur**, qui permet son amplification. Le clonage d'un gène entier permettra la production de la protéine recombinante correspondante.

Dans le laboratoire le protocole général pour la réalisation d'un clonage comprend différentes étapes :

- a) préparation de l'insert
- b) préparation du vecteur
- c) ligation de l'insert et le vecteur
- d) préparation de l'ADN plasmidique

Toutes ces étapes son critiques pour la réussite du clonage. Les détails de la procédure seront donnés par la suite.

##### a.) Préparation de l'insert par PCR

L'insert est le fragment d'ADN d'intérêt codant pour une protéine, ou un fragment de celle-ci, qui sera introduit dans le vecteur de clonage. La préparation de l'insert nécessite de l'utilisation d'une technique de biologie moléculaire appelé **PCR (Polymerase Chain Reaction)** (Mullis et al. 1986). La PCR est un méthode d'amplification génique *in vitro*. Cette technique permet de copier en grand nombre une séquence d'ADN connue, à partir d'une faible quantité d'acide nucléique contenant la séquence d'intérêt (matrice) et d'amorces spécifiques constitués d'oligonucléotides de synthèse.

Pour la réalisation d'une PCR plusieurs éléments sont nécessaires (*tableau 7*). Les quantités montrées sur le tableau sont indicatives d'un protocole général mais dans certains occasions elles ont du être modifiées. Pour chaque PCR, deux amorces sont dessinées comportant environ vingt/trente nucléotides. Les séquences nucléotidiques des amorces doivent être spécifiques des séquences complémentaires de l'ADN simple-brin auxquelles elles vont s'apparier. Les **oligonucléotides** contiennent un **codon d'initiation ou de terminaison de la traduction** ainsi

qu'un **site de restriction enzymatique** (en général *NdeI*, *BamHI*, *XhoI*) qui va permettre l'insertion du fragment d'ADN amplifié dans le vecteur. Les sites de restriction des amorces sont précédés d'une séquence destinée à augmenter l'efficacité de digestion de l'enzyme sur ce site (Cleavage Close to the End of DNA Fragments, *NEW ENGLAND BioLabs*<sup>®</sup>).

Composant de la réaction	Quantité	Description
Oligonucléotide 1 (100µM)	1µl (soit ≈1µg)	<ul style="list-style-type: none"> <li>- spécifique des séquences complémentaires</li> <li>- taux d'appariement entre elles faible</li> <li>- sans structure secondaire</li> <li>- T° de fusion proches</li> </ul>
Oligonucléotide 2 (100µM)	1µl (soit ≈1µg)	
Matrice	10ng	<ul style="list-style-type: none"> <li>- cDNA contenant le fragment d'intérêt</li> <li>- plasmide contenant le fragment d'intérêt (pour le clonage d'une région protéique)</li> </ul>
DNA polymerase	10U (1 µl)	- dans le laboratoire nous disposons de différentes enzymes ( <i>Phusion</i> , <i>Deep Vent</i> , <i>Pfu</i> , <i>Taq polymérase</i> ...)
dNTP	200µM final	- nucléotides libres pour la polymérisation
Tampon polymérase (10x)	10 µl	- dépend de l'enzyme et des caractéristiques de séquences
Autres : DMSO, MgCl <sub>2</sub>		- dans certains cas, ils servent à optimiser la réaction
H <sub>2</sub> O	qsp 50 µl	

**Tableau 7.** Mélange réactionnel utilisé pour la réalisation d'une PCR typique.

La PCR est une technique basée sur la répétition de cycles de transition de température. Chaque cycle inclus trois étapes détaillées ci-dessous (*tableau 8*). La réaction est effectuée dans un appareil adapté, un thermocycleur (*vapo.protect*, *Eppendorf*) permettant la variation de température souhaitée.

La PCR sert à l'amplification de la région d'ADN d'intérêt. Pour la préparation de l'insert, ce fragment d'ADN est digéré par les **enzymes de restriction** appropriées (5U/ µg d'ADN) pour une durée de 16 à 24h dans le tampon adéquat, à 37°C. Ensuite, une partie du produit de digestion (2%) est déposé sur un gel 1% d'agarose contenant du Bromure d'éthidium, BrEt (1µl d'une solution à 10mg/ml pour 100ml). Pour la migration du gel, il est soumis à un champ électrique de 100V dans du TBE 0,5x (89mM Tris, 89mM acide borique, 2mM EDTA). Le gel est analysé par rayonnement UV et, si on observe une bande à la taille attendue, l'insert est ensuite purifié à l'aide des kits commerciaux (*Machery-Nagel*<sup>TM</sup>, *Qiagen*<sup>TM</sup>), à partir d'un gel où tout l'insert est chargé ou directement à partir de la solution de PCR.

Etapes	Température	Temps	Description des évènements
Dénaturation initiale	98°C	2'	- dénaturation de l'ADN double brin - élimination des structures secondaires - homogénéisation du milieu réactionnel par agitation thermique - activation des polymérases type « Hot start »
Dénaturation	98°C	30''	- dénaturation de l'ADN double brin - homogénéisation du milieu à chaque cycle - retrait des polymérases encore liées à la matrice
Hybridation	45 à 65 °C (dépendante du $T_m$ des amorces <sup>1</sup> )	30''	- hybridation des amorces sens et anti-sens à l'ADN matrice
Elongation	72°C	30'' à 2' (dépendant de la longueur de l'amplicon <sup>2</sup> )	- la polymérase synthétise le brin complémentaire de l'ADN matrice à une $T^\circ$ qui lui est optimale à partir des dNTPs libres
Dernière polymérisation	72°C	10'	- étape suivie pour assurer l'élongation après les 30 cycles
Stockage	4-20°C	$\infty$	- l'ADN peut être stocké à -20°C

**Tableau 8.** Phases de la PCR. <sup>1</sup>Normalement la température utilisée est supérieure de quelques degrés par rapport au  $T_m$  des amorces. <sup>2</sup>Le temps de cette étape est d'approximativement 1 minute pour chaque kilobase amplifiée. Le nombre de cycles peut aussi être modifié.

### b.) Préparation du vecteur

L'insert doit faire partie d'un plasmide qui sera introduit dans la bactérie et qui permettra sa répllication. Les **vecteurs** sont des fragments d'acide nucléique circulaire (plasmides) capables de se répliquer de manière autonome possédant un site multiple de clonage et supportant l'insertion d'un fragment d'ADN plus ou moins grand. De plus ils possèdent un gène de résistance afin de sélectionner l'organisme ayant intégré le vecteur. Nous utilisons des vecteurs dérivés de vecteurs commerciaux (tableau 9).

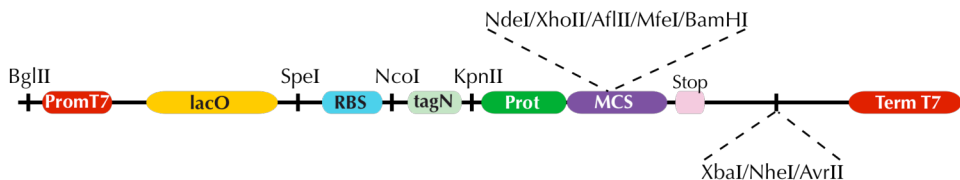
Vecteurs	Caractéristiques
pETMCN-EA-vH (His / H <sub>6</sub> )	Dérivé du vecteur pET15b (Novagen) ; MCN (Multi-Cloning and expressionN): Ori colE1 ; promoteur T7 ; étiquette six-histidine N-terminal; Ampicilline <sup>R</sup>
pETMCN-EK (Ka)	Provient du pET-28b : Ori colE1, (Kan <sup>R</sup> ) Kanamycine <sup>R</sup>
pMCN-YC (Cam)	Synthétisé à partir de pACYC11b : Ori p15A Chloramphénicol <sup>R</sup>
pETMCN-CS (Stp)	Dérivé du vecteur pCDF-DuET; Streptomycine <sup>R</sup>
pETMCN-EA-vGst (GST)	Provient du vecteur pET15b; contient une étiquette GST; Ampicilline <sup>R</sup>
pGAD-T7	Utilisé pour les études de double hybride ; lie le domaine activateur Gal4; Ampicilline <sup>R</sup>
pGBK-T7	Utilisé pour les études de double hybride ; contient le domaine de liaison à l'ADN Gal4 (« binding domain ») ; Kanamycine <sup>R</sup>

**Tableau 9.** Types de vecteurs utilisés au cours de ma thèse. Les noms entre parenthèses correspondent aux abréviations données pour chaque vecteur dans ce manuscrit.



Les vecteurs pETMCN-EA-vH et pETMCN-EA-vGst (fig. 88) contiennent les étiquettes, 6 His et protéine GST, respectivement, étiquettes qui peuvent être clivées par la protéase **TEV**. Le nom « **TEV** » lui est donné car elle correspond au domaine catalytique (27kDa) de la protéine d'inclusion nucléaire A (Nia protein) encodée par le virus de la gravure du tabac (*Tobacco etch virus*). En raison de sa spécificité de séquence, plus stricte que celle de la thrombine, la **TEV** est une enzyme de choix pour le clivage des protéines de fusion. Cette enzyme reconnaît un épitope linéaire de forme générale E-X<sub>aa</sub>-X<sub>aa</sub>-Y-X<sub>aa</sub>-Q-(G/S) avec un site de clivage situé entre les acides aminés Q et (G/S). Dans le laboratoire, nous utilisons la séquence ENLYFQG.

Le vecteur est digéré par les enzymes de restriction adéquats pendant aux moins 2 heures. La coupure peut être vérifiée sur gel. Après digestion, les extrémités 5' phosphate du vecteur « ouvert » sont déphosphorylées grâce à l'enzyme **antartic phosphatase** pendant 2h à 37°C et le vecteur digéré est finalement purifié via un kit commercial (Qiagen<sup>TM</sup>).



**Figure 88.** Site multiple de clonage des vecteurs MCN avec une étiquette en position N-terminale.

### c.) Ligation de l'insert et le vecteur

La ligation implique l'insertion du fragment d'ADN obtenu par PCR dans un vecteur. Pour la ligation, le rapport vecteur : insert est établi à partir des quantités observées dans un gel d'agarose, et est le plus souvent de 1 : 3. Les conditions de la réaction sont généralement les suivantes :

Composant de la réaction	Quantité
Vecteur	1 µl
Insert	3 µl
T4 DNA ligase	1 µl
Tampon T4 ligase	1 µl
H <sub>2</sub> O	qsp 10µl

**Tableau 10.** Réaction de ligation.

La réaction de ligation s'effectue pendant 16h à 20°C. Le plasmide, qui aura intégré l'insert de la protéine d'étude, peut maintenant être stocké à -20°C ou utilisé pour transformer des bactéries électrocompétentes.



#### **d.) Préparation de l'ADN plasmidique**

La mini-préparation de l'ADN plasmidique permet l'extraction du plasmide d'intérêt à partir des bactéries. Le plasmide est incorporé dans la bactérie par **électroporation**. Ce système de transformation est une méthode mécanique de transfection dans laquelle on applique un champ électrique intense afin de perméabiliser temporairement la paroi et la membrane bactériennes. Pour le clonage nous utilisons généralement la souche d'*E. coli* **XII-blue** (*Stratagene*, Tetracycline<sup>R</sup>). Les bactéries transformées sont reprises dans 450µl de milieu LB et incubées pendant 1h à 37°C. Elles sont ensuite étalées sur boîte de pétri LB agar contenant l'antibiotique adéquat pour la sélection et incubées à 37°C sur la nuit.

Les colonies bactériennes ayant poussées sur boîte ont intégré le plasmide d'intérêt mais nous devons nous assurer que l'insert est bien présent dans ce plasmide. Pour cela, certaines colonies sont récupérées et inoculées dans 4ml de LB pendant 16h. Le lendemain, à partir de cette culture on prélève 950µl dans un nouveau tube pour effectuer une préparation d'ADN. Après centrifugation 5' à 4000xg, le LB est éliminé et le pellet est repris avec 150µl de la **solution I** (Glucose 50mM, Tris 25mM pH 8, EDTA 10mM, RNase 50µg/ml). Puis 150µl de la **solution II** (NaOH 200mM, SDS 1%) sont ajoutés et la solution est homogénéisée. Enfin, 200µl de la **solution III** (Acétate de Potassium 3M pH 4.8) sont additionnés et la solution est de nouveau homogénéisée. La solution est alors centrifugée 10' à 13000xg et le surnageant prélevé est remplacé dans un microtube neuf. 2 volumes d'isopropanol sont ajoutés afin de précipiter l'ADN plasmidique. Après centrifugation 10' à 13000xg et élimination de l'isopropanol, le culot est lavé avec 200µl d'éthanol à 70% puis séché 20' à 37°C. Le culot d'ADN est alors resuspendu dans 30µl du tampon d'éluion (5mM Tris HCl pH 8.5).

L'ADN plasmidique extrait est alors digéré avec les enzymes correspondantes et le résultat de la digestion est analysé sur gel. Les colonies dont l'ADN plasmidique digéré présentent une bande à la taille correspondant à notre insert sont sélectionnées. Le reste de la culture (3ml) est alors purifié par la méthode de minipréparation d'ADN plasmidique commerciale (Machery-Nagel<sup>TM</sup>, Qiagen<sup>TM</sup>) avant d'être envoyé pour son séquençage (MilleGen<sup>TM</sup>).

#### **e.) Application des techniques du clonage pour la génération de mutants**

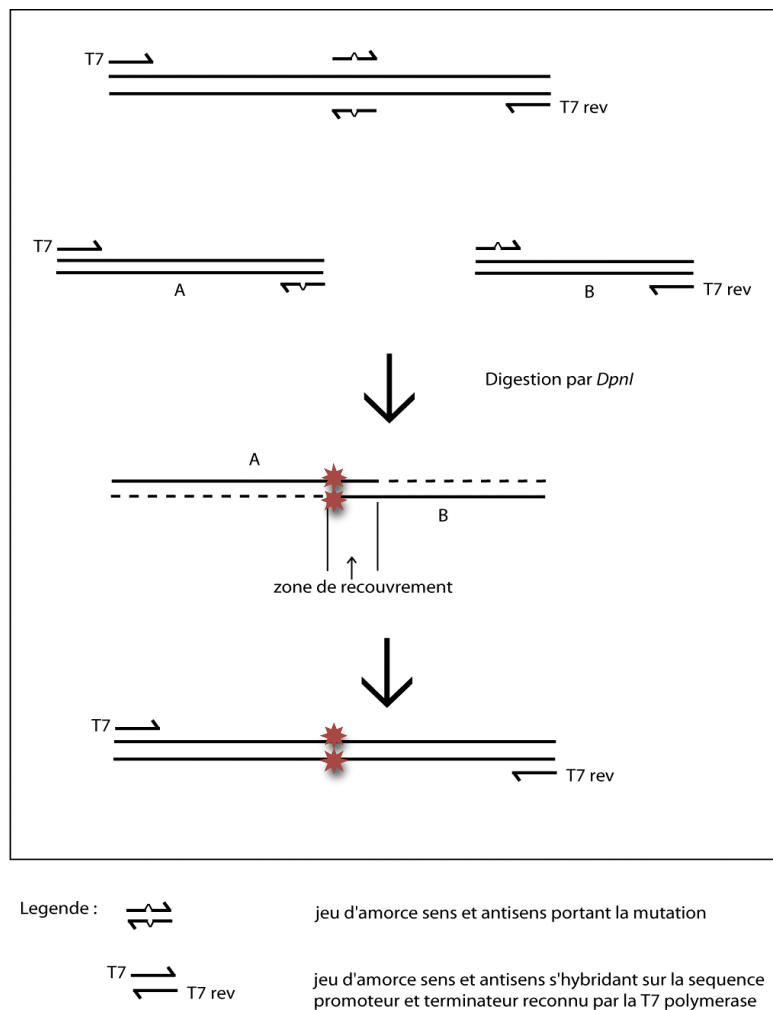
Dans le *Chapitre II : Résultats*, nous avons présenté l'utilisation de mutants de différentes protéines pour des expériences précises. En plus de ces mutants, d'autres ont été créés pour la vérification des structures obtenues en RMN ou pour des expériences d'interaction. Dans la plupart de cas, les mutations étaient ponctuelles, un seul résidu est substitué par un autre (pour CstF-50hs un quadruple mutant a été aussi réalisé mais le principe est similaire).

La mutagenèse dirigée par PCR peut s'effectuer par une méthode directe où l'on amplifie directement le plasmide à partir d'une matrice qui contient l'insert non muté et deux amorces contenant la mutation. La réaction est réalisée avec une enzyme ayant un très faible taux d'erreur

(*Pfu* ou *Deep Vent*). Les amorces contiennent en plus une séquence d'un minimum de 10 bases s'appariant strictement à la matrice de part et d'autre de la mutation.

Cependant, la méthode qui a permis l'obtention de mutants dans les cas de ma thèse est « **l'overlap extention** ». Cette méthode se déroule en deux étapes schématisées sur la *figure 89*. Dans la première étape on réalise en parallèle deux réactions de PCR, où l'on amplifie le fragment d'ADN contenant la mutation pour les brins sens et antisens, dans notre cas à l'aide des amorces T7 et T7 reverse respectivement. On obtient alors deux fragments d'ADN ayant une zone de recouvrement de la longueur de l'amorce utilisée pour introduire la mutation. On digère alors l'ADN parental par l'enzyme *DpnI*. Dans la deuxième étape on mélange les produits de PCR de l'étape précédente. A ce stade, les fragments d'ADN serviront à la fois de matrice et d'amorce. On réalise alors une troisième PCR où les oligonucléotides utilisés sont les amorces externes des premières PCRs.

L'insert portant la mutation est alors digéré avec les enzymes de restriction et inséré dans le vecteur souhaité à travers une réaction de ligation. Pour vérifier la présence de la mutation, le plasmide final doit être séquencé (dans notre cas en MilleGen<sup>TM</sup>).

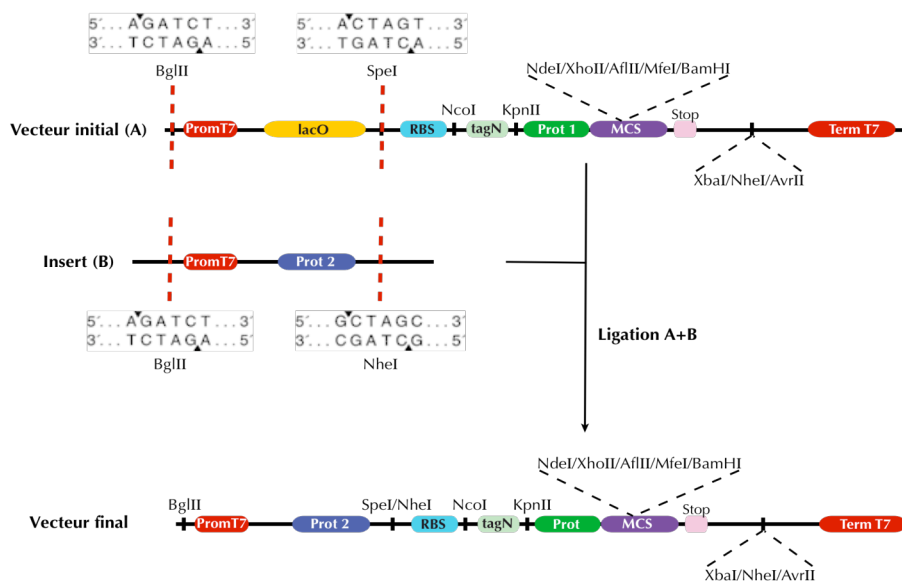


**Figure 89.** Méthode « overlap extension ». L'étoile rouge représente la mutation.

### f.) Un cas particulier : les plasmides bicistroniques

On dit qu'un plasmide est bicistronique quand il contient deux « cistrons » qui vont coder pour deux protéines, ou fragments de protéines, différentes. Nous avons trouvé un intérêt à son utilisation quand la transformation avec plus de trois plasmides a été nécessaire pour nos expériences. Par exemple, dans le cas de la reconstitution du complexe CF IA, la co-transformation des bactéries avec 4 plasmides différents impliquait l'utilisation 4 antibiotiques différents pour la sélection de bactéries portant l'ensemble de plasmides. Ceci rendait la croissance de colonies très difficile dans le milieu. De cette façon, plusieurs vecteurs intégrant 2 fragments d'ADN qui codent pour deux protéines différentes ont été créés. Un de ces plasmides bicistroniques a permis la reconstitution de CF IA. Il s'agit de **H<sub>6</sub>Rna15\_Rna14** (clonage réalisé par Cécile Monfoulet, technicienne du laboratoire), un vecteur *pETMCN\_γH* qui intègre au même temps la protéine Rna15p avec une étiquette de 6 His sur son extrémité N-terminale et la protéine Rna14p.

Pour la formation de ce type de vecteurs, on part d'un plasmide initial qui contient la première protéine avec une étiquette (p. ex. H<sub>6</sub>-Rna15p) ou sans elle (p. ex. Ka-Rna15p). Ce vecteur constitue la base pour l'intégration de la séquence d'ADN codant pour la deuxième protéine. Le vecteur est digéré avec les enzymes *BglIII* et *SpeI* pendant au moins 2h à 37°C et il est ensuite déphosphorylé grâce à l'antartic phosphatase comme indiqué (*section V.A.1b*). L'insert qui sera incorporé dans le vecteur (p. ex. Rna14p) est préparé par PCR et ensuite digéré avec les enzymes *BglIII* et *NheI* comme il a été montré (*section V.A.1a*). La ligase permet l'insertion du fragment d'ADN obtenu dans le vecteur préalablement digéré. Il faut noter que, avec les vecteurs utilisés, la ligation ne fonctionnerait pas si les séquences du site reconnu par *BglIII* n'étaient pas palindromiques et si les sites de restriction pour *NheI* et pour *SpeI* ne coïncidaient pas (*fig. 90*).



**Figure 90.** Schéma représentant les étapes du clonage pour la formation des vecteurs bicistroniques.

## 2. Systèmes d'expression

### a.) Expression chez *Escherichia coli*

#### a.1.) Souches bactériennes utilisées pour l'étude des niveaux d'expression

Le système général d'expression utilisé dans mes travaux a été *E. coli*. Pour les tests de surexpression, différentes souches de cette bactérie sont systématiquement transformées par électroporation avec le plasmide correspondant. Le choix de la souche a été fait à partir d'une analyse sur un gel SDS-PAGE.

Dans la plupart de cas, des différences importantes d'expression ont été observées selon la diversité de souche commerciale employée (tableau 11).

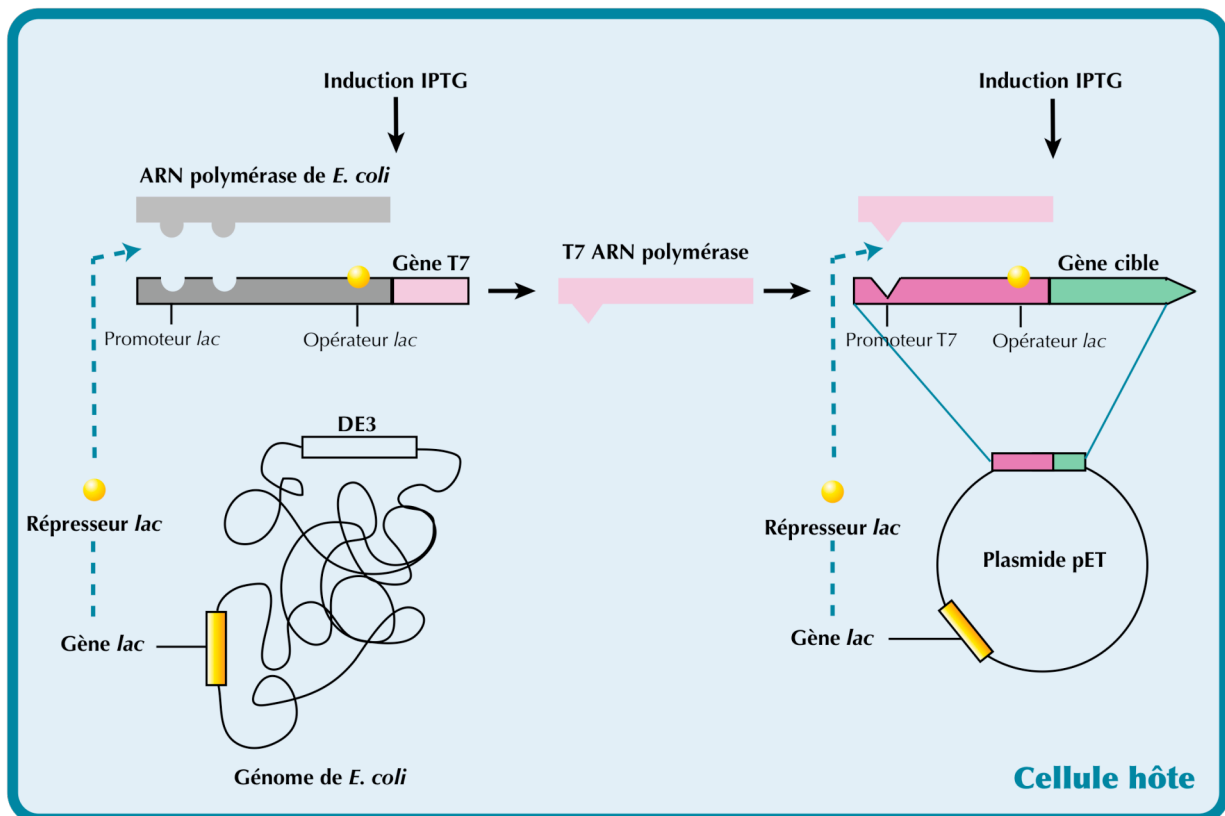
Souche	Caractéristiques principales	Résistance
BL21 (DE3) (Novagen)	Haut niveau d'expression. Facile induction	-
DL41 (DE3) (C Mueller)	Auxotrophe pour Met (pour marquage avec Se-Met)	-
B834 (DE3) (Novagen)	Auxotrophe pour Met (pour marquage avec Se-Met)	-
Rosetta <sup>TM</sup> (DE3) (Novagen)	Expression de protéines avec codons rares chez <i>E. coli</i>	Cam <sup>R</sup>
BL21 Start <sup>TM</sup> (DE3) pRare (Invitrogen)	Augmente la stabilité de l'ARNm	Cam <sup>R</sup>
BL21 Gold (DE3) pLysS (Stratagene)	Expression de protéines toxiques pour <i>E. coli</i>	Cam <sup>R</sup> Tet <sup>R</sup>
BL21 CodonPlus <sup>®</sup> (DE3) (Stratagene)	Expression de protéines avec codons rares chez <i>E. coli</i>	Cam <sup>R</sup> Tet <sup>R</sup>
ArticExpress <sup>TM</sup> (DE3) (Stratagene)	Active le repliement de protéines et leur solubilité	Tet <sup>R</sup> Gent <sup>R</sup>
Rosetta-gami (DE3) (Novagen)	Expression de protéines avec codons rares chez <i>E. coli</i>	Cam <sup>R</sup> Kan <sup>R</sup> Tet <sup>R</sup>

**Tableau 11.** Principales souches d'*E. coli* utilisées au laboratoire pour l'expression de protéines. Cam<sup>R</sup> indique une résistance au chloramphénicol, Tet<sup>R</sup> à la tétracycline, Kan<sup>R</sup> à la kanamycine et Gent<sup>R</sup> à la gentamycine.

En plus de la grande variété de souches disponibles au laboratoire, celles n'ayant pas de résistance au chloramphénicol (par exemple BL21, B834 ou DL41), peuvent être transformées au choix avec un des deux plasmides, **pLysS** ou **pRare**, qui confèrent, en plus d'une résistance au chloramphénicol, des propriétés additionnelles qui peuvent améliorer les niveaux d'expression. Le plasmide pLysS (*Stratagene*) est employé dans les bactéries  $\lambda$ DE3 lysogènes pour supprimer l'expression basale du promoteur T7 en produisant le lysozyme T7, un inhibiteur naturel de l'ARN polymérase T7. Le plasmide pRare (*Novagen*) permet la transcription de certains ARNs de transfert dont l'expression est faible chez *E. coli*.

### a.2) Surexpression des protéines dans la bactérie

La surexpression de la protéine d'intérêt est possible parce que le gène d'étude (« gène cible » sur la figure) est placé dans un vecteur (pET) sous contrôle d'un **promoteur fort de la T7 ARN polymérase**. Dans la bactérie le gène de la T7 ARN polymérase est régulé par le **promoteur lacUV5** apporté par recombinaison à partir de l'ADN du prophage  $\lambda$ DE3. Le promoteur du gène T7 et celui du gène cible sont bloqués par le **répresseur lac** dans des conditions « normales ». L'adjonction d'**IPTG** (Isopropyl-beta-thio-galactosido) au milieu de culture inactive ce répresseur permettant la libération des deux opérateurs *lac*, celui du gène d'intérêt et celui de la T7 polymérase. L'ARN polymérase bactérienne transcrit le gène de la T7 polymérase qui à son tour transcrit le gène cible. L'effet final est l'induction de l'expression de la protéine cible (fig. 91).



**Figure 91.** Mécanisme de fonctionnement du système de surexpression chez *E. coli*. Les formes représentant des protéines, pour les différencier de celles des gènes, n'ont pas été contournées en noir.

**b.) Un autre système d'expression : TNT, traduction 'in vitro'**

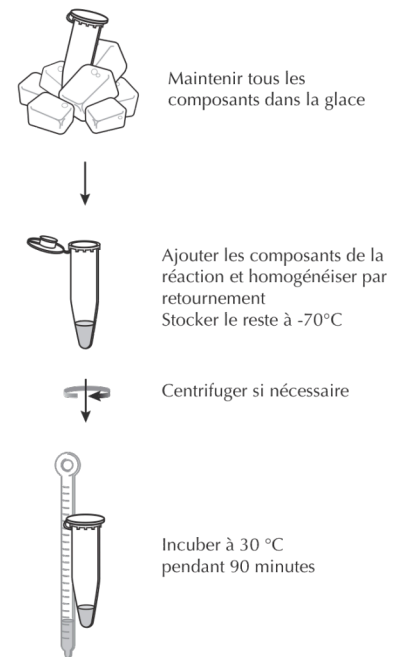
Pour la réalisation des expériences d'interaction entre CstF-50 et le CTD de l'ARN pol II nous nous sommes servis d'un kit commercial (TNT<sup>®</sup> Rabbit Reticulocyte Lysate System, *Promega*) qui permet la traduction *in vitro* d'un gène cible dans un système couplé au lysat de réticulocytes. Ce système présente des avantages : i) le processus est très simple, ii) le temps d'exécution pour l'obtention des protéines est minime comparé à d'autres systèmes d'expression, iii) le système est eucaryote, iv) le kit permet le marquage des différents acides aminés avec un isotope radioactif, v) en dehors du kit, nous ne requérons que nos plasmides et des acides aminés marqués. Néanmoins, le kit est très cher et la quantité de protéine obtenue est faible et implique sa détection par radioactivité.

Nous avons suivi le protocole général présenté sur la *figure 92*. Un détail de composants est donné sur le tableau ci-dessous :

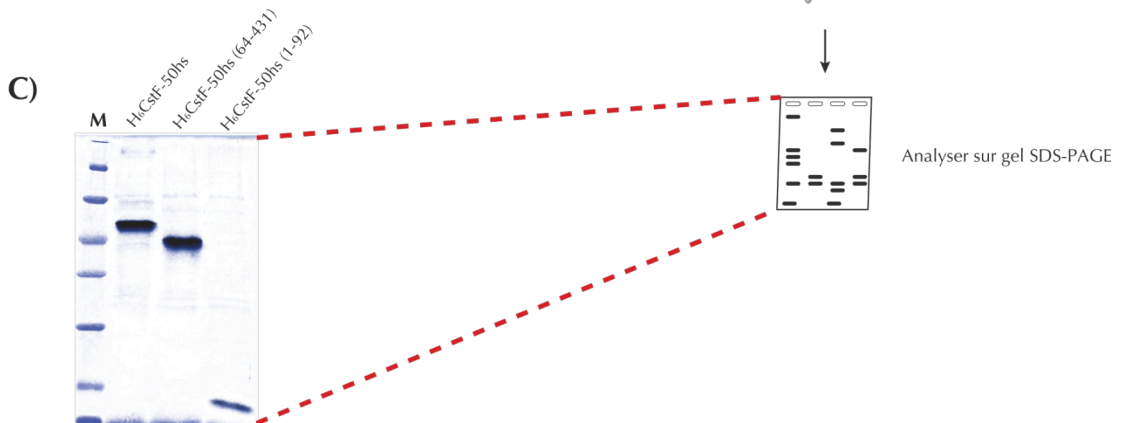
A)

Composants	Quantités
TNT <sup>®</sup> Lysat de Réticulocyte de lapin	25µl
Buffer de réaction TNT <sup>®</sup>	2µl
TNT <sup>®</sup> T7 RNA Polymérase	1µl
Mélange d'acides aminés sans Méthionine, 1mM	1µl
[ <sup>35</sup> S]methionine (1,000Ci/mmol à 10mCi/ml)*	2µl
Inhibiteur de Ribonucléase (40u/µl)*	1µl
Plasmide(s) (0.5µg/µl)*	2µl
H <sub>2</sub> O pour un volume final de :	50µl

B)



**Tableau 12.** Les quantités ont été calculées pour une réaction standard avec <sup>35</sup>S-Met. \*indique les composants qui ne sont pas fournis dans le kit commercial



**Figure 92.** A) Tableau 12, composants nécessaires pour les expériences de TNT. B) Schéma du protocole du TNT. C) Analyse des résultats obtenus pour différentes constructions de [<sup>35</sup>S]CstF-50hs. Pour la figure, le film autoradiographique a été superposé au gel SDS-PAGE 18% coloré au Coomassie.

### 3. Etudes d'interaction

#### a.) Western Blot

Le transfert de protéines (« Western Blot ») est une méthode de biologie moléculaire permettant la détection et l'identification de protéines spécifiques dans un échantillon biologique.

Nous avons utilisé cette technique pour des études d'interaction de protéines. Dans nos expériences, une des protéines contenant une étiquette (de 6 His ou de la protéine GST) est immobilisée sur la résine correspondante et mise en incubation avec un extrait bactérien, total ou soluble. Cet extrait contient la protéine potentiellement impliquée dans une interaction avec la première. Après plusieurs lavages, la résine est chargée sur gel et si une interaction existe, la protéine présente dans l'extrait pourra être détectée dans cette résine. Le « Western Blot » nous permet une identification plus sensible et plus précise des interactions par rapport à une détection avec le colorant bleu de Coomassie.

Le protocole est présenté ci-dessous:

- Electrophorèse des protéines en conditions dénaturantes. Le marqueur de taille est précoloré (PageRuler™ Prestained Protein Ladder, *Fermentas*) afin d'être visible lors du transfert sur la membrane.
- Transfert des protéines contenues dans le gel d'acrylamide (« blottées ») sur une **membrane de nitrocellulose**. Nous utilisons un système BioRad (Mini Trans-Blot®) capable d'effectuer le transfert en 60' à 100 Volt. Le tampon contient Tris Base 25mM pH=7.5, Glycine 192mM et éthanol 20%.
- Blocage de la membrane dans une solution de PBS-tween (PBS 1x, tween 20 0,05%) additionnée de 5% de lait pendant 1h.
- Lavage de la membrane trois fois 10' au PBS-tween.
- Incubation avec l'anticorps dirigé contre la protéine (**premier anticorps**) toute la nuit, à 4°C, sous agitation (voir *tableau 13* pour le type d'anticorps primaires utilisés et la composition de la solution qui les contient).
- Lavage trois fois 10' au PBS-tween.
- Incubation de la membrane avec **l'anticorps secondaire** (anti IgG) dirigé contre le premier anticorps, 2h, à température ambiante, sous agitation (*tableau 13*).
- Lavage trois fois 10' au PBS-tween.
- Pour la révélation nous utilisons des solutions commerciales (SuperSignal® West Pico Chemiluminescent Substrate, *Thermo Scientific*). La membrane est recouverte de ces solutions pendant 5' et une fois qu'elle est sèche, on la scelle dans une enveloppe de film étirable ('Saran') et on l'introduit dans la cassette. Dans une chambre noire, le film est placé dans la cassette sur la membrane. Après un temps allant de quelques minutes à

plusieurs heures, le film est plongé dans un liquide révélateur et un fixateur qui permettent l'apparition des bandes correspondantes aux protéines reconnues par les anticorps.

Anticorps	Quantité	Solution
Anti-His Monoclonal (anticorps primaire) (GE Healthcare)	1 : 3000	PBS 1x, Tween, BSA 3%
Anti-GST Monoclonal (anticorps primaire) (Covance)	1 : 1000	PBS 1x, Tween, BSA 3%
Anti-mouse IgG (anticorps secondaire) (GE Healthcare)	1 : 5000	PBS 1x, Tween, Lait 5%

Tableau 13. Anticorps utilisés pour les expériences de « Western Blot ».

La **révélation** s'effectue de manière indirecte grâce au couplage de l'anticorps secondaire à l'enzyme **HRP** (Horse Radish Peroxidase). L'enzyme HRP catalyse l'oxydation du luminol en présence de peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>) ce qui entraîne l'émission de lumière (à 430nm). Ce type de **détection** du complexe antigène/anticorps est appelée **chimioluminescence**, c'est-à-dire, effectuée par une émission de lumière due à une réaction chimique (fig. 93). La lumière est alors détectée par le film photographique, ou plus récemment par des appareils photographiques de CCD (Charge-Coupled Device) disponibles à l'IECB.

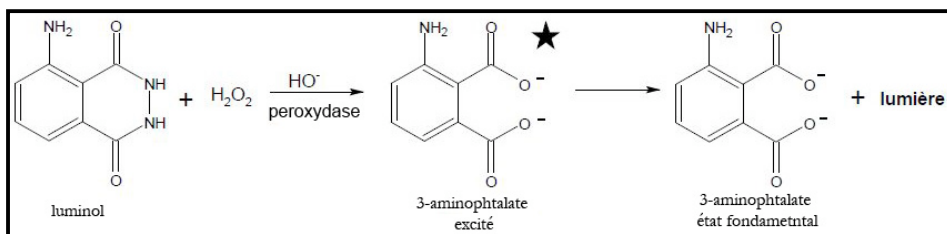
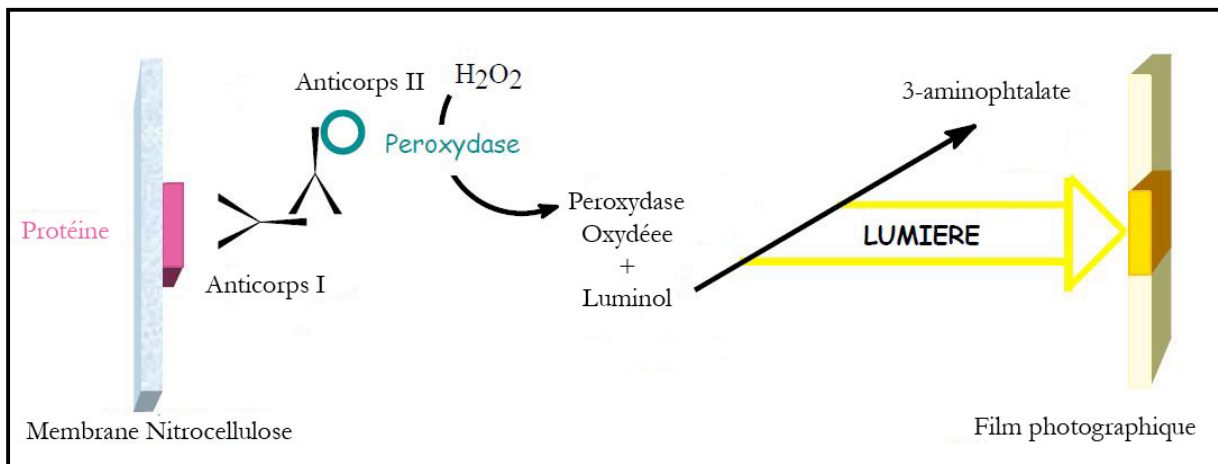


Figure 93. Principe et réaction de chimioluminescence lors d'un Western Blot.



## b.) Double hybride

La découverte du système du double hybride en 1989 par Stanley Fields et ses collaborateurs (Fields and Song 1989) a révolutionné la manière dont les interactions protéine-protéine pouvaient être détectées. La technique du double hybride chez la levure permet d'identifier l'interaction physique entre deux protéines X et Y grâce à l'activation d'un **gène rapporteur**.

Le principe est basé sur la reconstitution d'un facteur de transcription fonctionnel, généralement **GAL4**. Le facteur GAL4 contient deux domaines, un de **liaison à l'ADN (DBD ou DNA Binding Domain)** et un autre **activateur (AD ou Activation Domain)**. Le gène codant pour la **protéine « appât »**, dans notre cas GST-CTDhs, est fusionné au domaine de liaison DBD intégré dans le vecteur **pGBK T7**. DBD reconnaît une séquence spécifique de l'ADN, promotrice du gène rapporteur. La **protéine « proie »**, CstF-50hs ou des fragments de celle-ci, est fusionnée au domaine activateur de la machinerie transcriptionnelle (AD). Ces deux constructions hybrides sont introduites dans une levure diploïde qui exprimera les deux protéines fusion (BD-appât et AD-proie). La transcription du gène rapporteur n'aura lieu que si la protéine appât et la protéine proie interagissent (*fig. 94*).

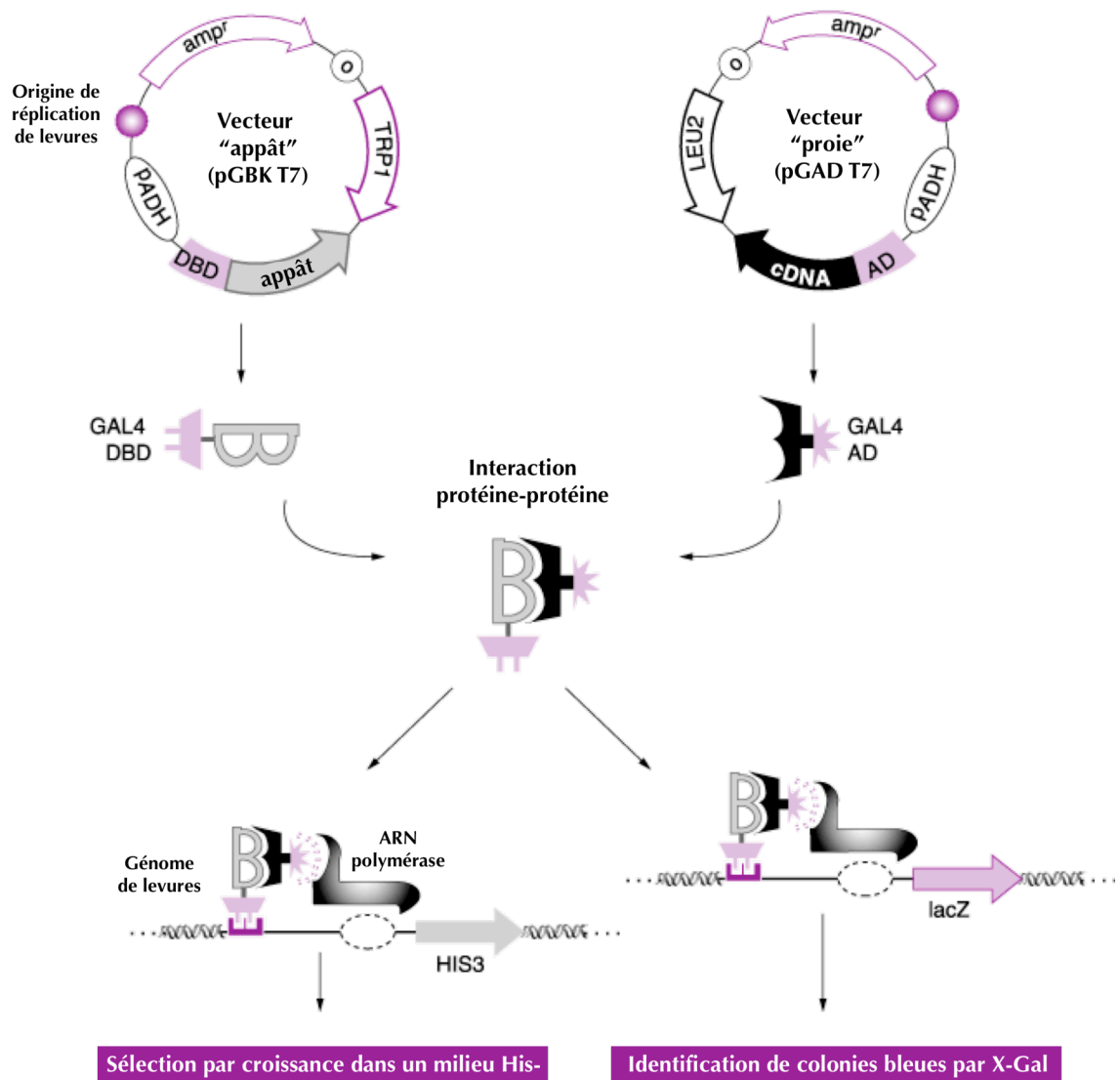
Nous avons utilisé deux types de gènes rapporteurs. La reconstitution du facteur de transcription fonctionnel GAL4 peut activer la production du **marqueur d'auxotrophie HIS3**, ce qui permet à son tour aux cellules de levure His<sup>-</sup> de se développer dans un milieu sélectif dépourvu d'histidine. L'autre gène rapporteur est **lacZ**. Ce gène code pour la protéine  $\beta$ -galactosidase, une enzyme capable d'hydrolyser le lactose en galactose et glucose. La présence de l'enzyme est facilement détectable en présence de X-Gal puisqu'elle donne une coloration bleue aux colonies.

Le protocole suivi implique plusieurs étapes qui sont résumées ci-dessous:

### b.1.) Préparation de plasmides

Le clonage a été réalisé comme présenté précédemment. Nous avons utilisé les vecteurs **pGBK T7** et **pGAD T7**. CstF-50hs, et les régions de cette protéine (1-63), (64-431), (92-431) ont été insérées dans le vecteur pGAD T7. Les enzymes de restriction utilisées ont été *NdeI* et *BamHI*. La protéine recombinante GST-CTDhs a été clonée dans pGBK T7 grâce aux mêmes enzymes de restriction.

Comme **contrôle négatif** nous avons eu besoin de pGBK T7 et pGAD T7 seuls et comme **contrôle positif**, pGBK53 et pGAD.T, deux plasmides qui codent pour DBD-p53 murine et pour l'antigène T du virus SV40, respectivement. Ces deux protéines sont connues pour interagir dans des essais de double hybride.



**Figure 94.** Technique de double hybride. Le vecteur « appât » porte le domaine de GAL4 de liaison à l'ADN et le vecteur « proie » le domaine activateur. L'interaction entre les deux protéines provoque l'interaction indirecte de ces deux domaines. Ceci permet l'activation de l'ARN pol II qui transcrit le gène de résistance à l'histidine ou de la protéine lacZ. pADH : promoteur ADHI ; TRP1 : gène de résistance au tryptophane ; LEU2 : gène de résistance à la leucine ; HIS3 : gène de résistance à l'histidine.

### b.2) Transformation de *S. cerevisiae*

La souche de levure **Y187** sera transformée avec les plasmides pGBK T7 et la souche **AH109** avec les pGAD T7.

**Jour 1, pré-cultures** : 3ml de culture de cellules Y187 et 5ml de culture de AH109 en milieu YPD liquide (5g d'extrait de levure, 10g de bactopeptone, 10g de glucose dans 500ml d'eau) sont mis en croissance pendant une nuit sous agitation à 30°C. Le volume de culture dépend du nombre de constructions avec lesquelles doivent être transformées les deux souches.

**Jour 2, transformation**: un plus grand volume de culture estensemencé à partir des pré-cultures et mis en croissance pendant 4h à 30°C. Les cellules sont alors centrifugées et lavées deux fois avec de l'eau stérile. Un dernier lavage avec du tampon TEAcLi est réalisé. Ce tampon est préparé à partir de deux solutions stock aux quelles nous ajoutons 50% du PEG<sub>4000</sub> :

- TE 10x : 0,1M Tris HCl pH 7.5, 0,01M EDTA ;
- LiAc 10x : 1M Lithium acétate pH ajustée à 7.5 avec acide acétique ;

Le culot est finalement repris dans 300 à 500µl du tampon TEAcLi (le volume est dépendant du nombre des constructions avec lesquelles chaque souche a été transformée au début).

Dans des microtubes, on prépare 3µl d'ADN de saumon, 3µl de chaque plasmide, 150µl de la solution contenant les levures (pour pGBK on ajoute Y187 et pour pGAD, AH109). A ce mélange il faut aussi additionner 700µl du PEGAcLi (même tampon sans TE). Incuber 1h à 30°C.

Les levures sont alors soumises à un **choc thermique**, 10' à 42°C. Après centrifugation 30'' à 5000rpm et la solution PEGAcLi est retirée. Nous pouvons maintenant ajouter 150µl d'eau stérile dans chaque tube et étaler sur des boîtes de milieu minimum SD gel (3,35g de YNB ou Yeast Nitrogen Base, 0,32g de CSM qui est un mélange d'acide aminés sans Leu et sans Trp, 10g de D-Glucose et 10g d'agar sur 500ml d'eau). Pour la souche Y189 transformée avec les plasmides pGBK on additionnera de la Leu, étant donné que ce plasmide apporte le Trp. Pour sélectionner les cellules AH109 qui ont incorporé le plasmide pGAD il faut ajouter dans le milieu du Trp. Nous laissons pousser 2 jours à 30°C.

**Jour 4, conjugaison (formation de diploïdes)** : on prélève avec une pointe des levures Y187 et des levures AH109 pour les mélanger dans 2ml d'un milieu YDP. Le but est de former des diploïdes contenant la combinaison des plasmides adéquate. On laisse 24h à 30°C sous agitation.

**Jour 5, sélection de diploïdes** : 100µl de chaque culture liquide sont étalés sur des boîtes SD (Leu<sup>-</sup>, Trp<sup>-</sup>). Nous laissons pousser les levures contenant les deux types de plasmide (pGBK et pGAD), 48h à 30°C.

**Jour 7, préparation de patches** : on réalise sur une boîte avec un milieu SD (Leu<sup>-</sup>, Trp<sup>-</sup>), 6 patches de chaque diploïde obtenu. Ces boîtes sont incubées 24h à 30°C.

**Jour 8, test d'activité β-galactosidase** : nous devons préparer dans un « bain marie » à 60°C une solution qui contient :

- 5ml agarose 1%
- 5ml tampon Potassium phosphate 1M à pH 7
- 600µl NN-diméthylformamide
- 100µl SDS 10%
- 200µl X-Gal 2% (dilué dans NN-diméthylformamide)

Les composants doivent être incorporés dans l'ordre indiqué ci-dessus. Les quantités indiquées ici sont pour une boîte de pétri. Cette préparation est alors versée sur la boîte qui est immédiatement mise à incuber à 37°C pendant 24h.

**Jour 9, analyses de résultats :** les patches colorés en bleu sont indicatifs de l'existence d'une interaction.

Nous avons effectué un test avec le **gène rapporteur HIS3** après avoir pris connaissance des résultats négatifs, mais ces expériences peuvent être réalisées à partir du jour 7, quand nous avons sélectionné les diploïdes.

**Jour 7', étalement sur des boîtes SD Leu<sup>-</sup>, Trp<sup>-</sup>, His<sup>-</sup> :** cette fois-ci nous pouvons étaler en strie sur une boîte SD Leu<sup>-</sup>, Trp<sup>-</sup>, His<sup>-</sup> des levures de chaque diploïde obtenu. En parallèle, on réalise la même opération sur un milieu SD Leu<sup>-</sup>, Trp<sup>-</sup> qui servira de contrôle. On laisse pousser 24h.

**Jour 8', analyse de résultats :** les levures poussent sur le milieu sans His seulement si une interaction existe entre les deux protéines d'étude.

## B. Protocoles de Biochimie

### 1. Tests d'expression

Dans notre protocole général, le plasmide contenant le gène d'étude est intégré dans les différentes souches bactériennes par électroporation. Les bactéries poussent dans un milieu LB agar contenant l'antibiotique adéquat toute la nuit à 37°C. Le lendemain, après incubation dans un milieu liquide (10ml/souche) jusqu'à une  $DO_{600nm} = 0.6-0.8$ , la production de la protéine est induite par l'addition de **1mM d'IPTG** à **15°C** pendant la nuit.

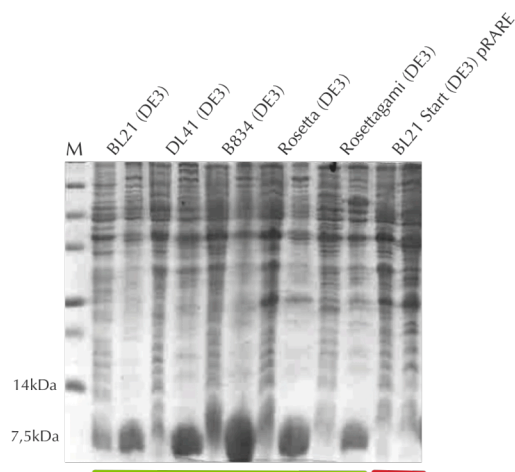
Le lendemain les cultures sont centrifugées, pour éliminer le milieu et elles sont reprises avec 1ml du tampon de lyse (PBS 1,5x, MgAcétate 1mM, NP-40 0,1%, Imidazole 20mM, Glycerol 10%). La méthode employée pour la lyse est la sonication (Sonifer 150, *Branson*). A partir du lysat, 50µl serviront pour contrôler l'expression de la protéine dans l'extrait total (les protéines surexprimées mais non-solubles peuvent ainsi être remarquées) et le reste est centrifugé à 13000rpm, 10' à 4°C. Le surnageant résultant est incubé pendant 30' sous agitation avec 30µl d'une **résine  $Co^{2+}$**  (HIS-Select® Cobalt Affinity Gel, *Sigma*), si la protéine contient une étiquette  $H_6$  ou avec 30µl d'une **résine glutathionne-sepharose** (Gluthatione Sepharose™ 4 Fast Flow, *Amersham Biosciences*), si la protéine recombinante porte la protéine GST comme étiquette.

Pour les essais de « pull-down », nous effectuons 3 lavages de la résine, avec le même tampon de lyse. Les niveaux d'expression sont analysés sur un gel SDS-PAGE où 10µl de la résine, à laquelle on a ajouté le tampon de charge (Laemmli Buffer), sont directement chargés dans les puits. Le gel migre à 200-220V, ≈45' puis il est coloré avec une solution de Bleu de Coomassie (*fig. 95*).

Une optimisation peut être réalisée en variant les paramètres suivants :

- température d'induction
- tampon de lavage
- concentration finale d'IPTG
- temps d'incubation avec la résine
- nombre de lavages réalisés

Les tests d'expression nous permettent d'identifier la souche la plus adéquate pour la surexpression de la/les protéine/s d'étude. De plus, le même protocole a été adapté pour des expériences d'interaction entre la protéine CstF-50 et le CTD de l'ARN pol II. Pour ces expériences, les protéines GST et GST-CTDhs, liées à la résine par la méthode indiquée ci-dessus, étaient incubés avec les extraits totaux de bactéries contenant les différents fragments de la protéine CstF-50hs (*voir Chapitre II A.2.g*).



**Figure 95.** Exemple d'un test d'expression ( $H_6CstF-50dm$  (1-65)). La couleur verte indique les souches dans lesquelles la protéine est surexprimée. On observe une différence importante pour la souche située à droite sur le gel. Pour chaque souche l'extrait bactérien total est chargé sur le gel suivi de la résine.

## 2. Production de protéines à grande échelle

### a.) Production de protéines dans des conditions natives

La souche bactérienne choisie à partir des tests d'expression, dans la plupart de nos cas BL21 (DE3), est transformée avec l'ADN plasmidique contenant le gène d'intérêt. Les cultures à grande échelle sont réalisées dans du milieu LB liquide à 37°C sous agitation. Le nombre de litres utilisés varie en fonction des niveaux de surexpression de la protéine et de la finalité de l'expérience. Pour les essais de cristallisation nous avons besoin de quantités importantes des facteurs protéiques. Un minimum de 2 litres et un maximum de 6 ont été utilisés pour la production des différentes protéines. Lorsque la densité optique à 600nm atteint une valeur d'environ 0.6 ( $DO_{600nm} = 0.6-0.8$ ), les cultures sont refroidies à 15°C pendant 2h puis l'induction avec 1mM final d'IPTG est effectuée sur la nuit à 15°C. Le lendemain, les cultures sont centrifugées à 4200 rpm pendant 15' et à 4°C. Le culot est repris avec un tampon contenant du Tris HCl 50mM pH 7.5 et du NaCl 150mM et peut être conservé à -20°C ou utilisé pour une purification immédiate de la protéine.

### b.) Production de protéines marquées avec du Se pour des études cristallographiques

L'étude structurale de  $CstF-50dm$  (1-65) a requis la substitution de certains atomes de soufre de la protéine par des atomes de Sélénium. Ce marquage peut être effectué par incorporation dans le milieu de culture de Se-Met ou Se-Cys. Pour nos expériences nous avons choisi la méthode plus souvent utilisée, l'introduction de **Se-Met** dans la séquence de la protéine. Pour le déroulement de la méthode nous disposons au laboratoire de deux souches bactériennes auxotrophes pour la Met, B834 (DE3) et DL41 (DE3). Nous avons choisi **B834 (DE3)** où les niveaux d'expression étaient plus importants.

Ainsi, la souche B834 (DE3) est transformée avec notre plasmide et étalée sur une boîte de Pétri avec du milieu LB agar contenant l'antibiotique de résistance du plasmide (Amp). Le lendemain, les bactéries qui ont incorporé le plasmide H<sub>6</sub>CstF-50dm (1-65) sont transférées à un milieu LB liquide jusqu'à l'obtention d'une DO<sub>600nm</sub> = 1.0-1.4, le temps de croissance est d'approximativement 6-7h. Quand les cultures atteignent cette DO, elles sont centrifugées à 4200 rpm 15' à 4°C. Le culot est alors repris dans un **milieu minimum** dont la composition est représentée sur le tableau ci-dessous (tableau 14).

Les cellules sont incubées toute la nuit à 15°C sous agitation. La Se-Met (0,05mg/ml) est alors ajoutée et la culture est à nouveau incubée à 15°C pendant 45', temps nécessaire pour l'incorporation de la Se-Met par les bactéries. L'induction avec 1mM d'IPTG (1mM) est effectuée à 15°C pendant une durée minimum de 4h-6h. Les cultures sont alors centrifugées de la même façon que pour les conditions natives.

Composants	Quantités pour 1litre de culture
Solution M9 (10x): - Na <sub>2</sub> HPO <sub>4</sub> : 80g - KH <sub>2</sub> PO <sub>4</sub> : 40g - NaCl : 5g - NH <sub>4</sub> Cl : 5g - Eau Millipore : 1litre (Note : stériliser et conserver à 4°C)	100ml
Solution d'éléments trace : EDTA, FeCl <sub>2</sub> , ZnCl <sub>2</sub> , CuCl <sub>2</sub> , CoCl <sub>2</sub> , H <sub>3</sub> BO <sub>3</sub> , MnCl <sub>2</sub>	-
Glucose 20% m/v (stériliser)	20ml
MgSO <sub>4</sub> 1 M (filtrer)	1ml
CaCl <sub>2</sub> 1 M (stériliser/filtrer)	0,3ml
Thiamine 1mg/mL (filtrer)	1ml
Biotine 1mg/mL (filtrer)	1ml
Solution d'acide aminés 50x - 0,5g de chaque acide aminé (sauf la Méthionine) sont ajoutés à 250ml de H <sub>2</sub> O. Ajusté pH à 7.5	20ml
Se-Méthionine (10mg/ml)	5ml
Eau Millipore	Qsp 1l

**Tableau 14.** Composition du milieu minimum pour la préparation des protéines marquées avec Se-Met. La Se-Met est incorporée au dernier moment au milieu comme vu ci-dessus. Généralement l'ajout des éléments trace n'est pas nécessaire.

**c.) Production de protéines marquées avec du  $^{15}\text{N}$ ,  $^{13}\text{C}$  pour des études en RMN**

Pour les expériences de RMN les protéines ont été marquées à l'azote  $^{15}\text{N}$  et, dans certains cas, un marquage additionnel avec les isotopes carbone  $^{13}\text{C}$  et/ou deutérium ( $^2\text{H}$ ) a été effectué. En effet, pour obtenir les empreintes RMN des protéines seules ou en présence de ligand (dans le cas du CTD), nous avons réalisé un marquage  $^{15}\text{N}$  et/ou  $^{13}\text{C}$ . Le marquage  $^{15}\text{N}$  permet d'observer par RMN la région des protons amides de la protéine en enregistrant des spectres à deux dimensions (HSQC ou TROSY  $^1\text{H} - ^{15}\text{N}$ ). Le marquage avec  $^{13}\text{C}$  permet de retrouver tous les carbones de la molécule grâce à la connaissance empirique des déplacements chimiques des carbones faisant partie de divers groupements fonctionnels.

Le protocole ne diffère pas énormément de celui utilisé pour le marquage avec Se-Met, cependant, cette fois-ci nous partons d'une pré-culture de 15ml de LB de la souche **BL21 (DE3)** transformée. Nous laissons pousser les bactéries la nuit à  $37^\circ\text{C}$  et le lendemain nous centrifugeons pour récupérer le culot dans un **milieu minimum M9 complété** (tableau 14). Pour le marquage, la composition du milieu doit être modifiée. L'utilisation du  $^{15}\text{NH}_4\text{Cl}$  à la place de  $\text{NH}_4\text{Cl}$  détermine la production d'une protéine avec les acides aminés marqués à  $^{15}\text{N}$  (on ajoute 10g par litre de culture). Le  $^{13}\text{C}$  est incorporé avec un **glucose** commercial spécial qui substitue le glucose non marqué dans le milieu. 2g de ce glucose peuvent être additionnés directement en poudre au milieu. Dans ce type de marquage un des éléments traces, le  $\text{FeCl}_2$ , est aussi additionné (1ml de  $\text{FeCl}_2$  1M). Quand la  $\text{DO}_{600\text{nm}}=1$ , nous pouvons induire la culture avec 1mM d'IPTG. La température optimale pour toutes les expériences réalisées était de  $25^\circ\text{C}$ , probablement parce que les bactéries qui ont poussé très lentement, à cause du milieu minimum, peuvent continuer à le faire doucement pendant la nuit. Le lendemain les cultures sont centrifugées.

### ***3. Purification de protéines***

La purification des protéines présentées a nécessité l'utilisation de quatre types de chromatographie. Les premières étapes sont communes à toutes les protéines présentées. La lyse a lieu dans un **homogénéisateur à haute pression** (EmulsiFlex<sup>®</sup>-C3) qui permet la rupture des parois des bactéries par effet de la pression. Pour une lyse efficace de *E. coli*, 3 ou 4 passages à une pression de 1000-1200 bars sont nécessaires. Après centrifugation à 50000 x g, pendant 1h à  $4^\circ\text{C}$ , le surnageant est récupéré et incubé avec la résine  $\text{Co}^{2+}$ , 30' sous agitation à  $4^\circ\text{C}$ .

Ensuite, la résine est centrifugée 2' à 1000rpm à  $4^\circ\text{C}$  et le surnageant est prélevé. Un premier lavage avec un tampon Tris HCl 50mM, NaCl 150mM-1M peut être effectué à ce moment. Nous disposons au laboratoire de deux systèmes de purification semi-automatique Äktabasic (*GE HealthCare*) situés dans une chambre tempérée à  $4^\circ\text{C}$ . La première étape de purification a été similaire pour toutes les protéines d'étude. Il s'agit d'une **chromatographie d'affinité sur ions métalliques immobilisés (IMAC)**. Tout d'abord, la résine est coulée manuellement dans une colonne de chromatographie (XK16/60 column GE Healthcare Bio-sciences). L'ensemble est alors



connecté au système et le programme de lavage et d'élution est déclenché. Pour l'élution de la protéine on utilise un gradient d'imidazole de 0mM à 250mM. La composition du tampon d'élution est donc : NaCl 150mM, Tris HCl 50mM pH 7.5 et imidazole 250mM. A chaque étape la pureté de la protéine est vérifiée sur un gel SDS-PAGE. Les fractions correspondant au pic de la protéine d'intérêt sont récupérées à approximativement 160mM d'imidazole.

Les fractions d'élutions sont alors réunies. A cette étape de la purification nous pouvons ajouter la **protéase TEV** qui permet l'élimination de l'étiquette de 6 His. On ajoute 10µg de TEV pour chaque milligramme de protéine. L'incubation peut être réalisée à 4°C toute la nuit. Cependant, pour certains cas où le clivage s'est avéré difficile et si la stabilité des protéines le permettait, j'ai ajouté une étape d'incubation à température ambiante de 2h, la TEV étant trois fois plus active à cette température. Dans la solution nous additionnons généralement 0,5mM EDTA (acide éthylène diamine tétra acétique) et/ou 1mM DTT (dithiothréitol).

Avant la deuxième étape de purification, on vérifie sur gel que la digestion par la TEV est totale. Après optimisation, pour les différentes constructions des protéines CstF-50dm et pour la formation du complexe Rna14p/Rna15p, cette deuxième étape correspond à une chromatographie d'exclusion stérique. Cependant, pour CstF-64hs (1-211)/ CstF-77hs (626-717) une étape d'échange ionique précède celle-ci et pour la purification de CF IA il y a en plus de la colonne d'échange ionique, une héparine (fig. 96).

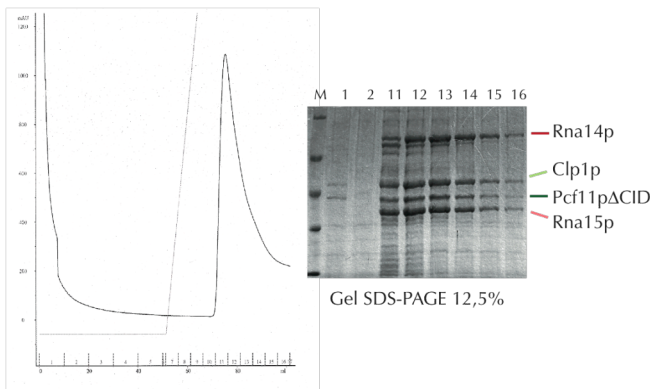
Pour le cas où la **chromatographie d'échange ionique** était nécessaire j'ai utilisé une colonne d'échange d'ions chargée positivement **HiTrap™ Q Sepharose HP de 5ml** (GE Healthcare Bio-sciences). Le tampon de course est le même que pour la première étape (Tris HCl 50mM pH 7.5 et NaCl 150mM). Le tampon d'élution contient Tris HCl 50mM pH 7.5 et NaCl 1M. Les protéines sont généralement éluées à une concentration de NaCl autour de 350 mM. Cependant, le complexe CF IA est élué à 220mM et cette concentration de NaCl doit être maintenue pendant le reste des étapes de purification afin de stabiliser le complexe. Pour ce complexe, une étape additionnelle avec une colonne d'affinité **HiTrap™ Heparin HP, 5ml** (GE Healthcare Bio-sciences) est nécessaire pour l'élimination de contaminants qui ne sont pas exclus par la colonne d'exclusion stérique. Les tampons utilisés pour cette colonne sont les mêmes que pour l'étape d'échange ionique.

L'ensemble des fractions d'élution correspondant à la protéine est alors concentré sur **membrane Amicon® Ultra-15 ultracel** (Millipore) avec un seuil de coupure adapté à la taille de la protéine d'intérêt (**30kDa, 10kDa ou 5kDa**). La protéine est concentrée jusqu'à un volume de 1,5 ml, correspondant au volume maximal acceptable pour le type de colonnes d'exclusion stérique utilisées. Les colonnes disponibles au laboratoire sont adaptées pour la séparation de protéines de grand ou de faible poids moléculaire. Pour toutes les protéines présentées j'ai utilisé une colonne **HiLoad 16/60 Superdex™ 75 pg** (Amhershan) sauf pour le complexe CF IA où nous avons choisi une colonne **HiLoad 16/60 Superdex™ 200 pg** (Amhershan). Dans les deux cas, la colonne permet la séparation des protéines en fonction de leur taille. Le tampon utilisé pour cette étape de

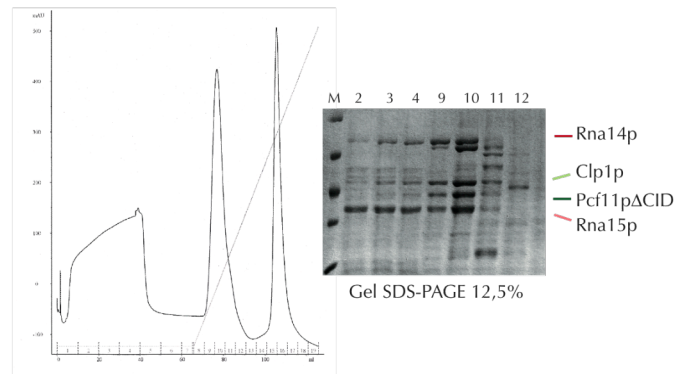
purification varie en fonction des expériences réalisées avec la protéine. Pour les essais de cristallisation, le tampon utilisé était Tris HCl 50mM pH 7.5 et NaCl 150mM avec ou sans DTT 1mM.

Les fractions correspondant à la protéine sont à nouveau concentrées par la même méthode. En fonction de la protéine de travail la concentration finale peut varier entre 3 et 50mg/ml. La méthode générale de mesure de concentration utilisée est celle du Bradford (*Sigma Aldrich*). La mesure de l'absorption à 280nm dans un NanoDrop™ 1000 Spectrophotometer a été employée dans certaines occasions pour avoir une idée plus précise de la concentration protéique (*Thermo Scientific*).

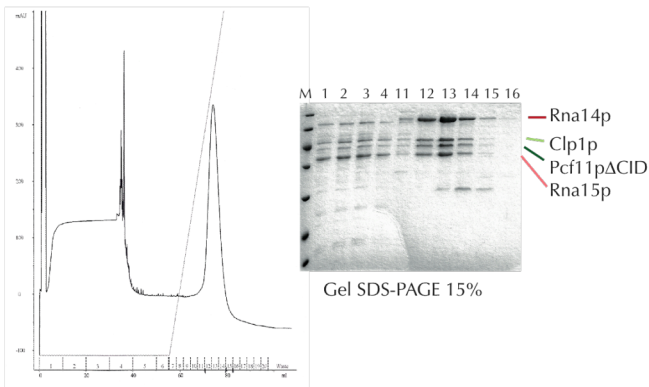
### Chromatographie d'affinité sur ions métalliques immobilisés (IMAC)



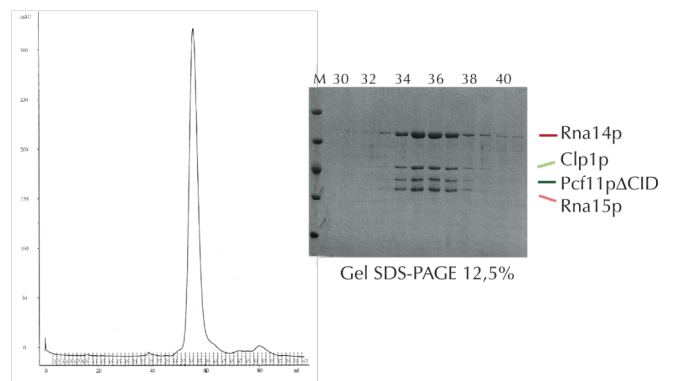
### Chromatographie d'échange d'ions



### Chromatographie d'affinité : colonne d'héparine



### Chromatographie d'exclusion stérique



**Figure 96.** Exemple d'une purification. Le protocole représenté a été l'utilisé pour complexe CFIA. Pour l'obtention du complexe pure, quatre étapes de purification étaient nécessaires commençant pour une chromatographie d'affinité (panneau en haut à gauche) au Cobalt qui est suivie d'une chromatographie d'échange d'anions (panneau en haut à droite), d'une autre d'affinité pour finir avec une chromatographie de filtration sur gel.

#### 4. Purification de l'ARN

L'ARN a été employé pour des études d'interaction avec des protéines, principalement, avec le complexe CstF-64hs (1-211)/ CstF-77hs (626-717) ou son homologue chez la levure, Rna15p/Rna14p. De plus, l'ARN a été utilisé pour des travaux qui n'ont pas été détaillés dans cette thèse, comme les expériences réalisées pour l'étude de l'interaction entre le complexe CF Im25/CF Im68 et une séquence d'ARN contenant le motif UGUA.

La séquence des ARNs a été déterminée au laboratoire pour chaque expérience. Ensuite les ARNs sélectionnés ont été synthétisés par N. Pierre au sein de notre unité Inserm (U869). Ceux-ci sont protégés en 2' avec 2'-O-t-butylidimethylsilyl (TBDMS) et accrochés à un support solide sous forme de billes. Les oligonucléotides doivent être déprotégés et purifiés avant utilisation.

Après optimisation, le protocole suivi a été le suivant :

- Déprotection des bases et décrochage du support solide

La CPG (phase solide qui inclus les billes et les oligonucléotides) est récupérée dans un tube à vis de 2ml. Pour sortir les billes du support nous pouvons nous servir d'une aiguille qui permet enlever le filtre de protection. Ensuite :

- ajouter 1ml de NH<sub>4</sub>OH :EtOH (3 :1) et incuber 16h à 55°C (note : NH<sub>4</sub>OH est conservé à 4°C),
- laisser refroidir et récupérer l'ARN qui se trouve maintenant dans le surnageant,
- rincer les billes 3 fois avec 250µl de EtOH :ACN :H<sub>2</sub>O (3 :1 :1) (ACN :acétonitrile). Pour chaque rinçage centrifuger quelques secondes à 3000rpm et reprendre les surnageants dans le même tube,
- évaporer la solution au « speed vac » (Concentrator 5301, *Eppendorf*) (si on utilise le programme 2, destiné à l'évaporation de solvants, cette étape peut prendre 2-3h).

L'ARN est maintenant séparé du support solide. A partir de ce moment les étapes de déprotection doivent être réalisées dans des conditions sans RNases (« RNase free »).

- ajouter 250µl de la solution suivante fraîchement préparée (sous la hotte):
  - o 750µl N-méthylpurrolidinone
  - o 350µl TEA (Triéthylamine)
  - o 500µl TEA.3HF (Triéthylamine trihydrofluoride),
- incuber 1h30' à 65°C sous agitation, on peut vortexer toutes les 15' pour solubiliser plus facilement,
- laisser refroidir et ajouter 500µl de n-butanol, un précipité se forme\*,
- vortexer, centrifuger 1' à 14500rpm et éliminer le butanol,
- resuspendre le culot dans 200µl d'H<sub>2</sub>O,
- ajouter 1,8ml de n-butanol pour reprécipiter l'oligonucléotide,

- vortexer, centrifuger 1' à 14500rpm et éliminer le butanol (laisser seulement les 200µl d'eau qui doivent rester au fond après la centrifugation),
- sécher au « speed vac » quelques heures,
- reprendre dans 1-2ml d'eau et stocker à -20°C.

\*note : pour les petits ARNs (moins de 10 nucléotides), la précipitation se fait à l'éthanol et non au butanol. On ajoute 25µl de NaOAc (sodium acétate) 3M et le tube est rempli avec de l'éthanol 100%. Après avoir mélangé au vortex, l'ARN est précipité toute la nuit à -20°C. Le lendemain, on centrifuge 15', à 15°C. Le culot contenant l'ARN peut être séché et il est enfin prêt pour la purification.

- Purification des ARNs

La purification était d'abord réalisée sur gel d'acrylamide mais cette étape a été fortement optimisée grâce à l'utilisation d'une **colonne d'échange cationique** qui permet l'obtention d'ARN pur plus rapidement et avec un meilleur rendement.

Le protocole demande des précautions particulières étant donné que le système dont nous avons besoin est le même que celui utilisé pour la purification de protéines. Les étapes de lavage sont très importantes pour éviter la contamination des échantillons par des RNases.

Pour la purification proprement dite, nous nous sommes servis d'une colonne HiTrap™ Q Sepharose HP de 5ml (*GE Healthcare Bio-sciences*), réservée exclusivement pour ce type de manipulations. Avec cette colonne connectée au système nous effectuons des lavages avec 25ml des solutions suivantes et dans cet ordre :

- NaOH 1M
- H<sub>2</sub>O mQ stérile
- Ethanol 20%
- H<sub>2</sub>O mQ stérile

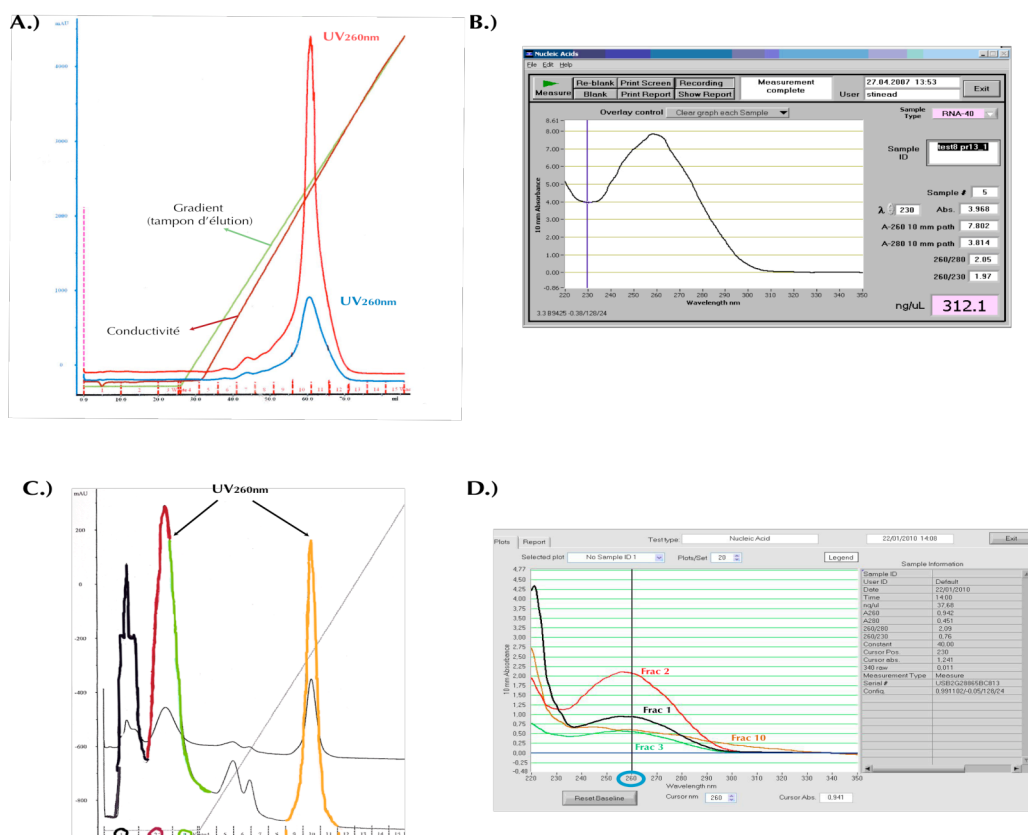
Ensuite, la solution contenant l'ARN est injectée avec une nouvelle seringue stérile et une aiguille destinée seulement à ce type d'échantillons. Le tampon utilisé contient Tris HCl 50mM pH 7.5, NaCl 150mM et EDTA 0,5mM et celui d'élution Tris HCl 50mM pH 7.5, NaCl 1M et EDTA 1mM. Le programme informatique est le même que pour la purification de protéines. Un gradient de sel permet l'élution de l'ARN à une concentration d'environ 400-500mM.

L'inconvénient de ce système de purification par rapport à la purification sur gel est que l'on se retrouve avec un volume important, normalement de ≈25ml. Cependant, la suite du protocole a pu être adapté au laboratoire de deux façons différentes :

- Les petits ARNs (de moins de 1,5kDa) doivent être précipités, pour cela :
  - on additionne 1/10 du volume total d'acétate de sodium 3M et 3 fois le volume total d'éthanol 100% froid (il est conservé à -20°C). Par exemple, pour un volume d'ARN sortant de la colonne de 22ml, on ajoute 2,2ml de NaAc et 66ml de EtOH,
  - la solution est homogénéisée et stockée au moins 2h à -80°C ou toute la nuit à -20°C,

- centrifuger 30' à 14000rpm et éliminer le surnageant,
  - lavage avec EtOH 70% froid,
  - centrifuger 10' à 14000rpm et éliminer le surnageant,
  - laisser sécher à 37°C quelques heures et reprendre avec le volume souhaité (généralement 200-400  $\mu$ l) d'H<sub>2</sub>O stérile.
- Pour les autres ARNs, dont la taille est plus importante, nous utilisons des **concentrateurs MACROSEP 1K omega (Pall Corporation)**. Après plusieurs centrifugations à 4500rpm, nous pouvons obtenir l'ARN concentré sur un volume de 500 $\mu$ l-1ml.

L'ARN pur est prêt à l'usage ou au stockage (-20°C). Pour mesurer la concentration des ARNs purifiés nous utilisons systématiquement un **spectrophotomètre NanoDrop™**. Ce type d'instrument permet en plus de vérifier que les fractions obtenues dans le pic d'élution correspondent à l'ARN. En effet, au cours de la mesure, nous pouvons obtenir un profil UV et nous pouvons facilement contrôler s'il coïncide avec le profil typique de ce type de molécules. Ainsi, le profil d'élution obtenu à partir de la purification de l'ARN GU-5 montrait différents pics. La mesure au NanoDrop™ des différentes fractions, m'a donné une indication du pic correspondant à l'ARN (*fig. 97*). Pour les autres ARNs purifiés, seulement un pic a été obtenu et le profil affiché au NanoDrop™ a permis simplement de corroborer la correspondance avec des molécules d'ARN.



**Figure 97.** A) Profil d'élution typique des ARNs purifiés. Celui-ci correspond à la purification du GU-20. B) Profil typique obtenu après la mesure de la concentration des molécules au NanoDrop™. C) Profil d'élution obtenu pour GU-5. D) Profils superposés de différentes fractionnes d'élution du GU-5.

## C. Techniques biophysiques

### 1. Cristallographie aux rayons X

La **cristallographie** est la science consacrée à l'étude des substances cristallines à l'échelle atomique. Les propriétés physico-chimiques d'un cristal sont étroitement liées à l'arrangement spatial des atomes dans la matière. De cette façon, la cristallographie peut être utilisée pour l'étude des positions tridimensionnelles précises des atomes dans une protéine. L'ensemble du travail cristallographique pour déterminer la structure d'une nouvelle protéine est conditionné par l'obtention de cristaux de taille et de qualité de diffraction suffisantes. Ce phénomène est connu comme **cristallogénèse** et constitue une étape critique dans la résolution des structures protéiques.

#### a.) Cristallogénèse

La cristallogénèse est la formation d'un cristal, soit en milieu naturel, soit de façon expérimentale. La **cristallisation** est le passage d'un état désordonné liquide, gazeux ou solide à un état ordonné solide. Le principe de la cristallisation est donc d'induire un changement d'état de la protéine de sa phase soluble à une phase solide et ordonnée, ces deux états étant en équilibre. L'état cristallin est défini par un caractère périodique et ordonné à l'échelle atomique ou moléculaire.

Pour les essais de cristallisation nous partons d'une protéine pure. La concentration protéique initiale est très variable en fonction de la protéine d'étude. Nous essayons de commencer à une concentration de 10mg/ml, une concentration statistiquement optimale pour la cristallisation (*Hampton Research*). Néanmoins, dans certains cas nous avons été limités en quantité de protéine et cette concentration n'a pas pu être atteinte.

Trouver les conditions dans lesquelles une protéine cristallise peut s'avérer difficile et requérir beaucoup de temps. Avec chaque nouvelle protéine, nous commençons par réaliser un criblage à travers l'utilisation de diverses matrices (« **screening** »). L'intérêt des criblages est de pouvoir identifier des premières pistes de cristallisation en utilisant peu de matériel mais en testant un grand nombre de paramètres.

L'identification des premières conditions de cristallisation s'effectue à partir des **kits commerciaux** chacun contenant **96** solutions différentes de cristallisation. Pour les tests, nous disposons des boîtes de 96 puits (*Greiner*) qui sont remplies manuellement ou à l'aide d'un robot de cristallisation (Honeybee 961 Cartesian Dispensing System Genomic Solutions). Dans le remplissage manuel, le volume des gouttes est d'1µl (0,5µl de protéine + 0,5µl de solution commerciale), tandis que le robot permet la réalisation de gouttes d'un volume plus faible où seulement 0,2µl de protéine sont employés pour chaque condition essayée.

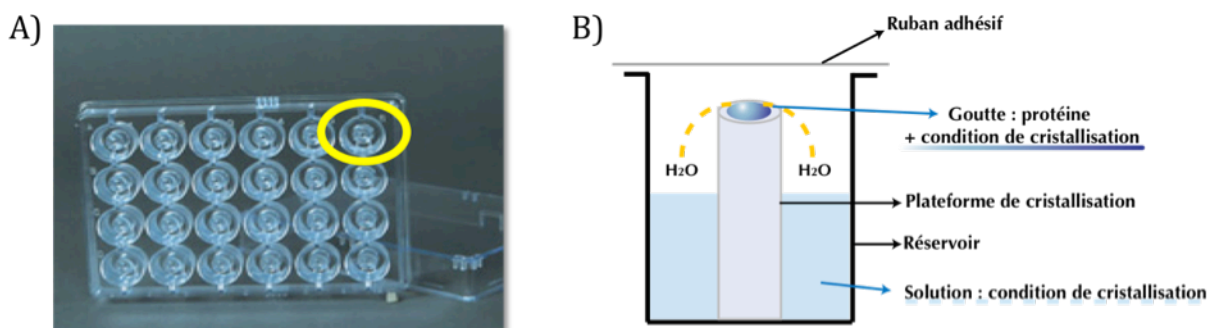
Les kits disponibles au laboratoire sont nombreux (*tableau 15*) et une sélection logique doit être réalisée quand on commence à travailler avec une nouvelle protéine et spécialement, en fonction des résultats obtenus à partir des premiers kits utilisés.

<b>Hampton Research (<a href="http://hamptonresearch.com/menus.aspx?id=2&amp;cid=1">http://hamptonresearch.com/menus.aspx?id=2&amp;cid=1</a>)</b>		
Matrix	PEG / Ion Screen	Index
Crystal Screen 2	Crystal Screen	Salt RX
<b>Qiagen (<a href="http://www1.qiagen.com/Products/Protein/Crystallization/">http://www1.qiagen.com/Products/Protein/Crystallization/</a>)</b>		
The Classics Suite	The PEGs Suite	The PEGs II Suite
The PACT Suite	The Cryo Suite	The JCSG + Suite
The Mb Class Suite	The Mb Class II Suite	The Protein Complex Suite
Anions suite	Cations suite	
<b>Emerald Biosystem (<a href="http://www.emeraldbiosystems.com/">http://www.emeraldbiosystems.com/</a>)</b>		
Cryo I et II	Wizard I, II et III	

*Tableau 15. Kits commerciaux disponibles au laboratoire.*

Pour l'optimisation des conditions nous utilisons généralement des boîtes de 24 puits (*Grenier*). Dans ces plaques, les gouttes réalisées généralement contiennent 1µl de protéine et 1µl de solution de cristallisation. Dans le réservoir, le volume de solution est de 500µl.

Les essais de cristallisation, dans les deux types de boîtes, ont été réalisés en utilisant la technique de **diffusion de vapeur en goutte assise** (*fig 98*). L'emploi de cette technique nécessite des plaques de cristallisation où un « siège », appelé plateforme de cristallisation sur la figure, est placé à l'intérieur du puits. Le mélange, composé de la solution de cristallisation et de la solution de la protéine, est situé sur ce « siège » et le réservoir est rempli avec la même solution de cristallisation.



*Figure 98. A) Plaque de cristallisation de 24 puits. B) Détail d'un puits. Schéma représentatif de la méthode de diffusion de vapeur en goutte assise.*

L'équilibre s'établit entre la goutte et le réservoir par diffusion des composés volatiles jusqu'à ce que la tension de vapeur soit la même dans les deux compartiments. La goutte, qui est plus diluée que le réservoir, perd du volume (H<sub>2</sub>O) ce qui a pour effet de concentrer la protéine présente dans la goutte. Pour induire la nucléation (germination), il faut amener la protéine dans un état soluble instable de **sursaturation** (fig. 99). Cet état s'établit progressivement en augmentant simultanément la concentration de la protéine et de l'agent précipitant, ce qui peut conduire à la formation de cristaux. La variation de différents paramètres thermodynamiques (pH, température, concentration et pureté de la protéine, force ionique) peut contribuer à la nucléation et à la croissance cristalline.

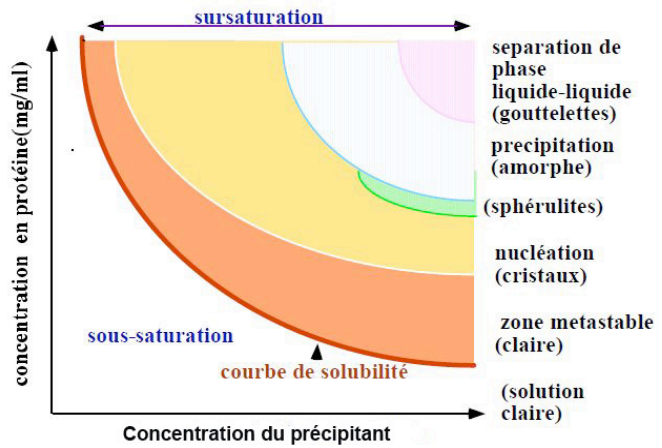


Figure 99. Diagramme de phase d'une protéine

La variation de différents paramètres thermodynamiques (pH, température, concentration et pureté de la protéine, force ionique) peut contribuer à la nucléation et à la croissance cristalline.

Pour la cristallisation de CstF-50dm (1-65), nous nous sommes servis en plus d'une des techniques de micro-ensemencement, le « **streak seeding** » (fig. 100). Cette méthode se base sur l'introduction de germes de cristaux préformés dans une nouvelle goutte pour contrôler la nucléation et modifier la manière dans laquelle les cristaux croissent. Pour effectuer cette expérience nous avons utilisé une **moustache de chat** que nous faisons passer sur un cristal déjà formé. Ensuite, la même moustache est délicatement plongée dans une autre goutte récente qui ne contient pas de cristaux. Cette technique a permis d'améliorer le temps de formation des cristaux et d'augmenter leur taille, de sorte qu'en seulement quelques heures, des cristaux de 700µm pouvaient être observés.

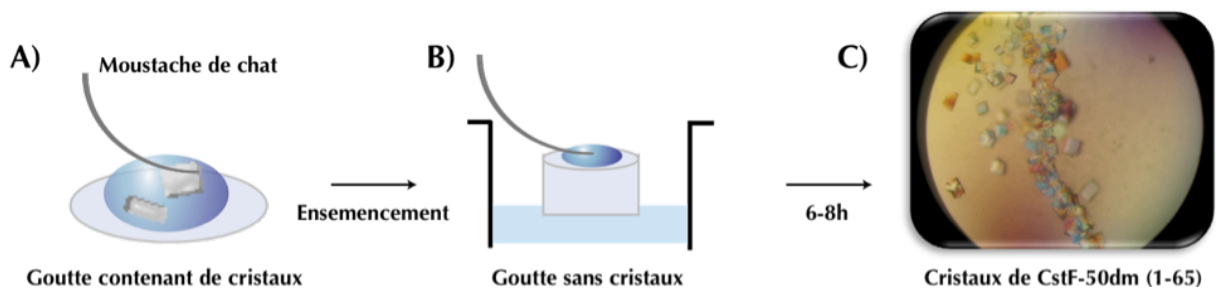


Figure 100. Etapes pour la réalisation de la technique de "streak seeding". A) Laisser glisser la moustache de chat sur un cristal préformé. B) Ensemencement : introduire la même moustache dans une nouvelle goutte contenant la protéine en solution avec une condition de cristallisation. La moustache apporte des germes cristallins du premier cristal. C) Photographie d'une goutte où l'on apprécie la trajectoire du passage de la moustache.



## b.) Diffraction aux rayons X

### b.1) Principe de la diffraction

La diffraction est le comportement des ondes lorsqu'elles rencontrent un obstacle. Pour qu'un objet diffracte une onde, la longueur d'onde incidente doit être de l'ordre de grandeur de l'objet étudié. Pour observer une macromolécule à la résolution atomique, il faut donc utiliser les **rayons X**, dont les longueurs d'onde sont de l'ordre de l'angström Å ( $100 \text{ \AA} < \lambda < 0.1 \text{ \AA}$ ).

L'interaction d'un faisceau de rayons X avec un objet va générer un phénomène de diffusion, c'est-à-dire l'émission de rayons X de façon continue dans toutes les directions de l'espace. Dans le cas d'une macromolécule isolée, le signal émis est trop faible pour être détecté, car la majorité des rayons vont la traverser sans être diffractés. Ce problème peut être contourné grâce à l'utilisation des cristaux. Un **cristal** est un solide polyédrique à structure régulière et périodique, formée d'un empilement ordonné d'un grand nombre de molécules, atomes ou ions (« motifs ») identiques (de l'ordre de  $10^{15}$ ). L'optique des rayons X ne permet pas d'obtenir une image directe de l'objet. La seule information disponible est la diffusion des rayons X par le cristal. De cette façon, quand les rayons X incident sur cristal, le signal sera collecté selon un réseau périodique et amplifié proportionnellement au nombre de molécules identiques.

- Facteurs de structure

Lorsqu'un cristal est irradié par un faisceau de rayons X, chacun des atomes du cristal réfléchit une onde de faible amplitude, qui se propage dans toutes les directions. Ces ondes vont interférer pour la plupart de manière **destructive** et s'annuler, mais certaines vont interférer de manière **constructive** et donner un nouveau signal X dans des directions privilégiées de l'espace. Ainsi, au niveau d'un écran (détecteur de rayons X) on observe une figure de diffraction composée de tâches correspondantes aux maximums des ondes diffusées en phase. Cette figure correspond à la **transformée de Fourier** de la densité électronique de l'objet. La transformée de Fourier du réseau présent dans l'espace réel correspond au **réseau réciproque**. L'équation représentée ci-dessous établit la relation entre le **facteur de structure**  $F_{hkl}$  et la **densité électronique**  $\rho$ .  $V$  est le volume de la maille cristalline,  $x, y, z$ , sont les coordonnées dans l'espace direct et  $h, k, l$ , les **indices de Miller** du réseau réciproque.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i (hx+ky+lz)}$$

$F_{hkl}$  est décrit par son module  $|F_{hkl}|$  et sa **phase**  $\varphi_{hkl}$  :

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

En effet, les détecteurs ne sont sensibles qu'à l'intensité  $I_{hkl}$  des ondes diffractées. L'intensité mesurée, obtenue après intégration des données, ne dépend que de l'amplitude du facteur de structure.

$$I_{hkl} = |F_{hkl}|^2$$

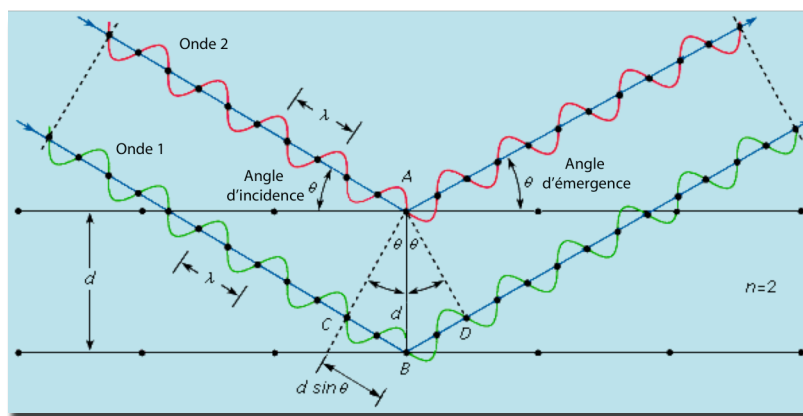
En revanche, il n'est pas possible d'en déduire la phase ni de reconstruire directement la densité électronique par la seule transformée de Fourier inverse. Ceci constitue l'un des problèmes majeurs en cristallographie des macromolécules (décrit plus en détail dans le point b.6. de la même section).

- Les directions de diffractions

Lorsque le cristal est soumis à un faisceau de rayons X, la famille de plans réticulaires d'indices de Miller (hkl) donne lieu à un faisceau diffracté si, et seulement si, la **loi de Bragg** est respectée :

$$2d_{hkl} \sin(\theta) = n\lambda$$

; où  $d_{hkl}$  est la distance interréticulaire de la famille de plans (hkl) et  $\theta$ , l'angle entre la direction du faisceau incident de longueur d'onde  $\lambda$  et la famille de plans réticulaires (fig. 101). La position des pics de diffraction peut se calculer en fonction des distances interréticulaires par la loi de Bragg. L'enregistrement d'un espace complet de diffraction nécessite de mesurer l'intensité d'un maximum d'ondes diffractées en tenant compte de la symétrie cristallographique. Le principe le plus couramment utilisé consiste à travailler en lumière monochromatique ( $\lambda$  constant), et à imposer un mouvement de rotation au cristal pour faire varier l'angle  $\theta$  (dans notre cas une rotation de  $180^\circ$  a permis la collection de 180 images).



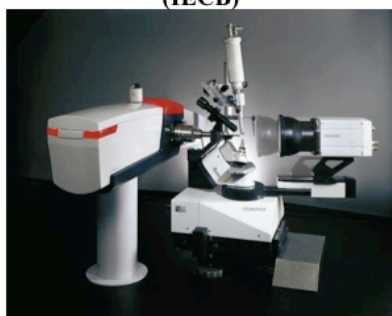
**Figure 101.** Démonstration de la Loi de Bragg. Réflexion des rayons X par une famille de plans réticulaires espacés d'une distance  $d$ . Deux rayons parallèles percutent sur deux atomes situés sur une même droite perpendiculaire à la surface. Le chemin supplémentaire parcouru par l'onde 2 est  $2d \cdot \sin(\theta)$ . Les interférences sont constructives si la différence de chemin introduit un déphasage multiple de  $2\pi$ , c'est-à-dire si le chemin supplémentaire est un multiple de  $\lambda$ .

### b.2) Sources de rayons X

Il existe plusieurs sources de rayons X. Dans l'IECB nous disposons d'un **diffractomètre à source conventionnelle** (Anode tournante Bruker Microstar 2.7kW microfoyer) dont nous nous sommes servis pour la collection des premiers jeux de données. Toutefois, la quasi-totalité des données de diffraction ont été obtenues à partir de deux sources **synchrotron** différentes, l'**ESRF** (*European Synchrotron Radiation Facility*, Grenoble) et **SOLEIL** (*Source Optimale de Lumière d'Énergie Intermédiaire de LURE*, Gif-sur-Yvette) (fig. 102). Le synchrotron est un accélérateur de particules qui génère des radiations de différentes longueurs d'onde ( $\lambda$ ). Le rayonnement synchrotron présente en effet plusieurs avantages :

- la brillance spectrale du faisceau est très élevée ce qui permet de réduire le temps d'exposition,
- la géométrie du faisceau permet une meilleure résolution spatiale des taches de diffraction,
- le rapport signal sur bruit est amélioré ce qui permet un gain en résolution,
- la longueur d'onde ajustable permet d'utiliser la diffusion anormale de certains éléments.

**Anode tournante Bruker  
Microstar 2.7kW microfoyer  
(IECB)**



- Premiers tests des cristaux
- Paramètres de maille:  
a=45.76Å b=51.77Å c=57.39Å  
 $\alpha=90^\circ \beta=90^\circ \gamma=90^\circ$
- Groupe d'espace: C222<sub>1</sub>

**ESRF (ID14-4) Grenoble**



- Phasage: **SAD**
- Résolution: 1,83Å
- Paramètres de maille :  
a=45.93Å b=51.87Å c=57.36Å  
 $\alpha=90^\circ \beta=90^\circ \gamma=90^\circ$
- Groupe d'espace: C222<sub>1</sub> (20)

**SOLEIL (PROXIMA1)- Paris**



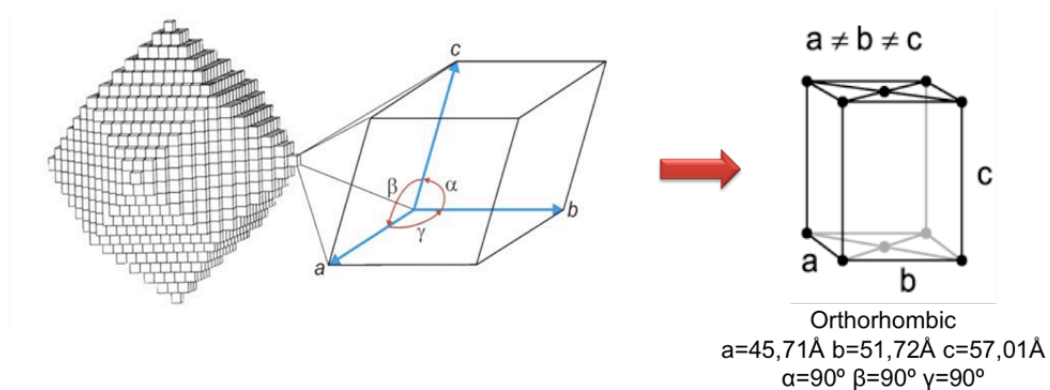
- Meilleure résolution : **1.3Å**
- Construction du modèle final
- Paramètres de maille:  
a=46.23 b=52.24 c=57.85  
 $\alpha=90^\circ \beta=90^\circ \gamma=90^\circ$
- Groupe d'espace: C222<sub>1</sub> (20)

**Figure 102.** Les trois sources utilisés pour la résolution de la structure de CstF-50dm (1-65). Les résultats principaux obtenus à partir de chaque source sont indiqués en dessous.

b.3) Caractéristiques des cristaux de protéines

Que le cristal soit ordonné signifie qu'un même motif est répété à l'identique selon un réseau régulier. La plus petite partie du réseau permettant de recomposer l'assemblage est appelée **maille**. Au même temps, le plus petit volume de la maille contenant des points non reliés entre eux par la symétrie cristallographique constitue l'**unité asymétrique**. Le motif (unité asymétrique) est répété à chaque nœud du réseau suivant les symétries du **groupe d'espace** (fig. 103). Autrement dit, l'ensemble de la maille est engendré par l'unité asymétrique à laquelle il est appliqué les opérateurs de symétrie du groupe d'espace.

CstF-50dm (1-65) cristallise dans le groupe d'espace **C222<sub>1</sub>** (orthorhombique à base centrée) avec une molécule par unité asymétrique. Bien qu'il n'y ait qu'une molécule par unité asymétrique, la protéine CstF-50 a cristallisé sous sa forme dimérique. L'axe de symétrie du dimère se situe sur un axe cristallographique.



**Figure 103.** Le panneau situé à gauche montre un schéma d'un cristal et un détail de l'unité la plus petite qui se répète dans celui-ci (maille). La figure à droite représente le type d'unité asymétrique obtenue pour CstF-50dm (1-65). Cette unité asymétrique suit les symétries d'un groupe d'espace orthorhombique. Les paramètres de maille sont indiqués.

b.4) Traitement de données

La détermination d'une structure par diffraction des rayons X consiste à :

- mesurer les intensités de la plupart des faisceaux diffractés par un cristal,
- obtenir à partir de ces intensités les facteurs de structure à travers un procédé d'assignation des phases de chacun de ces facteurs,
- reconstruire la distribution électronique dans l'unité cellulaire dont les maxima correspondront aux positions des atomes.

La **collecte des données** de diffraction consiste en l'enregistrement d'images de diffraction, comme celles montrées sur la *figure 104*, qui par la suite sont analysées pour obtenir les distances, la symétrie et les intensités dans l'espace réciproque. Les taches de diffraction observées

correspondent aux pixels du détecteur (dispositif à couplage de charges, **CCD**) excités par les rayons diffractés. Le nombre de photons reçus par l'ensemble des pixels permet de déterminer l'intensité de la tache. La position de chaque tache de diffraction dépend de la taille et de la forme de la maille, et de la symétrie dans le cristal.

La première étape du traitement des clichés de diffraction consiste à déterminer les **paramètres de la maille cristalline** et le **groupe d'espace**. Pour ceci, nous avons utilisé principalement le programme **XDS** (**X**-ray **D**etector **S**oftware) (Kabsch 1993) et nous nous sommes servis aussi de **MOSFLM** (Powell 1999), qui permet une étude plus visuelle des données.

A l'aide de ces programmes, les données sont **indexées** image par image, en affinant un certain nombre de paramètres comme la distance entre le cristal et le détecteur, la position du centre du détecteur, l'orientation du cristal, les paramètres de maille, etc. Les profils de diffraction prédits sont ajustés aux taches observées. Chaque tache est alors indexée en  $h$ ,  $k$  et  $l$ , et l'intensité de chaque tache de diffraction est mesurée en corrigeant le bruit de fond.

Enfin, la dernière étape du traitement consiste en la **mise à l'échelle** et à la **réduction** du jeu de données. Pour la mise à l'échelle, les données sont moyennées sur les réflexions équivalentes. Ces données moyennées peuvent alors être réduites à l'unité asymétrique. Cette dernière étape (mise à l'échelle et réduction) a été effectuée avec la suite du programme **CCP4** (**C**ollaborative **C**omputational **P**roject, **N**umber **4**, 1994).

La **résolution** donne la précision des données, c'est-à-dire, détermine la finesse des détails dans l'image de la densité électronique. La limite haute de la résolution des données mesurées peut être imposée par la configuration de l'expérience (taille du détecteur, distance cristal-détecteur,  $\lambda$ ) ou par la qualité de diffraction du cristal.

Pour l'évaluation de la **qualité d'un jeu de données** il faut tenir en compte différents critères (un exemple est exposé sur la *figure 104*):

- Le facteur  $R_{sym}$ : ce facteur mesure la différence entre les intensités de chaque réflexion équivalente et la valeur moyenne de ces réflexions. Il donne une estimation de l'erreur relative des mesures d'intensités.

$$R_{sym} = \frac{\sum_{hkl} |I_{hkl} - I_{sym}|}{\sum_{hkl} I_{hkl}}$$

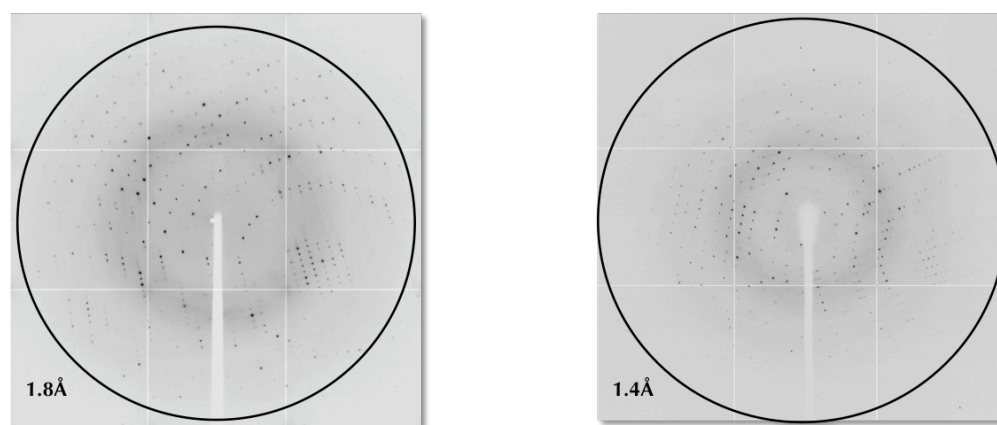
; où  $I_{hkl}$  est l'intensité d'une réflexion d'indices  $h$ ,  $k$  et  $l$  corrigés et mis à l'échelle et  $I_{sym}$  est la moyenne de l'ensemble des intensités équivalentes. Des valeurs de  $R_{sym}$  acceptables sont comprises entre 4 et 15%.

- La **complétude**: ce paramètre mesure le nombre de réflexions observées par rapport au nombre de réflexions théoriquement mesurables. Elle est fonction du groupe d'espace, des

paramètres de maille et de la résolution. La complétude est exprimée en pourcentage et pour des jeux de données de qualité elle doit se rapprocher de 100%.

- Le rapport signal sur bruit  $I/\sigma(I)$  :  $I$  correspond à l'intensité du signal des réflexions mesurées et le  $\sigma(I)$  à l'écart type des intensités des réflexions. Plus l'intensité du signal est élevée par rapport au bruit de fond, meilleures sont les estimations des intensités. Ce facteur rend compte de la précision des mesures et permet de déterminer, avec la complétude, le seuil de coupure en résolution d'un jeu de données.

- La **redondance** : ce facteur exprime le nombre de fois où l'intensité d'une réflexion unique a été mesurée. Plus une réflexion et ses symétriques sont quantifiées meilleure sera l'estimation de l'intensité moyenne  $I$ , plus petite sera celle de son écart type  $\sigma(I)$  et meilleure la qualité des données.



RESOLUTION LIMIT	NUMBER OF REFLECTIONS			COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	Rmrgd-F	Anomal Corr	SigAno	Nano	
	OBSERVED	UNIQUE	POSSIBLE										
5.58	992	290	413	70.2%	1.7%	2.0%	970	59.42	2.1%	1.2%	89%	3.104	97
4.00	2272	659	722	91.3%	2.0%	2.0%	2238	60.21	2.4%	1.4%	79%	2.150	261
3.28	3291	905	921	98.3%	2.0%	2.0%	3274	61.35	2.3%	1.3%	57%	1.429	396
2.84	4030	1068	1070	99.8%	1.7%	2.1%	4020	59.34	2.0%	1.2%	72%	1.725	477
2.55	4672	1227	1230	99.8%	1.8%	2.1%	4657	54.84	2.1%	1.3%	68%	1.594	555
2.33	5141	1350	1353	99.8%	1.9%	2.2%	5123	51.72	2.2%	1.4%	61%	1.421	609
2.15	5509	1448	1452	99.7%	2.1%	2.4%	5490	47.01	2.5%	1.7%	44%	1.285	657
2.02	6035	1596	1598	99.9%	2.6%	2.7%	6017	40.51	3.0%	2.3%	39%	1.283	730
1.90	4463	1537	1671	92.0%	3.5%	3.3%	4258	27.65	4.2%	3.7%	29%	1.195	594
total	36405	10080	10430	96.6%	2.0%	2.1%	36047	48.43	2.4%	1.6%	55%	1.484	4376

**Figure 104.** Deux images de diffraction obtenues pour CstF-50dm (1-65) à une résolution maximale de 1.8 Å (à gauche) et de 1.4 Å (à droite). En dessous, un tableau montrant des statistiques acquises à l'aide du programme XDS. Ce jeu de données correspond à la protéine CstF-50dm (1-65) « sélénée » et on peut apprécier la présence du signal anormal (2<sup>ème</sup> colonne en commençant par la droite).

### *b.5) Détermination des structures : le phasage*

Les amplitudes (intensités des faisceaux diffractés) des facteurs de structure ont été définies, cependant, afin d'obtenir une carte de densité électronique interprétable, en plus de l'amplitude, l'information de **déphasage des ondes diffractées** (les unes par rapport aux autres) doit être connue. Il n'existe pas actuellement de méthode expérimentale permettant d'enregistrer les phases, les détecteurs bidimensionnels ne sont sensibles qu'à l'intensité des faisceaux diffractés. Pourtant, pour construire l'image de la densité par transformée de Fourier, il est indispensable de posséder ces deux informations à la fois. L'incapacité à enregistrer directement les phases pendant une expérience de diffraction est connue comme « **problème des phases** ». Le problème peut être résolu par quatre techniques :

- **Les méthodes directes** : s'applique en cristallographie des petites molécules ;
- **Le remplacement moléculaire** : utilisent comme source de phase un modèle atomique homologue ;
- **Le remplacement isomorphe (SIR, MIR)** : consiste à mesurer les spectres de diffraction de cristaux de la protéine marquée aux atomes « lourds ». Ces atomes doivent posséder un nombre d'électrons plus grand que celui des atomes de la biomolécule, afin que l'intensité des taches de diffraction des cristaux contenant des atomes lourds fixés à la protéine (**cristaux dérivés**) diffère notablement de celle des taches correspondant aux cristaux de la protéine seule (**cristaux natifs**). Ainsi, il est possible d'estimer la phase de chaque réflexion à partir des différentes intensités observées. Les cristaux dérivés sont obtenus à partir de cristaux natifs trempés ou co-cristallisés dans une solution contenant l'atome lourd d'intérêt. L'une des conditions essentielles de la réussite de la méthode est que la structure des cristaux dérivés ressemble à celle du cristal natif (**isomorphisme**) ;
- **La diffusion anormale (MAD ou SAD)**: cette méthode utilise la modification des intensités diffusées lorsque la longueur d'onde varie autour du seuil d'absorption des atomes lourds présents dans le cristal. Le phénomène de diffusion anormale fait référence à la contribution supplémentaire qui est alors ajouté au facteur de diffusion « normal » (loin du seuil d'absorption). L'un des avantages de la méthode MAD par rapport au SIR/MIR est qu'il est possible d'obtenir des jeux de données natifs ou dérivés à partir d'un seul cristal. De cette façon, le problème de l'isomorphisme rencontré avec la méthode MIR est évité.

Dans la section suivante, seulement la technique de diffusion anormale à une longueur d'onde (SAD, Single wavelength Anomalous Dispersion), utilisée pour la résolution de la structure de CstF-50dm (1-65), sera décrite.



- Principe de la diffusion anormale

Le phénomène de diffusion anormale est dû au fait que le **facteur de diffusion atomique ( $f$ )** d'un atome varie fortement, en module et en phase, dans la région du seuil d'absorption de l'atome. Dans les conditions « normales », c'est-à-dire lorsque la longueur d'onde incidente est loin du seuil d'absorption d'un atome, il n'y a pas de différence de phase entre l'onde diffusée par cet atome et celle diffusée par un électron ( $e^-$ ). Le rayonnement incident n'est pas absorbé car il ne possède pas assez d'énergie pour exciter les électrons. Un atome « j » est caractérisé alors par son facteur de diffusion  $f_j$ , qui est corrélé au nombre d'électrons qui lui est associé. Dans ces conditions, **la loi de Friedel**, qui énonce l'égalité des amplitudes de réflexions ( $hkl$ ) et  $(-h-k-l)$ , est respectée :

$$|F(hkl)| = |F(-h-k-l)|$$

$$\varphi_{hkl} = -\varphi_{-h-k-l}$$

La loi de Friedel n'est plus respectée lorsque la longueur d'onde incidente est proche du seuil d'absorption d'un atome. Il y a alors absorption du rayonnement incident par les électrons des couches inférieures qui passent à des états d'énergie plus élevés. Le résultat est un changement de phase entre le rayonnement incident et le rayonnement diffracté (**phénomène de résonance**). Dans ces conditions, le facteur de diffusion  $f$  dépend de la longueur d'onde  $\lambda$  du rayon incident :

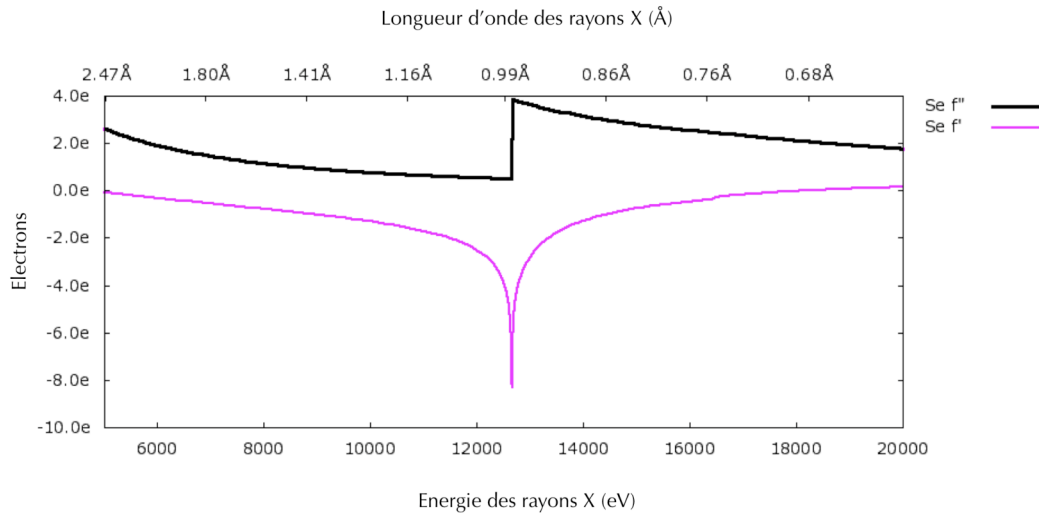
$$f_j = f^o + f'(\lambda) + i f''(\lambda)$$

; où  $f^o$  est la composante anormale sur l'axe réel et  $f''$  la composante anormale sur l'axe imaginaire (déphasage de  $\pi/2$ ).

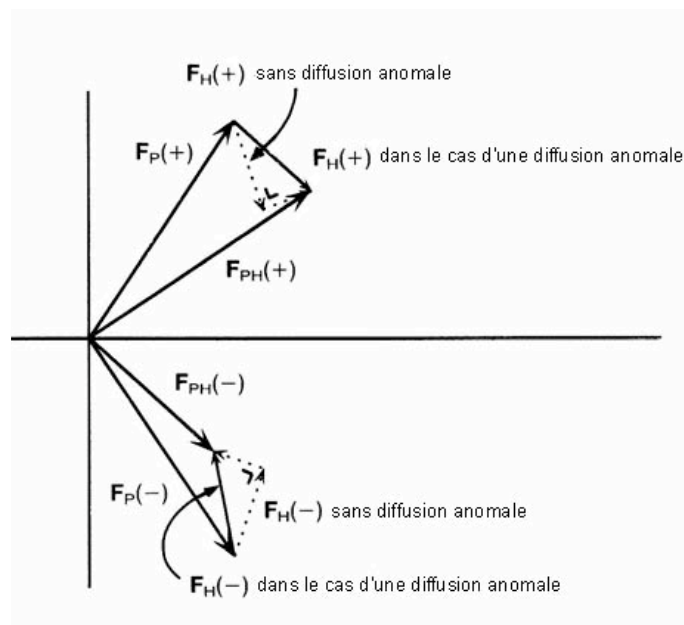
Ainsi, l'effet de résonance est fonction de la longueur d'onde incidente et de la nature de l'atome « j ». Cet effet est plus important pour les atomes de numéro atomique élevé, dits atomes lourds (comme le **sélénium** qui possède 34  $e^-$ ) que pour les atomes légers (majorité des atomes constituant les protéines, C, N, O entre 6 et 8  $e^-$ ). Dans notre étude, nous avons utilisé le sélénium comme diffuseur anormal. Les valeurs de  $f'$  et  $f''$  ont été déterminées à partir d'un spectre de fluorescence. La *figure 105* montre les variations de  $f'$  et  $f''$  en fonction de la longueur d'onde pour le sélénium.

En conclusion, le non respect de la loi de Friedel permet d'introduire une information supplémentaire pour le phasage. Si le dérivé lourd présente de la diffusion anormale,  $F_{PH}^+$  sera distinct de  $F_{PH}^-$  (*fig. 106*).





**Figure 105.** Graphique représentant les facteurs de diffusion  $f'$  et  $f''$  théoriques du sélénium. Le seuil d'absorption du sélénium est situé à  $\lambda=0,98 \text{ \AA}$ . Il est à noter que  $f'$  redevient faible aux longueurs d'onde supérieures à celle du seuil.



**Figure 106.** Non-respect de la loi de Friedel due à la diffusion anormale.  $P$  désigne la contribution des atomes qui diffractent « normalement » et  $H$  celle des atomes anormaux.  $F_P(+)$  et  $F_P(-)$  sont symétriques par rapport à l'axe horizontal (pas de diffuseur anormal dans la protéine). En l'absence de diffusion anormale,  $F_H(+)$  et  $F_H(-)$  sont également symétriques. Dans le cas d'une diffusion anormale, la contribution imaginaire implique le non-respect de la loi de Friedel, ici  $|F_{PH}(+)| > |F_{PH}(-)|$ .

- *La méthode du SAD (Single-wavelength Anomalous Dispersion)*

Cette méthode requière l'incorporation des atomes de dispersion anormale dans des protéines. La technique communément employée est la substitution des atomes de soufre des méthionines et des cystéines par des atomes de **sélénium** ((Doublie 1997) (Strub et al. 2003)). Ceci permet d'utiliser le signal anomal au seuil de cet élément ( $\lambda=0,98 \text{ \AA}$ ) pour obtenir des informations sur les phases. Les protéines contiennent en moyenne une méthionine pour 58 acides aminés. L'amplitude modérée de  $f''$  (une dizaine d'électrons) à ce seuil est compensée par le taux de présence du diffuseur anomal dans la protéine.

Un nouveau procédé dans cette méthode propose aussi l'utilisation du **soufre** présent dans les protéines. En effet, le soufre est l'atome le plus lourd naturellement présent dans les acides aminés. Cette stratégie consiste à augmenter la longueur d'onde pour se rapprocher du seuil du soufre ( $\lambda=5,02 \text{ \AA}$ ), afin d'augmenter sa contribution anormale.

Pour la technique du **MAD** (Multi-wavelength Anomalous Dispersion), trois ou quatre ensembles de mesures sont réalisés pour optimiser les différences d'intensité dues au signal anomal. Les points de mesure incluent la longueur d'onde correspondant au maximum de  $f''$  et la longueur d'onde dans laquelle  $f'$  atteint son minimum. Les extremums de  $f'$  et  $f''$  ne sont éloignés que de quelques eVs du seuil. Les autres points de mesure sont plus éloignés de ce point.

Contrairement à la méthode de MAD, pour le **SAD** seulement un jeu de données est collecté à une longueur d'onde adéquate. La méthode du SAD repose sur la précision des mesures de différences anormales, ainsi que sur leur intensité. De cette façon, il est préférable d'obtenir un jeu de données à une longueur d'onde où la contribution  $f''$  du diffuseur est maximale. La multiplicité des mesures est également un facteur essentiel pour la réussite de la méthode.

L'avantage du SAD par rapport au MAD est que le temps d'exposition du cristal au rayonnement X est minimisé. Ceci réduit les dommages causés par la radiation sur les molécules du cristal pendant l'enregistrement de données. Néanmoins, cette méthode aboutit à une **ambiguïté sur le choix de la phase**. Cette ambiguïté est absente dans la méthode **MAD** grâce à la mesure de jeux de diffraction à des longueurs d'onde différentes.

Pour atteindre une solution unique, la méthode **SIR** (Single Isomorphous Replacement) nécessite d'une source d'information supplémentaire (deuxième dérivé dans la méthode MIR, Multiple Isomorphous Replacement, ou diffusion anormale d'un atome lourd présent dans la structure dans la méthode SIRAS, Single Isomorphous Replacement with Anomalous Scattering). Dans ces cas, deux phases possibles et **équiprobables** sont obtenues. Une des procédures adoptée consiste à calculer une carte faisant intervenir les deux phases à poids égaux. Cette solution implique la superposition de la vraie solution et de la fausse solution pour la phase. La contribution des phases erronées introduit un bruit élevé dans la carte, ce qui rend son interprétation difficile.

Dans le cas du **SAD**, cependant, la situation est légèrement différente. Les deux phases trouvées comme solutions possibles ne sont **pas équiprobables**. En effet, il existe une probabilité plus

importante pour que la bonne solution pour la phase de la structure totale soit la solution plus proche de la phase de la sous-structure des atomes lourds. Cette probabilité augmente quand la contribution des atomes lourds au facteur de structure totale est plus importante. Les phases de la sous-structure des atomes lourds peuvent alors être prises comme phases de départ.

La détermination de structures par SAD utilise des méthodes de calculs de probabilités sur les phases comme **SHARP** (de La Fortelle and Bricogne 1997). Parmi les outils que propose **SHARP**, la carte résiduelle permet de vérifier les erreurs de positionnement des atomes lourds utilisés lors du phasage. Ce programme a été choisi pour la résolution du problème des phases de CstF-50dm (1-65). Sur 7 possibles Se-Met présentes dans la séquence de la protéine, 4 sites du Se ont pu être localisés et affinés.

#### *b.6) Construction et affinement de la structure*

La résolution de la structure implique l'obtention d'un modèle approximatif de la structure que nous cherchons à déterminer par modélisation dans une **carte de densité électronique interprétable**. Les facteurs de structure calculés sont plus ou moins en accord avec les facteurs de structures observés par les expériences de diffraction des rayons X. Pour l'amélioration de cet accord une étape d'affinement est nécessaire.

**L'affinement** est un processus d'alternance entre la construction du modèle dans l'espace réel, de façon manuelle (dans notre cas via le programme **Coot** (Emsley and Cowtan 2004)) ou automatique (**ARP/wARP** (Perrakis et al. 1999)), et l'affinement dans l'espace réciproque (programme **Refmac** (Murshudov et al. 1997)). Cette étape vise : 1) à modifier le modèle lorsque celui-ci est trop éloigné de la structure recherchée, donc à réduire l'écart entre les facteurs de structure observés et calculés, et 2) à compenser par l'affinement les erreurs induites par la reconstruction manuelle.

Pour affiner l'ensemble des paramètres de position et d'agitation de la structure atomique, le nombre d'observations doit être bien supérieur au nombre de ces paramètres. En effet, pour chaque méthode d'affinement, un rayon de convergence peut être défini. Ce rayon de convergence dépend de la résolution des données. En général, l'affinement numérique converge plus facilement avec des données ayant une résolution maximale proche de l'Angström que celui avec des données ayant une résolution de l'ordre de 3 Å pour lesquelles il n'y a qu'une observation par paramètre affiné.

Les programmes informatiques utilisés pour la construction et l'affinement de la structure de la protéine CstF-50hs (1-65) sont :

- **Coot** (Emsley and Cowtan 2004) : nous avons utilisé ce logiciel pour visualiser et manipuler les cartes de densités électroniques, ainsi que les modèles moléculaires.
- **ARP/wARP** : ce logiciel a été créé pour l'amélioration et l'interprétation objective des cartes de densité électronique et permet la reconstruction automatique et l'affinement des modèles

macromoléculaires. Le placement des acides aminés dans la carte de densité a été ainsi effectué manuellement et corroboré à l'aide de ce programme.

- **REFMAC5** (*Macromolecular Refinement program* (Murshudov et al. 1997) (Winn et al. 2001)): Ce programme, disponible avec la suite de programmes CCP4, a spécialement été développé pour affiner les structures par les méthodes de **minimisation**. Grâce à différentes options, le programme propose soit d'affiner la structure en corps rigides (le modèle est divisé en morceaux considérés comme rigides), ou d'affiner l'ensemble des coordonnées avec ou sans le facteur d'agitation de chaque atome. Le programme, à la fin de cette procédure, calcule les nouvelles coordonnées du modèle affiné et ses facteurs de structures associés.

- **phénix.refine** : ce programme a été utilisé de façon sporadique dans l'affinement de la structure. Phenix (Python-based Hierarchical ENvironment for Integrated Xtallography, (Adams et al. 2002) inclut un ensemble complet de logiciels pour la résolution des structures cristallographiques. Le programme **phénix.refine** effectue l'amélioration par le maximum de vraisemblance avec la possibilité d'affiner les paramètres de position et de déplacement (ADP) ainsi que d'agitation thermique ( $B_{\text{factor}}$ ) de manière individuelle ou groupée.

Pour contrôler l'affinement au fur et à mesure et pour la validation finale du modèle, nous disposons de plusieurs indices qui peuvent être vérifiés à chaque cycle d'affinement. Ainsi, il existe plusieurs **critères pour évaluer la qualité d'un modèle affiné** :

- **Le facteur R** peut donner une idée de la qualité du modèle en mesurant l'accord entre les facteurs de structure observés et les facteurs de structure calculés :

$$R = \frac{\sum_{h,k,l} \left| |F_{obs}^{hkl}| - K' |F_{calc}^{hkl}| \right|}{\sum_{h,k,l} |F_{obs}^{hkl}|}$$

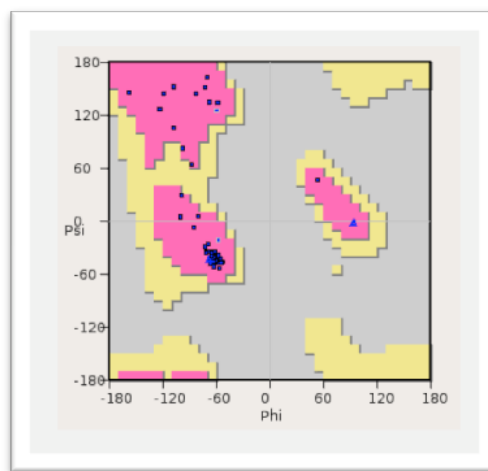
Toutefois, le facteur R n'est pas très objectif étant donné que, les méthodes d'affinement tentent de minimiser au maximum ce facteur et la modification des paramètres pendant l'affinement va influencer sa valeur. Une estimation moins biaisée de la qualité du modèle peut être définie par le facteur **R<sub>libre</sub>** ( $R_{\text{free}}$ ) calculé de la même façon que le facteur R, mais sur un ensemble de réflexions choisies aléatoirement (5 à 10 %) et non incluses dans l'affinement (Brunger 1992).

Les  $F_{\text{calc}}$  sont calculés globalement à partir du modèle, et sont comparés aux  $F_{\text{obs}}$  servant à l'affinement (R) et aux  $F_{\text{obs}}$  exclus ( $R_{\text{libre}}$ ). Une amélioration du modèle diminuera donc la valeur du facteur  $R_{\text{libre}}$ . Une diminution du R sans diminution du  $R_{\text{libre}}$  est le signe d'une modélisation de bruit à la place de l'amélioration du modèle. De cette manière, l'écart entre les facteurs R et  $R_{\text{libre}}$  est un indice que la qualité du modèle. Plus le rapport entre le nombre d'observations et le nombre de

paramètres à affiner est faible, plus cet écart est important. Pour CstF-50dm (1-65), le modèle a été affiné à  $1,4\text{\AA}$  de résolution avec un R et  $R_{\text{libre}}$  de **18,14%** et **21,74%**, respectivement.

- **Les déviations moyennes** des longueurs et des angles de liaisons entre atomes du modèle estiment sa qualité stéréochimique (root mean square déviation, r.m.s.d.). La qualité stéréochimique d'une structure est considérée comme acceptable quand ses déviations ne sont pas supérieures aux variations physiques des longueurs et des angles de liaison déjà observées pour les petites molécules (environ  $0.02\text{\AA}$  pour les longueurs et  $2^\circ$  pour les angles).

- **Le diagramme de Ramachandran** permet aussi de juger la qualité stéréochimique du modèle affiné (Ramakrishnan and Ramachandran 1965). En raison de l'encombrement stérique des chaînes latérales des aminoacides, les liaisons peptidiques ne peuvent adopter qu'un certain nombre limité de conformations. Pour chaque résidu, ces conformations sont caractérisées par les angles  $\Phi$  et  $\Psi$  de la liaison peptidique et sont représentées sur une graphique (le diagramme de Ramachandran) qui présente les régions énergétiquement favorables, les régions permises et celles non autorisées. Les chaînes latérales sont aussi soumises à des conformations énergétiquement favorables qui seront fortement représentées parmi les différentes conformations rencontrées dans les structures de protéines. Le même type de diagramme permet de valider leur stéréochimie (*fig. 107*).



**Figure 107.** Diagramme Ramachandran obtenu pour CstF-50. Il est observé sur l'image que tous les résidus sont situés dans les zones

- Enfin, **les contacts atomiques, la longueur et l'angle des liaisons hydrogènes potentiels** permettent d'identifier les régions erronées du modèle.



---

## Bibliographie

- Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff and G. Dreyfuss** (1986). "mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence." *Mol Cell Biol* **6**(8): 2932-2943.
- Adams, P. D., R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter and T. C. Terwilliger** (2002). "PHENIX: building new software for automated crystallographic structure determination." *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11): 1948-1954.
- Adamson, T. E. and D. H. Price** (2003). "Cotranscriptional processing of *Drosophila* histone mRNAs." *Mol Cell Biol* **23**(12): 4046-4055.
- Addepalli, B. and A. G. Hunt** (2007). "A novel endonuclease activity associated with the *Arabidopsis* ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor." *Nucleic Acids Res* **35**(13): 4453-4463.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman** (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* **25**(17): 3389-3402.
- Amrani, N., M. Minet, M. Le Gouar, F. Lacroute and F. Wyers** (1997). "Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro." *Mol Cell Biol* **17**(7): 3694-3701.
- Amrani, N., M. Minet, F. Wyers, M. E. Dufour, L. P. Aggerbeck and F. Lacroute** (1997). "PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I." *Mol Cell Biol* **17**(3): 1102-1109.
- Anderson, J. T., S. M. Wilson, K. V. Datar and M. S. Swanson** (1993). "NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability." *Mol Cell Biol* **13**(5): 2730-2741.
- Aravind, L.** (1999). "An evolutionary classification of the metallo-beta-lactamase fold proteins." *In Silico Biol* **1**(2): 69-91.
- Auweter, S. D., F. C. Oberstrass and F. H. Allain** (2006). "Sequence-specific binding of single-stranded RNA: is there a code for recognition?" *Nucleic Acids Res* **34**(17): 4943-4959.
- Awasthi, S. and J. C. Alwine** (2003). "Association of polyadenylation cleavage factor I with U1 snRNP." *RNA* **9**(11): 1400-1409.
- Bagga, P. S., G. K. Arhin and J. Wilusz** (1998). "DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro." *Nucleic Acids Res* **26**(23): 5343-5350.
- Bai, Y., T. C. Auperin, C. Y. Chou, G. G. Chang, J. L. Manley and L. Tong** (2007). "Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors." *Mol. Cell* **25**(6): 863-875.

- Barabino, S. M., W. Hubner, A. Jenny, L. Minvielle-Sebastia and W. Keller** (1997). "The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins." *Genes Dev* **11**(13): 1703-1716.
- Barabino, S. M., M. Ohnacker and W. Keller** (2000). "Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs." *Embo J* **19**(14): 3778-3787.
- Barbosa, E. and B. Moss** (1978). "mRNA(nucleoside-2'-)-methyltransferase from vaccinia virus. Characteristics and substrate specificity." *J Biol Chem* **253**(21): 7698-7702.
- Barbosa, E. and B. Moss** (1978). "mRNA(nucleoside-2'-)-methyltransferase from vaccinia virus. Purification and physical properties." *J Biol Chem* **253**(21): 7692-7697.
- Bard, J., A. M. Zhelkovsky, S. Helmling, T. N. Earnest, C. L. Moore and A. Bohm** (2000). "Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP." *Science* **289**(5483): 1346-1349.
- Barilla, D., B. A. Lee and N. J. Proudfoot** (2001). "Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*." *Proc Natl Acad Sci U S A* **98**(2): 445-450.
- Bell, S. A. and A. G. Hunt** (2010). "The Arabidopsis ortholog of the 77kDa subunit of the cleavage stimulatory factor (AtCstF-77) involved in mRNA polyadenylation is an RNA-binding protein." *FEBS Lett*.
- Berget, S. M.** (1995). "Exon recognition in vertebrate splicing." *J Biol Chem* **270**(6): 2411-2414.
- Bernstein, P. and J. Ross** (1989). "Poly(A), poly(A) binding protein and the regulation of mRNA stability." *Trends Biochem Sci* **14**(9): 373-377.
- Beyer, K., T. Dandekar and W. Keller** (1997). "RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA." *J. Biol. Chem.* **272**(42): 26769-26779.
- Bienroth, S., W. Keller and E. Wahle** (1993). "Assembly of a processive messenger RNA polyadenylation complex." *Embo J* **12**(2): 585-594.
- Birse, C. E., L. Minvielle-Sebastia, B. A. Lee, W. Keller and N. J. Proudfoot** (1998). "Coupling termination of transcription to messenger RNA maturation in yeast." *Science* **280**(5361): 298-301.
- Bloch, J. C., F. Perrin and F. Lacroute** (1978). "Yeast temperature-sensitive mutants specifically impaired in processing of poly(A)-containing RNAs." *Mol Gen Genet* **165**(2): 123-127.
- Boireau, S., P. Maiuri, E. Basyuk, M. de la Mata, A. Knezevich, B. Pradet-Balade, V. Backer, A. Kornblihtt, A. Marcello and E. Bertrand** (2007). "The transcriptional cycle of HIV-1 in real-time and live cells." *J Cell Biol* **179**(2): 291-304.
- Bond, U. M., T. A. Yario and J. A. Steitz** (1991). "Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro." *Genes Dev* **5**(9): 1709-1722.
- Bonneaud, N., L. Minvielle-Sebastia, C. Cullin and F. Lacroute** (1994). "Cellular localization of RNA14p and RNA15p, two yeast proteins involved in mRNA stability." *J Cell Sci* **107** ( Pt 4): 913-921.



- Bork, P., K. Hofmann, P. Bucher, A. F. Neuwald, S. F. Altschul and E. V. Koonin** (1997). "A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins." *FASEB J* **11**(1): 68-76.
- Both, G. W., Y. Furuichi, S. Muthukrishnan and A. J. Shatkin** (1975). "Ribosome binding to reovirus mRNA in protein synthesis requires 5' terminal 7-methylguanosine." *Cell* **6**(2): 185-195.
- Brodsky, A. S. and P. A. Silver** (2000). "Pre-mRNA processing factors are required for nuclear export." *RNA* **6**(12): 1737-1749.
- Brown, C. E. and A. B. Sachs** (1998). "Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation." *Mol Cell Biol* **18**(11): 6548-6559.
- Brown, K. M. and G. M. Gilmartin** (2003). "A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im." *Mol Cell* **12**(6): 1467-1476.
- Brunger, A. T.** (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**(6359): 472-475.
- Brzovic, P. S., J. R. Keefe, H. Nishikawa, K. Miyamoto, D. Fox, 3rd, M. Fukuda, T. Ohta and R. Klevit** (2003). "Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex." *Proc Natl Acad Sci U S A* **100**(10): 5646-5651.
- Bucheli, M. E., X. He, C. D. Kaplan, C. L. Moore and S. Buratowski** (2007). "Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI." *RNA* **13**(10): 1756-1764.
- Buratowski, S.** (2003). "The CTD code." *Nat Struct Biol* **10**(9): 679-680.
- Burkard, K. T. and J. S. Butler** (2000). "A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p." *Mol Cell Biol* **20**(2): 604-616.
- Callebaut, I., D. Moshous, J. P. Mornon and J. P. de Villartay** (2002). "Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family." *Nucleic Acids Res* **30**(16): 3592-3601.
- Calvo, O. and J. L. Manley** (2001). "Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination." *Mol Cell* **7**(5): 1013-1023.
- Cao, G. J., J. Pogliano and N. Sarkar** (1996). "Identification of the coding region for a second poly(A) polymerase in *Escherichia coli*." *Proc Natl Acad Sci U S A* **93**(21): 11580-11585.
- Cao, G. J. and N. Sarkar** (1992). "Identification of the gene for an *Escherichia coli* poly(A) polymerase." *Proc Natl Acad Sci U S A* **89**(21): 10380-10384.
- Castelo-Branco, P., A. Furger, M. Wollerton, C. Smith, A. Moreira and N. Proudfoot** (2004). "Polypyrimidine tract binding protein modulates efficiency of polyadenylation." *Mol Cell Biol* **24**(10): 4174-4183.
- Cevher, M. A., X. Zhang, S. Fernandez, S. Kim, J. Baquero, P. Nilsson, S. Lee, A. Virtanen and F. E. Kleiman** (2010). "Nuclear deadenylation/polyadenylation factors regulate 3' processing in response to DNA damage." *EMBO J*.

- Chanfreau, G., S. M. Noble and C. Guthrie** (1996). "Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF)." *Science* **274**(5292): 1511-1514.
- Chapman, R. D., M. Heidemann, C. Hintermair and D. Eick** (2008). "Molecular evolution of the RNA polymerase II CTD." *Trends Genet* **24**(6): 289-296.
- Chen, F., C. C. MacDonald and J. Wilusz** (1995). "Cleavage site determinants in the mammalian polyadenylation signal." *Nucleic Acids Res* **23**(14): 2614-2620.
- Chen, J. and C. Moore** (1992). "Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA." *Mol Cell Biol* **12**(8): 3470-3481.
- Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski** (2001). "Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain." *Genes Dev* **15**(24): 3319-3329.
- Chou, Z. F., F. Chen and J. Wilusz** (1994). "Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals." *Nucleic Acids Res* **22**(13): 2525-2531.
- Cole, C., J. D. Barber and G. J. Barton** (2008). "The Jpred 3 secondary structure prediction server." *Nucleic Acids Res* **36**(Web Server issue): W197-201.
- Colgan, D. F. and J. L. Manley** (1997). "Mechanism and regulation of mRNA polyadenylation." *Genes Dev* **11**(21): 2755-2766.
- Connelly, S. and J. L. Manley** (1988). "A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II." *Genes Dev* **2**(4): 440-452.
- Cooke, C. and J. C. Alwine** (1996). "The cap and the 3' splice site similarly affect polyadenylation efficiency." *Mol Cell Biol* **16**(6): 2579-2584.
- Cooke, C., H. Hans and J. C. Alwine** (1999). "Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal." *Mol Cell Biol* **19**(7): 4971-4979.
- Corden, J. L.** (1990). "Tails of RNA polymerase II." *Trends Biochem Sci* **15**(10): 383-387.
- Corden, J. L., D. L. Cadena, J. M. Ahearn, Jr. and M. E. Dahmus** (1985). "A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II." *Proc Natl Acad Sci U S A* **82**(23): 7934-7938.
- Coseno, M., G. Martin, C. Berger, G. Gilmartin, W. Keller and S. Doublie** (2008). "Crystal structure of the 25 kDa subunit of human cleavage factor Im." *Nucleic Acids Res* **36**(10): 3474-3483.
- Cosson, B., A. Couturier, S. Chabelskaya, D. Kiktev, S. Inge-Vechtomov, M. Philippe and G. Zhouravleva** (2002). "Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation." *Mol Cell Biol* **22**(10): 3301-3315.
- Cotten, M., O. Gick, A. Vasserot, G. Schaffner and M. L. Birnstiel** (1988). "Specific contacts between mammalian U7 snRNA and histone precursor RNA are indispensable for the in vitro 3' RNA processing reaction." *EMBO J* **7**(3): 801-808.

- Couttet, P., M. Fromont-Racine, D. Steel, R. Pictet and T. Grange** (1997). "Messenger RNA deadenylation precedes decapping in mammalian cells." *Proc Natl Acad Sci U S A* **94**(11): 5628-5633.
- Cramer, P.** (2002). "Multisubunit RNA polymerases." *Curr Opin Struct Biol* **12**(1): 89-97.
- Custodio, N., M. Vivo, M. Antoniou and M. Carmo-Fonseca** (2007). "Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site." *J Cell Biol* **179**(2): 199-207.
- Danckwardt, S., N. H. Gehring, G. Neu-Yilik, P. Hundsdoerfer, M. Pforsich, U. Frede, M. W. Hentze and A. E. Kulozik** (2004). "The prothrombin 3' end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations." *Blood* **104**(2): 428-435.
- Danckwardt, S., K. Hartmann, N. H. Gehring, M. W. Hentze and A. E. Kulozik** (2006). "3' end processing of the prothrombin mRNA in thrombophilia." *Acta Haematol* **115**(3-4): 192-197.
- Danckwardt, S., M. W. Hentze and A. E. Kulozik** (2008). "3' end mRNA processing: molecular mechanisms and implications for health and disease." *Embo J* **27**(3): 482-498.
- Danckwardt, S., I. Kaufmann, M. Gentzel, K. U. Foerstner, A. S. Gantzert, N. H. Gehring, G. Neu-Yilik, P. Bork, W. Keller, M. Wilm, M. W. Hentze and A. E. Kulozik** (2007). "Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals." *EMBO J* **26**(11): 2658-2669.
- Dantoni, J. C., K. G. Murthy, J. L. Manley and L. Tora** (1997). "Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA." *Nature* **389**(6649): 399-402.
- Darnell, J. E., L. Philipson, R. Wall and M. Adesnik** (1971). "Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA." *Science* **174**(8): 507-510.
- Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer** (2007). "In vivo dynamics of RNA polymerase II transcription." *Nat Struct Mol Biol* **14**(9): 796-806.
- Das, K., L. C. Ma, R. Xiao, B. Radvansky, J. Aramini, L. Zhao, J. Marklund, R. L. Kuo, K. Y. Twu, E. Arnold, R. M. Krug and G. T. Montelione** (2008). "Structural basis for suppression of a host antiviral response by influenza A virus." *Proc Natl Acad Sci U S A* **105**(35): 13093-13098.
- Dass, B., L. McDaniel, R. A. Schultz, E. Attaya and C. C. MacDonald** (2002). "The gene CSTF2T, encoding the human variant CstF-64 polyadenylation protein tauCstF-64, lacks introns and may be associated with male sterility." *Genomics* **80**(5): 509-514.
- Dass, B., S. Tardif, J. Y. Park, B. Tian, H. M. Weitlauf, R. A. Hess, K. Carnes, M. D. Griswold, C. L. Small and C. C. Macdonald** (2007). "Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility." *Proc Natl Acad Sci U S A* **104**(51): 20374-20379.
- de La Fortelle, E. and G. Bricogne** (1997). "Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods." *Methods Enzymol* **Volume 276**: 472-494.
- de Vries, H., U. Ruegsegger, W. Hubner, A. Friedlein, H. Langen and W. Keller** (2000). "Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors." *Embo J* **19**(21): 5895-5904.

**Decker, C. J. and R. Parker** (2002). "mRNA decay enzymes: decappers conserved between yeast and mammals." *Proc Natl Acad Sci U S A* **99**(20): 12512-12514.

**Deka, P., M. E. Bucheli, C. Moore, S. Buratowski and G. Varani** (2008). "Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals." *J Mol Biol* **375**(1): 136-150.

**Deka, P., P. K. Rajan, J. M. Perez-Canadillas and G. Varani** (2005). "Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein." *J Mol Biol* **347**(4): 719-733.

**Delaney, K. J., R. Xu, J. Zhang, Q. Q. Li, K. Y. Yun, D. L. Falcone and A. G. Hunt** (2006). "Calmodulin interacts with and regulates the RNA-binding activity of an Arabidopsis polyadenylation factor subunit." *Plant Physiol* **140**(4): 1507-1521.

**Deo, R. C., J. B. Bonanno, N. Sonenberg and S. K. Burley** (1999). "Recognition of polyadenylate RNA by the poly(A)-binding protein." *Cell* **98**(6): 835-845.

**Dettwiler, S., C. Aringhieri, S. Cardinale, W. Keller and S. M. Barabino** (2004). "Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization." *J Biol Chem* **279**(34): 35788-35797.

**Dheur, S., T. A. Vo le, F. Voisinet-Hakil, M. Minet, J. M. Schmitter, F. Lacroute, F. Wyers and L. Minvielle-Sebastia** (2003). "Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation." *Embo J* **22**(11): 2831-2840.

**Dichtl, B., R. Aasland and W. Keller** (2004). "Functions for *S. cerevisiae* Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation." *RNA* **10**(6): 965-977.

**Dichtl, B., D. Blank, M. Ohnacker, A. Friedlein, D. Roeder, H. Langen and W. Keller** (2002). "A role for SSU72 in balancing RNA polymerase II transcription elongation and termination." *Mol Cell* **10**(5): 1139-1150.

**Dichtl, B., D. Blank, M. Sadowski, W. Hubner, S. Weiser and W. Keller** (2002). "Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination." *Embo J* **21**(15): 4125-4135.

**Dichtl, B. and W. Keller** (2001). "Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor." *Embo J* **20**(12): 3197-3209.

**Dominski, Z. and W. F. Marzluff** (1999). "Formation of the 3' end of histone mRNA." *Gene* **239**(1): 1-14.

**Dominski, Z., X. C. Yang and W. F. Marzluff** (2005). "The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing." *Cell* **123**(1): 37-48.

**Dominski, Z., X. C. Yang, M. Purdy, E. J. Wagner and W. F. Marzluff** (2005). "A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100." *Mol Cell Biol* **25**(4): 1489-1500.

**Doublet, S.** (1997). "Preparation of selenomethionyl proteins for phase determination." *Methods Enzymol* **276**: 523-530.

- Dower, K. and M. Rosbash** (2002). "T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export." *RNA* **8**(5): 686-697.
- Dubrovskaya, V., A. C. Lavigne, I. Davidson, J. Acker, A. Staub and L. Tora** (1996). "Distinct domains of hTAFIII100 are required for functional interaction with transcription factor TFIIF beta (RAP30) and incorporation into the TFIID complex." *EMBO J* **15**(14): 3702-3712.
- Edmonds, M.** (2002). "A history of poly A sequences: from formation to factors to function." *Prog. Nucleic Acid. Res. Mol. Biol.* **71**: 285-389.
- Edmonds, M. and R. Abrams** (1960). "Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei." *J Biol Chem* **235**: 1142-1149.
- Edmonds, M., M. H. Vaughan, Jr. and H. Nakazato** (1971). "Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship." *Proc Natl Acad Sci U S A* **68**(6): 1336-1340.
- Edwards-Gilbert G, V. K., Milcarek C.** (1997). "Alternative poly(A) site selection in complex transcription units: means to an end?" *Nucleic Acids Res. Jul 1;25(13):2547-61.* **25**(13): 2547-2561.
- Edwards, R. A., M. S. Lee, S. E. Tsutakawa, R. S. Williams, I. Nazeer, F. E. Kleiman, J. A. Tainer and J. N. Glover** (2008). "The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50." *Biochemistry* **47**(44): 11446-11456.
- Egloff, S. and S. Murphy** (2008). "Cracking the RNA polymerase II CTD code." *Trends Genet* **24**(6): 280-288.
- Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy** (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." *Science* **318**(5857): 1777-1779.
- Elliott, B. J., T. Dattaroy, L. R. Meeks-Midkiff, K. P. Forbes and A. G. Hunt** (2003). "An interaction between an Arabidopsis poly(A) polymerase and a homologue of the 100 kDa subunit of CPSF." *Plant Mol Biol* **51**(3): 373-384.
- Emes, R. D. and C. P. Ponting** (2001). "A new sequence motif linking lissencephaly, Treacher Collins and oral-facial-digital type 1 syndromes, microtubule dynamics and cell migration." *Hum Mol Genet* **10**(24): 2813-2820.
- Emsley, P. and K. Cowtan** (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.
- Ensinger, M. J., S. A. Martin, E. Paoletti and B. Moss** (1975). "Modification of the 5'-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus." *Proc Natl Acad Sci U S A* **72**(7): 2525-2529.
- Eperon, L. P., I. R. Graham, A. D. Griffiths and I. C. Eperon** (1988). "Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase?" *Cell* **54**(3): 393-401.
- Fields, S. and O. Song** (1989). "A novel genetic system to detect protein-protein interactions." *Nature* **340**(6230): 245-246.

- Fitzgerald, M. and T. Shenk** (1981). "The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs." *Cell* **24**(1): 251-260.
- Fong, N. and D. L. Bentley** (2001). "Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD." *Genes Dev* **15**(14): 1783-1795.
- Fong, N., G. Bird, M. Vigneron and D. L. Bentley** (2003). "A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing." *EMBO J* **22**(16): 4274-4282.
- Forbes, K. P., B. Addepalli and A. G. Hunt** (2006). "An Arabidopsis Fip1 homolog interacts with RNA and provides conceptual links with a number of other polyadenylation factor subunits." *J Biol Chem* **281**(1): 176-186.
- Furuichi, Y., A. LaFiandra and A. J. Shatkin** (1977). "5'-Terminal structure and mRNA stability." *Nature* **266**(5599): 235-239.
- Garcia-Gimeno, M. A., I. Munoz, J. Arino and P. Sanz** (2003). "Molecular characterization of Ypi1, a novel *Saccharomyces cerevisiae* type 1 protein phosphatase inhibitor." *J Biol Chem* **278**(48): 47744-47752.
- Gaudet, R., A. Bohm and P. B. Sigler** (1996). "Crystal structure at 2.4 angstroms resolution of the complex of transducin betagamma and its regulator, phosducin." *Cell* **87**(3): 577-588.
- Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga** (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." *Nature* **415**(6868): 141-147.
- Gerlitz, G., E. Darhin, G. Giorgio, B. Franco and O. Reiner** (2005). "Novel functional features of the Lis-H domain: role in protein dimerization, half-life and cellular localization." *Cell Cycle* **4**(11): 1632-1640.
- Getz, M. J., P. K. Elder, E. W. Benz, Jr., R. E. Stephens and H. L. Moses** (1976). "Effect of cell proliferation on levels and diversity of poly(A)-containing mRNA." *Cell* **7**(2): 255 -265.
- Ghazy, M. A., X. He, B. N. Singh, M. Hampsey and C. Moore** (2009). "The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing." *Mol Cell Biol* **29**(8): 2296-2307.
- Gilbert, W. and C. Guthrie** (2004). "The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA." *Mol Cell* **13**(2): 201-212.
- Gilbert, W., C. W. Siebel and C. Guthrie** (2001). "Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation." *RNA* **7**(2): 302-313.
- Gilmartin, G. M. and J. R. Nevins** (1989). "An ordered pathway of assembly of components required for polyadenylation site recognition and processing." *Genes Dev* **3**(12B): 2180-2190.

- Gilmartin, G. M. and J. R. Nevins** (1991). "Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA." *Mol Cell Biol* **11**(5): 2432-2438.
- Glover-Cutter, K., S. Kim, J. Espinosa and D. L. Bentley** (2008). "RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes." *Nat Struct Mol Biol* **15**(1): 71-78.
- Gouet, P., X. Robert and E. Courcelle** (2003). "ESPrpt/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins." *Nucleic Acids Res* **31**(13): 3320-3323.
- Graber, J. H., C. R. Cantor, S. C. Mohr and T. F. Smith** (1999). "Genomic detection of new yeast pre-mRNA 3'-end-processing signals." *Nucleic Acids Res* **27**(3): 888-894.
- Green, M. R.** (1991). "Biochemical mechanisms of constitutive and regulated pre-mRNA splicing." *Annu Rev Cell Biol* **7**: 559-599.
- Gross, S. and C. Moore** (2001). "Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I." *Proc. Natl. Acad. Sci. U. S. A.* **98**(11): 6080-6085.
- Gross, S. and C. L. Moore** (2001). "Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation." *Mol. Cell. Biol.* **21**(23): 8045-8055.
- Guhaniyogi, J. and G. Brewer** (2001). "Regulation of mRNA stability in mammalian cells." *Gene* **265**(1-2): 11-23.
- Gunderson, S. I., K. Beyer, G. Martin, W. Keller, W. C. Boelens and L. W. Mattaj** (1994). "The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase." *Cell* **76**(3): 531-541.
- Gunderson, S. I., M. Polycarpou-Schwarz and I. W. Mattaj** (1998). "U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase." *Mol Cell* **1**(2): 255-264.
- Guo, Z. and F. Sherman** (1995). "3'-end-forming signals of yeast mRNA." *Mol Cell Biol* **15**(11): 5983-5990.
- Guo, Z. and F. Sherman** (1996). "3'-end-forming signals of yeast mRNA." *Trends Biochem Sci* **21**(12): 477-481.
- Hammell, C. M., S. Gross, D. Zenklusen, C. V. Heath, F. Stutz, C. Moore and C. N. Cole** (2002). "Coupling of termination, 3' processing, and mRNA export." *Mol Cell Biol* **22**(18): 6441-6457.
- Hastings, M. L. and A. R. Krainer** (2001). "Pre-mRNA splicing in the new millennium." *Curr Opin Cell Biol* **13**(3): 302-309.
- Hatton, L. S., J. J. Eloranta, L. M. Figueiredo, Y. Takagaki, J. L. Manley and K. O'Hare** (2000). "The *Drosophila* homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene." *Nucleic Acids Res* **28**(2): 520-526.

- He, X., A. U. Khan, H. Cheng, D. L. Pappas, Jr., M. Hampsey and C. L. Moore** (2003). "Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1." *Genes Dev* **17**(8): 1030-1042.
- He, X. and C. Moore** (2005). "Regulation of yeast mRNA 3' end processing by phosphorylation." *Mol Cell* **19**(5): 619-629.
- Hector, R. E., K. R. Nykamp, S. Dheur, J. T. Anderson, P. J. Non, C. R. Urbinati, S. M. Wilson, L. Minvielle-Sebastia and M. S. Swanson** (2002). "Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export." *Embo J* **21**(7): 1800-1810.
- Helmling, S., A. Zhelkovsky and C. L. Moore** (2001). "Fip1 regulates the activity of Poly(A) polymerase through multiple interactions." *Mol Cell Biol* **21**(6): 2026-2037.
- Herr, A. J., A. Molnar, A. Jones and D. C. Baulcombe** (2006). "Defective RNA processing enhances RNA silencing and influences flowering of Arabidopsis." *Proc Natl Acad Sci U S A* **103**(41): 14994-15001.
- Hirose, Y. and J. L. Manley** (1998). "RNA polymerase II is an essential mRNA polyadenylation factor." *Nature* **395**(6697): 93-96.
- Ho, C. K. and S. Shuman** (1999). "Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme." *Mol Cell* **3**(3): 405-411.
- Hockert, J. A., H. J. Yeh and C. C. Macdonald** (2010). "The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation." *J Biol Chem*.
- Hollingworth, D., C. G. Noble, I. A. Taylor and A. Ramos** (2006). "RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11." *Rna* **12**(4): 555-560.
- Holm, L. and P. Rosenstrom** (2010). "Dali server: conservation mapping in 3D." *Nucleic Acids Res* **38** Suppl: W545-549.
- Hook, A. G. and R. E. Kellems** (1988). "Localization and sequence analysis of poly(A) sites generating multiple dihydrofolate reductase mRNAs." *J Biol Chem* **263**(5): 2337-2343.
- Hoshino, S., M. Imai, T. Kobayashi, N. Uchida and T. Katada** (1999). "The eukaryotic polypeptide chain releasing factor (eRF3/GSPT) carrying the translation termination signal to the 3'-Poly(A) tail of mRNA. Direct association of erf3/GSPT with polyadenylate-binding protein." *J Biol Chem* **274**(24): 16677-16680.
- Hu, J., C. S. Lutz, J. Wilusz and B. Tian** (2005). "Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation." *RNA* **11**(10): 1485-1493.
- Huang, Y. and G. G. Carmichael** (1996). "Role of polyadenylation in nucleocytoplasmic transport of mRNA." *Mol Cell Biol* **16**(4): 1534-1542.
- Huez, G., G. Marbaix, D. Gallwitz, E. Weinberg, R. Devos, E. Hubert and Y. Cleuter** (1978). "Functional stabilisation of HeLa cell histone messenger RNAs injected into *Xenopus* oocytes by 3'-OH polyadenylation." *Nature* **271**(5645): 572-573.
- Hunt, A. G., R. Xu, B. Addepalli, S. Rao, K. P. Forbes, L. R. Meeks, D. Xing, M. Mo, H. Zhao, A. Bandyopadhyay, L. Dampanaboina, A. Marion, C. Von Lanken and Q. Q. Li** (2008).



"Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling." *BMC Genomics* **9**: 220.

**Irminger-Finger, I. and W. C. Leung** (2002). "BRCA1-dependent and independent functions of BARD1." *Int J Biochem Cell Biol* **34**(6): 582-587.

**Irniger, S. and G. H. Braus** (1994). "Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in *Saccharomyces cerevisiae*." *Proc Natl Acad Sci U S A* **91**(1): 257-261.

**Jacobson, A. and M. Favreau** (1983). "Possible involvement of poly(A) in protein synthesis." *Nucleic Acids Res* **11**(18): 6353-6368.

**Jacobson, A. and S. W. Peltz** (1996). "Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells." *Annu Rev Biochem* **65**: 693-739.

**Jenny, A., L. Minvielle-Sebastia, P. J. Preker and W. Keller** (1996). "Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I." *Science* **274**(5292): 1514-1517.

**Johnson, S. A., G. Cubberley and D. L. Bentley** (2009). "Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11." *Mol Cell* **33**(2): 215-226.

**Juge, F., A. Audibert, B. Benoit and M. Simonelig** (2000). "Tissue-specific autoregulation of Drosophila suppressor of forked by alternative poly(A) site utilization leads to accumulation of the suppressor of forked protein in mitotically active cells." *RNA* **6**(11): 1529-1538.

**Kabsch, W.** (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." *J. Appl. Cryst.* **26**: 795-800.

**Kashiwabara, S., T. Zhuang, K. Yamagata, J. Noguchi, A. Fukamizu and T. Baba** (2000). "Identification of a novel isoform of poly(A) polymerase, TPAP, specifically present in the cytoplasm of spermatogenic cells." *Dev Biol* **228**(1): 106-115.

**Kaufmann, I., G. Martin, A. Friedlein, H. Langen and W. Keller** (2004). "Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase." *Embo J* **23**(3): 616-626.

**Kazerouninia, A., B. Ngo and H. G. Martinson** (2009). "Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro." *RNA*.

**Keller, R. W., U. Kuhn, M. Aragon, L. Bornikova, E. Wahle and D. G. Bear** (2000). "The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail." *J Mol Biol* **297**(3): 569-583.

**Kennedy, S. A., M. L. Frazier, M. Steiniger, A. M. Mast, W. F. Marzluff and M. R. Redinbo** (2009). "Crystal structure of the HEAT domain from the Pre-mRNA processing factor Symplekin." *J Mol Biol* **392**(1): 115-128.

**Kerwitz, Y., U. Kuhn, H. Lilie, A. Knoth, T. Scheuermann, H. Friedrich, E. Schwarz and E. Wahle** (2003). "Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA." *Embo J* **22**(14): 3705-3714.

- Kessler, M. M., M. F. Henry, E. Shen, J. Zhao, S. Gross, P. A. Silver and C. L. Moore** (1997). "Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast." *Genes Dev* **11**(19): 2545-2556.
- Kessler, M. M., J. Zhao and C. L. Moore** (1996). "Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends." *J Biol Chem* **271**(43): 27167-27175.
- Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedeja, J. F. Greenblatt and S. Buratowski** (2004). "The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II." *Nature* **432**(7016): 517-522.
- Kim, M. H., D. R. Cooper, A. Oleksy, Y. Devedjiev, U. Derewenda, O. Reiner, J. Otlewski and Z. S. Derewenda** (2004). "The structure of the N-terminal domain of the product of the lissencephaly gene *Lis1* and its functional implications." *Structure* **12**(6): 987-998.
- Kleiman, F. E. and J. L. Manley** (1999). "Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50." *Science* **285**(5433): 1576-1579.
- Kleiman, F. E. and J. L. Manley** (2001). "The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression." *Cell* **104**(5): 743-753.
- Knaus, R., R. Pollock and L. Guarente** (1996). "Yeast SUB1 is a suppressor of TFIIB mutations and has homology to the human co-activator PC4." *EMBO J* **15**(8): 1933-1940.
- Kolev, N. G. and J. A. Steitz** (2005). "Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs." *Genes Dev*.
- Kolev, N. G., T. A. Yario, E. Benson and J. A. Steitz** (2008). "Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation." *EMBO Rep* **9**(10): 1013-1018.
- Komarnitsky, P., E. J. Cho and S. Buratowski** (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." *Genes Dev* **14**(19): 2452-2460.
- Kornblihtt, A. R., M. de la Mata, J. P. Fededa, M. J. Munoz and G. Nogues** (2004). "Multiple links between transcription and splicing." *RNA* **10**(10): 1489-1498.
- Korner, C. G. and E. Wahle** (1997). "Poly(A) tail shortening by a mammalian poly(A)-specific 3'-exoribonuclease." *J Biol Chem* **272**(16): 10448-10456.
- Korner, C. G., M. Wormington, M. Muckenthaler, S. Schneider, E. Dehlin and E. Wahle** (1998). "The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes." *Embo J* **17**(18): 5427-5437.
- Kozlov, G., J. F. Trempe, K. Khaleghpour, A. Kahvejian, I. Ekiel and K. Gehring** (2001). "Structure and function of the C-terminal PABC domain of human poly(A)-binding protein." *Proc Natl Acad Sci U S A* **98**(8): 4409-4413.
- Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey** (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." *Mol Cell* **14**(3): 387-394.

- Kyburz, A., A. Friedlein, H. Langen and W. Keller** (2006). "Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing." *Mol Cell* **23**(2): 195-205.
- Kyburz, A., M. Sadowski, B. Dichtl and W. Keller** (2003). "The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation." *Nucleic Acids Res* **31**(14): 3936-3945.
- Lee, J. H. and D. G. Skalnik** (2005). "CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex." *J Biol Chem* **280**(50): 41725-41731.
- Leeper, T. C., X. Qu, C. Lu, C. Moore and G. Varani** (2010). "Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1." *J Mol Biol* **401**(3): 334-349.
- Legrand, P., N. Pinaud, L. Minvielle-Sebastia and S. Fribourg** (2007). "The structure of the CstF-77 homodimer provides insights into CstF assembly." *Nucleic Acids Res* **35**(13): 4515-4522.
- Lemay, J. F., A. D'Amours, C. Lemieux, D. H. Lackner, V. G. St-Sauveur, J. Bahler and F. Bachand** (2010). "The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs." *Mol Cell* **37**(1): 34-45.
- Lemieux, C. and F. Bachand** (2009). "Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs." *Nucleic Acids Res* **37**(10): 3418-3430.
- Li, Q. and A. G. Hunt** (1995). "A near-upstream element in a plant polyadenylation signal consists of more than six nucleotides." *Plant Mol Biol* **28**(5): 927-934.
- Li, T., X. Chen, K. C. Garbutt, P. Zhou and N. Zheng** (2006). "Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase." *Cell* **124**(1): 105-117.
- Libri, D.** (2010). "Nuclear poly(a)-binding proteins and nuclear degradation: take the mRNA and run?" *Mol Cell* **37**(1): 3-5.
- Libri, D., K. Dower, J. Boulay, R. Thomsen, M. Rosbash and T. H. Jensen** (2002). "Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation." *Mol Cell Biol* **22**(23): 8254-8266.
- Licatalosi, D. D. and R. B. Darnell** (2010). "RNA processing and its regulation: global insights into biological networks." *Nat Rev Genet* **11**(1): 75-87.
- Licatalosi, D. D., G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil and D. L. Bentley** (2002). "Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II." *Mol. Cell* **9**(5): 1101-1111.
- Logan, J., E. Falck-Pedersen, J. E. Darnell, Jr. and T. Shenk** (1987). "A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene." *Proc Natl Acad Sci U S A* **84**(23): 8306-8310.
- Loke, J. C., E. A. Stahlberg, D. G. Strenski, B. J. Haas, P. C. Wood and Q. Q. Li** (2005). "Compilation of mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential secondary structures." *Plant Physiol* **138**(3): 1457-1468.

- Lunde, B. M., C. Moore and G. Varani** (2007). "RNA-binding proteins: modular design for efficient function." *Nat Rev Mol Cell Biol* **8**(6): 479-490.
- Lutz, C. S.** (2008). "Alternative polyadenylation: a twist on mRNA 3' end formation." *ACS Chem Biol* **3**(10): 609-617.
- Lutz, C. S., K. G. Murthy, N. Schek, J. P. O'Connor, J. L. Manley and J. C. Alwine** (1996). "Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro." *Genes Dev* **10**(3): 325-337.
- MacDonald, C. C., J. Wilusz and T. Shenk** (1994). "The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location." *Mol. Cell. Biol.* **14**(10): 6647-6654.
- Mandel, C. R., Y. Bai and L. Tong** (2008). "Protein factors in pre-mRNA 3'-end processing." *Cell Mol Life Sci* **65**(7-8): 1099-1122.
- Mandel, C. R., S. Kaneko, H. Zhang, D. Gebauer, V. Vethantham, J. L. Manley and L. Tong** (2006). "Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease." *Nature* **444**(7121): 953-956.
- Mangus, D. A., M. C. Evans and A. Jacobson** (2003). "Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression." *Genome Biol* **4**(7): 223.
- Maniatis, T. and R. Reed** (1987). "The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing." *Nature* **325**(6106): 673-678.
- Maniatis, T. and R. Reed** (2002). "An extensive network of coupling among gene expression machines." *Nature* **416**(6880): 499-506.
- Manley, J. L.** (1983). "Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate." *Cell* **33**(2): 595-605.
- Manley, J. L.** (1988). "Polyadenylation of mRNA precursors." *Biochim Biophys Acta* **950**(1): 1-12.
- Maris, C., C. Dominguez and F. H. Allain** (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." *FEBS J* **272**(9): 2118-2131.
- Martin, G. and W. Keller** (1996). "Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases." *EMBO J* **15**(10): 2593-2603.
- Martin, G. and W. Keller** (1998). "Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides." *RNA* **4**(2): 226-230.
- Martin, G., W. Keller and S. Doublet** (2000). "Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP." *Embo J.* **19**(16): 4193-4203.
- Martin, G., A. Moglich, W. Keller and S. Doublet** (2004). "Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase." *J. Mol. Biol.* **341**(4): 911-925.
- Martincic, K., R. Campbell, G. Edwalds-Gilbert, L. Souan, M. T. Lotze and C. Milcarek** (1998). "Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition." *Proc Natl Acad Sci U S A* **95**(19): 11095-11100.

- Mateja, A., T. Cierpicki, M. Paduch, Z. S. Derewenda and J. Otlewski** (2006). "The dimerization mechanism of LIS1 and its implication for proteins containing the LisH motif." *J Mol Biol* **357**(2): 621-631.
- McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman and D. L. Bentley** (1997). "5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Genes Dev* **11**(24): 3306-3318.
- McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D. Patterson, M. Wickens and D. L. Bentley** (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." *Nature* **385**(6614): 357-361.
- McCracken, S., D. Longman, I. L. Johnstone, J. F. Caceres and B. J. Blencowe** (2003). "An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation." *J Biol Chem* **278**(45): 44153-44160.
- McLauchlan, J., D. Gaffney, J. L. Whitton and J. B. Clements** (1985). "The consensus sequence YGTGTTY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini." *Nucleic Acids Res* **13**(4): 1347-1368.
- McLennan, A. G.** (2006). "The Nudix hydrolase superfamily." *Cell Mol Life Sci* **63**(2): 123-143.
- Meinhart, A. and P. Cramer** (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* **430**(6996): 223-226.
- Meinhart, A., T. Kamenski, S. Hoepfner, S. Baumli and P. Cramer** (2005). "A structural perspective of CTD function." *Genes Dev* **19**(12): 1401-1415.
- Meinke, G., C. Ezeokonkwo, P. Balbo, W. Stafford, C. Moore and A. Bohm** (2008). "Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein." *Biochemistry* **47**(26): 6859-6869.
- Meyer, S., C. Urbanke and E. Wahle** (2002). "Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths." *Biochemistry* **41**(19): 6082-6089.
- Millevoi, S., A. Decorsiere, C. Loulergue, J. Iacovoni, S. Bernat, M. Antoniou and S. Vagner** (2009). "A physical and functional link between splicing factors promotes pre-mRNA 3' end processing." *Nucleic Acids Res* **37**(14): 4672-4683.
- Minvielle-Sebastia, L., K. Beyer, A. M. Krecic, R. E. Hector, M. S. Swanson and W. Keller** (1998). "Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP." *Embo J.* **17**(24): 7454-7468.
- Minvielle-Sebastia, L. and W. Keller** (1999). "mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription." *Curr. Opin. Cell. Biol.* **11**(3): 352-357.
- Minvielle-Sebastia, L., P. J. Preker and W. Keller** (1994). "RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor." *Science* **266**(5191): 1702-1705.
- Minvielle-Sebastia, L., P. J. Preker, T. Wiederkehr, Y. Strahm and W. Keller** (1997). "The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation." *Proc. Natl. Acad. Sci. U. S. A.* **94**(15): 7897-7902.

- Minvielle-Sebastia, L., B. Winsor, N. Bonneaud and F. Lacroute** (1991). "Mutations in the yeast RNA14 and RNA15 genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA-binding domain in the RNA15 protein." *Mol. Cell. Biol.* **11**(6): 3075-3087.
- Mirkin, N., D. Fonseca, S. Mohammed, M. A. Cevher, J. L. Manley and F. E. Kleiman** (2008). "The 3' processing factor CstF functions in the DNA repair response." *Nucleic Acids Res* **36**(6): 1792-1804.
- Monarez, R. R., C. C. MacDonald and B. Dass** (2007). "Polyadenylation proteins CstF-64 and tauCstF-64 exhibit differential binding affinities for RNA polymers." *Biochem J* **401**(3): 651-658.
- Montell, C., E. F. Fisher, M. H. Caruthers and A. J. Berk** (1983). "Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA." *Nature* **305**(5935): 600-605.
- Moore, C. L., J. Chen and J. Whoriskey** (1988). "Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding." *EMBO J* **7**(10): 3159-3169.
- Moore, C. L. and P. A. Sharp** (1984). "Site-specific polyadenylation in a cell-free reaction." *Cell* **36**(3): 581-591.
- Moore, C. L. and P. A. Sharp** (1985). "Accurate cleavage and polyadenylation of exogenous RNA substrate." *Cell* **41**(3): 845-855.
- Mosavi, L. K., T. J. Cammett, D. C. Desrosiers and Z. Y. Peng** (2004). "The ankyrin repeat as molecular architecture for protein recognition." *Protein Sci* **13**(6): 1435-1448.
- Mowry, K. L. and J. A. Steitz** (1987). "Identification of the human U7 snRNP as one of several factors involved in the 3' end maturation of histone pre-messenger RNA's." *Science* **238**(4834): 1682-1687.
- Muller, B. and D. Schumperli** (1997). "The U7 snRNP and the hairpin binding protein: Key players in histone mRNA metabolism." *Semin Cell Dev Biol* **8**(6): 567-576.
- Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich** (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." *Cold Spring Harb Symp Quant Biol* **51 Pt 1**: 263-273.
- Munroe, D. and A. Jacobson** (1990). "mRNA poly(A) tail, a 3' enhancer of translational initiation." *Mol Cell Biol* **10**(7): 3441-3455.
- Murshudov, G. N., A. A. Vagin and E. J. Dodson** (1997). "Refinement of macromolecular structures by the maximum-likelihood method." *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-255.
- Murthy, K. G. and J. L. Manley** (1992). "Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus." *J. Biol. Chem.* **267**(21): 14804-14811.
- Murthy, K. G. and J. L. Manley** (1995). "The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation." *Genes Dev.* **9**(21): 2672-2683.
- Nag, A., K. Narsinh and H. G. Martinson** (2007). "The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase." *Nat Struct Mol Biol* **14**(7): 662-669.

- Nedea, E., X. He, M. Kim, J. Pootoolal, G. Zhong, V. Canadien, T. Hughes, S. Buratowski, C. L. Moore and J. Greenblatt** (2003). "Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends." *J Biol Chem* **278**(35): 33000-33010.
- Nedea, E., D. Nalbant, D. Xia, N. T. Theoharis, B. Suter, C. J. Richardson, K. Tatchell, T. Kislinger, J. F. Greenblatt and P. L. Nagy** (2008). "The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes." *Mol Cell* **29**(5): 577-587.
- Neilson, J. R. and R. Sandberg** (2010). "Heterogeneity in mammalian RNA 3' end formation." *Exp Cell Res*.
- Nemeroff, M. E., S. M. Barabino, Y. Li, W. Keller and R. M. Krug** (1998). "Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3'end formation of cellular pre-mRNAs." *Mol Cell* **1**(7): 991-1000.
- Nemeth, A., S. Krause, D. Blank, A. Jenny, P. Jenó, A. Lustig and E. Wahle** (1995). "Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II." *Nucleic Acids Res* **23**(20): 4034-4041.
- Nesic, D. and L. E. Maquat** (1994). "Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation." *Genes Dev* **8**(3): 363-375.
- Neuwald, A. F. and A. Poleksic** (2000). "PSI-BLAST searches using hidden markov models of structural repeats: prediction of an unusual sliding DNA clamp and of beta-propellers in UV-damaged DNA-binding protein." *Nucleic Acids Res* **28**(18): 3570-3580.
- Nevins, J. R. and J. E. Darnell, Jr.** (1978). "Steps in the processing of Ad2 mRNA: poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing." *Cell* **15**(4): 1477-1493.
- Niwa, M. and S. M. Berget** (1991). "Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns." *Genes Dev* **5**(11): 2086-2095.
- Noble, C. G., B. Beuth and I. A. Taylor** (2007). "Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor." *Nucleic Acids Res.* **35**(1): 87-99.
- Noble, C. G., D. Hollingworth, S. R. Martin, V. Ennis-Adeniran, S. J. Smerdon, G. Kelly, I. A. Taylor and A. Ramos** (2005). "Key features of the interaction between Pcf11 CID and RNA polymerase II CTD." *Nat. Struct. Mol. Biol.* **12**(2): 144-151.
- Noble, C. G., P. A. Walker, L. J. Calder and I. A. Taylor** (2004). "Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA." *Nucleic Acids Res.* **32**(11): 3364-3375.
- Nunes, N. M., W. Li, B. Tian and A. Furger** (2010). "A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence." *EMBO J.*
- O'Hare, K.** (1995). "mRNA 3' ends in focus." *Trends Genet* **11**(7): 255-257.
- Ohnacker, M., S. M. Barabino, P. J. Preker and W. Keller** (2000). "The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex." *Embo J.* **19**(1): 37-47.

- Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler and P. A. Sharp** (1986). "Splicing of messenger RNA precursors." *Annu Rev Biochem* **55**: 1119-1150.
- Pan, Z., H. Zhang, L. K. Hague, J. Y. Lee, C. S. Lutz and B. Tian** (2006). "An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism." *Gene* **366**(2): 325-334.
- Pancevac, C., D. C. Goldstone, A. Ramos and I. A. Taylor** (2010). "Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors." *Nucleic Acids Res.*
- Pandey, N. B., N. Chodchoy, T. J. Liu and W. F. Marzluff** (1990). "Introns in histone genes alter the distribution of 3' ends." *Nucleic Acids Res* **18**(11): 3161-3170.
- Perales, R. and D. Bentley** (2009). "'Cotranscriptionality': the transcription elongation complex as a nexus for nuclear transactions." *Mol Cell* **36**(2): 178-191.
- Perez Canadillas, J. M. and G. Varani** (2003). "Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein." *Embo J.* **22**(11): 2821-2830.
- Perez-Canadillas, J. M.** (2006). "Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1." *Embo J.* **25**(13): 3167-3178.
- Perrakis, A., R. Morris and V. S. Lamzin** (1999). "Automated protein model building combined with iterative structure refinement." *Nat Struct Biol* **6**(5): 458-463.
- Peterlin, B. M. and D. H. Price** (2006). "Controlling the elongation phase of transcription with P-TEFb." *Mol Cell* **23**(3): 297-305.
- Phatnani, H. P. and A. L. Greenleaf** (2006). "Phosphorylation and functions of the RNA polymerase II CTD." *Genes Dev* **20**(21): 2922-2936.
- Phillips, C., N. Pachikara and S. I. Gunderson** (2004). "U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions." *Mol Cell Biol* **24**(14): 6162-6171.
- Powell, H. R.** (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." *Acta Crystallogr D Biol Crystallogr* **55**(Pt 10): 1690-1695.
- Preiss, T. and M. W. Hentze** (1998). "Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast." *Nature* **392**(6675): 516-520.
- Preker, P. J. and W. Keller** (1998). "The HAT helix, a repetitive motif implicated in RNA processing." *Trends Biochem. Sci.* **23**(1): 15-16.
- Preker, P. J., J. Lingner, L. Minvielle-Sebastia and W. Keller** (1995). "The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase." *Cell* **81**(3): 379-389.
- Preker, P. J., M. Ohnacker, L. Minvielle-Sebastia and W. Keller** (1997). "A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor." *Embo J.* **16**(15): 4727-4737.
- Proudfoot, N.** (2004). "New perspectives on connecting messenger RNA 3' end formation to transcription." *Curr. Opin. Cell. Biol.* **16**(3): 272-278.



- Proudfoot, N. J. and G. G. Brownlee** (1976). "3' non-coding region sequences in eukaryotic messenger RNA." *Nature* **263**(5574): 211-214.
- Proudfoot, N. J., A. Furger and M. J. Dye** (2002). "Integrating mRNA processing with transcription." *Cell* **108**(4): 501-512.
- Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm and B. Seraphin** (2001). "The tandem affinity purification (TAP) method: a general procedure of protein complex purification." *Methods* **24**(3): 218-229.
- Qu, X., S. Lykke-Andersen, T. Nasser, C. Saguez, E. Bertrand, T. H. Jensen and C. Moore** (2009). "Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation." *Mol Cell Biol* **29**(19): 5327-5338.
- Qu, X., J. M. Perez-Canadillas, S. Agrawal, J. De Baecke, H. Cheng, G. Varani and C. Moore** (2007). "The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing." *J. Biol. Chem.* **282**(3): 2101-2115.
- Raabe, T., K. G. Murthy and J. L. Manley** (1994). "Poly(A) polymerase contains multiple functional domains." *Mol Cell Biol* **14**(5): 2946-2957.
- Ramakrishnan, C. and G. N. Ramachandran** (1965). "Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units." *Biophys J* **5**(6): 909-933.
- Rappsilber, J., U. Ryder, A. I. Lamond and M. Mann** (2002). "Large-scale proteomic analysis of the human spliceosome." *Genome Res* **12**(8): 1231-1245.
- Reyes-Reyes, M. and M. Hampsey** (2007). "Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation." *Mol Cell Biol* **27**(3): 926-936.
- Rigo, F. and H. G. Martinson** (2009). "Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing." *RNA* **15**(5): 823-836.
- Roberts, G. C., C. Gooding, H. Y. Mak, N. J. Proudfoot and C. W. Smith** (1998). "Co-transcriptional commitment to alternative splice site selection." *Nucleic Acids Res* **26**(24): 5568-5572.
- Rosonina, E., S. Kaneko and J. L. Manley** (2006). "Terminating the transcript: breaking up is hard to do." *Genes Dev.* **20**(9): 1050-1056.
- Rottman, F., A. J. Shatkin and R. P. Perry** (1974). "Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing." *Cell* **3**(3): 197-199.
- Rougemaille, M., G. Dieppois, E. Kisseleva-Romanova, R. K. Gudipati, S. Lemoine, C. Blugeon, J. Boulay, T. H. Jensen, F. Stutz, F. Devaux and D. Libri** (2008). "THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association." *Cell* **135**(2): 308-321.
- Rouillard, J. M., C. Brendolise and F. Lacroute** (2000). "Rna14p, a component of the yeast nuclear cleavage/polyadenylation factor I, is also localised in mitochondria." *Mol. Gen. Genet.* **262**(6): 1103-1112.

- Rozenblatt-Rosen, O., T. Nagaike, J. M. Francis, S. Kaneko, K. A. Glatt, C. M. Hughes, T. LaFramboise, J. L. Manley and M. Meyerson** (2009). "The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors." *Proc Natl Acad Sci U S A* **106**(3): 755-760.
- Ruby, S. W. and J. Abelson** (1991). "Pre-mRNA splicing in yeast." *Trends Genet* **7**(3): 79-85.
- Ruegsegger, U., K. Beyer and W. Keller** (1996). "Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors." *J Biol Chem* **271**(11): 6107-6113.
- Russell, R. B., P. D. Sasiemi and M. J. Sternberg** (1998). "Supersites within superfolds. Binding site similarity in the absence of homology." *J Mol Biol* **282**(4): 903-918.
- Ryan, K.** (2007). "Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation." *RNA Biol* **4**(1): 26-33.
- Ryan, K. and D. L. Bauer** (2008). "Finishing touches: post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation." *Int J Biochem Cell Biol* **40**(11): 2384-2396.
- Ryan, K., O. Calvo and J. L. Manley** (2004). "Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease." *Rna* **10**(4): 565-573.
- Sachs, A.** (1990). "The role of poly(A) in the translation and stability of mRNA." *Curr Opin Cell Biol* **2**(6): 1092-1098.
- Sachs, A. and E. Wahle** (1993). "Poly(A) tail metabolism and function in eucaryotes." *J Biol Chem* **268**(31): 22955-22958.
- Sachs, A. B. and R. W. Davis** (1989). "The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation." *Cell* **58**(5): 857-867.
- Sadowski, M., B. Dichtl, W. Hubner and W. Keller** (2003). "Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination." *Embo J* **22**(9): 2167-2177.
- Saguez, C. and T. H. Jensen** (2009). "Assembly of export-competent mRNP: it's all about being connected." *Mol Cell* **33**(2): 139-140.
- Saguez, C., M. Schmid, J. R. Olesen, M. A. Ghazy, X. Qu, M. B. Poulsen, T. Nasser, C. Moore and T. H. Jensen** (2008). "Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation." *Mol Cell* **31**(1): 91-103.
- Salisbury, J., K. W. Hutchison and J. H. Graber** (2006). "A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif." *BMC Genomics* **7**: 55.
- Scharl, E. C. and J. A. Steitz** (1994). "The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP." *EMBO J* **13**(10): 2432-2440.
- Schumperli, D.** (1988). "Multilevel regulation of replication-dependent histone genes." *Trends Genet* **4**(7): 187-191.

- Scully, R., A. Xie and G. Nagaraju** (2004). "Molecular functions of BRCA1 in the DNA damage response." *Cancer Biol Ther* **3**(6): 521-527.
- Shankarling, G. S., P. W. Coates, B. Dass and C. C. Macdonald** (2009). "A family of splice variants of CstF-64 expressed in vertebrate nervous systems." *BMC Mol Biol* **10**: 22.
- Sharp, P. A.** (1985). "On the origin of RNA splicing and introns." *Cell* **42**(2): 397-400.
- Shatkin, A. J.** (1976). "Capping of eucaryotic mRNAs." *Cell* **9**(4 PT 2): 645-653.
- Sheets, M. D., S. C. Ogg and M. P. Wickens** (1990). "Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro." *Nucleic Acids Res* **18**(19): 5799-5805.
- Shi, Y., D. C. Di Giammartino, D. Taylor, A. Sarkeshik, W. J. Rice, J. R. Yates, 3rd, J. Frank and J. L. Manley** (2009). "Molecular architecture of the human pre-mRNA 3' processing complex." *Mol Cell* **33**(3): 365-376.
- Shimotohno, K., Y. Kodama, J. Hashimoto and K. I. Miura** (1977). "Importance of 5'-terminal blocking structure to stabilize mRNA in eukaryotic protein synthesis." *Proc Natl Acad Sci U S A* **74**(7): 2734-2738.
- Siebel, C. W., L. Feng, C. Guthrie and X. D. Fu** (1999). "Conservation in budding yeast of a kinase specific for SR splicing factors." *Proc Natl Acad Sci U S A* **96**(10): 5440-5445.
- Simons, A., C. Melamed-Bessudo, R. Wolkowicz, J. Sperling, R. Sperling, L. Eisenbach and V. Rotter** (1997). "PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb." *Oncogene* **14**(2): 145-155.
- Simpson, G. G., P. P. Dijkwel, V. Quesada, I. Henderson and C. Dean** (2003). "FY is an RNA 3' end-processing factor that interacts with FCA to control the Arabidopsis floral transition." *Cell* **113**(6): 777-787.
- Skaar, D. A. and A. L. Greenleaf** (2002). "The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p." *Mol Cell* **10**(6): 1429-1439.
- Spencer, E., D. Loring, J. Hurwitz and G. Monroy** (1978). "Enzymatic conversion of 5'-phosphate-terminated RNA to 5'-di- and triphosphate-terminated RNA." *Proc Natl Acad Sci U S A* **75**(10): 4793-4797.
- Spies, N., C. B. Nielsen, R. A. Padgett and C. B. Burge** (2009). "Biased chromatin signatures around polyadenylation sites and exons." *Mol Cell* **36**(2): 245-254.
- Stirnimann, C. U., E. Petsalaki, R. B. Russell and C. W. Muller** (2010). "WD40 proteins propel cellular networks." *Trends Biochem Sci*.
- St. Pierre, B., X. Liu1, L. T. Kha, X. Zhu, O. Ryan, Z. Jiang and E. Zacksenhaus** (2005). "Conserved and specific functions of mammalian *ssu72*." *Nucleic Acids Res* **33**(2): 464-477.
- Strub, M. P., F. Hoh, J. F. Sanchez, J. M. Strub, A. Bock, A. Aumelas and C. Dumas** (2003). "Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing." *Structure* **11**(11): 1359-1367.

- Stutz, F., A. Bachi, T. Doerks, I. C. Braun, B. Seraphin, M. Wilm, P. Bork and E. Izaurralde** (2000). "REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export." *RNA* **6**(4): 638-650.
- Takahashi, Y., S. Helmling and C. L. Moore** (2003). "Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor." *Nucleic Acids Res* **31**(6): 1744-1752.
- Tai, C. Y., D. L. Dujardin, N. E. Faulkner and R. B. Vallee** (2002). "Role of dynein, dynactin, and CLIP-170 interactions in LIS1 kinetochore function." *J Cell Biol* **156**(6): 959-968.
- Takagaki, Y. and J. L. Manley** (1994). "A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein." *Nature* **372**(6505): 471-474.
- Takagaki, Y. and J. L. Manley** (1997). "RNA recognition by the human polyadenylation factor CstF." *Mol. Cell. Biol.* **17**(7): 3907-3914.
- Takagaki, Y. and J. L. Manley** (1998). "Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation." *Mol Cell* **2**(6): 761-771.
- Takagaki, Y. and J. L. Manley** (2000). "Complex protein interactions within the human polyadenylation machinery identify a novel component." *Mol. Cell. Biol.* **20**(5): 1515-1525.
- Takagaki, Y., J. L. Manley, C. C. MacDonald, J. Wilusz and T. Shenk** (1990). "A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs." *Genes Dev.* **4**(12A): 2112-2120.
- Takagaki, Y., L. C. Ryner and J. L. Manley** (1989). "Four factors are required for 3'-end cleavage of pre-mRNAs." *Genes Dev.* **3**(11): 1711-1724.
- Tarun, S. Z., Jr. and A. B. Sachs** (1996). "Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G." *EMBO J* **15**(24): 7168-7177.
- Thompson, J. D., D. G. Higgins and T. J. Gibson** (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res* **22**(22): 4673-4680.
- Tian, B., J. Hu, H. Zhang and C. S. Lutz** (2005). "A large-scale analysis of mRNA polyadenylation of human and mouse genes." *Nucleic Acids Res* **33**(1): 201-212.
- Torchet, C., C. Bousquet-Antonelli, L. Milligan, E. Thompson, J. Kufel and D. Tollervey** (2002). "Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs." *Mol. Cell* **9**(6): 1285-1296.
- Tresaugues, L., P. Stenmark, H. Schuler, S. Flodin, M. Welin, T. Nyman, M. Hammarstrom, M. Moche, S. Graslund and P. Nordlund** (2008). "The crystal structure of human cleavage and polyadenylation specific factor-5 reveals a dimeric Nudix protein with a conserved catalytic site." *Proteins* **73**(4): 1047-1052.
- Tsai, T. F., M. J. Wu and T. S. Su** (1998). "Usage of cryptic splice sites in citrullinemia fibroblasts suggests role of polyadenylation in splice-site selection during terminal exon definition." *DNA Cell Biol* **17**(8): 717-725.

- Tsuzuki, M., W. Wu, H. Nishikawa, R. Hayami, D. Oyake, Y. Yabuki, M. Fukuda and T. Ohta** (2006). "A truncated splice variant of human BARD1 that lacks the RING finger and ankyrin repeats." *Cancer Lett* **233**(1): 108-116.
- Uchida, N., S. Hoshino, H. Imataka, N. Sonenberg and T. Katada** (2002). "A novel role of the mammalian GSPT/eRF3 associating with poly(A)-binding protein in Cap/Poly(A)-dependent translation." *J Biol Chem* **277**(52): 50286-50292.
- Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadmodar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg** (2000). "A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*." *Nature* **403**(6770): 623-627.
- Vagner, S., C. Vagner and I. W. Mattaj** (2000). "The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing." *Genes Dev* **14**(4): 403-413.
- Valentini, S. R., V. H. Weiss and P. A. Silver** (1999). "Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation." *Rna* **5**(2): 272-280.
- Venkataraman, K., K. M. Brown and G. M. Gilmartin** (2005). "Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition." *Genes Dev* **19**(11): 1315-1327.
- Venkatesan, S., A. Gershowitz and B. Moss** (1980). "Modification of the 5' end of mRNA. Association of RNA triphosphatase with the RNA guanylyltransferase-RNA (guanine-7-)methyltransferase complex from vaccinia virus." *J Biol Chem* **255**(3): 903-908.
- Veraldi, K. L., G. K. Arhin, K. Martincic, L. H. Chung-Ganster, J. Wilusz and C. Milcarek** (2001). "hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells." *Mol Cell Biol* **21**(4): 1228-1238.
- Villa, T., M. Rougemaille and D. Libri** (2008). "Nuclear quality control of RNA polymerase II ribonucleoproteins in yeast: tilting the balance to shape the transcriptome." *Biochim Biophys Acta* **1779**(9): 524-531.
- Vinciguerra, P. and F. Stutz** (2004). "mRNA export: an assembly line from genes to nuclear pores." *Curr Opin Cell Biol* **16**(3): 285-292.
- Vo, L. T., M. Minet, J. M. Schmitter, F. Lacroute and F. Wyers** (2001). "Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA." *Mol Cell Biol* **21**(24): 8346-8356.
- Wahle, E.** (1991). "Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors." *J Biol Chem* **266**(5): 3131-3139.
- Wahle, E.** (1995). "Poly(A) tail length control is caused by termination of processive synthesis." *J Biol Chem* **270**(6): 2800-2808.
- Wahle, E. and W. Keller** (1996). "The biochemistry of polyadenylation." *Trends Biochem Sci* **21**(7): 247-250.
- Wahle, E. and U. Ruegsegger** (1999). "3'-End processing of pre-mRNA in eukaryotes." *FEMS Microbiol Rev* **23**(3): 277-295.

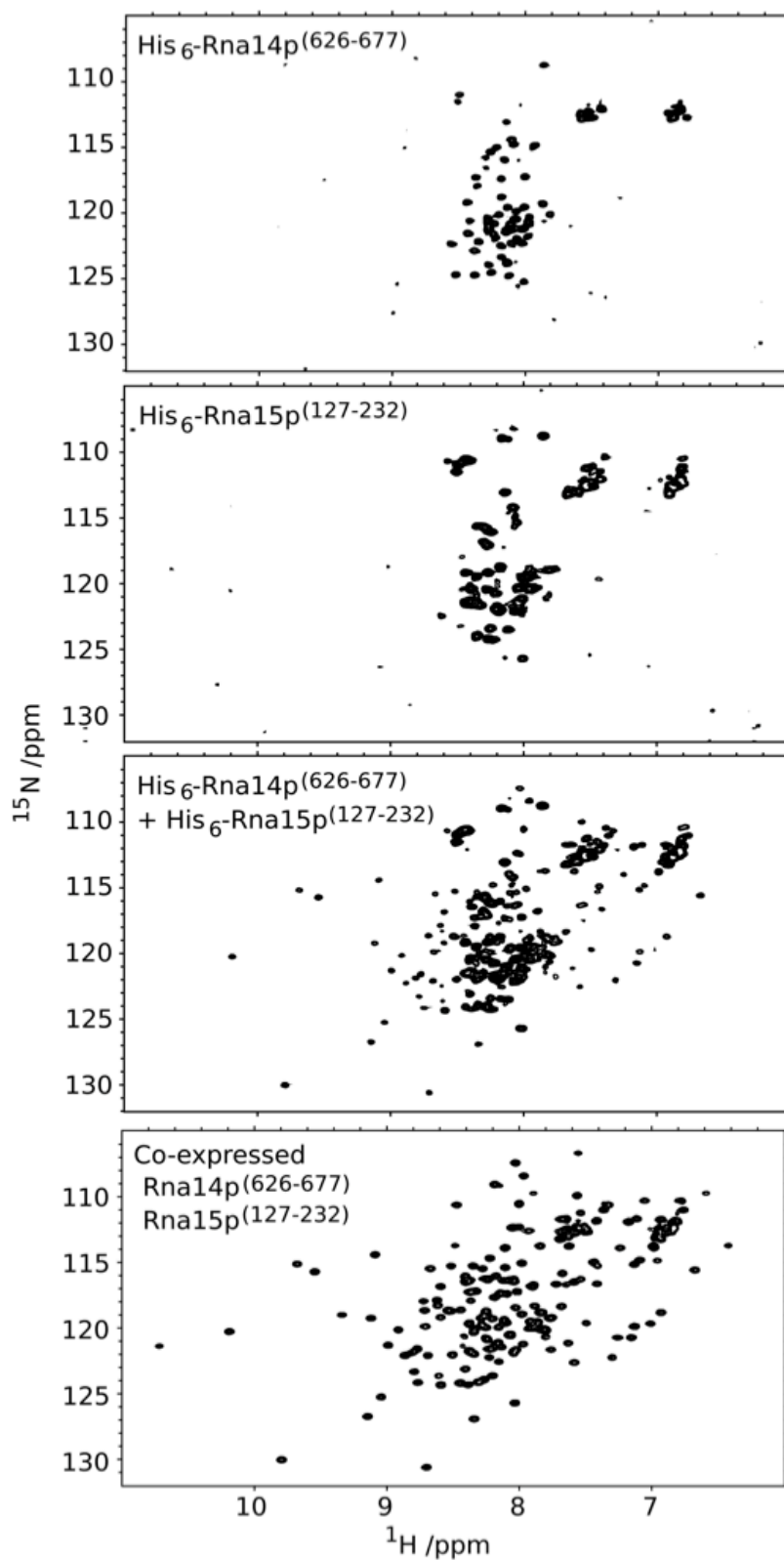
- Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay** (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." *EMBO J* **1**(8): 945-951.
- Wallace, A. M., B. Dass, S. E. Ravnik, V. Tonk, N. A. Jenkins, D. J. Gilbert, N. G. Copeland and C. C. MacDonald** (1999). "Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells." *Proc Natl Acad Sci U S A* **96**(12): 6763-6768.
- Wang, Y., J. A. Fairley and S. G. Roberts** (2010). "Phosphorylation of TFIIB links transcription initiation and termination." *Curr Biol* **20**(6): 548-553.
- Wang, Z. F., R. Tisovec, R. W. Debry, M. R. Frey, A. G. Matera and W. F. Marzluff** (1996). "Characterization of the 55-kb mouse histone gene cluster on chromosome 3." *Genome Res* **6**(8): 702-714.
- Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton** (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." *Bioinformatics* **25**(9): 1189-1191.
- Weiss, E. A., G. M. Gilmartin and J. R. Nevins** (1991). "Poly(A) site efficiency reflects the stability of complex formation involving the downstream element." *EMBO J* **10**(1): 215-219.
- Weitzer, S. and J. Martinez** (2007). "The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs." *Nature* **447**(7141): 222-226.
- Wells, D. and L. Kedes** (1985). "Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs." *Proc Natl Acad Sci U S A* **82**(9): 2834-2838.
- Wells, S. E., P. E. Hillner, R. D. Vale and A. B. Sachs** (1998). "Circularization of mRNA by eukaryotic translation initiation factors." *Mol Cell* **2**(1): 135-140.
- West, S., N. Gromak and N. J. Proudfoot** (2004). "Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites." *Nature* **432**(7016): 522-525.
- West, S. and N. J. Proudfoot** (2008). "Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination." *Nucleic Acids Res* **36**(3): 905-914.
- West, S., N. J. Proudfoot and M. J. Dye** (2008). "Molecular dissection of mammalian RNA polymerase II transcriptional termination." *Mol Cell* **29**(5): 600-610.
- Wickens, M. and P. Stephenson** (1984). "Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation." *Science* **226**(4678): 1045-1051.
- Wilusz, J. and T. Shenk** (1988). "A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif." *Cell* **52**(2): 221-228.
- Winn, M. D., M. N. Isupov and G. N. Murshudov** (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." *Acta Crystallogr D Biol Crystallogr* **57**(Pt 1): 122-133.

- Wu, W. H., I. Pinto, B. S. Chen and M. Hampsey** (1999). "Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB." *Genetics* **153**(2): 643-652.
- Xiang, K., T. Nagaike, T. Kilic, M. M. Beh, J. L. Manley, L. Tong** (2010). "Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex." *Nature* **467**, 729-733.
- Xing, D., H. Zhao, R. Xu and Q. Q. Li** (2008). "Arabidopsis PCFS4, a homologue of yeast polyadenylation factor Pcf11p, regulates FCA alternative processing and promotes flowering time." *Plant J* **54**(5): 899-910.
- Xu, R., H. Zhao, R. D. Dinkins, X. Cheng, G. Carberry and Q. Q. Li** (2006). "The 73 kD subunit of the cleavage and polyadenylation specificity factor (CPSF) complex affects reproductive development in Arabidopsis." *Plant Mol Biol* **61**(4-5): 799-815.
- Yan, J. and T. G. Marr** (2005). "Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat." *Genome Res* **15**(3): 369-375.
- Yang, Q., G. M. Gilmartin and S. Double** (2010). "Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing." *Proc Natl Acad Sci U S A* **107**(22): 10062-10067.
- Yang, X. C., K. D. Sullivan, W. F. Marzluff and Z. Dominski** (2009). "Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing." *Mol Cell Biol* **29**(1): 31-42.
- Yao, Y., L. Song, Y. Katz and G. Galili** (2002). "Cloning and characterization of Arabidopsis homologues of the animal CstF complex that regulates 3' mRNA cleavage and polyadenylation." *J Exp Bot* **53**(378): 2277-2278.
- Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill** (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.
- Zhang, Z., J. Fu and D. S. Gilmour** (2005). "CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11." *Genes Dev* **19**(13): 1572-1580.
- Zhao, J., L. Hyman and C. Moore** (1999). "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis." *Microbiol. Mol. Biol. Rev.* **63**(2): 405-445.
- Zhao, J., M. Kessler, S. Helmling, J. P. O'Connor and C. Moore** (1999). "Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor." *Mol Cell Biol* **19**(11): 7733-7740.
- Zhelkovsky, A., Y. Tacahashi, T. Nasser, X. He, U. Sterzer, T. H. Jensen, H. Domdey and C. Moore** (2006). "The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*." *RNA* **12**(3): 435-445.



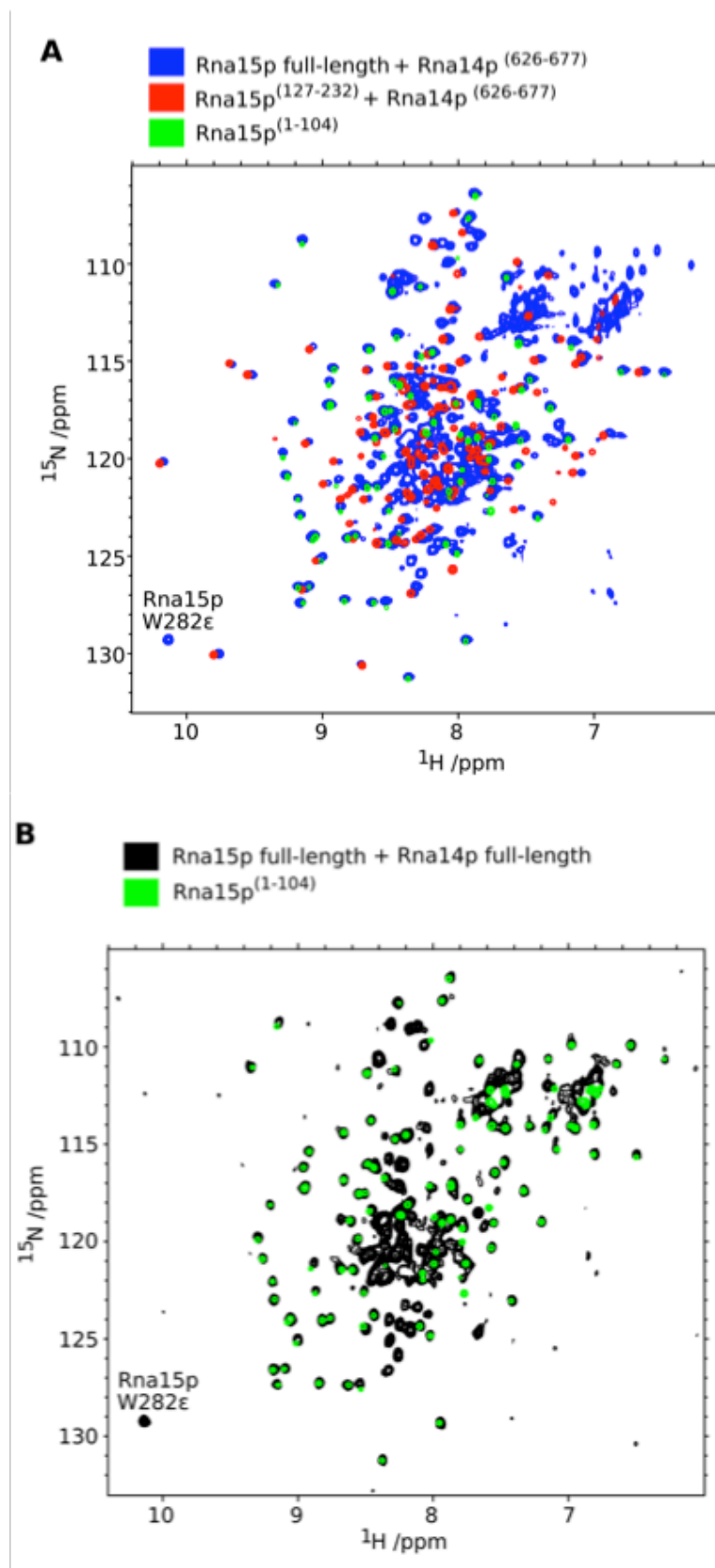


## Annexes

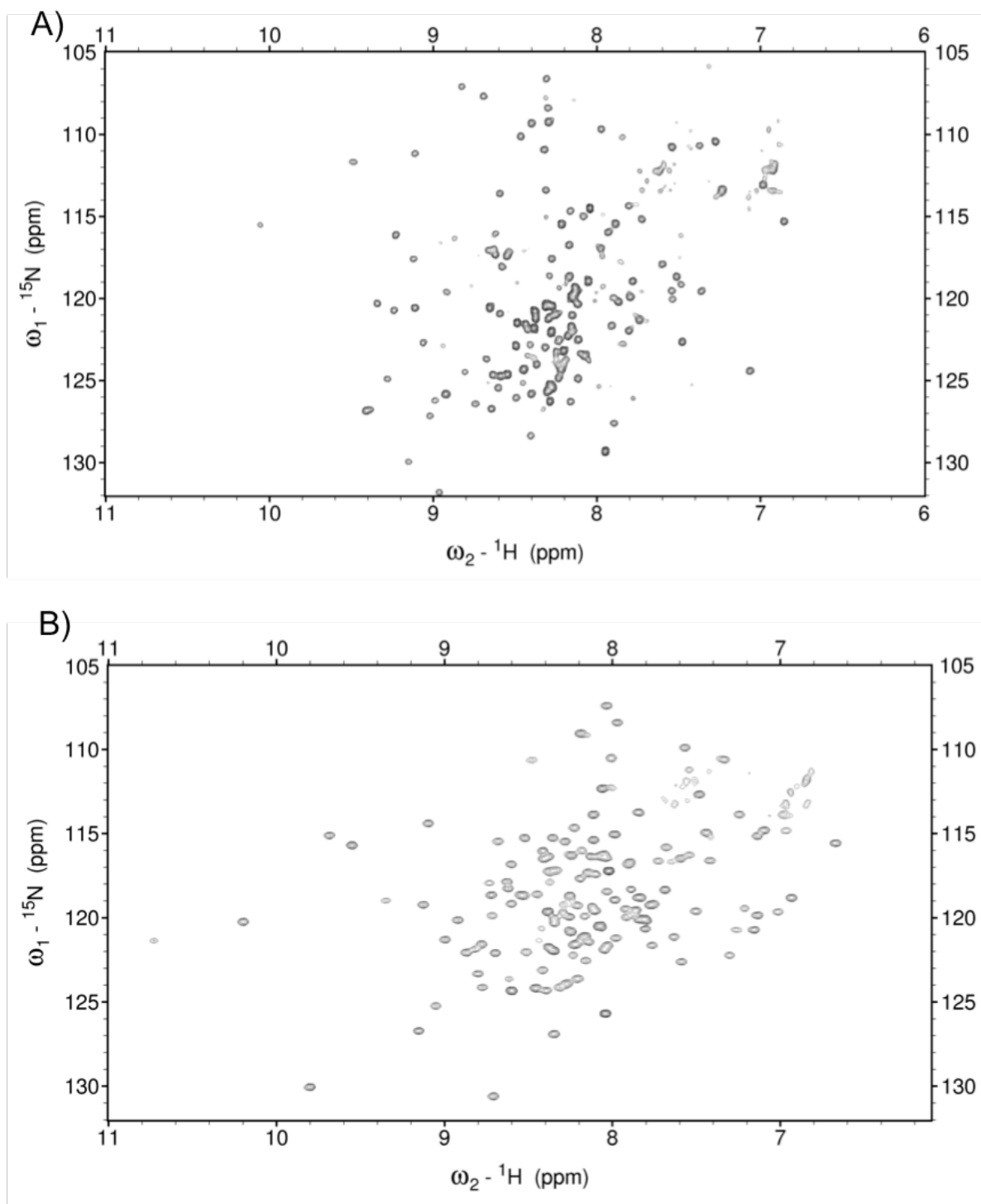


**Annexe 1.** Spectre bidimensionnel  $^1\text{H}^{15}\text{N}$ -HSQC à 700 Mhz. Les constructions protéiques sont indiquées sur chaque spectre. Le tampon utilisé pour les expériences contient : 20mM sodium phosphate (pH 6.5), 50mM NaCl, 2mM dithiothreitol et 10% (v/v)  $^2\text{H}_2\text{O}$ . Les complexes étaient concentrés à 200 $\mu\text{M}$  sauf pour la combinaison équimolaire H<sub>6</sub>-Rna14p (626-677) + H<sub>6</sub>Rna15p (127-232) (3<sup>ème</sup> spectre) où 100 $\mu\text{M}$  de chaque protéine étaient employée.





*Annexe 3. Superposition des spectres  $^1\text{H}^{15}\text{N}$ -TROSY des constructions indiquées. Le pic correspondant au Trp putatif est marqué par le nom du résidu.*



**Annexe 4.** A.) Spectre  ${}^1\text{H}, {}^{15}\text{N}$ -TROSY obtenu pour le complexe CstF-64 (1-211)/ CstF-77-(626-717). B.) Spectre  ${}^1\text{H}, {}^{15}\text{N}$ -TROSY obtenu pour le complexe Rna14p (626-677)/ Rna15p (127-232). Le profil des deux spectres ne diffère pas énormément ce qui peut être un signe d'un arrangement similaire. Il faut tenir en compte que certains pics du spectre situé à gauche correspondent aux résidus du RRM de CstF-64.









### **Title: Structural studies of the homologous metazoan CstF and yeast CFIA complexes essential for 3'-processing of pre-mRNA**

The removal of the 3' region of pre-mRNA followed by polyadenylation is a key step in mRNA maturation. In metazoa, Cleavage stimulation Factor (CstF) recognizes U and G/U rich cis-acting RNA sequence elements through its 64kDa subunit and helps stabilize the Cleavage Polyadenylation Stimulating Factor (CPSF) complex at the polyadenylation site. We describe the crystal structure of the N-terminal domain of the CstF-50 subunit. Through highly conserved residues, CstF-50 forms a compact homodimer that exposes two geometrically opposite and identical conserved surfaces. Together with prior data, the structure of the CstF-50 homodimerization domain supports a hexameric model of CstF. The yeast homologue of CstF is the Cleavage/polyadenylation Factor IA (CF IA) complex and is involved in both the cleavage and polyadenylation of pre-mRNA. We have reconstituted 'in vitro' the overall complex and also solved the solution structure of one of the inter-subunit regions, specifically the heterodimer involving peptides from Rna14p and Rna15p. Upon binding, a short C-terminal region from Rna14p wraps intimately within the central hinge domain from Rna15p. Conservation of residues reveals that the structural tethering is preserved in the homologous mammalian proteins.

**Keywords:** 3' end maturation; crystallography; dimerization; CstF; CstF-50; CF IA; Hinge; Rna14p; Rna15p



### **Titre : Etude structurale du complexe CstF et de son homologue chez la levure CF IA, deux facteurs indispensables pour la maturation 3' des pré-ARN messagers**

Une étape clé dans la maturation des pré-ARNms est le clivage et la polyadénylation que ceux-ci subissent sur leur extrémité 3'. Chez les métazoaires, le complexe CstF (Cleavage stimulation Factor) reconnaît une région de l'ARNm riche en U et U/G et stabilise le complexe CPSF (Cleavage Polyadenylation Stimulating Factor) sur le site de polyadénylation. Nous avons déterminé la structure cristallographique du domaine N-terminal d'une des trois sous-unités de CstF, CstF-50. Ce domaine forme un homodimère compact et présente deux surfaces identiques conservées dérivées de la formation du dimère. La structure dimérique de CstF-50 est en accord avec le modèle hexamérique du complexe. L'homologue de CstF chez la levure, CF IA (Cleavage/polyadenylation Factor IA), est impliqué dans les réactions de clivage et polyadénylation de la maturation 3'. Nous avons reconstitué le complexe entier 'in vitro' et résolu la structure en solution par RMN des régions minimales impliquées dans l'interaction des sous-unités Rna14p et Rna15p. Pour la formation de l'hétérodimère, la région C-terminale de Rna14p, que nous avons appelé domaine « monkeytail », s'entrelace intimement avec la région « hinge » de Rna15p. La présence de ces deux domaines chez leurs homologues de mammifères, CstF-77 et CstF-64, suggère la conservation de ce type d'organisation entre ces deux sous-unités à travers les espèces.

**Mots clés :** maturation 3' des pré-ARNms ; cristallographie ; dimère ; CstF ; CstF-50 ; CF IA ; Rna14p ; Rna15p



### **Título : Estudio estructural del complejo CstF y de su homólogo de levaduras CF IA, dos factores indispensables para la maduración 3' del pre-ARN mensajero.**

La maduración 3' del pre-ARNm es un proceso clave de la expresión génica que incluye el corte y la poliadenilación del extremo 3' libre del pre-ARNm. En metazoos, el complejo CstF (Cleavage stimulation Factor) reconoce una secuencia del pre-ARNm rica en U y G/U y permite la estabilización del complejo CPSF (Cleavage Polyadenylation Stimulating Factor) en el sitio de poliadenilación. Hemos descrito la estructura cristalina del dominio N-terminal de una de las tres subunidades de CstF, CstF-50. La estructura ha revelado la organización de la proteína en un dímero compacto y conservado entre las especies. Dos zonas idénticas conservadas se encuentran expuestas a ambos lados de la superficie estructural. Nuestros resultados corroboran así la hipótesis sobre el modelo hexamérico del complejo CstF. CF IA (Cleavage/ polyadenylation Factor IA), el homólogo de CstF en levaduras, interviene en las dos etapas de la maduración 3'. Las bases para la reconstitución del factor CF IA 'in vitro' han sido establecidas. Al mismo tiempo, hemos resuelto la estructura del subcomplejo formado por las regiones de interacción de Rna14p y de Rna15p en solución mediante RMN. En el heterodímero, las dos proteínas forman una entidad única a través de la región C-terminal de Rna14p, dominio "monkeytail", y el dominio "hinge" de Rna15p, quedando las hélices de la dos proteínas entrelazadas. La localización de estos dominios en sus homólogos mamíferos, CstF-77 et CstF-64, sugiere que este tipo de organización está conservada entre las especies.

**Palabras clave:** maduración 3' del pre-ARN; cristalografía; dímero; CstF; CstF-50; CF IA; Rna14p; Rna15p