



**HAL**  
open science

**Etude de l'ARN polymérase I et du rôle de ses  
sous-unités spécifiques chez la levure *Saccharomyces  
cerevisiae***

Tommy Darrière

► **To cite this version:**

Tommy Darrière. Etude de l'ARN polymérase I et du rôle de ses sous-unités spécifiques chez la levure *Saccharomyces cerevisiae*. Génétique. Université Paul Sabatier - Toulouse III, 2017. Français. NNT : 2017TOU30293 . tel-01988584

**HAL Id: tel-01988584**

**<https://theses.hal.science/tel-01988584>**

Submitted on 21 Jan 2019

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

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



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

---

**Présentée et soutenue par :**

**Tommy Darriere**

le jeudi 21 décembre 2017

**Titre :**

Etude de l'ARN polymérase I et du rôle de ses sous-unités spécifiques chez la levure *Saccharomyces cerevisiae*

---

**École doctorale et discipline ou spécialité :**

ED BSB : Génétique moléculaire

**Unité de recherche :**

UMR5099

**Directeur/trice(s) de Thèse :**

Olivier GADAL

**Jury :**

M. Patrick Schultz, DR CNRS, rapporteur  
M. Bertrand Daignan-Fornier, DR CNRS, rapporteur  
M. Cosmin Saveanu, PI Institut Pasteur, rapporteur  
Mme Célia Plisson-Chastang, CR1 CNRS, examinatrice  
Mme Kerstin Bystricky, Professeur UPS, examinatrice  
M. Olivier Gadal, DR CNRS, directeur de thèse



# Remerciements

Je ferai allègrement des remerciements à tous les membres du bâtiment, mais je manquerai de place... Je serai bref (j'essaierai), mais je n'en pense pas moins.

Plus particulièrement, je tiens à remercier les membres de mon équipe, qui ont travaillé de près ou de loin sur ce projet, expérimentalement ou théoriquement, mais surtout pour la bonne humeur et l'ambiance générale dans le groupe.

Un grand merci aux deux Christophe, dits "Le Normand" et "CriCri" pour leur aide toutes ces années dans le labo. Que ce soit sur du Renaud ou du David Guetta, la clope au bec ou prêt pour un 100 Km, j'ai toujours pu compter sur vous, dans la bonne humeur (oui, Normand qui râle sur Tom ou Bob, c'est de la bonne humeur...). Quant à toi Dez je te revois à Ibiza l'an prochaine, en forme...

Merci à Marta, le petit beignet (Pon tchou check en VO, je ne saurais pas l'écrire, mais c'est visuel comme ça), pour ta bonne humeur et ton "hihihi" (mais pas pour tes rhumes).

Fred, mon ami Gastropode, merci pour ton enthousiasme et tes chants/danses/reste (hmm hmm) tout au long de ces années, ne change pas, garde cet esprit, j'ai pris beaucoup de plaisir à travailler à la paillasse à côté d'un "veilleur de nuit noire". Mais prends toujours bien garde aux latrines du fond, elles recèlent encore quelques dangers et/ou surprises...

Mention spéciale à Isabelle, qui a investi un temps considérable à m'aider, notamment dans l'explication, l'obtention et l'analyse des fameux et magnifiques "Arbres de Noël", ainsi qu'à la relecture et correction de ce manuscrit, un grand merci à toi pour ta patience et ta gentillesse, et les bières bavaroises.

Enfin seconde mention spéciale à Olivier bien sûr, qui m'a tout appris ces trois années, autant du point de vue expérimental que de la réflexion, et qui a passé beaucoup de temps avec moi, ne serait-ce qu'à me montrer comment déposer une goutte sur une boîte. Je n'oublie même plus les contrôles, c'est pour dire le chemin parcouru! (quasi plus). Les "murs" à prendre, bien que parfois bien hauts, n'auront jamais été vraiment infranchissables, et c'est avec beaucoup de plaisir que j'ai manipulé/discuté avec toi. J'espère que l'on pourra continuer à travailler ensemble, quelque soit le chemin que je prendrai par la suite.

Une pensée en particulier pour ma MAMAN du laboratoire, Célia, qui me suit maintenant depuis le Master 2, et qui m'a permis de rencontrer la Gadal team. Ta gentillesse et ta bonne humeur (et la microscopie électronique) ont largement contribué à ma venue dans ce laboratoire, et j'espère garder contact avec toi autant que possible. Tommychou pourra

peut être repartir courir avec toi sur une longue distance, mais ça y est j'ai grandi, alors je te laisserai passer devant bien plus tôt!

En parlant de MET, je remercie aussi grandement les deux super demoiselles, les deux super commères, de la plate-forme METI, qui m'ont beaucoup formé et aidé au microscope, quelque soit la date ou l'heure (oui, l'aprem du 23 décembre aussi...), que j'ai des ragots sous la main ou non...

Une pensée également pour Cécile Bon, qui m'a encadré et formé durant mon M2 à l'IPBS, et qui m'a donné envie de continuer dans cette voie. J'espère que Pks13 n'aura bientôt plus de secret à cacher, j'y crois fort !

Je pense aussi au Club des Renés, soyez toujours forts et courageux pour les défis sportifs à venir, rien n'est insurmontable, mais bon ça fait mal des fois...

Merci aux différents services de l'IBCG, et notamment à Cathy, Safi, Alain, Bertrand et les autres, qui font que ce bâtiment et ce labo sont ce qu'ils sont au quotidien.

Et bien évidemment, un grand merci à tous mes amis étudiants et thésards (qui ne le sont déjà plus pour certains), avec qui j'ai tissé de nombreux liens et que j'espère garder, ces années auront réellement été fructueuses en termes d'amitiés.

Merci notamment à Tic et Tac, Natacha et Clément (ou l'inverse, je n'ai jamais su), qui étaient à mes côtés et que je connais depuis le M2, et qui m'ont permis de découvrir le laboratoire sous son meilleur angle, celui de la bonne humeur. Merci à Lise, ma confidente d'équipe, j'ai vraiment adoré travaillé à tes côtés, à celui de ton rire communicatif et des clins d'oeil entendus à certains moments de la journée... J'ai hâte de voir ta thèse maintenant, tu es la prochaine! Merci à Thomas et Sylvain, les deux compères de l'équipe d'à côté, que j'ai connu durant la thèse, et avec qui j'ai passé de supers moments, en faisant et buvant une (124578...) bières, ou devant l'ordinateur. On se suivra les prochaines années sur Star Citizen je n'en doute pas (Thomas achète toi un PC...). Et un grand merci à l'ensemble des étudiants et thésards du bâtiment et d'ailleurs, et en particulier à Ramtin, Cécilia, Duy Khanh, Alex, Roxanne, Lydia, Titouan, Barbara, Elise, Alix, Lud, et bien sûr Emilie pour tous ces moments, que ce soit au bar, dans le labo, au bar, à l'Upsi, au bar, ou ailleurs...

Et pour finir, un infini merci à ma famille, ma mère, mon père, ma soeur, mes grands-parents, et tous les autres, qui m'ont soutenu de tous points de vue, tout au long des ces années, de thèse certes, mais d'études tout court, et dieu sait si c'est long... Merci de m'avoir permis de réaliser ce que je voulais faire.

Un grand merci à tous...

## ABSTRACT

In eukaryotes cells, there are 3 nuclear RNA Polymerases (Pol I, II and III), each having a particular RNA synthesis function. The RNA Polymerase I (Pol I) produces a single transcript: the precursor of the large ribosomal RNA (rRNA), which correspond to a massive transcription activity in the cell. Structural data of this 14-subunits enzyme is now available. This allows a better understanding of its operating mode, and confirmed that the Pol I has 3 specific subunits, also called "Built-in transcription factors", capable of regulatory activities. Two of them, Rpa49/Rpa34, are forming a heterodimer structurally related to the Pol II transcription factors TFIIF and TFIIE, implicated in both initiation and elongation of the transcription. The last one, Rpa12, is known to have a role in Pol I stability and the cleavage activity of the paused Pol I (via its C-terminus part), like its homologous TFIIS in the Pol II. We performed extensive genetic studies of Pol I mutants lacking one of these subunits: Rpa49 (*rpa49Δ*). This depletion is viable, but results in initiation and elongation problems. Here, we report the cloning and characterization of extragenic suppressors mapping point mutations in three subunits of Pol I restoring efficient Pol I activities in absence of Rpa49. All suppressor mutations identified were structurally mapped firstly in the two largest subunits Rpa190 and Rpa135 very closed to Rpa12, and then in Rpa12 itself, indicating a possible interplay between the Rpa49 and Rpa12 subunits, and the area around. Notably, a RNA pol I specific element in Rpa190, called "DNA Mimicking Loop", is structurally very closed to the region in which we find all of our suppressor mutations. The genetic, structural but also biochemical and functional characterizations of these suppressors allow us to propose roles of these Rpa49 and Rpa12 subunits, but also of the small area of Rpa190, which has never been highlighted before.

## RESUME

Dans les cellules eucaryotes, il existe 3 ARN Polymérase nucléaires (Pol I, II et III), chacune ayant une fonction de synthèse d'ARN qui lui est propre. L'ARN polymérase I (Pol I) est responsable de la synthèse du précurseur des grands ARN ribosomiques (ARNr), ce qui correspond à une activité de transcription massive dans la cellule. Des données structurales sur cette enzyme de 14 sous-unités sont disponibles. Ceci permet une meilleure compréhension de son mode de fonctionnement, et a confirmé que l'ARN Pol I possède 3 sous-unités spécifiques, aussi appelées «Built-in Transcription Factors», responsables d'activités régulatrices. Deux d'entre elles, Rpa49 / Rpa34, forment un hétérodimère structurellement proche des facteurs de transcription TFIIF et TFIIE de la Pol II, impliqués tant dans l'initiation que dans l'élongation de la transcription. La dernière, Rpa12, est connue pour avoir un rôle dans la stabilité de l'ARN Pol I et l'activité de clivage du transcrit de la polymérase en cours de pause (via son extrémité C-terminale), comme son homologue TFIIS, facteur de transcription de l'ARN Pol II. Nous avons effectué des études génétiques sur des mutants de l'ARN Pol I dépourvus de la sous-unité Rpa49 (*rpa49Δ*). Cette délétion est viable mais entraîne des problèmes d'initiation et d'élongation. Nous étudions dans ce travail le clonage et la caractérisation de "suppresseurs" extragéniques correspondant à des mutations ponctuelles de trois sous-unité de la Pol I qui rétablissent une synthèse d'ARNr en l'absence de la sous-unité Rpa49. Toutes ces mutations suppressives identifiées ont été localisées premièrement dans les deux grandes sous-unités Rpa190 et Rpa135, structurellement très proches de Rpa12, mais également au sein même de Rpa12, indiquant une possible interaction entre les sous-unités Rpa49 et Rpa12 ou la région autour. Notamment, un élément spécifique de l'ARN Pol I dans Rpa190, le "DNA Mimicking Loop", est structurellement très proche de la région où l'on trouve ces mutations suppresseur. Les caractérisations génétiques, structurales, mais aussi biochimiques et fonctionnelles de ces suppresseurs nous permettent de proposer des hypothèses sur les rôles de Rpa49 et de Rpa12, mais aussi de cette petite région de Rpa190, en l'initiation de la transcription, ce qui n'a jamais été mis en évidence auparavant.

# Sommaire :

|   |    |
|---|----|
| <b>ABREVIATIONS</b> .....   | 4  |
| <b>PREAMBULE</b> .....  | 6  |
| <b>INTRODUCTION</b> .....   | 7  |
| <b>Chapitre I - Caractéristiques des eucaryotes uni- et multicellulaires, et de la levure</b>   |    |
| <b><i>Saccharomyces cerevisiae</i></b> .....  | 7  |
| I-1- Introduction aux cellules eucaryotes et à la levure <i>Saccharomyces cerevisiae</i> .....  | 7  |
| I-2- Le noyau chez les eucaryotes multicellulaires et <i>S. cerevisiae</i> . ....   | 10 |
| I-2-a- Chez <i>S.cerevisiae</i> , Le pore nucléaire et le centre organisateur des microtubules/SPB sont ancrés dans l'enveloppe nucléaire .....   | 15 |
| I-2-b- Introduction aux ARN polymérases.....  | 18 |
| I-2-c- Le nucléole, "l'usine à transcription" de la cellule.....  | 18 |
| <b>Chapitre II - Une étape limitante dans la prolifération cellulaire : la production des ARNr, composants essentiels des ribosomes</b> .....   | 22 |
| II-1- la biogénèse des ribosomes, machineries moléculaires responsables de la synthèse des protéines cellulaires.....   | 22 |
| II-2- Le locus d'ADNr et l'unité de transcription du précurseur des grands ARNr.....  | 27 |
| <b>Chapitre III - Différentes ARN polymérases dans le règne du vivant et leur conservation évolutive : généralités sur les ARN polymérases multimériques, et introduction par l'ARN polymérase bactérienne.</b> ..... | 30 |
| III-1- La synthèse des ARNs .....   | 35 |
| III-2- L'activité transcriptionnelle de l'ARN polymérase bactérienne: l'ARN polymérase multimérique modèle.....   | 38 |
| III-3- Le Core enzyme bactérien.....  | 39 |
| <b>Chapitre IV - Le cas des archées : une autre ARN polymérase ADN dépendante cellulaire unique</b> .....   | 45 |
| IV-1- La transcription chez les archées .....   | 45 |
| IV-2- Structure de l'ARN polymérase archéenne .....   | 48 |

## **Chapitre V - Les différentes ARN polymérases ADN dépendantes chez les eucaryotes .51**

|   |    |
|---|----|
| V -1- Présentation générale des ARN polymérases chez les eucaryotes .....   | 51 |
| V -2- L'activité de transcription par l'ARN pol II, la polymérase la plus étudiée et caractérisée .....                       | 54 |
| V-2-a- L'initiation par l'ARN polymérase II et ses facteurs associés: formation du complexe de pré-initiation .....           | 54 |
| <i>Les séquences d'ADN importantes pour la formation du complexe de pré-initiation pour la transcription par l'ARN pol II</i> |    |
| <i>Les facteurs généraux de transcription de la pol II, structures et rôles</i>   |    |
| V-2-b- Elongation, "Backtrack" et terminaison de la transcription par la pol II .....   | 64 |
| V-3- L'ARN polymérase III .....   | 67 |
| V-3-a- Activité de transcription et facteurs associés .....   | 67 |
| V-3-b- Données structurales .....   | 67 |

## **Chapitre VI - L'ARN polymérase I, responsable de la synthèse du précurseur des grands ARN ribosomiques : éléments fonctionnels et structuraux de la transcription de l'ADN ribosomique. ....**

|   |     |
|---|-----|
| .....   | 70  |
| VI-1- Définition des éléments structuraux de l'ARN polymérase I .....   | 73  |
| VI-1-a- Etats oligomériques de l'ARN pol I .....  | 73  |
| VI -1-b- Le Stalk de l'ARN pol I .....  | 78  |
| VI-1-c- Le facteur essentiel d'initiation Rrn3 .....  | 81  |
| VI -1-d- Les éléments caractéristiques du Core enzyme de l'ARN pol I .....  | 85  |
| VI-1-e- L'ARN polymérase I possède des sous-unités spécifiques :<br>les "Built-in Transcription factors".....                             | 90  |
| <i>Rpa34- Rpa49</i>   |     |
| <i>Rpa12</i>  |     |
| VI-2- La transcription par l'ARN pol I .....  | 98  |
| VI-2-a- La transcription par l'ARN pol I : régions promotrices et terminatrices de la transcription chez <i>S.cerevisiae</i> . ....       | 99  |
| VI -2-b- La formation du complexe d'initiation de l'ARN pol I requiert des facteurs d'initiation qui lui sont spécifiques .....           | 102 |
| <i>Des éléments clés dans la formation du PIC de l'ARN polymérase I chez S.cerevisiae : "Core Factor" et "Upstream Activating Factor"</i> |     |
| <i>Données structurales sur le PIC de l'ARN pol I</i>   |     |
| VI-2-c- L'équivalence de la formation du complexe d'initiation de l'ARN pol I chez  |     |



|   |            |
|---|------------|
| les mammifères .....  | 111        |
| VI-2-d- La transcription par l'ARN pol I : l'élongation de la transcription.....  | 113        |
| VI-2-e- La transcription par l'ARN pol I : la terminaison de la transcription .....   | 122        |
| <br>  |            |
| <b>Chapitre VII - Analyse fonctionnelle <i>in vivo</i> des modules structuraux de l'ARN pol I.</b>                                  | <b>126</b> |
| VII-1- Analyses des phénotypes associés à des mutants de certaines sous-unités de l'ARN pol I. ...                                  | 126        |
| VII-1-a- Le Stalk Rpa14-Rpa43 et Rrn3 .....   | 126        |
| <i>Rpa14</i>  |            |
| <i>Rpa43</i>  |            |
| <i>Rrn3</i>   |            |
| VII-1-a- Les Built-in Transcriptions factors Rpa34-Rpa49 et Rpa12.....  | 128        |
| <i>Rpa34 et Rpa49</i>   |            |
| - L'hétérodimère "TFIIF like module "   |            |
| - Analyse du mutant <i>rpa49Δ</i> et de la partie C-terminale de <i>Rpa49</i> : le domaine <i>tWH</i>                               |            |
| <i>La sous-unité Rpa12</i>  |            |
| <br>  |            |
| <b>PARTIE EXPERIMENTALE .....</b>   | <b>138</b> |
| Ebauche du manuscrit : " Genetic analysis of RNA polymerase I unveils roles of Rpa12 subunit during transcription initiation" ..... | I-XXXII    |
| <br>  |            |
| <b>DISCUSSIONS/PERSPECTIVES.....</b>  | <b>175</b> |
| I – De la génétique aux modules structuraux sur l'ARN polymérase I chez <i>S.cerevisiae</i> ....                                    | 175        |
| I - 1 - Analyse d'une polymérase sans la sous-unité Rpa49 ou sans le C-ter de Rpa49. ....   | 175        |
| I - 2 - Les positions possibles du domaine C-terminal de Rpa49. ....  | 178        |
| II - Un lien avec le facteur d'initiation Rrn3 ? .....  | 183        |
| III - La région à l'interface entre Rpa190, Rpa135 et Rpa12 aurait un rôle dans l'initiation de la transcription .....              | 186        |
| IV – Rpa12, un élément mobile de l'ARN pol I .....  | 190        |
| <br>  |            |
| <b>REFERENCES.....</b>  | <b>192</b> |

## Abréviations

|                  |   |
|------------------|---|
| 35S              | Précurseur des ARNribosomiques 25S, 18S et 5,8S |
| a.a.             | acide aminé                                     |
| ADN              | Acide DésoxyriboNucléique                       |
| ADNr             | ADN ribosomique                                 |
| AMPc             | Adenosine Monophosphate cyclique                |
| ARN              | Acide RiboNucléique                             |
| ARNm             | ARN messenger                                   |
| ARNr             | ARN ribosomal                                   |
| ARNt             | ARN de transfert                                |
| ARS              | "Autonomous Replicating Sequence"               |
| ATP              | Adénosine TriPhosphate                          |
| BH               | "Bridge Helix"                                  |
| BTFs             | "Built-in Transcription factors"                |
| CC               | "Closed Complex"                                |
| CE               | "Controle Element"                              |
| CF               | "Core Factor"                                   |
| Cs               | "Cryo-sensitive", Cryosensible                  |
| CTD              | "Carboxy-Terminal Domain"                       |
| EC               | "Elongating Complex"                            |
| <i>E.coli</i>    | <i>Escherichia coli</i>                         |
| ETS              | "External Transcribed Sequence"                 |
| eWH              | "extended Winged Helix"                         |
| GIM              | "Genetic Interaction Mapping"                   |
| HMG              | High Mobility Group"                            |
| IGS              | "Intergenic Spacer"                             |
| ITC              | "Initially Transcribing Complex"                |
| ITS              | "Internal Transcribed Sequence"                 |
| kb               | kilobase  |
| kDa              | kiloDalton                                      |
| MET              | Microscopie Electronique à Transmission         |
| Nsi1             | "NTS1 silencing protein 1"                      |
| nt               | nucléotide                                      |
| NTP              | Nucleoside Triphosphate                         |
| OC               | "Open Complex"                                  |
| ORF              | "Open Reading Frame"                            |
| PAD              | "Propeller/Promoter-Associated Domain"          |
| PIC              | "Pre-Initiation Complex"                        |
| PCR              | "Polymerase Chain Reaction"                     |
| Pol I - II - III | Polymerase I - II - III                         |
| pb               | paire(s) de base                                |
| RBS              | "Ribosome Binding Sequence"                     |
| RdDM             | "RNA-directed DNA Methylation"                  |

|                     |  |
|---------------------|--|
| RFB                 | "Replication Fork Barrier"                     |
| RNPs                | "RiboNucleoprotein Particles"                  |
| RX                  | Rayons X                                       |
| S                   | Svedberg (unité)                               |
| <i>S.cerevisiae</i> | <i>Saccharomyces cerevisiae</i>                |
| snoARN              | "small nucleolar ARN"                          |
| SPB                 | "Spindle Pole Body"                            |
| SL                  | "Synthetic Lethal"                             |
| SL1                 | "Selectivity Factor 1"                         |
| s.u.                | sous-unité                                     |
| TAF                 | "TBP associated Factors"                       |
| TBP                 | "TATA Binding Protein"                         |
| TFIIB               | "Transcription Factor of RNA polymerase II, B" |
| TFIID               | "Transcription Factor of RNA polymerase II, D" |
| TFIIE               | "Transcription Factor of RNA polymerase II, E" |
| TFIIF               | "Transcription Factor of RNA polymerase II, F" |
| TFIIH               | "Transcription Factor of RNA polymerase II, H" |
| TFIIS               | "Transcription Factor of RNA polymerase II, S" |
| TOR                 | "Target Of Rapamycine"                         |
| Ts                  | "Thermosensitive", Thermosensible              |
| TSS                 | "Transcription Start Site"                     |
| TTF1                | "Transcription release Factor for Pol I"       |
| tWH                 | "tandem Winged Helix"                          |
| UAF                 | "Upstream Activating Sequence"                 |
| UBF                 | "Upstream Binding Factor"                      |
| UCE                 | "Upstream Control Element"                     |
| WH                  | "Winged Helix"                                 |
| WT                  | "Wild-Type" = sauvage, de référence            |
| Zn                  | Zinc   |

# Préambule

Lors de la rédaction de ce manuscrit, j'ai souhaité aborder les différents aspects, théoriques comme techniques, qui m'ont accompagnés tout au long de ces trois années ; ceci afin de proposer un aperçu à la fois général du contexte d'étude, mais aussi un point de vue plus détaillé sur les éléments précis qui ont guidés mes travaux. Dans cette optique, je présenterai ce manuscrit selon trois grands axes. Premièrement, l'introduction, qui sera divisée en sept chapitres, selon une organisation précise et décroissante, comparable à un entonnoir : du plus global au plus détaillé, de l'organisme à la molécule le composant. L'idée sera en premier lieu de présenter brièvement l'organisme modèle que j'ai utilisé durant toute ma thèse, la levure à bourgeon *Saccharomyces cerevisiae*. A partir de cette description physiologique et cellulaire, je me concentrerai ensuite sur l'un des aspects fondamentaux de la croissance cellulaire, l'étude des ribosomes, et de leur biogénèse. Cette dernière, débute notamment par la synthèse du précurseur des grands ARN ribosomiques (ARNr) qui composeront ensuite les ribosomes matures. Cette activité est effectuée par une enzyme bien particulière : l'ARN polymérase I (ARN pol I), dont l'étude représente le cœur de mon sujet de thèse. La structure en trois dimensions à haute résolution de cette protéine multimérique a été obtenue dès 2013, en faisant un outil de choix pour l'étude mécanistique de cette enzyme et de son activité. D'autres structures plus récentes représentant des états transcriptionnels transitoires de cette enzyme, ainsi que les structures d'autres ARN polymérases multimériques similaires (ARN pol bactérienne, ARN pol II,...), seront également largement abordées, et nous permettront de proposer un aperçu détaillé des bases moléculaires connues régissant la transcription par l'ARN pol I. Ces aspects structuraux et *in vitro*, même si bien caractérisés, ne permettent cependant pas d'expliquer l'intégralité du mécanisme exact s'effectuant *in vivo*.

Ainsi dans un second temps, à partir de cette présentation initiale et notamment des structures 3D de l'ARN pol I, j'introduirais la partie centrale de mes travaux, qui consiste en l'étude génétique de certaines sous-unités spécifiques de l'ARN pol I. L'outil génétique nous permet de cibler et d'étudier les fonctions de ces sous-unités, de suggérer des interactions possibles avec certains facteurs, ou même entre les sous-unités de l'ARN pol I elle-même. Grâce à ce ciblage, il est possible de mettre en valeur des zones régulées/régulables de l'enzyme, d'apporter un point de vue *in vivo* à la mécanistique des études structurales effectuées, de proposer/vérifier des hypothèses *in vivo* et *in silico*, afin d'imaginer des études cellulaires, fonctionnelles et structurales complémentaires. Toutes ces données permettent de mieux comprendre le mécanisme de la transcription par l'ARN pol I, allant de l'échelle moléculaire à l'échelle de l'organisme. J'aborderai ainsi ces aspects dans un troisième axe, incluant la fin des résultats obtenus et la partie "Discussions" de ce manuscrit, tous deux liés.

# INTRODUCTION

## Chapitre I - Caractéristiques des eucaryotes uni- et multicellulaires, et de la levure *Saccharomyces cerevisiae*

### I-1- Introduction aux cellules eucaryotes et à la levure *Saccharomyces cerevisiae*

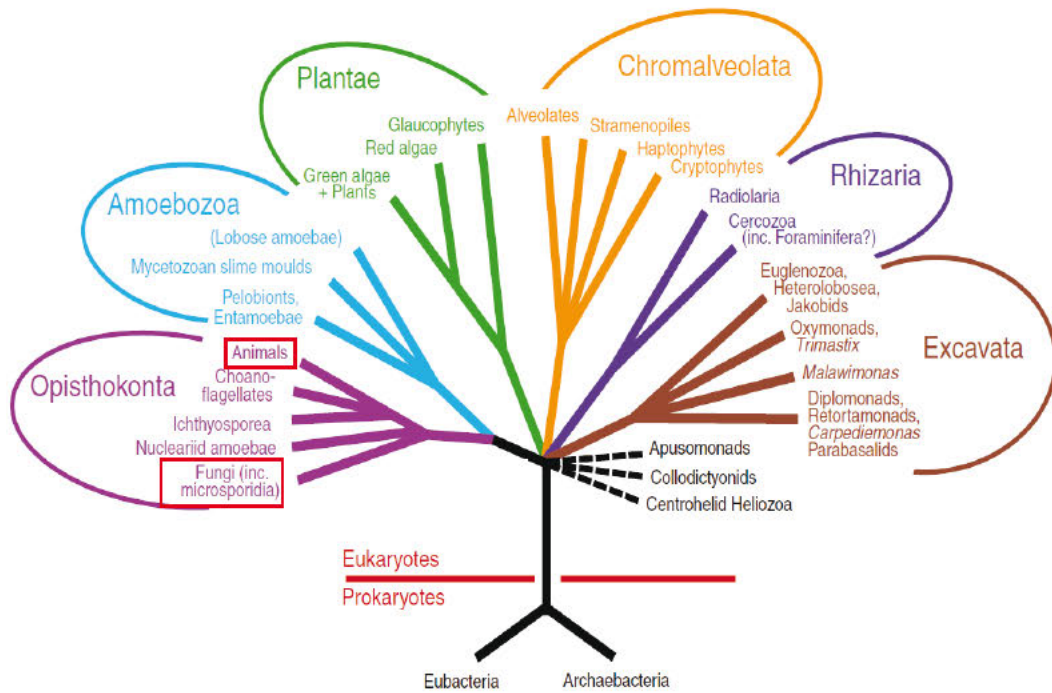
La dichotomie entre cellules eucaryotes et procaryotes, à savoir entre cellules avec et sans noyau, est un concept introduit dès les années 1920 par le microbiologiste Edouard Chatton, notamment grâce aux énormes progrès en microscopie. Un troisième domaine du vivant fut décrit plus tardivement, représentant le groupe des Archées (Woese & Fox, 1977). Nommées anciennement archéobactéries, et longtemps considérées comme des bactéries extrêmophiles, ce sont des organismes unicellulaires sans noyau ni organites, (Figure 1.A). Les progrès en séquençage génomique, notamment sur les séquences des gènes des ARNs ribosomiques, ont permis dès le milieu des années 1970 de dédoubler l'arbre procaryote en deux domaines distincts, celui des archées et celui des bactéries (Hug *et al*, 2016). Il a d'ailleurs été proposé que les eucaryotes découleraient d'un scénario où une cellule hôte archéenne et un endosymbiote alphaproteobactérien (mitochondrial) auraient fusionnés résultant en la première cellule eucaryote (Zaremba-Niedzwiedzka *et al*, 2017)).

Mais outre la distinction majeure qu'est la présence ou non d'un noyau, le fait que les cellules eucaryotes possèdent des organites cytoplasmiques les différencient des bactéries/archées qui n'en possèdent pas. On retrouve ainsi des organites qui possèdent leur propre génome, comme les mitochondries (ou les chloroplastes chez les plantes). Un système vésiculaire de sécrétion (réticulum endoplasmique et appareil de Golgi, endosomes et lysosome), des filaments intermédiaires, un cytosquelette d'actine et des microtubules, et enfin une organisation linéaire

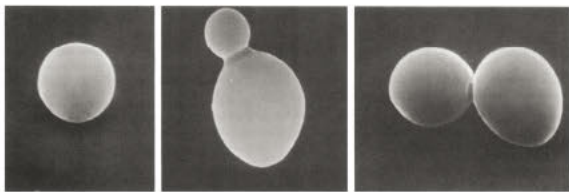
de plusieurs chromosomes basée sur leur transmission méiotique et mitotique, terminent de caractériser la structure fortement organisée des cellules eucaryotes.

Dans ce travail, nous utiliserons comme outil l'organisme modèle eucaryote unicellulaire le plus étudié et caractérisé en laboratoire: la levure à bourgeon, *Saccharomyces cerevisiae* (*S.cerevisiae*), ou plus communément appelée levure de boulanger (Herskowitz, 1988) (Figure 1.B ; Figure 4.A et C.). Premièrement, il faut noter que les levures appartiennent au groupe des champignons, dénués d'activité de photosynthèse comme les végétaux chlorophylliens, et tirent leur énergie de la combustion des hydrates de carbone par respiration ou fermentation. Les levures sont des champignons particuliers, caractérisés par leur développement unicellulaire. La levure *Saccharomyces cerevisiae*, fait partie des ascomycètes, en référence à leur méiose qui a lieu dans un asque. Là, les spores sont assemblées en tétrades, et ne peuvent pas se séparer, migrer vers l'extérieur ou se détacher individuellement. Une distinction particulière parmi les ascomycètes peut être faite, classant *S.c.* dans le groupe des hémiascomycètes (reproduction par bourgeonnement) (Figure 1.C.), en opposition par exemple aux levures à fission comme *Saccharomyces pombe* (archiascomycètes) ou aux champignons filamenteux (euascomycètes) (voir pour référence "Les organismes modèles, La levure", de Pierre Thuriaux, 2004).

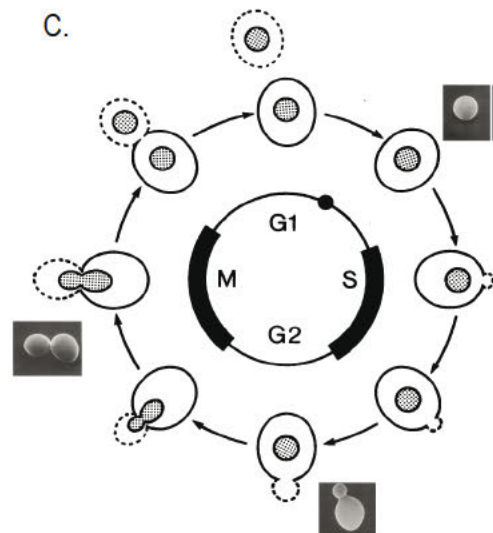
A.



B.



C.



**Figure.1 - La levure à bourgeons, *Saccharomyces cerevisiae* (*S.cerevisiae*).**

A. Arbre phylogénétique sous forme de diagramme représentant l'organisation de la majorité des eucaryotes en 6 groupes distincts. C'est dans celui des Opisthokontes que nous retrouverons les animaux, mais aussi la levure, avec les champignons et de nombreux autres groupes eucaryotes unicellulaires. Par clarté les animaux et les fungi ont de plus été encadrés en rouge. Ce schéma est adapté de la revue de Simpson and Roger, 2004.

B. Ici sont présentés des clichés (Mag x6000) représentant différentes morphologies de levures *S.c.*, donc à différentes étapes du cycle cellulaire: en haut à gauche sans bourgeon; au milieu et en haut à droite en cours de division, avec bourgeons de différentes tailles. Les clichés du bas correspondent à des zygotes, issus de la reproduction entre deux cellules de sexes différents (Mata et Mata). D'après Herskovitz et al, 1988.

C. Vue d'ensemble des différentes phases du cycle mitotique de *S.c.*, représentant une cellule mère (en traits pleins) bourgeonnant pour donner naissance à une cellule fille (en traits pointillés). Le noyau au sein de chaque cellule est représenté par une surface avec des points, délimitée par une ligne continue. Les différentes phases du cycle sont indiquées: S, phase S, synthèse d'ADN ; M, Mitose, division cellulaire; G1/2, phases de croissance cellulaire. Par ailleurs, les trois premiers clichés de la Fig1.B. ont été remplacés par rapport à leur place dans ce cycle.

*S.cerevisiae* est le premier organisme eucaryote dont le génome ait été entièrement séquencé (Goffeau *et al*, 1996). Les travaux génétiques et cellulaires sur cette levure en ont fait le modèle unicellulaire de référence pour les eucaryotes. En effet, leur croissance rapide (environ 1h30 de temps de doublement pour une levure sauvage, dite "Wild-Type", ou WT) sur des milieux définis, l'aisance avec laquelle on peut obtenir et étudier génétiquement des mutants, mais surtout leur fort rapprochement évolutif avec le règne animal (Opisthokonta, incluant levures et métazoaires; Figure 1A), en font des candidats modèles de choix pour l'étude des cellules humaines. Tous les travaux génétiques, moléculaires et cellulaires présentés ultérieurement au cours de cette thèse ont été réalisés chez la levure *S.cerevisiae*. Une présentation plus poussée du fonctionnement précis de cette levure (évolution des génomes de levure, cycles de reproduction et cellulaire, régulation métabolique,...) sortirait du cadre de ce manuscrit. Je me contenterai d'introduire plus en détails l'organite caractéristique des eucaryotes, le noyau, ainsi qu'un domaine particulier de ce dernier, le nucléole, lieu de formation précoce des ribosomes.

### **I -2- Le noyau chez les eucaryotes multicellulaires et *S.cerevisiae*.**

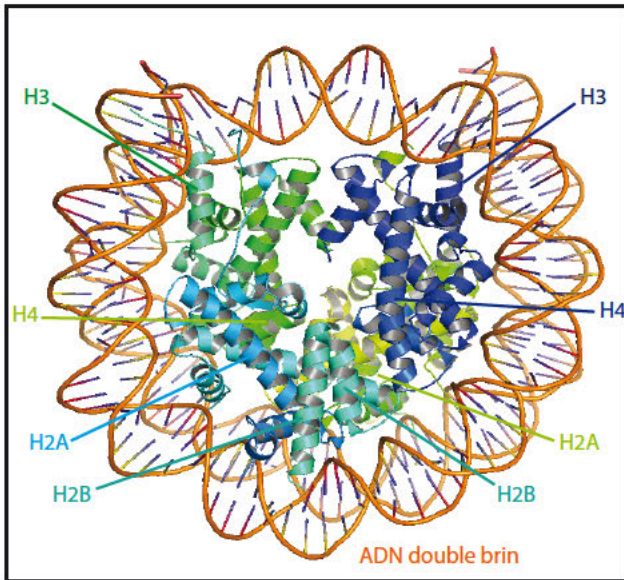
Le noyau est un organite spécifique des cellules eucaryotes, ayant une taille variable selon les types cellulaires et les organismes (rayon d'environ 5  $\mu\text{m}$  pour la majorité des métazoaires, contre environ 1 $\mu\text{m}$  pour les levures). Il est le lieu de stockage, d'expression et de réplication des chromosomes constitués de molécules d'ADN et supports de l'information génétique. Des protéines s'associent alors à l'ADN pour former la chromatine. Celle-ci se distribue et s'organise de façon à répondre aux contraintes physico-chimiques et environnementales locales du noyau, à savoir un espace restreint dans lequel on trouve un très grand nombre de molécules différentes telles que des ARNs, protéines, etc. On parle alors de la compaction de la chromatine, qui s'effectue sur plusieurs niveaux (Figure 2.C.), organisation variable selon le type cellulaire, l'étape du cycle cellulaire ou les conditions physiologiques de croissance.



Ainsi, le premier niveau de compaction correspond à l'enroulement de l'ADN, chargé négativement via ses phosphates, autour de l'octamère d'histones (deux exemplaires de chaque histone H2A, H2B, H3 et H4, toutes chargées positivement, permettant le repliement de la fibre d'ADN). La structure ainsi décrite a été initialement observée dès les années 1970 en microscopie électronique (Olins & Olins, 1974), et nommée nucléosome (Oudet *et al*, 1975). La structure du nucléosome a de plus été obtenue à haute résolution par cristallographie aux Rayons X (RX), notamment chez *S. cerevisiae* (Davey *et al*, 2002; White *et al*, 2001) (Figure 2.A.). Un nucléosome mesure 13 nm de diamètre, pour 11 nm de haut, et contient un segment de 146 paires de bases (pb) (Davey *et al*, 2002). Deux nucléosomes successifs sont séparés par des séquences de liaison, ou "linker", d'ADN "libre". Ce segment d'ADN peut toutefois être en liaison avec l'histone linker H1, ou d'autres protéines liées à l'ADN comme les protéines HMG. La dénomination de l'ADN en "collier de perles" vient de la compaction d'une fibre d'ADN par ces structures particulières, formant la fibre de 10nm, qui sera elle-même compactée à un niveau supérieur pour former la fibre chromatinienne de 30nm (Figure 2.B.). Les niveaux d'organisation largement décrits dans la littérature (30 nm, 120 nm) ne sont observés qu'*in vitro*, dans des conditions salines et avec des séquences répétées permettant un fort positionnement de nucléosomes. Le niveau d'organisation présent dans le noyau apparaît comme une structure de 10 à 24 nm, irrégulière, dont la nature biophysique reste largement inconnue (Ou *et al*, 2017).

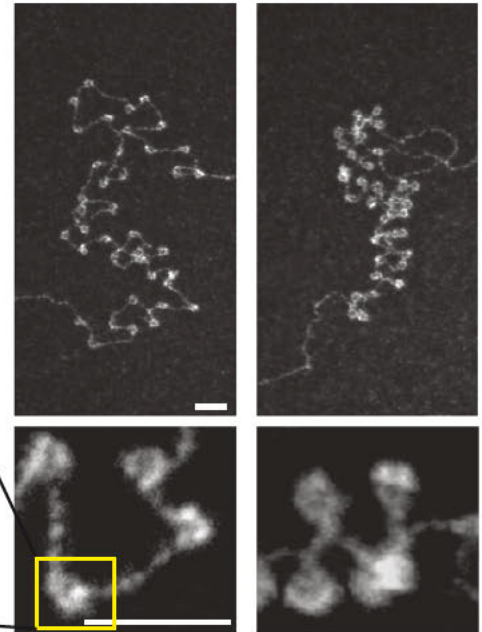
Le génome bactérien n'est pas organisé par des nucléosomes, mais par des NAP (Nucléoïdes Associated Proteins), comme HU, qui contactent et organisent l'ADN (Wang *et al*, 2011). Certaines Archées, par contre, possèdent également des protéines de type histones ayant des similarités avec les histones des eucaryotes, mais qui ne s'organisent pas en octamères. On les retrouve en deux dimères de H3 et H4, formant des tétramères, couvrant moins de paires de bases, mais présentant une plus grande flexibilité.

A.



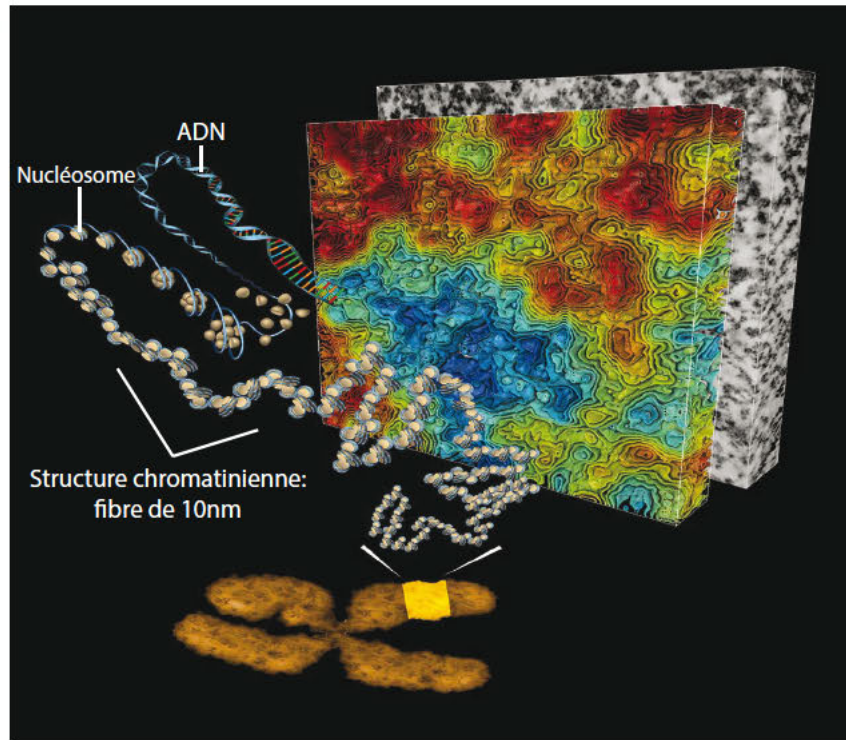
White et al, 2001 ; pdb: 1ID3

B.



Recouvreux et al, 2005

C.



Hornig et al, 2017

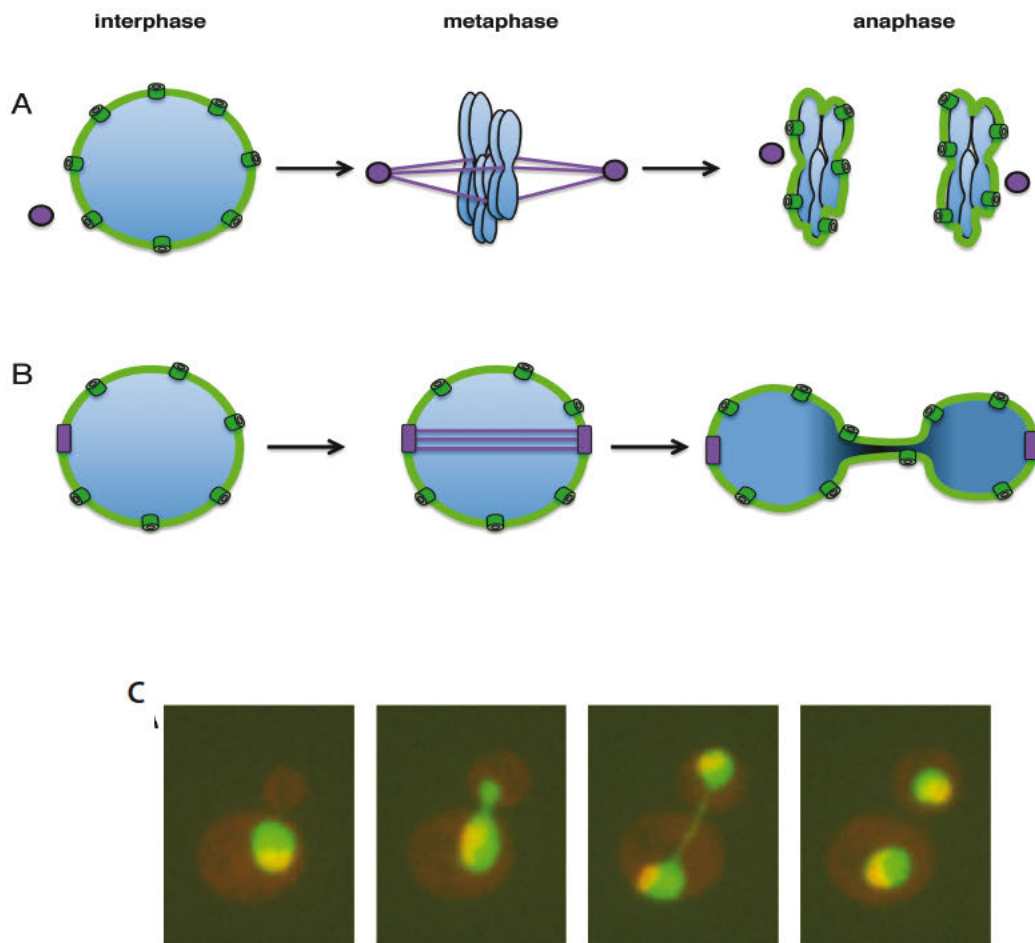
## Figure 2. Organisation de la chromatine dans le noyau.

A. Structure à haute résolution d'un nucléosome chez *S.cerevisiae*, composé de l'ADN double brin s'enroulant autour d'un octamère d'histones (2 exemplaires de H2A, H2B, H3 et H4).

B. Images en microscopie électronique de la fibre de chromatine, représentant l'enchaînement des nucléosomes, permettant le repliement de la fibre. On remarque que le niveau de repliement n'est pas le même dans les deux images. En effet, l'absence (cliché de gauche) ou la présence (cliché de droite) de «l'histone linker», ou histone H1, permet un niveau de repliement différent. Les clichés inférieurs sont des grossissements correspondants aux images du haut.

C. L'ADN de 2 nm de diamètre s'enroule autour d'une structure protéique particulière, formant le «nucléosome». L'enchaînement des nucléosomes permet l'organisation de la chromatine en différents niveaux de compaction, afin de rentrer dans l'espace restreint qu'est le noyau. La technique de ChromEMT (Tomographie en Microscopie Electronique couplée à un marquage fluorescent) permet de visualiser l'organisation et l'ultrastructure 3D de la chromatine (Hornig et al, 2017)

L'une des différences fondamentales entre les cellules des métazoaires et les hémiascomycètes tels que *S.c.* réside dans le fait que ces derniers effectuent une "mitose fermée", c'est à dire sans rupture de l'enveloppe nucléaire, contrairement aux cellules animales ou végétales où l'intégrité de l'enveloppe nucléaire est temporairement perdue lors de la mitose (Arnone *et al*, 2013) (Figure 3 A-C). Il existe cependant une forte conservation évolutive des protéines nucléaires entre les levures et les cellules des eucaryotes multicellulaires, ce qui permet une forte extrapolation lors de leurs études génétiques et biochimiques. Le noyau possède certaines caractéristiques conservées que nous énumérerons ici.



(Arnone et al, 2013 )

**Figure 3. La mitose fermée, l'une des caractéristiques des hémiascomycètes.**

A. Représentation schématique d'une mitose de type ouverte. La chromatine en interphase (en bleu) est contenue dans le noyau, délimitée par l'enveloppe nucléaire (en vert). Celle-ci se destructure alors en mitose pour permettre aux microtubules (lignes violettes) rattachés aux centrosomes (points violets) d'aligner les chromosomes sur un plan, appelé plan équatorial de la cellule, pour leur séparation ultérieure en anaphase. En fin d'anaphase, après ségrégation des chromosomes, l'enveloppe nucléaire de chaque cellule se reforme.

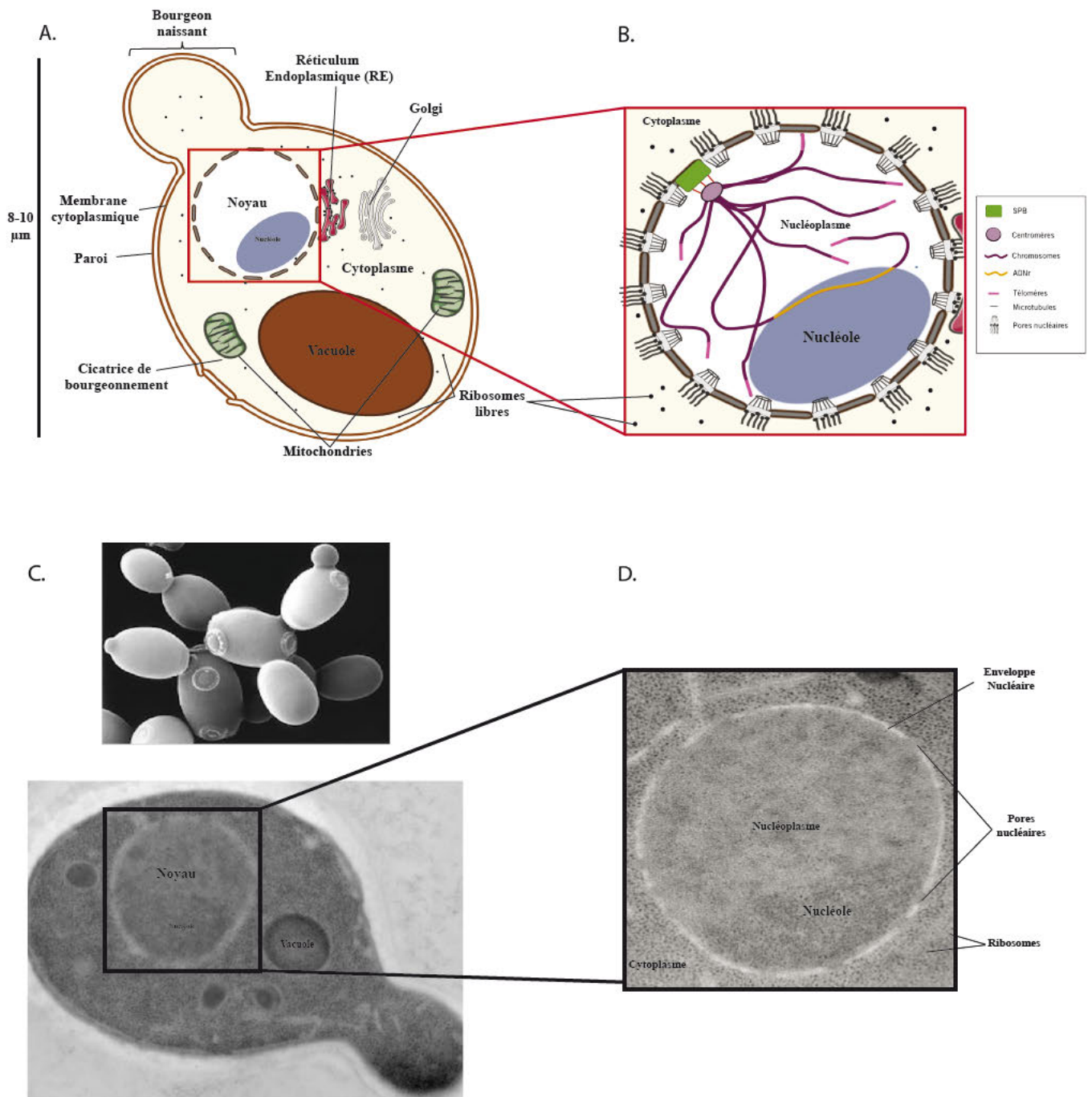
B. En comparaison, on présente ici le cas d'une mitose fermée, comme c'est le cas pour la levure *S.cerevisiae*. Là, le SPB (Spindle Pole Body, en violet), est rattaché à l'enveloppe nucléaire tout au long du cycle. Il se duplique au début de phase S. Durant la phase G2, et début de mitose, le noyau s'allonge vers le bourgeon naissant. Durant l'anaphase, les deux SPB sont diamétralement opposés, dans un noyau en forme de sablier, avec une constriction au niveau du site de bourgeonnement. L'enveloppe nucléaire s'étend, sans rupture, tandis que les chromatides soeurs migrent chacune en sens opposés dans chaque cellule (cellule mère et bourgeon), qui contiennent alors chacune un SPB.

C. Visualisation en microscopie optique à fluorescence d'une levure à bourgeon *S.c.*, au cours du cycle cellulaire. Les clichés successifs sont pris toutes les 20 minutes. Les cellules expriment un marqueur nucléoplasmique étiqueté GFP (Pus1, en vert) et un marqueur du nucléole (Nsr1, en rouge). On discerne effectivement la migration du matériel nucléaire de la cellule mère vers le bourgeon et la séparation des deux cellules mère et fille après l'anaphase (Arnone et al, 2013 ).

## I-2-a- Chez *S.cerevisiae*, Le pore nucléaire et le centre organisateur des microtubules/SPB sont ancrés dans l'enveloppe nucléaire

Tout d'abord, étant un organite spécifique des cellules eucaryotes, le noyau est donc compartimenté par une enveloppe nucléaire. Elle est constituée d'une double bicouche lipidique l'isolant du cytoplasme, les deux membranes étant séparées par l'espace périnucléaire. La membrane externe, recouverte de ribosomes, est en directe continuité avec le réticulum endoplasmique rugueux. La face interne de l'enveloppe nucléaire (INM, pour "Inner Nuclear Membrane") en revanche est en contact avec la chromatine. Chez les métazoaires, un organisateur majeur de la chromatine périphérique sont les lamines nucléaires (Taddei & Gasser, 2012). Les lamines ne sont pas retrouvées chez les champignons unicellulaires, ni chez les plantes. Chez *S.cerevisiae*, des interactions entre la face interne de l'enveloppe et la chromatine sont toujours présentes. Des interactions chromatine – INM sont assurées par les pores nucléaires et par des protéines telles qu' Esc1, Scr1/Heh1.. (Albert *et al*, 2012). Les pores nucléaires sont les complexes protéiques au travers desquels s'effectue le trafic nucléocytoplasmique (protéines et ARN). Ils sont composés d'une trentaine de protéines appelées Nucléoporines (Aitchison & Rout, 2012). Ces complexes sont bien conservés, avec une masse moléculaire tout de même significativement différente entre vertébrés (125 MDa) et levures (60MDa). D'un diamètre d'environ 120 nm, ils permettent le passage par diffusion de petites molécules de moins de 40 kDa. Ceci est complété par d'autres systèmes actifs de passage pour les plus grosses molécules (tels que les précurseurs ribosomiques) via des "karyophérines", ainsi qu'une diffusion passive pour les plus petites (Beck *et al*, 2004; Hoelz *et al*, 2016). Le pore nucléaire peut être considéré comme un composant organisateur du noyau. Sa structure est asymétrique, avec des protéines spécifiques de la face cytoplasmique (Nup159), ou de la face nucléoplasmique, (comme Nup2). Les protéines de la face interne Nup1, Nup2, Mlp1, Mlp2.. servent d'ancrage à la chromatine (Dilworth *et al*, 2001; Galy *et al*, 2004). De même,

et en lien avec l'organisation chromatinienne, on peut citer un autre élément massif ( $>1$  GDa) et spécialisé du noyau, le centre organisateur des microtubules (inséré dans l'enveloppe nucléaire et appelé "Spindle Pole Body" (SPB) chez la levure, ou MTOC ("Microtubule Organizing Center"). Cette structure permet de maintenir les microtubules intra-nucléaires, et contribue à l'orientation des chromosomes pendant les différentes phases du cycle cellulaire. Du côté nucléoplasmique, le SPB permet l'ancrage des microtubules kinétochoriens, rattachés aux centromères des chromosomes. Le kinétochore est ancré à la région centromérique de chaque chromosome, définie par une séquence de seulement 125 paires de base (pb) (Figure 4.B.). Le SPB est indispensable à la ségrégation des chromosomes (Jin *et al*, 1998; Bystricky *et al*, 2004), puis montré par 3C (Duan *et al*, 2010). Il organise également le cytosquelette dans sa partie cytoplasmique .



**Figure.4 - Architecture d'une cellule de levure et de son noyau**

A. Vue schématique d'une coupe transversale de cellule de *Saccharomyces cerevisiae* (S.c.). On repère de nombreux éléments différents: la levure est délimitée par une paroi ainsi qu'une membrane plasmique interne. On trouve dans le cytoplasme les mitochondries, vacuole, appareil de Golgi et réticulum endoplasmique, ainsi que de nombreux ribosomes libres. Un autre compartiment de la levure est le noyau, lui-même délimité par une enveloppe nucléaire, qui contient l'ensemble des chromosomes de levure, et possède une région non-délimitée, le nucléole. On distingue également une cicatrice de bourgeonnement, témoin d'une division par bourgeonnement antérieure.

B. Zoom schématique sur une région particulière de la levure S.c., le noyau, et son organisation générale. Le centre organisateur des microtubules, le SPB, est situé à l'opposé du nucléole. Plusieurs chromosomes sont ici représentés, tous attaché via leurs centromères, et reliés au SPB par le réseau microtubulaire. L'ADNr, situé sur le chromosome XII, est représenté.

C. Image en microscopie électronique à balayage de levures à bourgeons (panel du dessus) et coupe de cellule de levure *S.cerevisiae* observée au microscope électronique à transmission (panel du dessous). On y distingue notamment la paroi et la membrane cytoplasmique délimitant la levure du reste du champ, des vacuoles, et le noyau, possédant une enveloppe nucléaire et une région plus dense, le nucléole.

D. Coupe de noyau de levure S.c. observée en MET après cryofixation et cryosubstitution. On observe notamment une grande quantité de petits points denses dans le cytoplasme correspondant aux ribosomes. Le noyau est délimité par l'enveloppe nucléaire, plus claire, la séparant du cytoplasme. Le nucléole est le compartiment nucléaire non-délimité par une membrane et plus dense aux électrons (photographies: Isabelle Léger-Silvestre)

## I -2-b- Introduction aux ARN polymérase

Le noyau est également le lieu de stockage et d'activité de trois enzymes particulières, qui permettent l'activité dite de "transcription", où un ARN est synthétisé à partir d'une matrice d'ADN. Les détails moléculaires et structuraux de l'activité de transcription seront largement abordés et présentés ultérieurement dans ce manuscrit. Il existe trois ARN polymérase nucléaires responsables de la transcription des gènes dans les cellules eucaryotes: les ARN Pol I, Pol II et Pol III.

Plus précisément, la Pol I est responsable de la production des grands ARN ribosomiques (ARNr) qui composent le complexe macromoléculaire qu'est le ribosome (les ARNr 25S, 5.8S et 18S chez *S.cerevisiae*). La Pol II permet la transcription de petits ARNs nucléaires et nucléolaires (small nuclear / nucleolar RNA, sn/snoRNA), du grand nombre d'ARN non-codants (courts <200 nt, ou longs >200 nt) (Tisseur *et al*, 2011) et de l'ensemble des ARN messagers (ARNm) qui seront ensuite traduits par le ribosome pour produire les protéines cellulaires. Enfin, la Pol III synthétise une variété d'ARNs non-codants (ncRNA), souvent courts et abondants, comprenant notamment les ARN de transfert (ARNt) et l'ARN ribosomique 5S.

La transcription par l'ARN pol I a lieu dans une région spécifique du noyau, le nucléole.

## I -2-c- Le nucléole, "l'usine à transcription" de la cellule

On trouve au sein du noyau un territoire facilement identifiable, le nucléole, qui nous intéressera plus particulièrement. Sa visualisation est aisée en microscopie électronique de par sa composition très élevée en particules ribonucléoprotéiques, possédant une forte densité aux électrons lorsqu'un contrastant est utilisé, tel que l'acétate d'uranyle (Figure 4 C.D.). Il est en effet le siège de des étapes précoces de la biogenèse des ribosomes. Historiquement, il a d'abord été défini comme l' "usine à ribosomes". De nombreuses autres fonctions lui sont



maintenant attribuées, telles que la biogenèse de la plupart des particules ribonucléoprotéiques cellulaires (RNPs) ou la rétention de certaines particules immatures.

Chez la levure *S.cerevisiae*, en croissance exponentielle, le nucléole possède une forme caractéristique de croissant accolé à l'enveloppe nucléaire, occupant 1/3 du noyau (Figure 4). Il est rapidement remanié lors de stress ou de carence nutritionnelle (Tsang *et al*, 2003).

C'est donc dans ce compartiment qu'a lieu la transcription de l'ADN ribosomique (ADNr) par l'ARN pol I, la maturation et les modifications des pré-ARN ribosomiques (ARNr), et les étapes initiatrices de l'assemblage des ribosomes (Thiry & Lafontaine, 2005). Aucune membrane ne sépare ce territoire du reste du noyau, et n'est donc pas considéré comme un organite. Cependant, il a été proposé que la compartimentalisation observée soit maintenue par des mécanismes de "Liquid-like phase separation" (Feric *et al*, 2016). On trouve de un à cinq nucléoles dans les cellules de mammifères, alors qu'un seul est présent chez la levure *S.cerevisiae*. Ceci est dû au fait que le locus d'ADNr n'est porté que par un seul chromosome chez *S.cerevisiae* (le chromosome XII), alors qu'il est réparti sur plusieurs chez l'homme par exemple. Ces régions chromosomiques particulières organisent le nucléole, et sont appelées les "Nucleolar Organizer Region" (NOR) (Bouteille & Hernandez-Verdun, 1979). Plusieurs sous-domaines nucléolaires ont pu être identifiés au cours du temps sur la base de leur morphologie, même si certains sont encore largement sujets à discussion, en particulier selon les organismes étudiés. Ils sont au nombre de trois: le Centre Fibrillaire (FC), le Composant Fibrillaire Dense (DFC) et le Composant Granulaire (GC). Chez les métazoaires, le FC est la partie que l'on peut considérer comme la moins contrastée du fait de sa composition, en majorité des ARN polymérase I et ses facteurs associés, ainsi que des transcrits naissants en moindre quantité (Schwarzacher & Wachtler, 1991). Chez *S.cerevisiae*, ce composant est le plus sujet à discussion. On peut citer notamment le nombre de FC variable, qui évolue selon la phase du cycle cellulaire, (plus important dans des cellules en phase de croissance rapide).

De plus, la présence même de FC chez tous les eucaryotes demeure encore débattue. En particulier, dans le cas de *S.c.* par exemple, des images en microscopie électronique révéleraient des structures similaires à des FC (Léger-Silvestre *et al*, 1999), même une plus récente décrirait qu'ils n'existeraient pas en tant que tels chez cet organisme (Thiry & Lafontaine, 2005). Le DFC, plus dense, entoure les FC et serait le lieu où les transcrits d'ARNr s'accumuleraient. Celui-ci serait enfin en contact direct avec le dernier composant, le GC (Cmarko *et al*, 2000). Le GC contient les pre-ARN sous forme de pré-ribosomes tardifs en cours de maturation (clivages endo/exo-nucléotidiques, pseudouridylations), qui seront par la suite libérés du nucléole, pour être à termes exportés via les pores nucléaires pour une maturation cytoplasmique finale.

Le nucléole est sujet à une forte dynamique en termes de tailles et de morphologies, dépendant des conditions de croissance, et reflétant l'activité transcriptionnelle intense qui s'y déroule (Schwarzacher & Wachtler, 1993). Ainsi, une souche de levure mutée pour l'ARN polymérase I (ARN pol I; allèle thermosensible à 37°C de la plus grande sous-unité, ou "s.u." , Rpa190) présente une forte restructuration du nucléole: il se décolle de la membrane et devient très compact (Trumtel *et al*, 2000). A l'opposé, lors d'un défaut transcriptionnel causé par l'absence d'une sous-unité spécifique de ARN pol I, Rpa49, largement décrite dans le chapitre VI, le nucléole est plus étendu et se désagrège dans le nucléoplasme (Albert *et al*, 2011). Il est également à noter qu'un flux important et constant de protéines s'effectue avec le nucléoplasme.

Même si l'activité majeure du nucléole est représentée par la biogénèse précoce des ribosomes, il faut donc souligner le fait qu'il peut avoir d'autres fonctions, telles que la biogénèse d'autres ribo-nucléoparticules (SRP, tRNA, ect..), la réparation des dommages à l'ADN, le contrôle du cycle cellulaire ou la réplication, ce qui a notamment été mis en

évidence par des études de spectrométrie de masse des protéines qui s'y concentrent (Andersen *et al*, 2005).

## **Chapitre II - Une étape limitante dans la prolifération cellulaire : la production des ARNr, composants essentiels des ribosomes**

### **II -1- la biogénèse des ribosomes, machineries moléculaires responsables de la synthèse des protéines cellulaires**

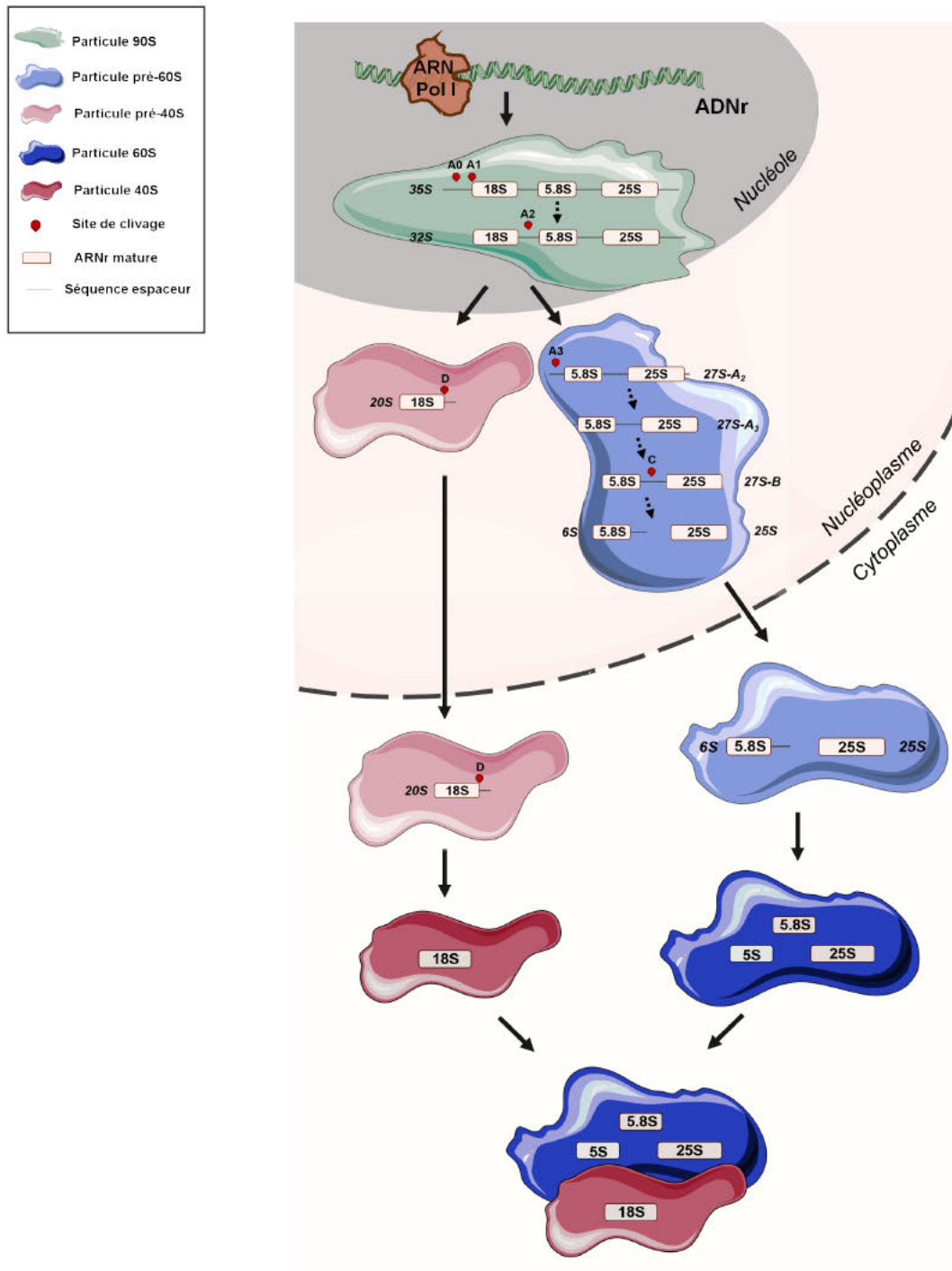
M'intéressant principalement au phénomène de transcription par la Pol I, ce chapitre présente les principales étapes de la biogénèse et la fonction des ribosomes. Cette activité cellulaire massive nécessite en effet un très grand nombre d'étapes co- et post-transcriptionnelles.

Les ribosomes sont les complexes moléculaires responsables de la production des protéines cellulaires. Leur biogénèse est une activité majeure et très coûteuse en énergie pour la cellule. Elle est finement régulée afin de s'adapter aux conditions de croissance et se coordonner aux autres processus cellulaires fondamentaux. La perte du contrôle de la synthèse des ribosomes est par ailleurs l'un des aspects essentiels du développement tumoral puisqu'une production accrue de ribosomes est nécessaire à la croissance et à la prolifération rapide et anarchique de nombreux types de cellules malignes (Drygin *et al*, 2010).

Les ribosomes sont des complexes ribonucléoprotéiques massifs, allant de 2.5 MDa chez les bactéries jusqu'à 4.2 MDa chez les eucaryotes. Le composant majeur des ribosomes sont les ARN ribosomiques, dont la synthèse s'effectue majoritairement via l'activité de transcription par l'ARN polymérase I dans le nucléole. Cette activité transcriptionnelle est réalisée au niveau d'un locus chromosomique précis, l'ADN ribosomique ou ADN<sub>r</sub>, que nous présenterons par la suite (Chapitre II-2). Les ARNs constituent plus de 50% de la masse d'un ribosome. De plus, son centre catalytique dépend de l'ARN, faisant du ribosome un ribozyme. Deux sous-unités composent un ribosome actif en traduction: une petite (40S), responsable du décodage de l'information génétique sur l'ARN<sub>m</sub>, et une grande (60S), portant l'activité peptidyl-transférase du ribosome (S = Svedberg: unité de coefficient de sédimentation), formant le

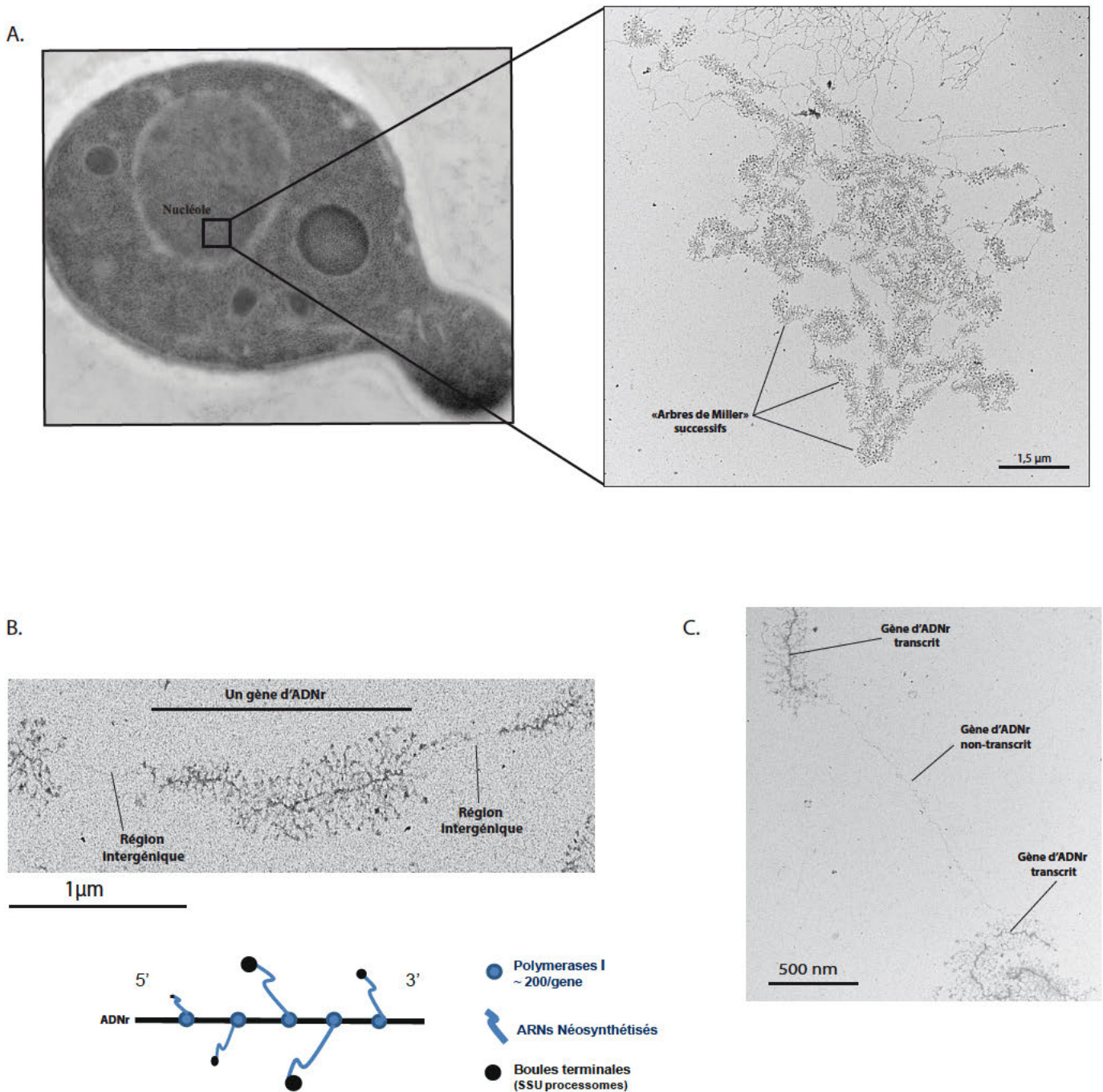
ribosome 80S. Dans le cas précis de *S.c.*, la sous-unité 40S est composée de l'ARNr 18S synthétisé par la pol I, et de 33 protéines ribosomiques. La sous-unité 60S, quant à elle, se compose des autres ARNr produits par la pol I: ARNr 25S (28S chez l'homme) et ARNr 5,8S, ainsi que de l'ARNr 5S, transcrit par la pol III. Ils sont couplés à 46 protéines ribosomiques. La structure 3D à haute résolution (3Å) du ribosome 80S de *S.c.* a par ailleurs été obtenue par cristallographie aux rayons X (Ben-Shem *et al*, 2011), permettant des études approfondies sur le sujet.

Chez les eucaryotes la biogenèse des ribosomes est un processus largement conservé. Elle débute dans le nucléole, où le précurseur des grands ARNr, appelé pré-ARNr 35S (pré-ARNr 47S chez les mammifères) est synthétisé par l'ARN pol I. Ce précurseur subit par la suite un ensemble de modifications post-transcriptionnelles, faisant intervenir des petits ARN guides: les "small nucleolar ARN" (snoARN). La majorité de ces modifications sont des méthylations de l'atome d'oxygène en position 2' des riboses, ainsi que des conversions d'uridines en pseudouridines (Lafontaine & Tollervey, 1998). En plus de ces snoARN, une série d'exo et d'endonucléases interviennent dans la maturation et permettent la production des formes matures des ARNr 18S, 5,8S et 25S. Un certain nombre de protéines ribosomiques, ainsi que plus de 200 facteurs non-ribosomiques, s'associent à l'ARNr dès les étapes initiales de sa maturation, formant une particule 90S. Suite aux clivages précoces (co-transcriptionnels chez la *S.c.*), cette particule sera ensuite scindée en deux précurseurs des sous-unités 40S et 60S. Enfin, les dernières étapes de maturation du pré-complexe 40S prendront place dans le cytoplasme, tandis que le pré-60S sera majoritairement mûré dans le noyau avant son export dans le cytoplasme (Henras *et al*, 2015) (Figure 5).



**Figure 5. Présentation schématique de la maturation des ARN ribosomiques composant les sous-unités des ribosomes chez *S.cerevisiae*.** Sur ce schéma sont indiquées les différentes formes des deux sous-unités qui composeront les ribosomes matures, et les ARN les composant à chaque étape de leur maturation. La petite et la grande sous-unité ribosomiques sont maturées de façon indépendante pour finalement s'assembler dans le cytoplasme. Le précurseur 35S des grands ARNr 25S, 18S et 5.8S est maturé suivant trois zones cellulaires différentes qui accueillent ces étapes successives de la maturation: le nucléole et le nucléoplasme (tous deux situés dans le noyau), et le cytoplasme. La production de l'ARNr 5S est indépendante de la maturation du pré-ARNr 35S. Les ARNr sont encadrés en noir, et les sites de clivages successifs sont indiqués en rouge.

La synthèse et la maturation de l'ARNr sont deux phénomènes concomitants. Ceci est révélé par la présence de boules terminales situées à l'extrémité des ARN naissants synthétisés par l'ARN pol I dans le nucléole, que l'on a pu observer dès 1969 en MET sur les fameux "Miller spreads", ou "arbres de Miller" (Figure 6). Cette technique repose sur l'étalement de la chromatine à partir de levures ayant subies une digestion partielle de leur paroi, sur des grilles de microscopie, puis contrastées par de l'acétate d'uranyle. Elles sont finalement observées au microscope électronique à transmission (Osheim *et al*, 2009). Sur ces clichés, on peut nettement observer une ou plusieurs unités d'ADNr (Figure 6.A. panel de droite), actives (Figure 6.B.) ou inactives (Figure 6.C.) en transcription. Sur les gènes actifs en transcription ("copies ouvertes") on distingue aisément les ARNs naissants, ainsi que les boules terminales au bout de ces ARNs (cliché et schéma Figure 6.C.) (Mougey *et al*, 1993). Bien qu'identifiées dès 1969, la nature moléculaire de ces boules terminales n'a été déterminée que 30 ans plus tard, renommées alors "SSU processome" (Dragon *et al*, 2002). Elles contiennent le snoARN U3, qui est impliqué dans le repliement du transcrit naissant, et non dans les modifications de bases comme la majorité des snoRNAs. U3 est essentiel pour les étapes de clivage précoces, et permet la formation de l'ARNr 18S. En plus de ce snoARN U3, ces boules terminales contiennent des protéines nucléolaires UTP ("U Three Protein"). Ces dernières se répartissent en 3 sous-complexes: UTPA (ou "t-UTP", pour "UTP requis pour la transcription"), UTPB et UTPC. Les UTPA seraient essentiels à la transcription et la maturation des ARNr (Gallagher *et al*, 2004), ce qui reste sujet à discussion dans le cas de la transcription (Wery *et al*, 2009; Dez *et al*, 2007).



**Figure 6. Région nucléolaire et «arbres de Miller» chez *S.cerevisiae*, en microscopie électronique à transmission (MET).**

A. A droite, cliché d'une coupe de levure *S.cerevisiae* en MET, où l'on distingue clairement le noyau, ainsi que le nucléole, plus dense aux électrons. A gauche, cliché de chromatine d'une région nucléolaire étalée et contrastée avant d'être observée au microscope électronique (méthode de «Miller spread»).

B. Cliché (panel du dessus) et schéma (panel du dessous) d'un gène actif en transcription, entouré par les régions intergéniques. On distingue les polymérase successives sur la région transcrite, ainsi que les ARN naissants, au bout desquels sont visibles les boules terminales, qui témoignent de la maturation co-transcriptionnelle.

C. Exemple d'une «copie fermée» d'ADNr, donc inactive en transcription, entourée par deux «copies ouvertes» actives en transcription.

Clichés A et C: Albert et al, 2011

Cliché B: communication personnelle.

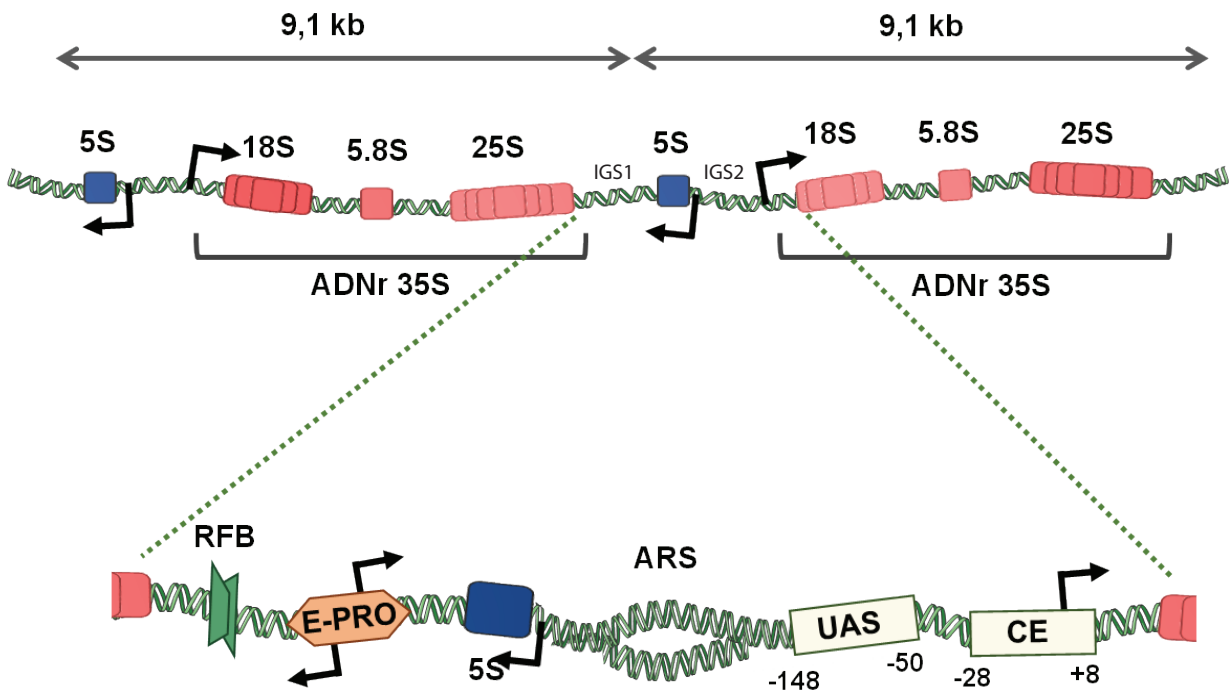


## II -2- Le locus d'ADNr et l'unité de transcription du précurseur des grands ARNr

Dans le cas de la levure *S.c.*, le locus d'ADNr est organisé en répétitions en tandem d'unités de 9,1 kb chacune (Figure 7). On trouve environ 150 copies de cette unité dans une souche sauvage, espacées par des séquences intergéniques de 2.4 kb, représentant environ 10% du génome. Les transcrits E-Pro, le 5S, et le transcrit primaire précurseur des grands ARNr ribosomaux 18S, 5.8S et 25S, le 37S, sont synthétisés à ce locus. Nous nous intéressons plus particulièrement à l'ARN 35S, d'une taille de 6,7 kb, qui est le produit de maturation du pré-ARNr 37S, clivé par l'endonucléase Rnt1. En comparaison, chez les mammifères, la taille de l'unité répétée d'ADNr est de 43 kb, avec une séquence codante de 13,7 kb et une séquence intergénique de 29,3 kb. Le nombre de répétitions de gènes d'ADNr est variable au sein du règne eucaryote. Ainsi, chez les mammifères, on évalue leur nombre à 400, réparties sur 5 paires de chromosomes (XIII, XIV, XV, XXI, et XXII), au niveau des loci particuliers que sont les NORs, cités précédemment. Chez *S.cerevisiae*, en revanche, leur nombre est de 150 à 200 copies (Schweizer *et al*, 1969; Pasero & Marilley, 1993). Un fait remarquable est que dans le cas précis des hémiascomycètes, le gène codant pour l'ARNr 5S est retrouvé entre deux unités d'ADNr, cas rare, seulement documenté chez l'amibe *Dictyostelium discoideum* (Haeusler & Engelke, 2006).

La structure globale des gènes d'ADNr peut être divisée en deux principales régions: la région codante pour le pré-ARNr 37S, et les régions intergéniques, IGS1 et IGS2, entourant le 5S (Figure 7, panel du haut: deux unités répétées de gènes d'ADNr ; panel du bas: zoom sur régions intergéniques). Une fois transcrit, le précurseur suit un ensemble d'étapes de maturation afin d'aboutir aux formes d'ARNr matures. La taille du transcrit primaire varie selon l'organisme étudié, allant jusqu'à plus de 13 kb dans le cas des mammifères. On peut attribuer cette variabilité à la différence de taille des espaceurs internes (ITS) et externes (ETS) des régions transcrites.

Les IGS1 et IGS2 séparent donc deux unités de transcription du précurseur de l'ARNr 35S, et sont situées de part et d'autre du gène codant pour l'ARNr 5S, lui-même synthétisé par l'ARN pol III. Ces régions intergéniques contiennent notamment une origine de réplication et des régions 5' et 3', respectivement promotrice et terminatrice, régulant la transcription par l'ARN pol I. De même que la région transcrite, la taille de la région intergénique varie selon les espèces, allant de presque 5 kb chez la levure, à plus de 30 kb chez l'Homme.



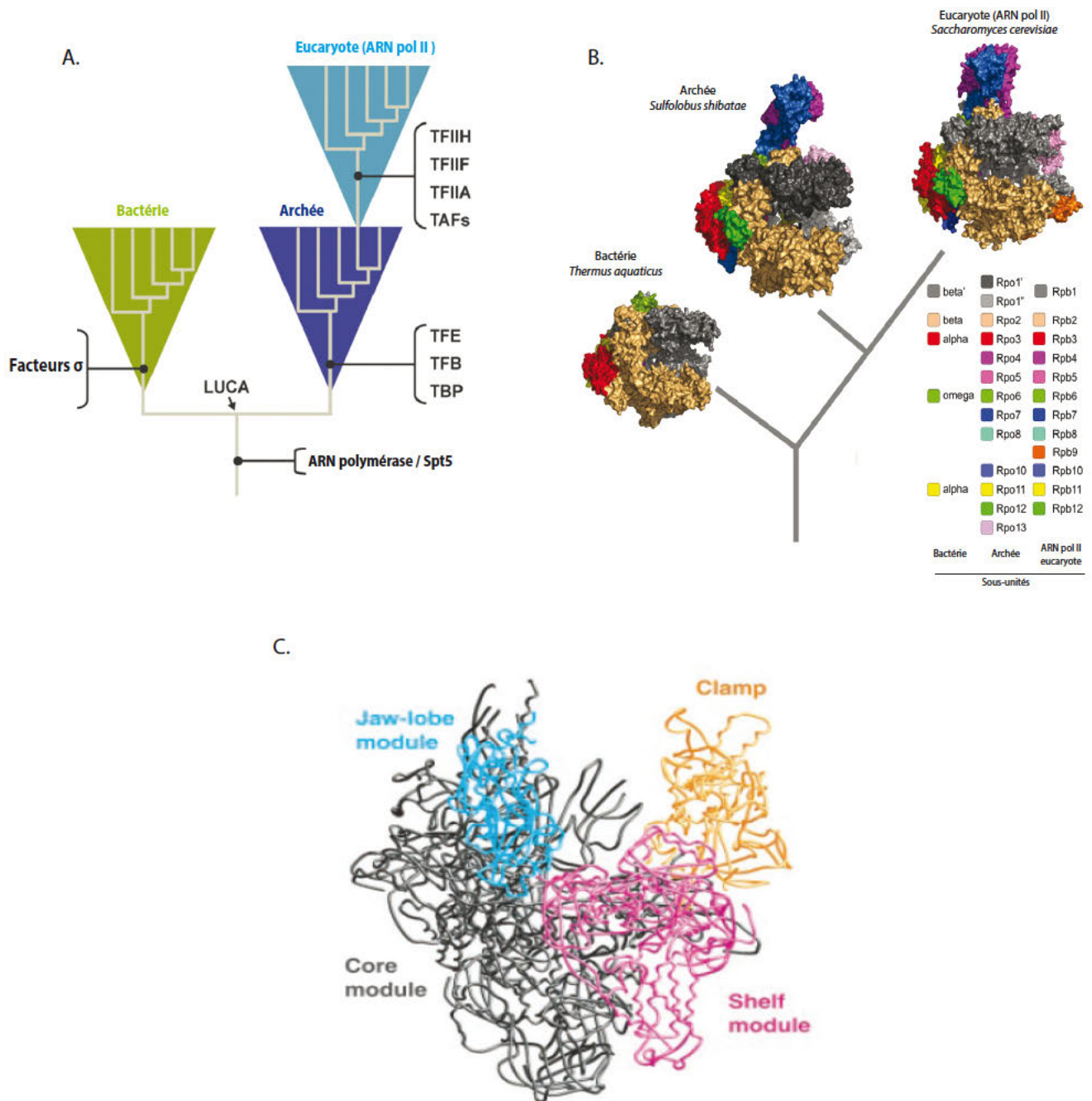
**Figure 7. Organisation d'une répétition en tandem de gènes d'ADN ribosomique (ADNr) de *S.c.***  
 Un gène d'ADNr contient la séquence codant pour le précurseur des grands ARN ribosomiques 18S, 5.8S et 25S (rouges). La région entre deux gènes d'ADNr contient le gène codant pour l'ARNr 5S transcrit par la pol III (bleu), une origine de répliation (ARS = Autonomously Replicating Sequence), les séquences initiatrices de la transcription par la pol I (UAS = Upstream Activating Sequence ; CE = Core Element), et le promoteur bi-directionnel pol II E-Pro (Enhancer Promoter) (orange) ; (RFB = Replication Fork Barrier). Les gènes d'ADNr sont répétés environ 150 à 200 fois dans le génome de *S.cerevisiae*.

### **Chapitre III - Différentes ARN polymérases dans le règne du vivant et leur conservation évolutive : généralités sur les ARN polymérases multimériques, et introduction par l'ARN polymérase bactérienne.**

Il existe deux types d'ARN polymérases ADN dépendantes. Les premières sont les ARN polymérases monomériques, que l'on retrouve dans les mitochondries et dans les phages T3 ou T7, et dont nous ne parlerons pas dans ce manuscrit (Cermakian *et al*, 1997). Les deuxièmes sont les ARN polymérases multimériques, présentes dans toutes cellules, classées en deux groupes.: les polymérases qui reconnaissent directement l'ADN, de type  $\alpha\beta\beta'\omega + \sigma$ , telle que l'ARN polymérase bactérienne, et les enzymes des archées et eucaryotiques, qui reconnaissent des facteurs généraux de transcription en complexe avec l'ADN à transcrire (Figure 8.A.), formant un complexe de pré-initiation (PIC pour "Pre-Initiation Complex") (Buratowski *et al*, 1989; Blombach *et al*, 2016). En observant la conservation évolutive de ces facteurs de transcription, il est intéressant de mentionner qu'un seul facteur général de transcription associé à l'enzyme est commun à toutes les ARN polymérases, et ce depuis avant l'ancêtre commun LUCA ("Last Universal Common Ancestor", organisme le plus récent dont seraient issus les trois règnes bactériens, archéens et eucaryotes) : le facteur d'élongation de la transcription Spt5 (NusG chez les bactéries) (Blombach *et al*, 2016). L'activité transcriptionnelle des ARN polymérases multimériques ADN dépendantes s'effectue en plusieurs étapes successives, qui sont classiquement décrites en trois phases majeures. L'initiation est caractérisée par le recrutement des ARN polymérases au niveau d'un complexe multi-protéique permettant sa fixation sur l'ADN et le début de son activité catalytique. La transition entre initiation et élongation est caractérisée par un changement dans l'affinité de la polymérase à l'ADN et la dissociation du PIC. L'élongation est la phase où la polymérisation dite processive d'ARN à partir de la matrice ADN s'effectue. Cette phase, fondamentalement discontinue, est une succession de synthèses d'ARNs et de pauses (voir chapitre V-2.). Enfin,

l'étape de terminaison, initiée par une pause programmée, permet le relargage du transcrit néo-synthétisé. Une étape de recyclage de l'ARN polymérase est également à mentionner, permettant une ré-initiation efficace du cycle transcriptionnel.

D'un point de vue structural global, les ARN polymérases cellulaires sont des complexes protéiques possédant de nombreuses sous-unités différentes, et adoptant une forme générale de "pince de crabe". On peut distinguer quatre modules principaux dans ce type d'enzyme, que l'on citera selon leur nomination anglaise pour plus de clarté par la suite : le "core," le "shelf," le "clamp," et le module "jaw-lobe" (Cramer *et al*, 2001; Zhang *et al*, 1999) (Figure 8.C.) . Les modules core et shelf forment la partie centrale de l'enzyme, où l'on peut retrouver deux canaux principaux. Le premier, le plus large, permet la fixation des acides nucléiques. On y trouve également les deux ions magnésium catalytiques ( $Mg^{2+}$ ). Le second, plus petit, aussi nommé "secondary channel", permet l'entrée des substrats nucléotidiques (NTPs) nécessaires à la réaction enzymatique à proprement parler, ainsi que la sortie des pyrophosphates produits par la réaction. On trouve également dans cette partie centrale deux éléments essentiels que nous étudierons en détails par la suite: la "Bridge Helix" (BH) et la "Trigger Loop" (TL), proches du site catalytique. Les modules clamp et jaw-lobe, enfin, sont respectivement des extensions des modules shelf et core, et complètent le canal le plus large pour maintenir l'hybride ADN-ARN et l'ADN en aval. (Gnatt *et al*, 2001; Vassylyev *et al*, 2007).



**Figure 8. Modules et sous-unités des ARN polymérases multimériques.**

A. arbre évolutif simplifié des machineries d'initiation de la transcription dans différents règnes du vivant (bactéries, archées, eucaryotes) et les facteurs de transcription associés. Dans le cas des eucaryotes, seule l'ARN pol II a été représentée. Seuls l'ARN polymérase et Spt5 étaient présents avant le dernier ancêtre commun (LUCA = «Last Universal Common Ancestor»). Des facteurs de transcription additionnels ont été rajoutés par la suite dans les différents règnes du vivant, requis pour leurs activités respectives (D'après Blombach et al, 2016).

B. Représentation simplifiée d'un arbre évolutif des ARN polymérases multimériques des trois domaines du vivant, avec leurs structures. Les sous-unités sont indiquées selon le code couleur en bas à droite. D'après la revue de Finn Werner, 2013 (*Thermus aquaticus*: pdb 116V ; *Sulfolobus shibatae*: pdb 4B1O ; *Saccharomyces cerevisiae* ARN pol II: pdb 1NT9).

C. Structure simplifiée mettant en évidence les quatre modules principaux formant l'ARN polymérase II de *S.c.* D'après Cramer et al, 2001.

En termes de nomenclature, les sous-unités des ARN polymérase des 3 règnes du vivant possèdent des nominations différentes. De multiples dénominations ont été proposées par différents laboratoires, je ne présenterai ici que la nomenclature utilisée (Tableau 1 ; Figure 8.B.).

Premièrement, les sous-unités de l'ARN polymérase bactérienne sont nommées d'après l'alphabet grec ( $\alpha$ ,  $\beta$ ,  $\beta'$  et  $\omega$ ), et sont associées au facteur permettant la reconnaissance des promoteurs, le facteur  $\sigma$ . Ensuite, dans le cas de la polymérase archéenne, elles sont nommées Rpo (RNA polymerase) suivies d'un nombre (1,2, etc.), ou d'une lettre, selon l'ancienne nomenclature (Rpo A, B etc.). Dans les systèmes eucaryotes, les sous-unités sont nommées selon leur appartenance, A, B ou C, respectivement pour la pol I, II ou III. Les sous-unités partagées entre plusieurs complexes sont nommées AC pour celles communes à Pol I, Pol III, ou ABC pour les sous-unités communes aux trois polymérases. La taille en kDa est ensuite utilisée pour les distinguer, ce qui peut porter à confusion. Cette nomenclature est encore largement utilisée pour les sous-unités des ARN pol I et III. Les sous-unités de l'ARN polymérase II sont maintenant notées Rpb (pour RNA polymerase b, à savoir de type II), suivies de leur rang, généralement décrit par ordre décroissant de taille (Sayre *et al*, 1992). Les gènes codant pour les sous-unités communes sont maintenant uniquement nommés suivant la nomenclature Pol II (Rpb5, Rpb6, Rpb8, Rpb10 et Rpb12).

On retrouve donc dans la littérature des sous-unités de type "Rpa, Rpb ou Rpc" correspondant aux ARN pol I, II ou III, ainsi que des peptides A et C suivis de la taille. Une nomenclature unifiée, utilisant le rang, pour Pol I et Pol III est toujours absente.

Au cours de ce travail de thèse, j'ai choisi de me concentrer sur une forme d'ARN pol I dépourvue d'une de ses sous-unités. Je vais donc présenter en détail les compositions des polymérases multimériques et les rôles connus des sous-unités les constituant.

|   | <u>Bactéries</u>   | <u>Archées</u> | <u>Eucaryotes (<i>Saccharomyces cerevisiae</i>)</u> |                        |                        |
|---|--------------------|----------------|---|------------------------|------------------------|
|   |                    |                | <u>ARN Pol II</u>                                   | <u>ARN Pol I</u>       | <u>ARN Pol III</u>     |
| <u>"Core enzyme"</u>                            | $\beta$            | Rpo1 (A)       | Rpb1  | Rpa190                 | Rpc160                 |
|   | $\beta'$           | Rpo2 (B)       | Rpb2  | Rpa135                 | Rpc128                 |
|   | $\alpha$           | Rpo3 (D)       | Rpb3  | AC40 (Rpc40)           | AC40 (Rpc40)           |
|   | $\alpha$           | Rpo11 (L)      | Rpb11   | AC19 (Rpc19)           | AC19 (Rpc19)           |
|   | $\omega$           | Rpo6 (K)       | Rpb6  | ABC23 (Rpb6)           | ABC23 (Rpb6)           |
| <u>Sous-unités<br/>paralogues</u>               |                    | Rpo5 (H)       | Rpb5  | ABC27 (Rpb5)           | ABC27 (Rpb5)           |
|   |                    | Rpo4 (F)       | Rpb4  | Rpa14                  | Rpc17                  |
|   |                    | Rpo 7 (E)      | Rpb7  | Rpa43                  | Rpc25                  |
|   |                    | Rpo M ou TFS   | Rpb9  | Rpa12                  | Rpc11                  |
|   |                    | Rpo10 (N)      | Rpb10   | ABC10 $\alpha$ (Rpb10) | ABC10 $\alpha$ (Rpb10) |
|   |                    | Rpo12 (P)      | Rpb12   | ABC10 $\beta$ (Rpb12)  | ABC10 $\beta$ (Rpb12)  |
|   |                    | Rpo8 (G)       | Rpb8  | ABC14.5 (Rpb8)         | ABC14.5 (Rpb8)         |
| <u>"Built in<br/>transcription<br/>factors"</u> |                    |                | TFIIF $\beta$                                       | Rpa34.5                | Rpc37                  |
|   |                    |                | TFIIF $\alpha$                                      | Rpa49                  | Rpc53                  |
|   |                    |                | TFIIE $\beta$                                       |                        | Rpc34                  |
|   |                    |                | TFIIE $\alpha$                                      |                        | Rpc82                  |
|   |                    | Rpo13          |   |                        | Rpc31                  |
| <u>Facteurs<br/>de transcription</u>            |                    | TBP            | TBP   | TBP                    | TFIIB (TBP)            |
|   |                    | TFB            | TFIIB   | Rm7                    | TFIIB (Brl1)           |
|   |                    |                | TFIIA   |                        |                        |
|   |                    |                | TFIIH   |                        |                        |
|   |                    | TFS            | TFIIS   | Rpa12 (C-ter)          |                        |
|   |                    | Spt4           | Spt4  |                        |                        |
|   |                    | Spt5           | Spt5  |                        |                        |
|   | NusG               | NusA           |   |                        |                        |
|   | sigma ( $\sigma$ ) |                |   |                        |                        |
|   | rho                |                |   |                        |                        |
| Gre   |                    |                |   |                        |                        |

**Tableau 1.** Classification des différentes sous-unités et de certains facteurs de transcription dans les trois règnes du vivant. Pour les eucaryotes, l'organisme modèle *Saccharomyces cerevisiae*, utilisé durant ces travaux de thèse, a été pris comme exemple. De plus, l'ARN polymérase I, étudiée en particulier, est indiquée en rouge.



### III-1- La synthèse des ARNs

La synthèse des ARNs est effectuée par les enzymes particulières que sont les ARN polymérase ADN dépendantes. Cette réaction catalytique s'effectue en effet par complémentarité à l'ADN matrice via la polymérisation du côté 3' de l'ARN naissant, nucléotide après nucléotide. Les substrats de cette réaction enzymatique sont donc des ribonucléotides triphosphate (rNTP), et la succession d'ARNs ( $\text{poly}_n(\text{NMP})$ ). Il y a alors libération d'un Pyrophosphate (PPi) après allongement de la chaîne d'ARN synthétisée (Figure 9.).



La forme générale d'une ARN polymérase multimérique peut être comparée à une pince de crabe, qui contient en son sein le site catalytique, et dans laquelle l'ADN vient se fixer entre les deux bords de la pince pour atteindre ce dernier. Les ARN pol multimériques sont ainsi organisées autour de leur site actif, formé par deux tonneaux- $\beta$  (" $\beta$ -barrels") situés dans les deux plus grandes sous-unités de l'enzyme eucaryote respective (correspondantes aux sous-unités  $\beta$  et  $\beta'$  chez la bactérie, Rpo1 et Rpo2 chez l'archée, ou Rpa190 et Rpa135 pour l'ARN pol I) (les deux plus grosses sous-unités forment la majeure partie de l'enzyme). Ces deux tonneaux- $\beta$  ont chacun des résidus d'acides aminés bien distincts qui sont organisés de façon asymétrique (de motif -NADFDGD-, qui chélatent les ions magnésium), et deux résidus lysines impliqués dans la liaison de l'ADN matrice. Concernant la réaction enzymatique, qui ressemble beaucoup à celle des ADN polymérase, deux ions magnésiums ( $\text{Mg}^{2+}$ ) sont requis, mais un seulement est lié de manière constitutive à l'enzyme (on parle alors d'un "mécanisme d'addition de nucléotide à deux ions magnésium"). Ces deux ions  $\text{Mg}^{2+}$  accompagnent le dNTP entrant, et sont liés à celui-ci *via* les phosphates du nucléotide et à deux résidus d'acides aspartiques, très conservés. Dans le mécanisme de polymérisation, le premier ion abaisse

l'affinité de l'hydrogène du groupement -OH en 3' du NTP initiateur, qui attaque alors ( $3' O^-$ ) le groupement phosphate  $\alpha$  du second NTP (par réaction de substitution nucléophile). Le deuxième ion entre dans le site actif en complexe avec le NTP et laisse l'ARN polymérase liée au pyrophosphate après la formation de la liaison phosphodiester. Les deux ions magnésiums stabilisent donc ce mécanisme (Steitz, 1998).



### III-2- L'activité transcriptionnelle de l'ARN polymérase bactérienne: l'ARN polymérase multimérique modèle

Quel que soit l'organisme étudié, le terme "expression génique", regroupe factuellement à la fois les notions de transcription (décodage de l'information génétique ADN et synthèse d'ARN par les ARN polymérase) et de traduction (utilisation des matrices ARN pour la production des protéines cellulaires par les ribosomes). Or, chez les procaryotes, il a été montré que ces mécanismes étaient liés (Byrne *et al*, 1964), alors qu'ils sont dissociés chez les eucaryotes. On parle alors d'un complexe alliant transcription et traduction chez les procaryotes, comprenant l'ARN polymérase et le ribosome en interaction avec l'ADN: "l'expressome", qui a notamment été montré/étudié chez *E.coli*. On notera tout particulièrement qu'une structure récente de l'expressome a été obtenue à haute résolution, montrant les différentes interactions entre l'ARN polymérase et le ribosome (Kohler *et al*, 2017). Entre autres, une interface ARN pol - ribosome importante s'effectuerait au niveau du domaine  $\alpha$  C-terminal de la polymérase ( $\alpha$ CTD).

Chez la bactérie, on ne retrouve qu'une seule ARN polymérase qui permet la transcription de l'ensemble du répertoire des gènes bactériens, caractérisée dès les années 1960 chez *Escherichia coli* (*E.coli*), créant ainsi l'un des premiers modèles étudiés (Chamberlin & Berg, 1962).

Or, lorsque l'ARN polymérase bactérienne est en cours de transcription et produit un ARNm contenant une séquence de liaison au ribosome (RBS, pour "Ribosome Binding Sequence"), la traduction peut être initiée et empêcher la terminaison de la transcription. La formation de ce complexe typiquement procaryote, l'expressome, ne peut donc se former que pendant la phase d'élongation de la transcription, ce qui explique pourquoi la traduction prévient les phénomènes de pause, "backtrack" et terminaison de la transcription.

Enfin, afin de reconnaître leurs promoteurs, les ARN polymérases bactériennes nécessitent un facteur spécialisé, appelé le facteur sigma,  $\sigma$ , que l'on ne retrouve pas chez les archées ou les eucaryotes. Il permet de contacter directement la séquence promotrice cible. Le complexe constitué du facteur  $\sigma$  et du reste des sous-unités formant la polymérase constitue ce que l'on appelle l'Holoenzyme,  $\alpha_2\beta\beta'\omega\sigma$ . Les bactéries possèdent des facteurs  $\sigma$  variés (Helmann & Chamberlin, 1988), qui reconnaissent spécifiquement des séquences promotrices différentes, déterminant quels sont les gènes qui seront transcrits. Cette spécificité s'explique par le fait que ces facteurs  $\sigma$  reconnaissent et se lient à des séquences d'ADN clefs que sont les éléments -10 et -35 du promoteur. Ainsi, l'efficacité de reconnaissance du promoteur est en partie dû à la force d'interaction des différents types de facteurs  $\sigma$  avec ces deux éléments. Ce facteur  $\sigma$  se détache de la polymérase une fois que la synthèse processive des ribonucléotides commence (Werner & Grohmann, 2011).

Les polymérases bactériennes (*E.coli* et *Thermus aquaticus*) ont été largement utilisées pour les premières études des mécanismes fondamentaux de la transcription (Bar-Nahum *et al*, 2005).

### **III-3- Le Core enzyme bactérien**

L'ARN polymérase bactérienne, d'un poids moléculaire d'environ 400 kDa, est composée d'un cœur enzymatique, que l'on nomme le "Core enzyme", nom qui sera souvent repris par la suite car commun à toutes les ARN polymérases multimériques que nous étudierons dans ce manuscrit. Une base de 5 sous-unités compose le Core enzyme :  $\alpha_2\beta\beta'\omega$  (Figure 10.A.), qui est conservé dans l'évolution de la bactérie à l'humain. Ces sous-unités possèdent de nombreuses interfaces d'interaction entre elles. De fait, la structure de l'ARN polymérase de *Thermus aquaticus*, une bactérie thermophile, a été obtenue en 1999 à haute résolution par cristallographie aux rayons X (Zhang *et al*, 1999), permettant ensuite de proposer un modèle

de complexe d'élongation où le centre catalytique est résolu à l'échelle atomique (Korzheva *et al*, 2000).

Ainsi, le Core enzyme révèle cette fameuse structure en forme de "pinces de crabe", avec un canal interne s'étendant le long des "pinces", et le site actif en son centre. On retrouve deux sous-unités monomériques de type  $\alpha$ , formant un homodimère  $\alpha_2$ , ne participant pas directement à l'activité catalytique de l'enzyme (aucun résidu de ce dimère n'accède au site actif de l'enzyme). Elles permettent l'ancrage et l'assemblage des deux sous-unités dites catalytiques,  $\beta$  et  $\beta'$ . Ces dernières sont les composantes des deux "pinces" de l'enzyme,  $\beta$  l'une,  $\beta'$  la seconde. Chacune contacte exclusivement l'un des monomères  $\alpha$ , et elles interagissent largement entre elles. La sous-unité  $\omega$  ne contacte que la s.u.  $\beta'$  en s'enroulant autour de son extrémité C-terminale, suggérant que  $\omega$  jouerait un rôle dans les étapes finales d'assemblage de l'ARN polymérase bactérienne (Zhang *et al*, 1999).

Les s.u.  $\beta$  et  $\beta'$  possèdent une zone d'interaction majeure située à la base du canal du site actif, positionnant le motif universellement conservé -NADFDGD- permettant de chélater les deux ions  $Mg^{2+}$  du site actif. C'est au niveau de ce domaine appelé le "Wall" de  $\beta\beta'$  que l'ADN, après être passé le long du canal des "pinces", change d'angle pour être positionné au niveau du centre catalytique de l'enzyme. Après la synthèse d'un court ARN, la transcription passe en phase d'élongation, où le brin de l'ARN naissant reste lié au brin d'ADN matrice, formant un hybride ADN-ARN de 8-9 pb. Dans un premier temps, de courts ARNs (3-8 nt) sont produits alors que le complexe reste au niveau du promoteur. On parle de transcription abortive. Ce n'est que pour des ARNs de 10 nt et plus que l'on entre dans une deuxième phase, dite de transcription processive. Cette étape permet que la polymérase se dissocie du promoteur ("promoter escape") pour entrer dans la phase d'élongation processive, entraînant la dissociation du facteur  $\sigma$  de façon stochastique, qui ne se refixe à l'enzyme que durant la terminaison ou lorsque la polymérase est en pause (Paget, 2015). L'hybride, assez long, et le

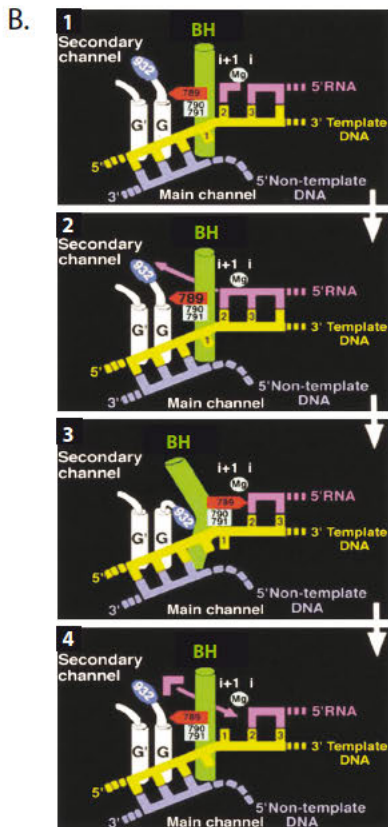
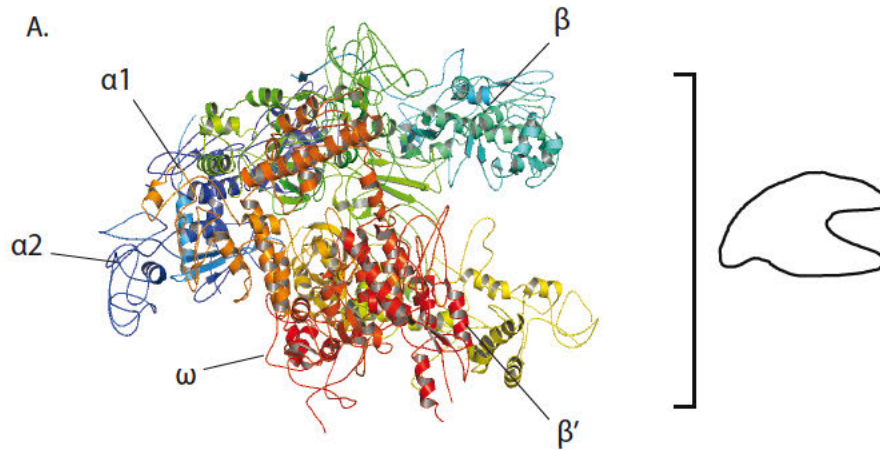
duplex d'ADN en aval, sont fermement maintenus dans le "primary channel", canal qui est formé par les pinces de la polymérase, permettant une élongation processive où l'ADN est transcrit avec fidélité. Le substrat nucléotidique entrant arrive dans le site actif par un autre canal appelé le "secondary channel", derrière les pinces. Le site catalytique où l'addition de nucléotides s'effectue est localisé à la jointure entre ces pinces. Là, la "Bridge Helix" (BH), connecte les deux pinces à la jonction entre le site de liaison de l'hybride ADN-ARN, le site de liaison de l'ADN en aval, et le "secondary channel". La BH peut adopter des états conformationnels différents: droit/raide, ou replié. Ces éléments structuraux sont au cœur du mécanisme catalytique des ARN polymérases multimériques en transcription: le mécanisme de "Brownian ratchet", ou "cliquet moléculaire", qui entre en jeu avant même la formation du lien Phosphodiester (voir Figure 10.B.1/2/3/4). Ce mécanisme est caractérisé par plusieurs étapes consécutives et unidirectionnelles, dues aux interactions entre les acides nucléiques (nucléotides substrats, ADN matrice, ARN produit) et les éléments protéiques de l'ARN polymérase ("Bridge Helix, Trigger Loop", au niveau du site actif). Pour plus de clarté, ces différentes étapes sont schématisées en 4 phases principales dans la Figure 10.B. : (1) Le nucléotide substrat complémentaire de l'ADN matrice entre dans le site actif par un canal secondaire ("secondary channel") et vient se placer au site que l'on nomme site "i+1", empêchant tout mouvement rétrograde de la polymérase. (2-3) La Trigger Loop induit le mouvement de la Bridge Helix (BH), ce qui a pour conséquence le repliement de cette dernière. En se repliant, des résidus de la BH entrent en contact avec le substrat et l'ADN au niveau du site "i+1", occupant physiquement ce site, et poussant le substrat du site "i+1" vers le site "i". On appelle ce phénomène la translocation. (4) Une fois que la translocation est effectuée, la BH se déplie et ré-adopte une conformation droite, laissant le site i+1 libre pour que le nouveau substrat complémentaire arrive. Ces étapes fonctionnent comme des cliquets

moléculaires, bloquant le système dans une conformation donnée à chaque étape et l'empêchant de revenir en arrière.

L'hybride ADN-ARN s'élève alors hors du site actif perpendiculairement au duplex ADN matrice entrant, et les brins sont séparés.

En 2005, Bar-Nahum *et al* proposent un modèle de complexe en élongation (EC) à "deux cliquets" (Figure 10.C.). Ce modèle explique les bases de l'élongation de l'enzyme en incluant le concept du cliquet moléculaire schématisé comme une roue dentelée avec un cliquet qui restreint son mouvement dans une seule direction. Il y aurait ainsi en réalité deux types de cliquet, qui coexisteraient durant l'élongation. Le premier, le "cliquet stationnaire" (violet dans la figure 10.C) représente le substrat nucléotidique complémentaire qui arrive au site  $i+1$ , et empêche mécaniquement l'extrémité 3' de l'ARN de s'y placer. Le mouvement inverse est de ce fait rendu impossible via l'insertion d'un élément dans la "dent" de la roue, représenté ici par le substrat qui s'insère dans le site actif. Le second, le "cliquet réciproque" (vert, figure 10.C) correspond au mouvement de la Bridge Helix qui peut adopter des conformations repliées ou raides, déplaçant le complexe vers l'avant depuis le site  $i+1$  en se repliant. En retournant à sa conformation droite, la BH permet alors au substrat complémentaire suivant d'entrer dans le site  $i+1$ . On trouve les 2 types dans l'EC, en compétition pour le même site de liaison.





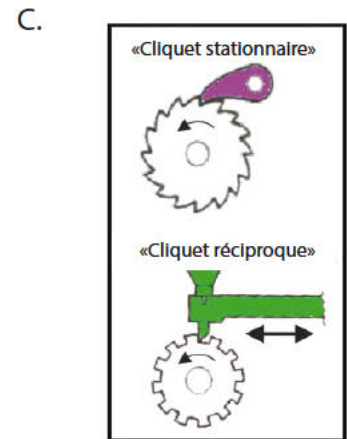
1. Etat du complexe de transcription avant que le lien phosphodiester ne soit établi. Un nucléotide substrat est entré via le secondary channel pour se placer au site actif (représenté par l'ion Mg en blanc), au niveau du site «i+1». La «Bridge Helix» (BH) (en vert) est dans une conformation droite. Les hélices de la «Trigger loop» ou «G loop», G et G' (en blanc) permettent de réguler la position de la BH.

2. Etape de pré-translocation. La flèche rose représente ici la trajectoire que prendrait l'ARN si un phénomène de «backtrack», ou de relargage de l'ARN abortif se produisait. La BH garde sa conformation raide.

3. Etape de translocation. La méthionine 932 (Met932, bleue) de la Trigger Loop entre en contact avec la lysine 789 (lys789, en rouge) de la BH qui occupe alors le site i+1. La BH est alors en position repliée. Les chaînes latérales des acides aminés 790 et 791 de la BH (en blanc) entrent en collision avec la base de l'ADN du site «i+1» (en jaune), induisant sa translocation.

4. La Bridge Helix se relâche et ré-adopte la conformation droite (1), laissant le site «i+1» libre pour le nucléotide substrat suivant (flèche rose), ou pour une translocation inverse (Backtrack).

D'après Epshtein et al, 2002



**Figure 10. L'ARN polymérase bactérienne et le mécanisme réactionnel de «cliquet moléculaire».**

A. Structure 3D de l'ARN polymérase bactérienne de *Thermus aquaticus* (pdb :1I6V) de type  $\alpha_2\beta\beta'\omega$ . On remarque cette forme caractéristique de «pince de crabe», schématisée en noir à droite de la structure.

B. Représentations structurales schématisées des différentes étapes lors de l'addition d'un nucléotide au sein du site actif d'une ARN polymérase bactérienne. Les éléments protéiques de l'enzyme sont schématisés sous forme de cylindres, les acides nucléiques sous forme de barres rectangulaires. L'ARN (rose) en interaction complémentaire avec le brin matrice d'ADN (jaune) effectue le phénomène de translocation du site «i+1» au site «i» après addition d'un nucléotide substrat, grâce aux ions  $Mg^{2+}$ . Le canal secondaire, ici «secondary channel», accueille les nucléotides substrats et l'ARN lors de l'activité de Backtrack ou de relargage abortif de l'ARN.

C. Représentation imagée des deux formes de «cliquet moléculaire» qui agissent dans un complexe en élongation (EC), comme proposé par Bar-Nahum et al, 2005.

On dit alors que l'ARN polymérase peut adopter deux états conformationnels (Sekine *et al*, 2015) (Figure 10.D.) :

1) un état dit resserré ("tight polymerase"), associé à l'élongation, dans lequel la BH de la polymérase est rigide, droite, mais peut se plier dans certaines conditions, et où l'enzyme peut réaliser l'activité d'addition de nucléotides au niveau de son site actif.

2) une forme repliée ("ratcheted polymerase"), où la BH adopte une conformation spécifique et se plie, pouvant mener à des fonctions transcriptionnelles telles que les activités de pause et de "Backtrack" de la polymérase.

Nous pouvons ainsi noter à ce niveau un point important dans le cadre de cette étape de la transcription, quelle que soit l'ARN polymérase multimérique étudiée : l'activité de "Backtrack" de la polymérase, basée sur le phénomène de Brownian ratchet ou cliquet moléculaire. En effet, des bactéries aux eucaryotes, en passant par les archées, les complexes en élongation de la transcription peuvent également adopter des états "inactifs" en transcription. Certaines séquences d'ADN particulières, des dommages à l'ADN ou l'incorporation de mauvais nucléotides substrats peuvent former des obstacles à la transcription. Il a été montré que s'arrêter à de tels sites ne bloquait pas totalement pour autant la polymérase dans une conformation donnée et à un endroit fini. Cela induit en revanche son glissement depuis sa position initiale vers l'arrière le long de l'ADN et de l'ARN sans dégradation du transcrit: c'est le phénomène de "Backtrack" (Komissarova & Kashlev, 1997; Nudler *et al*, 1997; Saeki & Svejstrup, 2009) (Figure 10.D et E.). Ce mouvement rétrograde fait que le site actif de l'enzyme n'est alors plus en contact direct avec l'extrémité 3'-OH de l'ARN, ce qui empêche temporairement l'enzyme d'effectuer son activité catalytique. Une polymérase en Backtrack possède des différences structurales mineures avec une polymérase en élongation, notamment au niveau de son site actif, où l'on remarque la position de la

Trigger Loop est modifiée (Figure 10.F.). Cet état de l'enzyme est en équilibre avec celui où elle peut retourner à sa position d'origine, faisant de la polymérase un objet qui oscille entre états catalytiquement actifs et inactifs. L'état inactif de "Backtrack" de l'enzyme constitue un intermédiaire important dans le mécanisme de clivage de l'ARN. Il existe des facteurs d'élongation qui permettent de surmonter ces obstacles, tels que les facteurs GreA ou GreB chez la bactérie.

## **Chapitre IV - Le cas des archées : une autre ARN polymérase ADN dépendante cellulaire unique**

Le fait que la machinerie de transcription par l'ARN polymérase archéenne se rapproche fortement des activités connues des complexes transcriptionnels de type eucaryote permet l'étude, dans un système simple comptant un nombre restreint de facteurs en trans, de processus tels que la reconnaissance du promoteur, le recrutement de l'ARN polymérase ou l'initiation de la transcription. Le système archéen permet une étude des mécanismes fondamentaux de la transcription très précise du point de vue moléculaire. C'est dans cette optique que nous allons brièvement présenter l'appareil de transcription chez l'Archée et l'ARN polymérase qui s'y réfère.

### **IV -1- La transcription chez les archées**

Les archées sont des procaryotes, mais possèdent une machinerie de transcription très similaire à l'ARN Pol II des eucaryotes, à la fois structuralement et fonctionnellement (figure 8A). La reconnaissance des promoteurs ne dépend pas de facteurs de type sigma comme chez les bactéries. Pour que la transcription chez l'archée soit effective, trois facteurs de transcription principaux, que sont la "TATA binding protein" (TBP), TFB et TFE sont

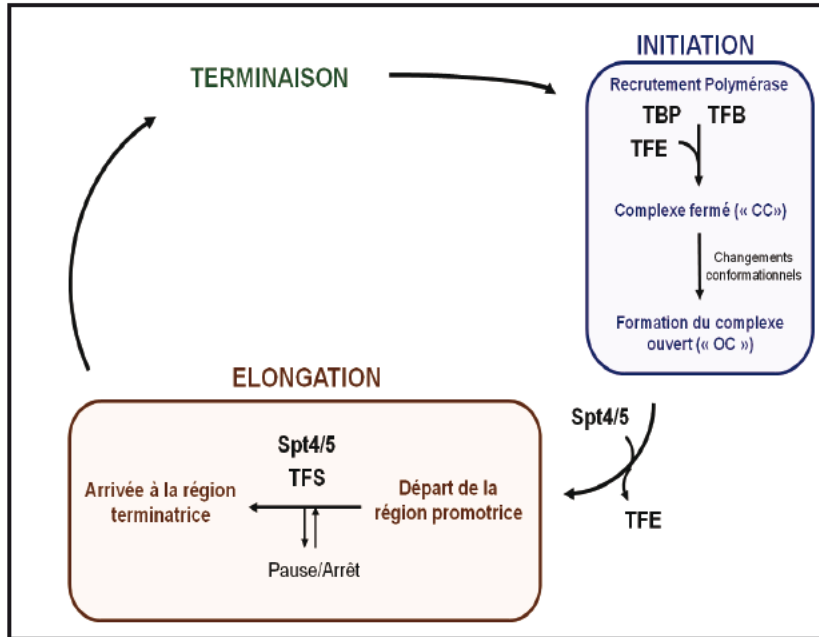
nécessaires. Tous ces facteurs ont une fonction similaire à un ou plusieurs facteurs de transcription eucaryotes (Werner, 2013).

La présence chez l'archée d'un petit nombre de facteurs de transcription, mais d'un mécanisme commun proche du système eucaryote, rend son étude et celle de son cycle transcriptionnel (Figure 11) importante pour mettre en lumière les événements clés dans la production d'un ARN *via* la formation d'un complexe de pré-initiation, ou PIC, ce qui diffère du système bactérien.

Les facteurs généraux de transcription, comprenant TBP et le "transcription factor B" (TFB) reconnaissent les éléments du Core promoter pour recruter ensuite l'ARN polymérase archéenne. La TBP reconnaît la TATA box du promoteur, ce qui amène au recrutement de TFB et de l'ARN polymérase. Ce complexe de pré-initiation (PIC) TBP-TFB-ARN polymérase est dit "fermé", ou "Close Complex" (CC), avec l'ADN sous forme double-brin. Puis des changements conformationnels s'effectuent, où l'ADN est déroulé et le brin matrice va jusqu'au site actif de l'enzyme, formant un complexe "ouvert" ("Open Complex", OC).

Chez l'Archée, cette transition du CC vers l'OC se fait de manière spontanée, c'est à dire sans nécessité d'hydrolyse d'ATP, uniquement stimulée par le facteur d'initiation TFE (Weinzierl, 2013).

La régulation de l'expression génique chez l'Archée peut s'effectuer de deux façons. Soit en empêchant aux facteurs de transcription généraux l'accès à la région du promoteur, soit en augmentant le recrutement des composants de la machinerie de transcription générale pour une activation gène-spécifique (Bell *et al*, 2001).



**Figure 11. Le cas de l'archée: cycle de transcription schématique de l'ARN polymérase archéenne.**

On retrouve les trois phases principales de la transcription: initiation, élongation et terminaison. Sont indiqués également, aux phases du cycle correspondantes, les facteurs de transcription archéens, à savoir TBP, TFB, TFE, Spt4/Spt5, et TFS.

#### **IV -2- Structure de l'ARN polymérase archéenne**

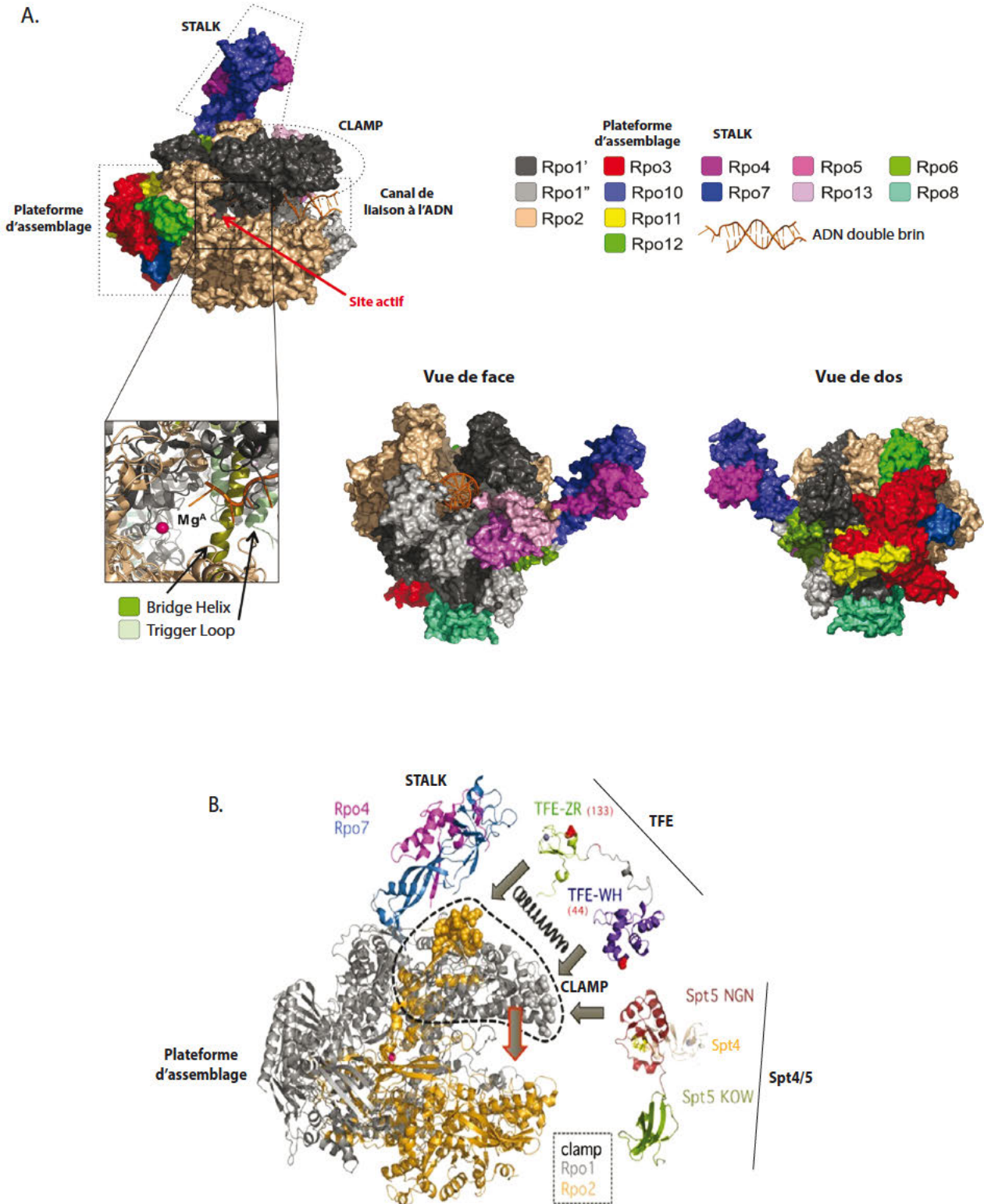
En termes de structure et de fonction, l'ARN polymérase archéenne et les enzymes eucaryotes (principalement l'ARN pol II) possèdent de fortes homologies, avec une composition en sous-unités directement comparable (voir Tableau 1). Les régions importantes du domaine catalytique entre polymérases bactériennes et archéennes sont conservées, telles que la Bridge Helix ou la Trigger loop, et de fortes similarités structurales sont retrouvées avec les polymérases eucaryotes. Les sous-unités Rpo1/2/5/13 interagissent avec l'ADN matrice, constituant les deux "pinces" du Clamp de l'ARN polymérase archéenne (Figure 12.A.). De même que les enzymes eucaryotes, l'élément appelé la "mâchoire", ou "jaw" de l'ARN polymérase de l'Archée est constitué des deux plus grandes s.u. Rpo1 et Rpo2 et de la petite s.u. Rpo5 (correspondante au C-ter de Rpb5), et est très proche de l'ADN promoteur en aval, comme son homologue eucaryote.

Le domaine du "Stalk", que l'on ne retrouve pas dans l'ARN pol bactérienne, mais qui est présent dans les polymérases des archées et des eucaryotes, est composé de deux s.u. : Rpo4 et Rpo7 chez l'Archée. Rpo7 permet la fixation du stalk au reste de la polymérase, et possède également un domaine de liaison à l'ARN. Rpo4 se lie à Rpo7 à l'opposé de ce domaine, et stabilise cette structure. Ainsi, lors de l'élongation, le transcrit naissant sort de l'enzyme et est dirigé vers le stalk, où il interagit avec Rpo7 via des liaisons électrostatiques (Werner, 2013).

Le facteur TFE de l'Archée possède quant à lui deux domaines d'interaction avec l'enzyme. Le domaine "Winged Helix" (WH) en N-ter se lie au Clamp de l'ARN polymérase, tandis que le domaine C-ter (à doigt de Zinc) interagit avec le stalk (Figure 12.B.). TFE permet la fusion de l'ADN et la formation du complexe ouvert OC. Ceci s'effectuerait de façon allostérique, TFE modulant la position du Clamp de l'enzyme. Spt4/5, un complexe hétérodimérique composé des deux facteurs Spt4 et Spt5, se lie au bout du Clamp de l'ARN polymérase, et la majeure partie du complexe est localisée proche du site de liaison à l'ADN. Durant l'élongation, cela

permet de verrouiller l'ADN matrice au sein de l'enzyme, et de prévenir la dissociation du complexe ADN-ARN-Polymérase, augmentant la processivité de la transcription, à savoir la propriété de la polymérase à catalyser des réactions successives en restant associée avec le brin d'ADN transcrit, sans dissociation. Les sites de liaison du facteur archéen d'initiation TFE et d'élongation Spt4/5 convergent dans le Clamp de l'enzyme archéenne.

Concernant les deux derniers facteurs évoqués, TFE et Spt4/5, il est intéressant de noter qu'il existe entre eux une interaction essentielle à l'activité de transcription. En effet, le facteur Spt4/5 peut bloquer l'accès au site d'entrée de l'ADN, lorsque lié à des ARN polymérases libres. Le complexe Spt4/5-polymérase ne peut alors pas engager de transcription sur l'ADN. Dans ce cas, le facteur TFE peut forcer le déplacement de Spt4/5 depuis l'ARN polymérase, par compétition sur les sites de liaison (TFE possède une plus forte affinité que Spt4/5 lors de la formation du PIC via son domaine WH) (Grohmann *et al*, 2011).



**Figure 12. Le cas de l'archée: structure de l'ARN polymérase unique.**

A. Structure de l'ARN polymérase archéenne en représentation de surface, où sont indiqués les principaux éléments caractéristiques de l'enzyme: le site actif, le canal d'insertion de l'ADN, la plateforme d'assemblage, et le Stalk. Les différentes sous-unités colorées sur la structure sont reportées sur la légende de droite. Le zoom en-dessous représente le centre catalytique, où l'on retrouve un ion magnésium, l'ADN, la Bridge Helix et la Trigger loop. Deux autres vues structurales de l'enzyme (face et dos) sont également représentées.

B. Structure de l'ARN pol archéenne selon la même orientation que dans A., où ont été rajoutées les positions structurales qu'adoptent deux facteurs de transcription indiqués en A : TFE et Spt4/5 (*S.shibatae*, pdb : 2WAQ, d'après Grohmann et al, 2011)



## Chapitre V - Les différentes ARN polymérases ADN dépendantes chez les eucaryotes

### V -1- Présentation générale des ARN polymérases chez les eucaryotes

Il est intéressant de souligner la notion de spécialisation parmi les ARN polymérases multimériques. Divers systèmes de transcription ont en effet vu le jour à travers les trois grands règnes du vivant, depuis le système bactérien, où une ARN polymérase basale unique, aidée de facteurs  $\sigma$  diversifiés, permet de reconnaître directement l'ADN, jusqu'à l'archée, qui nécessite la formation d'un complexe de pré-initiation (PIC) sur l'ADN, et donc la présence de facteurs de transcription additionnels. De même, de l'ARN polymérase de l'archée aux ARN polymérases eucaryotes, et en particulier l'ARN pol II, le système se complexifie, avec un plus grand nombre de facteurs généraux de transcription nécessaires. Les ARN polymérases I et III possèdent un nombre assez restreint de promoteurs, là où l'ARN pol II en possède un très grand nombre. Certains facteurs de transcription de l'ARN pol II deviennent des sous-unités à part entière dans le cas des ARN pol I/III, permettant de remplir les fonctions de transcription particulières de ces ARN polymérases.

Dans le cadre de la transcription chez les eucaryotes, nous nous intéressons aux trois polymérases les plus connues et caractérisées ces dernières décennies, à savoir les ARN pol I/II/III, chacune ayant une activité de transcription spécifique. Mais il est intéressant de mentionner qu'il existe d'autres polymérases multimériques eucaryotes, avec des fonctions spécifiques, telles que pol IV et pol V chez les plantes, ou les ARN polymérases spécifiques des mitochondries ou des chloroplastes, que nous présenterons brièvement.

Les ARN polymérases IV et V ont été décrites récemment chez des plantes telles qu'*Arabidopsis thaliana*, l'un des modèles de plantes le plus étudié et caractérisé à ce jour. Ces ARN polymérases spécifiques jouent un rôle dans les modifications de la chromatine médiées

par l'ARN, comme les méthylations de l'ADN (ou RdDM, pour "RNA-directed DNA Methylation"), processus répandus chez les eucaryotes. Les produits de transcription de ces deux polymérases ont cependant des rôles bien distincts dans le processus de RdDM. L'ARN pol IV, dans un premier temps, synthétise des précurseurs d'ARN, maturés ensuite en petits ARNs (smallRNAs), qui permettent l'initiation du RdDM. Pol V agit ensuite, plus en aval dans le processus, en produisant une matrice ARN qui permettra de recruter les petits ARN produits par la pol IV et associés à des protéines Argonautes, afin de diriger le complexe vers les sites de chromatine cibles (Wendte & Pikaard, 2016). Structuralement, ces deux ARN polymérases sont composées de 12 sous-unités, toutes étant homologues ou paralogues aux 12 s.u. de l'ARN pol II, formant ainsi deux formes évoluées et spécialisées de la pol II. Il a cependant été montré ultérieurement par spectrométrie de masse, chez le maïs, que la seule différence fondamentale entre ces deux ARN polymérases résidait dans la plus grande sous-unité des enzymes, qui n'auraient pas tout à fait les mêmes fonctions. Le centre catalytique de ces polymérases est formé par les 2 plus grandes sous-unités, et la deuxième plus grande s.u. est identique chez pol IV et pol V. Les différences d'activité entre les deux polymérases devraient donc s'expliquer par des différences dans la composition en acides aminés de leurs plus grandes sous-unités respectives. Notamment, chez Pol V, la plus grande s.u. est caractérisée par un domaine C-Terminal Domain (CTD) très grand (environ 700 acides aminés), ne possédant pas de similarités de séquences strictes avec celui de l'ARN pol II, et le fait qu'ils aient des fonctions analogues n'est pas expliqué à ce jour (Wendte *et al*, 2017).

On retrouve également une ARN polymérase dans les mitochondries, caractérisée notamment chez l'humain (Arnold *et al*, 2012). Cette enzyme, nommée POLRMT, ou h-mtRNAP, est une enzyme monomérique ADN-dépendante, que l'on peut rapprocher des enzymes monomériques telle que celle du bactériophage T7. Avec l'aide de nombreux facteurs protéiques, cette ARN polymérase est responsable de la transcription du génome

mitochondrial, qui aboutit à la synthèse de deux ARNr et 22 ARNt requis pour l'expression du génome mitochondrial, et la synthèse de 13 sous-unités des complexes de la chaîne respiratoire de la mitochondrie.

Enfin, les gènes chloroplastiques sont transcrits par deux ARN polymérase différentes: une ARN polymérase chloroplastique de type bactérienne, composée de quatre sous-unités  $\alpha$ ,  $\alpha'$ ,  $\beta$  et  $\beta'$  homologues aux quatre s.u. de l'ARN pol de la bactérie ; et une seconde, monomérique et plus proche des ARN polymérase des bactériophages T3 et T7.

Dans ce manuscrit, nous parlerons plus en détails des trois ARN polymérase eucaryotes nucléaires. En forme de fer-à-cheval, ou de pinces de crabe, elles partagent un cœur structural de 10 sous-unités. La structure du cœur de l'ARN pol II est celle qui a été obtenue le plus tôt, permettant de mieux comprendre les mécanismes de la transcription pol II (Cramer *et al*, 2001; Cheung & Cramer, 2012). Structuralement, dans le cœur enzymatique ("Core enzyme"), on retrouve 5 sous-unités communes, présentes dans toutes les ARN polymérase nucléaires (nommées d'après la nomenclature de l'ARN pol II: Rpb5/6/8/10/12), additionnées de cinq sous-unités homologues. Ces dernières sont représentées par les deux plus grandes s.u. Rpa190 et Rpa135 pour la pol I, respectivement Rpb1/Rpc160 et Rpb2/Rpc128 pour pol II/pol III, de l'hétérodimère AC40-AC19, partagé avec la pol III, et homologue à Rpb3-Rpb11 chez la pol II, ainsi que de la sous-unité Rpa12, ou A12.2 (Rpc11 chez pol III, Rpb9 + facteur TFIIIS chez la pol II).

En dehors de ce Core enzyme, on trouve les sous-unités composant l'élément le plus frappant structuralement des polymérase archéenne et eucaryotes, le Stalk, composé de l'hétérodimère Rpa14-Rpa43 (Rpb7-Rpb4 pour pol II, Rpc25-Rpc17 chez pol III), ainsi que l'hétérodimère Rpa34-Rpa49, possédant de nombreuses régions d'homologie avec les facteurs de la pol II

TFIIF et TFIIIE, et les sous-unités Rpc53/37/82/34 de la pol III, ce que nous évaluerons en détails par la suite.

## **V -2- L'activité de transcription par l'ARN pol II, la polymérase la plus étudiée et caractérisée**

### **V-2-a- L'initiation par l'ARN polymérase II et ses facteurs associés: formation du complexe de pré-initiation**

Dans le but d'effectuer le recrutement site-spécifique de l'ARN pol II, et donc d'initier la transcription d'un gène donné, des séquences d'ADN régulatrices, ainsi que des facteurs de transcription spécifiques sont requis, que nous présenterons dans cette partie.

#### ***Les séquences d'ADN importantes pour la formation du complexe de pré-initiation pour la transcription par l'ARN pol II***

Les séquences ADN des promoteurs de l'ARN pol II à proprement parler sont très variables, avec une grande hétérogénéité en taille et composition. Cependant, certains éléments sont fréquemment conservés. Brièvement, des séquences de gènes codant pour des protéines ont été fusionnées à des gènes rapporteurs, permettant une mesure simple et efficace de l'effet de ces séquences régulatrices clonées sur la transcription. Les régions actives ont pu être identifiées par la suite, et des expériences de mutagenèse *in vitro* ont permis de déterminer précisément le rôle de séquences spécifiques au sein de ces régions clonées (Zenke *et al*, 1986). Ainsi, un élément clé, appelé le "Core promoter", en 5' de la séquence codante, a été révélée. Il est composé de séquences d'ADN conservées (environ 60 nucléotides entourant le site d'initiation de la transcription), cruciales pour la fonction du promoteur, c'est-à-dire l'assemblage et l'orientation corrects du PIC, impactant directement l'efficacité de la transcription.

On retrouve ainsi au moins quatre éléments dans ce Core promoter. Premièrement, la "TATA-box", une séquence riche en A/T localisée environ 26 à 31 pb en amont du "Transcription Start Site" (TSS). On trouve également une séquence contenant une région riche en pyrimidines, entourant le TSS: l'initiateur (Inr). Une troisième région localisée 28 à 32 nt en aval du TSS est appelée DPE, pour "Downstream Promoter Element". Enfin, un élément de reconnaissance du facteur TFIIB, situé de 32 à 37 nt en amont du TSS, complète le Core promoter (BRE: TFIIB Recognition Element) (Juven-Gershon & Kadonaga, 2010). D'autres motifs, tels que le MTE ("Motif Ten Element"), peuvent également être retrouvés dans cette région, et peuvent promouvoir la transcription (Figure 13.A.).

Outre la région particulière du Core Promoter, d'autres régions régulatrices de la transcription sont retrouvées au niveau des régions 3' et 5', les "enhancers", ou régions activatrices de la transcription. Chez la levure, les enhancers sont relativement proches (500 pb du TSS); ils peuvent être situés à des méga-bases du TSS dans les génomes métazoaires. Ils peuvent contenir jusqu'à 10 sites reconnus par des facteurs "spécifiques" de transcription, c'est-à-dire n'agissant que sur une sous-classe de promoteur.

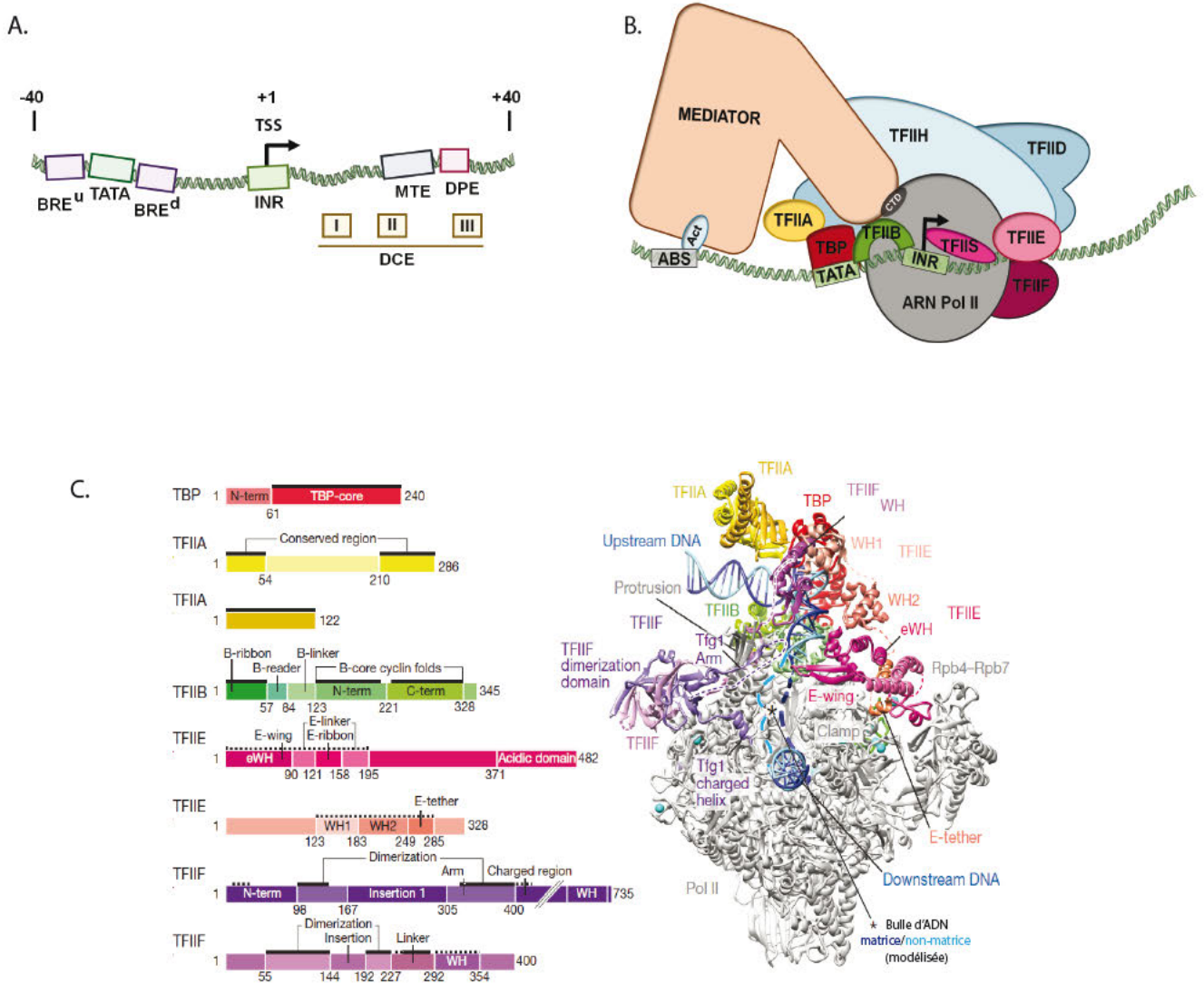
Ainsi, un gène eucaryote transcrit par l'ARN pol II contiendra plusieurs éléments "enhancers", qui pourront être localisés en 3' ou 5'. Ces séquences dites cis-régulatrices sont d'une importance capitale pour le contrôle de l'expression des gènes, et sont le lieu de fixation spécifiques des facteurs de transcription (TFs, pour "Transcription Factors"), permettant ensuite le recrutement de la machinerie de transcription au niveau du "Core Promoter".

### Les facteurs généraux de transcription de la pol II, structures et rôles

Les facteurs généraux de la transcription (GTFs) nous intéresseront plus particulièrement. Ils sont en effet requis pour la transcription de tous les gènes pol II. Ces facteurs généraux et l'ARN pol II sont requis au niveau du Core Promoter pour constituer le PIC, et on trouve

classiquement 6 GTFs associés dans le PIC (Figure 13.B. ; à noter que dans cette schématisation, on retrouve l'ensemble des GTFs. La structure résolue à ce jour du PIC a été déterminée avec un nombre plus restreint de GTFs (Plaschka *et al*, 2016) (Figure 13.C.)).

La formation du PIC s'effectue via l'interaction d'un complexe de grande taille, TFIID, (env. 1 MDa) sur le Core Promoter. Il est composé de la "TATA binding protein" (TBP) et de 13 "TBP Associated Factors", les TAFs (TAFs 1-13), qui concilient les interactions spécifiques avec d'autres facteurs du PIC, faisant de TFIID le facteur de reconnaissance primaire du Core Promoter. La TBP se lie notamment à la TATA box, et plusieurs TAFs se fixent sur des séquences du promoteur en aval de la TATA box: TAF1 et 2 sur l'initiateur (Inr), TAF 6 et 9 au niveau du DPE (Juven-Gershon & Kadonaga, 2010). La structure d'un complexe TFIID en liaison avec son promoteur a été obtenue récemment (Louder *et al*, 2016). Lorsque TFIID n'est pas présent, un autre complexe protéique contenant également la TBP, le complexe SAGA, prend le relais. Notamment, chez la levure, il est intéressant de noter que seulement 20% environ des promoteurs contiennent une TATA box. Or, les gènes qui en contiennent sont souvent associés à des réponses au stress, et sont fortement régulés. On remarque qu'ils utilisent alors préférentiellement SAGA plutôt que TFIID, comparés aux promoteurs qui ne contiennent pas de TATA box (Basehoar *et al*, 2004).



**Figure 13. L'ARN polymérase II chez *Saccharomyces cerevisiae* et ses facteurs de transcription.**

A. Représentation schématique du «Core promoter» de l'ARN pol II. Les différentes séquences ADN importantes pour l'initiation de la transcription sont indiquées. «BRE = TFIIB Recognition Element ; INR = Initiator ; TSS = Transcription Start Site ; MTE = Motif Ten Element ; DPE = Downstream Promoter Element ; DCE = Downstream Core Element».

B. Schéma simplifié de la machinerie de transcription de l'ARN pol II avec ses facteurs de transcription associés. Il est à noter que les différents facteurs n'agissent pas au même moment durant la transcription, cette représentation globale est donc simplifiée. Le complexe SAGA n'est pas indiqué dans ce schéma. Le médiateur, ou «Mediator», est un complexe multi-protéique qui interagit avec l'enzyme (notamment via le CTD), différents GTFs et des protéines activatrices de la transcription ( Act = protéine Activatrice ; ABS = Activator Binding Site, site de liaison des protéines activatrices, aussi appelé «Enhancer»).

C. Structure 3D (Plaschka et al, 2016) de l'ARN pol II de *S.c.* (en gris) en complexe dit ouvert («Open complex, OC»), en interaction avec les facteurs de transcription de base TBP, TFIIA, TFIIB, TFIIE et TFIIF. La bulle d'ADN (matrice et non-matrice) est également représentée. Sur le panel coloré de gauche sont reportés les différents facteurs placés sur la structure, et les domaines présents dans chaque sous-unité sont indiqués.

La formation de ce complexe TFIID sur le Core promoter amène au recrutement d'autres GTFs: TFIIA et TFIIB. Ils interagissent avec TFIID, stabilisent sa fixation à l'ADN et induisent le recrutement de l'ARN pol II hypo-phosphorylée via le domaine C-terminal de sa plus grande sous-unité Rpb1 (le CTD, "Carboxy-Terminal Domain"). TFIIB interagit notamment avec l'élément "Inr" ainsi qu'avec la TBP, et l'assiste dans le recrutement de la polymérase II au Core Promoter.

Dans le même temps, TFIIF, un autre facteur important pour le positionnement et la rétention de l'ADN, entre en jeu. TFIIF adopte un repliement complexe. Il est composé de deux modules distincts (Tfg1 et Tfg2), reliés entre eux par un "linker". Son module de dimérisation, ainsi qu'une hélice chargée, se placent au niveau du domaine du "Lobe" de la pol II, tandis qu'un bras de la plus large sous-unité Tfg1 se projette dans le Cleft de l'enzyme, où il stabiliserait l'ADN. Le linker dans Tfg2 émane du module de dimérisation, serpente le long de la base du domaine "protrusion" de la pol II et de TFIIB, en stabilisant ce dernier sur l'ARN pol II (en violet, Figure 13.C).

Enfin, deux autres facteurs entrent en jeu, TFIIE et TFIIH. Ils induisent notamment une phosphorylation de l'ARN pol II au niveau du domaine CTD. Le domaine CTD joue un rôle important dans toutes les étapes de la biogénèse des ARNm : l'ajout de la coiffe en 5', l'épissage dans le corps du transcrit et la maturation en 3' des précurseurs des ARNm. Il est conservé chez les eucaryotes, et est composé de répétitions (26 chez *S.cerevisiae*, 52 chez les mammifères) d'un heptapeptide de séquence consensus YSPTSPS (Millhouse & Manley, 2005). Cet heptapeptide contient des résidus qui peuvent être modifiés chimiquement, et notamment phosphorylables sur ses résidus Sérines. Au moins deux formes d'ARN pol II peuvent être retrouvées: une forme hypo-phosphorylée, qui permet l'ancrage de l'enzyme au complexe de pré-initiation en formation, et une forme hyper-phosphorylée, qui sera associée aux facteurs d'élongation de la transcription. Brièvement, la phosphorylation sur la sérine 5



est impliquée en initiation de la transcription, et celle de sur la sérine 2 en élongation, mais les autres résidus peuvent également être modifiés (hydroxy-tyrosines, etc...), dont les rôles et interactions exacts sont encore à caractériser. Notamment, la phosphorylation de la sérine 7 a été montré comme étant importante pour la transcripion des snARNs (Egloff *et al*, 2007).

Le facteur TFIIE est composé de 2 domaines: TFIIE-alpha (TFIIE $\alpha$ ) et TFIIE-béta (TFIIE $\beta$ ) chez l'humain (appelés Tfa1 et Tfa2 chez *S.c.*). TFIIE $\alpha$  contient notamment un domaine "Winged-Helix" (WH) en N-ter, un domaine à doigt de Zn, et un domaine regroupant des acides aminés acides en C-ter, tous trois étant essentiels à sa fonction dans la transcription. Plus en détails, les domaines WH et à doigt de Zn sont suffisants pour l'interaction avec TFIIE $\beta$  (à noter que la perturbation de cette interaction provoque de sévères défauts de croissance, chez la levure (Miwa *et al*, 2016), tandis que le domaine en C-ter recrute le facteur TFIIH pour compléter la formation du PIC. TFIIE $\beta$  contient quant à lui deux domaines WH, suivis par une région dite "Coiled-Coil" (ou superhélice, motif structural protéique dans lequel 2 à 7 hélices  $\alpha$  s'enroulent ensemble les unes autour des autres) ainsi qu'une dernière région appelée "basic helix" en C-ter, qui peut se lier à des GTFs et à l'ADN. Des études ont également suggéré que TFIIE aurait un site de fixation proche du domaine "Clamp" de Rpb1, la plus grande sous-unité de la pol II. TFIIE est un GTF qui interagirait avec l'ADN dans et juste en amont de la séparation des deux brins d'ADN qui forme une région d'ADN dite "déroulée" : la bulle de transcription (Miwa *et al*, 2016).

Des protéines ayant de l'homologie avec TFIIB, TFIIE et TFIIF sont présentes dans le cas des ARN pol I et III, ainsi que des homologues pour TBP, TFIIB et TFIIE dans le cas de la polymérase archéenne (Plaschka *et al*, 2016).

Enfin, TFIIH est un facteur connu pour être impliqué dans la transcription par l'ARN pol II, mais aussi dans la réparation de l'ADN par excision de nucléotides ("Nucléotides Excision

Repair" ou NER), permettant notamment la réparation des lésions dues aux ultraviolets (UV) (Compe & Egly, 2012). Dans la transcription, il possède plusieurs rôles, dont le déroulement de l'ADN, l'analyse du site d'initiation (TSS), le départ de la polymérase du promoteur et la phosphorylation du CTD de l'ARN pol II. Ce facteur hétéromultimérique de grande taille (510 kDa) est composé de 10 sous-unités, et possède notamment une double activité hélicase dépendante de l'ADN (5'=>3' et 3'=>5'), associées aux deux protéines Rad3 et Ssl2 (ou Rad25) chez *S.cerevisiae* (XPD et XPB respectivement chez l'Homme, dont les mutations sont à l'origine du Xeroderma pigmentosum). Seule Ssl2 est requise pour l'activité de déroulement de l'ADN dans la transcription, là où l'activité hélicase de Rad3 est requise dans le NER. Rad3 participe néanmoins à la stabilisation globale du complexe TFIIH, jouant donc un rôle important dans la transcription. Ssl2 est une translocase ATP-dépendante de l'ADN double brin. Elle permet l'insertion de l'ADN matrice au sein du site actif, entraînant le déroulement de l'ADN en aval du promoteur et aboutissant à la formation d'un complexe "ouvert" (OC pour "Open Complex") (Fishburn *et al*, 2015). L'initiation de la transcription par l'ARN pol II commençant en effet en aval du site de formation du PIC, Ssl2 permet de balayer l'ADN en aval du promoteur afin de trouver un site TSS approprié pour initier la transcription (Fazal *et al*, 2015).

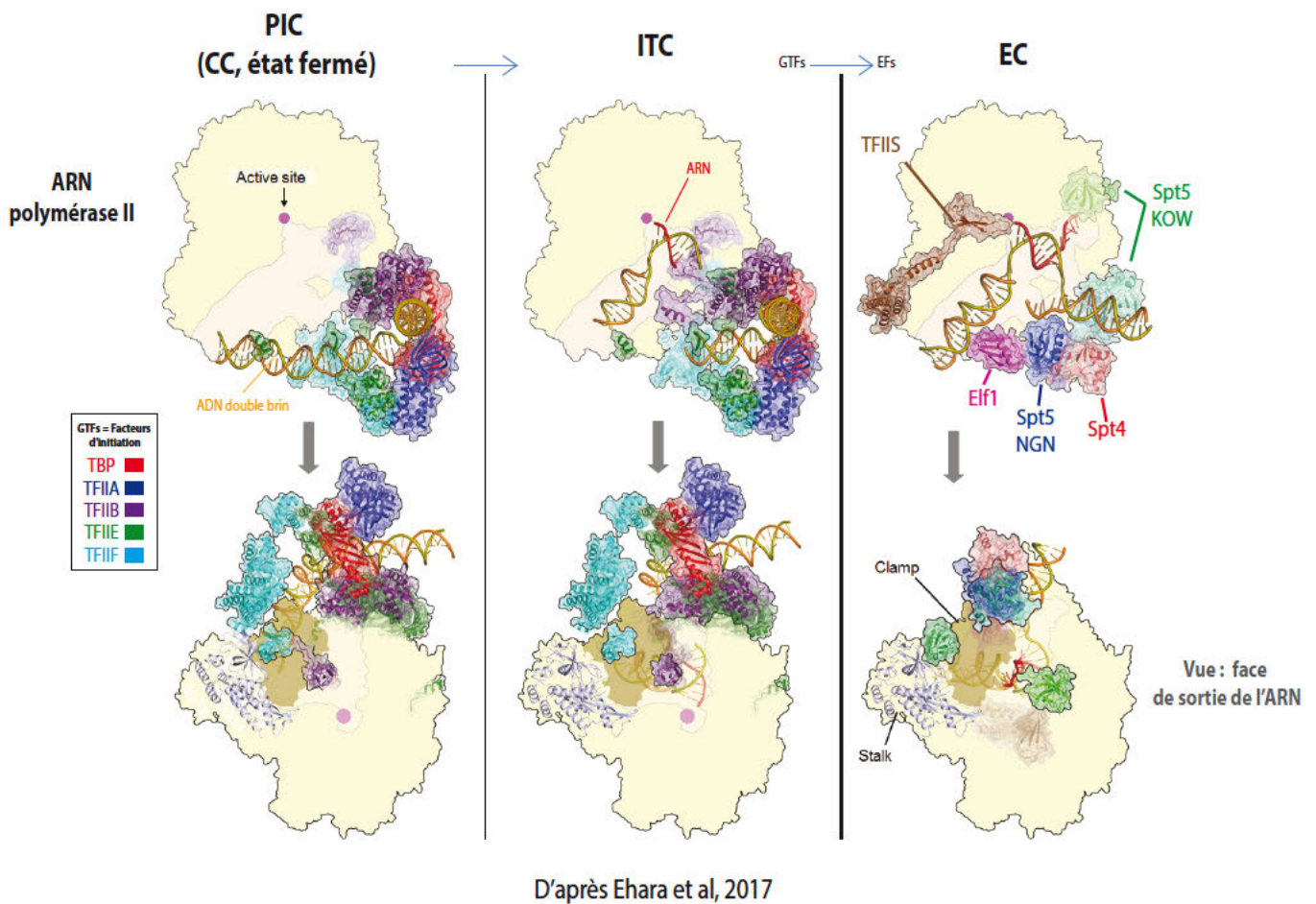
Les facteurs TFIIIE et TFIIH sont les deux facteurs basaux qui arrivent le plus tard au niveau du PIC. Il est important de mentionner que certains promoteurs ne nécessitent pas la présence du facteur TFIIH.

Dans le cadre de l'initiation pol II, nous devons également mentionner l'existence d'un complexe protéique, le médiateur ("Mediator"), découvert associé à l'ARN pol II dès les années 1990. Ce complexe fut historiquement découvert via l'étude de protéines particulières qui permettaient de réguler l'activité de l'ARN pol II en se liant au CTD de l'enzyme, les SRB ("Suppressors of RNA polymerase B") (Thompson *et al*, 1993), qui ont toutes par la suite été

incluses dans le médiateur. Ce complexe joue un rôle central dans la voie d'activation de la transcription pol II: enhancer => activateur => Médiateur => Pol II => promoteur. Le médiateur est un complexe massif de plus de 20 protéines requis pour la régulation de la transcription par l'ARN pol II, retrouvé chez la levure comme chez les mammifères (Kornberg, 2005). Il interagit avec les protéines activatrices de la transcription, mais aussi directement avec l'ARN pol II en complexe de pré-initiation avec les facteurs de transcription généraux TFIIA, -B, -D, -E, -F, -H et -S. Un cycle de phosphorylations / déphosphorylations du CTD via un sous complexe de TFIIH (= TFIIK) régule son interaction avec la pol II. Une structure d'un complexe PIC - médiateur (médiateur de 52 protéines, 2,5 MDa) a d'ailleurs été obtenue récemment (Robinson *et al*, 2016).

Une fois le complexe de pré-initiation formé sur les séquences du Core promoter, plusieurs protéines régulatrices telles que des activateurs transcriptionnels, co-activateurs, modificateurs de la chromatine et des facteurs d'élongation, sont recrutés à cet endroit. Cette étape est importante car elle permet la formation de boucles permettant aux enhanceurs plus distants d'interagir avec cette région promotrice et de jouer leur rôle de régulation. Ainsi, l'assemblage de l'enzyme et de ses facteurs de transcription généraux forme un complexe dit "fermé", (CC, pour "Closed Complex"), dans lequel l'ADN est positionné et retenu au niveau du Cleft de la pol II. Les deux brins d'ADN s'ouvrent alors, et le brin matrice rentre au niveau du site actif de l'enzyme, formant une "bulle" d'environ 20-30pb, et menant au complexe appelé complexe ouvert, ou "Open Complex (OC)". Le site d'initiation de la transcription, le TSS ("Transcription Start Site"), est alors reconnu, et la synthèse d'ARN peut commencer, amenant à la formation du complexe ITC ("Initially Transcribing Complex") (Plaschka *et al*, 2016). Une fois que la longueur de l'ARN synthétisé est suffisamment importante, les facteurs d'initiation se dissocient pour laisser place au complexe d'élongation proprement dit (EC, "Elongating Complex"). Les facteurs d'initiation sont alors échangés avec des facteurs

d'élongation pour permettre cette transition et une transcription processive. Ces différentes formes structurales successives sont résumées dans la Figure 14.



**Figure 14. Schéma structural récapitulatif reprenant les différents complexes que forme l'ARN polymérase II en interaction avec ses facteurs de transcription durant les différentes phases du cycle transcriptionnel.** Sont indiqués les facteurs d'initiation TBP, TFIIA, TFIIB, TFIIE, TFIIIF (voir panel couleur à gauche), les facteurs d'élongation Elf1, Spt4, Spt5-NGN / Spt5-KOW (NGN et KOW = domaines de Spt5), TFIIS, l'ADN, l'ARN, et le site actif (point violet). Les trois structures en-dessous sont les mêmes que celles du dessus respectivement, vues par le côté de sortie de l'ARN. De gauche à droite: Complexe de pré-initiation (PIC) fermé, ou «Close complex, CC» (pdb 5FZ5, Plaschka et al, 2016) ; complexe initiateur de la transcription («Initially Transcribing Complex, ITC») (pdb : 4V1N, Plaschka et al, 2014) ; complexe en élongation, avec les facteurs d'élongation de la transcription (Ehara et al, 2017). Le passage des complexes d'initiation (PIC, ITC) au complexe d'élongation se fait via l'échange des facteurs d'initiation (GTFs) par les facteurs d'élongation (EFs).

## V-2-b- Elongation, "Backtrack" et terminaison de la transcription par la pol II

Le passage du complexe fermé au complexe ouvert au niveau du promoteur est effectué via la séparation des deux brins d'ADN, pour former la bulle de transcription, et le positionnement de l'ADN simple brin au niveau du centre catalytique de l'ARN pol II, accompagné de l'échange des facteurs d'initiation pour les facteurs d'élongation. Ceci permet le début de la synthèse de l'ARN à partir du TSS. Dans cette partie, je traiterai des réarrangements de certains éléments de l'enzyme lors de cette étape de la transcription, ainsi que de certains facteurs basaux d'élongation, les facteurs Spt4/5 et TFIIS .

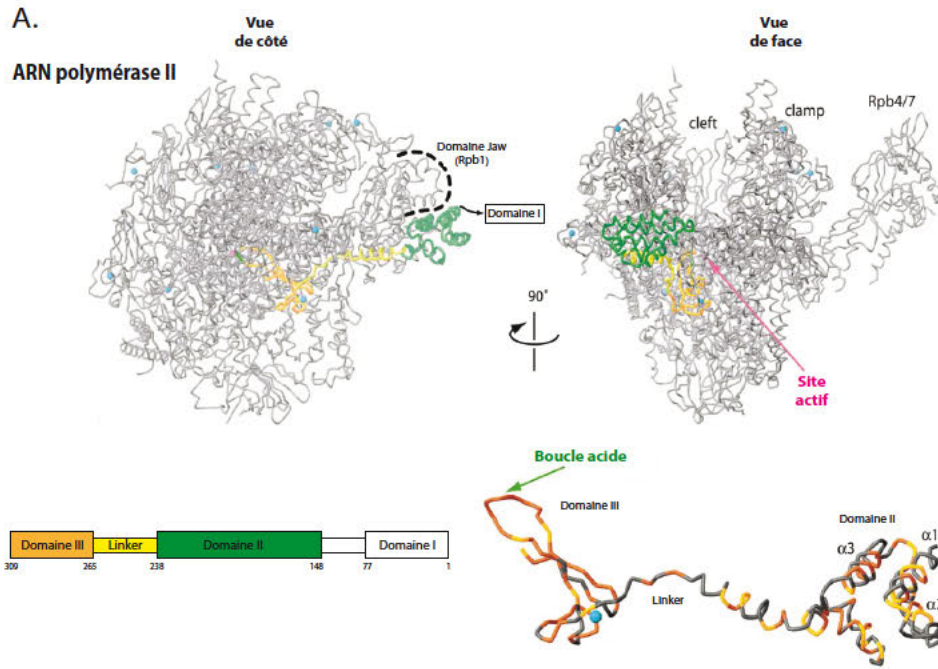
Le complexe Spt4/5, est responsable de l'augmentation significative de la processivité de l'ARN pol II. Spt4 est une protéine connue pour stabiliser le domaine NGN de Spt5. Spt5 est une protéine à plusieurs domaines, composée donc du domaine conservé NGN, de différents domaines KOW connectés par des "linkers" flexibles, et d'une région répétée en C-terminal (CTR) (Ponting, 2002). Ce complexe est lié à l'enzyme au niveau de son domaine "Clamp", et est juste adjacent au brin d'ADN non-transcrit dans la bulle de transcription. Spt4/5 ne peut se lier à son site de fixation que lorsque le complexe transcriptionnel n'est plus au niveau du promoteur. Le domaine N-terminal de Spt5 ferme le Cleft autour du site actif, verrouillant la position de l'ADN à l'intérieur, et augmentant la processivité de l'enzyme. Il est également intéressant de rappeler que des études de FRET (Förster Resonance Energy Transfer) chez l'archée ont montré que le facteur d'initiation TFE, (l'homologue de TFIIE $\alpha$  de l'ARN pol II) et Spt4/5 étaient en compétition pour la liaison au Clamp de la polymérase dans la transition de l'initiation à l'élongation de la transcription archéenne (Grohmann *et al*, 2011), de même que TFIIE et Spt4/5, montrant une certaine conservation évolutive de ces éléments.

On peut ici mentionner à nouveau les deux éléments structuraux importants pour le maintien de la liaison acides nucléiques - protéine pendant l'élongation de la transcription pol II,

notamment durant le cycle d'addition des nucléotides ("NAC"): la "Trigger loop" et la "Bridge helix". Le premier élément contrôle l'ouverture/fermeture du site actif, en coopération avec la libération du pyrophosphate lors de la réaction d'ajout d'un nucléotide. Ainsi, la Trigger loop se replie en élongation de façon à fermer le centre catalytique, permettant aux NTPs de légèrement se déplacer pour occuper leur site d'insertion, et interagit avec la Bridge Helix.

Aidée de la Trigger Loop, la Bridge Helix joue un rôle dans le déplacement de l'hybride ADN-ARN naissant en dehors du site actif. Elle permet notamment la translocation du prochain nucléotide de la matrice ADN dans une position de pré-transcription. S'ensuit une seconde étape de relâchement de cet élément structural, permettant à ce nucléotide substrat de rentrer au sein du centre catalytique (Cheung & Cramer, 2012).

Comme expliqué dans le chapitre III-3 dans le cadre l'ARN polymérase bactérienne, l'ARN pol II peut également adopter des état dits "inactifs" en transcription. Nous parlerons notamment ici de TFIIS, facteur qui permet de réactiver un complexe d'élongation en arrêt en stimulant le clivage de l'ARN (Kettenberger *et al*, 2003) (Figure 15.A.). L'ARN pol II possède une activité de clivage intrinsèque faible, qui requiert la sous-unité Rpb9 (Walmacq *et al*, 2009), et qui est fortement augmentée en présence du facteur TFIIS (Lisica *et al*, 2016). Le clivage de l'ARN est catalysé par une activité d'hydrolyse intrinsèque au site actif de l'ARN pol II, stimulé par l'extrémité 3' de l'ARN. TFIIS quant à lui possède une activité en deux temps. Premièrement, il se lie à l'ARN pol II au niveau de son domaine "funnel", en y causant un changement conformationnel qui repositionne l'ARN qui a effectué le "Backtrack". Puis, TFIIS s'insère dans le "pore" de l'enzyme proche de cet ARN et atteint le site actif avec une structure en épingle à cheveux (ou "hairpin") qui contient 3 résidus chargés qui stimulent le clivage de l'ARN par l'enzyme. Cela génère alors une nouvelle extrémité 3' de l'ARN, et permet le redémarrage de l'élongation de la transcription.



**Figure 15. Données structurales sur le facteur de stimulation du clivage de l'ARN pol II, TFIIIS.**

A. Structure du facteur d'élongation de l'ARN pol II, TFIIIS, stimulateur de l'activité de clivage, et son insertion dans l'enzyme. Trois domaines (I, II et III) composent ce facteur, avec un long linker entre les domaines II et III, comme indiqué sur le panel du dessous. Sur le panel du dessus est représentée l'insertion de TFIIIS dans la polymérase: le linker (jaune) s'étend depuis le domaine II (vert) le long de la surface de la polymérase (grise), prolongé par le domaine III (orange) qui s'insère dans l'ARN pol II jusqu'au site actif (point rose). Le domaine I en N-terminal, faiblement conservé, forme un paquet de 4 hélices, et n'est pas requis pour l'activité de TFIIIS. D'après Kettenberger et al, 2003



### V-3- L'ARN polymérase III

#### V-3-a- Activité de transcription et facteurs associés

L'ARN pol III est responsable de la synthèse de la totalité des ARN de transfert (ARNt), du snARN U6, de l'ARNr 5S et de nombreux autres petits ARNs qui n'excèdent pas 400 nt. (Dieci *et al*, 2007). Cette activité transcriptionnelle nécessite un petit nombre de facteurs généraux de transcription, moins nombreux que pour l'ARN pol II, qui sont spécifiques des différents types de promoteurs que l'enzyme peut rencontrer. Ainsi, chez *Saccharomyces cerevisiae*, le promoteur de l'ARNr 5S nécessite la présence des facteurs TFIIIA, TFIIIB et TFIIIC, là où les deux autres types de promoteurs n'ont besoin que de TFIIIB et C, au moins *in vitro* (Hoffmann *et al*, 2016). La TBP, essentielle à toute formation d'un complexe de pré-initiation interagit avec le sous-complexe Brf1, une protéine TFIIIB-like. Ces éléments sont des composants du facteur TFIIIB de l'ARN pol III.

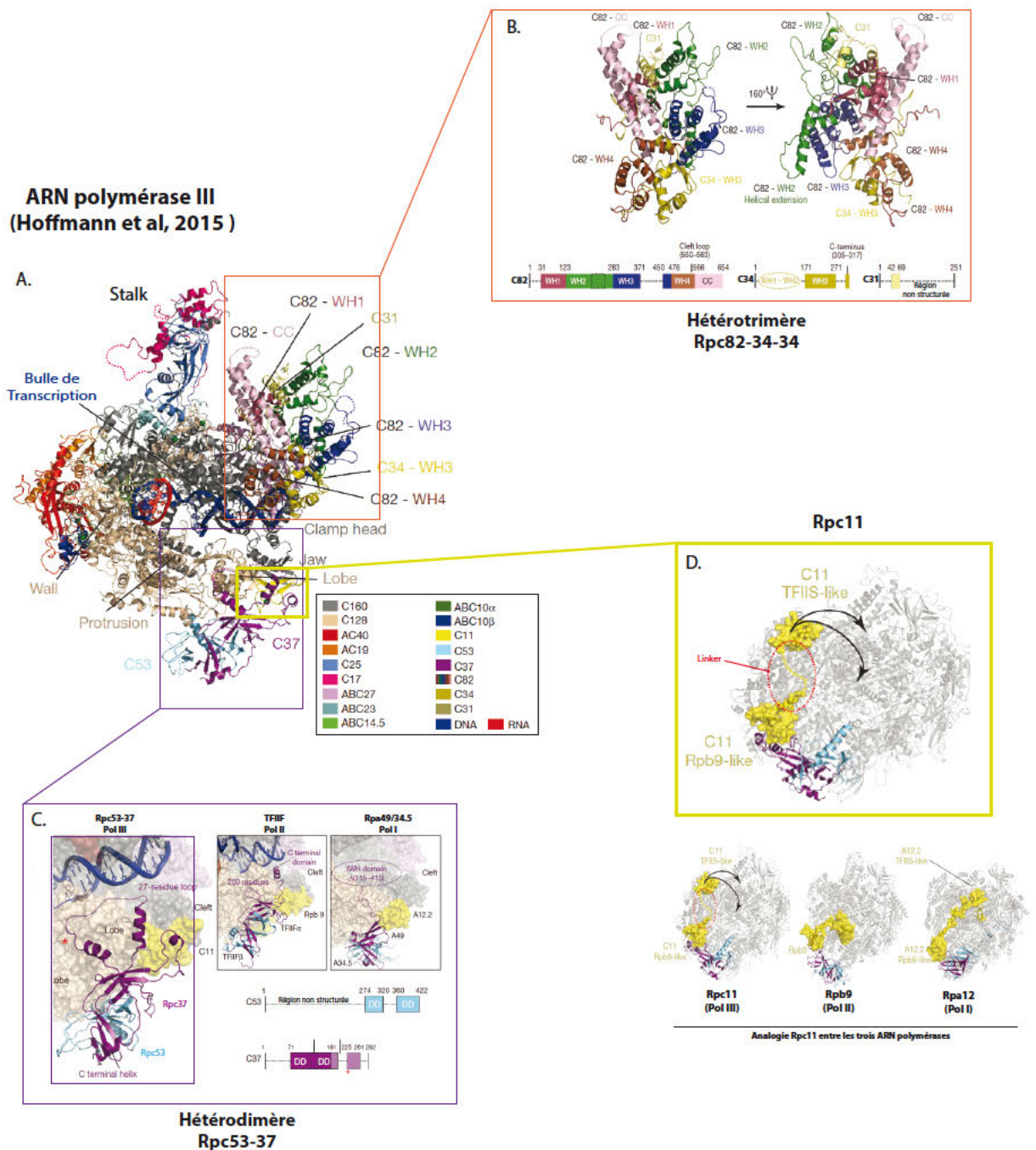
#### V-3-b- Données structurales

La structure en Cryo-microscopie électronique à transmission (Cryo-MET) de l'enzyme entière a été obtenue à une résolution allant jusqu'à 3.9Å (Hoffmann *et al*, 2015) (Figure 16.A.). Composée de 17 sous-unités, pour un poids moléculaire de 0.7 MDa, elle est donc la plus grande ARN polymérase multimérique eucaryote caractérisée jusqu'à présent. L'architecture de son cœur enzymatique de 10 sous-unités est conservée avec toutes les polymérases eucaryotes. On retrouve ainsi les deux plus grandes s.u. Rpc 160 et Rpc128, qui contiennent le site actif, cinq s.u. additionnelles du cœur partagées entre les trois polymérases (ABC27/23/14.5/10 $\alpha$  et ABC10 $\beta$ ), les s.u. AC19 et AC40 communes pour les ARN pol I et III, et homologues à Rpb11 et Rpb3 pour l'ARN pol II, et enfin Rpc11, impliquée dans la terminaison, et homologue à Rpa12 chez l'ARN pol I. Notamment, le facteur pol II impliqué dans la stimulation du clivage, TFIIIS, est homologue à la partie C-terminale de la sous-unité

Rpc11, elle-même connue pour être mobile durant la transcription. La partie N-ter de Rpc11 est comparable à la sous-unité Rpb9 de l'ARN pol II (Figure 16.D.).

Les s.u. Rpc17 et Rpc25, impliquées dans l'initiation et dans la liaison à l'ARN, composent quant à elles le stalk de l'ARN pol III.

Enfin, on remarque que les sous-unités restantes peuvent être comparées à des GTFs de l'ARN pol II, TFIIE/TFIIF. Ainsi, l'hétérodimère Rpc53-37 (Figure 16.C.), impliqué dans l'initiation et la terminaison (Landrieux *et al*, 2006), est similaire, en termes de structures secondaires et tridimensionnelles (mais pas en termes de séquences primaires), au facteur TFIIF de l'ARN pol II, et à l'hétérodimère Rpa49-34 de l'ARN pol I. Cependant, les fonctions exactes dans la transcription de chaque facteur TFIIF-like pour chaque type de polymérase pourraient être différentes (Wu *et al*, 2011). On notera que des interactions existent entre Rpc37 et Rpc11, suggérant un rôle de Rpc37 dans l'association de Rpc11 à la polymérase, comme dans les cas des ARN pol II et pol I (avec TFIIF $\alpha$  et Rpa49-Nter respectivement). L'hétérotrimère Rpc82-34-31 (Figure 16.B.) est retrouvé très proche du stalk de l'enzyme, et est également impliqué dans le processus d'initiation (Vannini & Cramer, 2012) et dans le recrutement de l'ARN pol III au niveau de TFIIB (Brun *et al*, 1997). En particulier, Rpc82 et Rpc34 sont similaires à des parties de TFIIE de l'ARN pol II (TFIIE $\alpha$  et TFIIE $\beta$  respectivement), et du Cter de Rpa49 pour l'ARN pol I: Rpc34 possède deux domaines WH (Winged Helix) en N-terminal qui ne sont pas visibles dans la densité de Cryo-MET de l'ARN pol III, comme c'est le cas également pour le domaine tWH (tandemWH) du C-ter de Rpa49 dans la structure cristalline de l'ARN pol I (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013). Des adaptations fonctionnelles ont donc eu lieu: les sous-unités de type TFIIE-/TFIIF-like sont incorporées dans l'enzyme, là où ces facteurs sont recrutés transitoirement au moment de l'assemblage du PIC dans le cas de l'ARN pol II.



**Figure 16. Données structurales sur l'ARN polymérase III de *S.cerevisiae*.**

A. L'ARN pol III en élongation dans sa totalité, avec les différentes sous-unités formant l'enzyme (code couleur selon l'encart en bas à droite) et les acides nucléiques (ADN en bleu, ARN en rouge) (Hoffmann et al, 2015 ; pdb: 5FJ8). Les modules spécifiques dont les sous-unités sont indiquées sur la structure globale sont détaillés dans les encarts B.C. et D. WH = Winged Helix domain.

B. Architecture et composition de l'hétérotrimère spécifique Rpc82-34-31. Deux vues structurales sont présentées, en rotation de 160° l'une par rapport à l'autre. Sur le panel du bas est également indiquée la représentation schématique des trois sous-unités, et les domaines les composant.

C. L'hétérodimère spécifique Rpc53-Rpc37. Comme en B., les données de structure sont complétées par la représentation schématique des domaines de chaque sous-unité. A gauche est présenté l'hétérodimère de la pol III, qui peut être comparé aux deux panels de droite représentant les équivalents chez la pol II (TFIIIF) et la pol I (hétérodimère Rpa49-Rpa34). L'astérisque rouge indique 5 acides aminés dont l'absence provoque un phénotype de transcription à travers un terminateur (Wu et al, 2011). DD = Domaine de Dimérisation.

D. Structure de la sous-unité mobile Rpc11 (jaune). Les deux parties de RPB11, N- et C-ter, sont reliées par un linker (en rouge). Sur le panel du bas sont présentées les sous-unités équivalentes à Rpb11 chez l'ARN pol II (Rpb9 + facteur de transcription TFIIIS) et l'ARN pol I (Rpa12, ou A12.2)

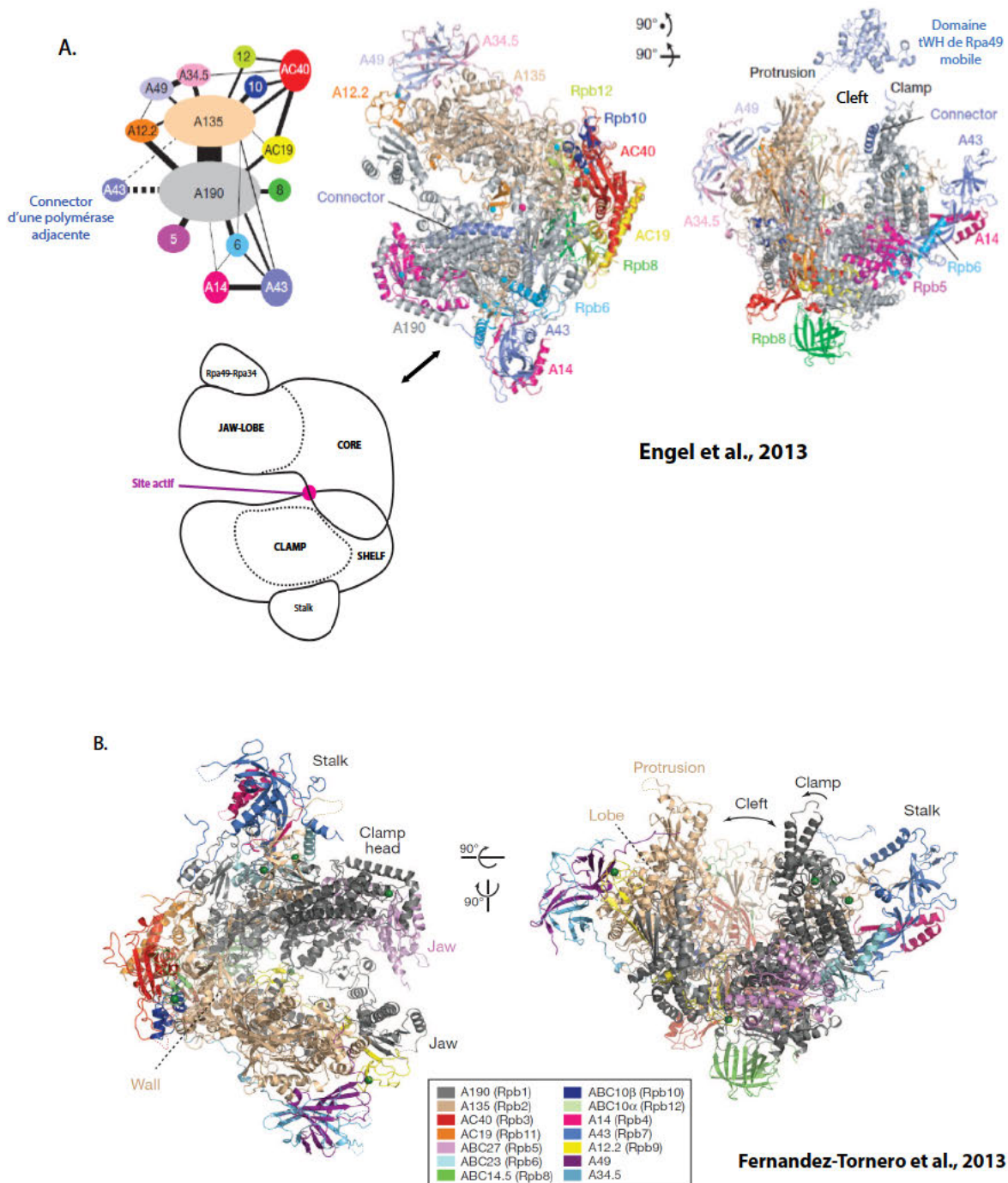
## **Chapitre VI - L'ARN polymérase I, responsable de la synthèse du précurseur des grands ARN ribosomiques : éléments fonctionnels et structuraux de la transcription de l'ADN ribosomique.**

C'est chez l'organisme modèle *Saccharomyces cerevisiae*, modèle utilisé tout au long de mes travaux de thèse, que l'ARN polymérase la plus étudiée. C'est une enzyme de 589 kDa, composée de 14 sous-unités, alors que l'ARN pol II et l'ARN pol III en possèdent 12 et 17 respectivement. Elle est peu sensible à l' $\alpha$ -amanitine avec 50% d'inhibition pour des concentrations de l'ordre de 600  $\mu\text{g/ml}$  d' $\alpha$ -amanitine, alors que 50% d'inhibition de la pol II est atteinte avec à peine 1.0  $\mu\text{g/ml}$  d' $\alpha$ -amanitine (Schultz & Hall, 1976).

Nous allons dans cette partie présenter en détails les éléments structuraux composant cette enzyme particulière, ainsi que les fonctions qui leurs sont associées. La forme et l'architecture globales de l'enzyme ont été révélées il y a plus de 20 ans par microscopie électronique (Schultz *et al*, 1993; Kuhn *et al*, 2007), et seules des reconstructions à basse résolution en MET étaient alors disponibles (Kuhn *et al*, 2007). La structure 3D de l'ARN pol I a été largement étudiée au cours de la dernière décennie, et a notamment été obtenue à haute résolution (de l'ordre de 3Å) en cristallographie aux Rayons X, par deux équipes, donnant lieu à deux publications de référence dans le domaine (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013) (Figure 17.A et B.). Suite à ces deux travaux, plusieurs études récentes ont permis une étude structure-fonction détaillée pour cette enzyme, qui constitue la ligne directrice de mes travaux de thèse (Pilsel *et al*, 2016; Engel *et al*, 2016; Tafur *et al*, 2016; Neyer *et al*, 2016; Han *et al*, 2017; Engel *et al*, 2017; Torreira *et al*, 2017; Sadian *et al*, 2017).

Certains composants de l'enzyme sont communs ou homologues à d'autres sous-unités/facteurs de transcription des autres ARN polymérases eucaryotes ou même archéenne

et bactérienne, comparaisons que nous ferons souvent. Certains autres éléments de la pol I sont dits "spécifiques", et nous les détaillerons plus amplement par la suite.



**Figure 17. Structures 3D à haute résolution de l'ARN pol I de *S.cerevisiae*, selon deux études parallèles.**  
 A. Structure obtenue par Engel et al., 2013. A gauche : schéma des sous-unités de l'ARN pol I, colorées différemment, et des interactions entre ces sous-unités. A droite : deux orientations de la structure entière de l'enzyme. Le code couleur est le même que pour le schéma. Le nom des sous-unités est indiqué directement sur la structure. Certains éléments caractéristiques sont mentionnés : le Connector de Rpa43, le Clamp, le domaine Protrusion, et le Cleft. Le domaine mobile tWH de Rpa49 (Cter), cristallisé à part, a été représenté sur l'orientation de gauche, mais sa position exacte reste inconnue (le linker la reliant au Nter de Rpa49 est représenté en pointillés). En bas: représentation schématique des quatre modules mobiles principaux de l'ARN pol I : Core, Shelf, Jaw-Lobe et Clamp, autour du site actif (rose), additionné des deux hétérodimères périphériques (Stalk = Rpa14-Rpa43) et Rpa49-Rpa34.  
 B. Structure obtenue par Fernandez-Tornero et al., 2013. Deux orientations de la polymérase sont également présentées. Les sous-unités sont colorées selon la légende dans l'encart du bas, dans lequel sont également indiquées les sous-unités de l'ARN pol II correspondantes. Les éléments Stalk, Clamp, Jaw, Wall, Cleft, Protrusion et Lobe sont indiqués sur la structure. Le domaine Cter de Rpa49 flexible n'est ici pas représenté.

## VI-1- Définition des éléments structuraux de l'ARN polymérase I

La structure cristallographique de l'ARN pol I révèle des motifs structuraux et fonctionnels spécifiques pour son activité transcriptionnelle, mais aussi des différences notables dans les éléments de structure comparés aux autres ARN polymérases multimériques ADN dépendantes.

### VI-1-a- Etats oligomériques de l'ARN pol I

Un fait important à remarquer est que l'ARN polymérase I a été décrite comme pouvant adopter deux états oligomériques différents: on la retrouve sous forme de monomère, active en transcription, et sous forme de dimère, inactive.

Il a ainsi été observé que les deux états pouvaient être retrouvés en solution sous certaines conditions, bien que la fonction cellulaire de cette dimérisation ne soit initialement pas établie (Milkereit *et al*, 1997). En effet, *in vitro*, il a été montré que la fraction contenant l'ARN pol I pouvant initier la transcription sur un promoteur de gène d'ADNr de façon optimale consistait en deux formes de l'enzyme: monomérique et dimérique. La majeure partie des complexes contenait l'ARN pol I sous forme dimérique, inactive. Cependant, les dimères pouvaient être dissociés en monomères de façon réversible. Le complexe contenant l'ARN pol I sous forme monomérique, potentiellement associée à d'autres facteurs de transcription alors non-caractérisés, permettait l'utilisation optimale du promoteur d'ADNr, et ces complexes d'initiation étaient résistants à de hautes concentrations en sel. Il avait alors été proposé que la forme dimérique servait de "stock" d'ARN polymérases I inactives, qui pouvaient, sous l'action/association de certains facteurs de transcription, devenir monomériques et actives pour la transcription, en maintenant un complexe d'initiation stable (Milkereit *et al*, 1997).

La dimérisation de l'ARN pol I purifiée de *S.cerevisiae* a ensuite été observée en MET en coloration négative (Bischler *et al*, 2002). Cette étude visait à localiser structurellement des

sous-unités associées à l'ARN pol I. Après purification des ARN pol I, les formes monomériques et dimériques étaient visibles en MET, et la forme dimérique pouvait être clairement discriminée au vu de sa forme plus allongée. Dans cette étude, un modèle de l'ARN pol I sous forme de dimère a ainsi pu être obtenu (Figure 18.A.), mais également sous forme de monomère (Figure 18.B.). Lors du remplacement moléculaire ("Docking") du modèle du monomère au sein de celui du dimère, les auteurs se sont aperçus que l'interaction entre les deux polymérases du dimère devait s'effectuer via le Stalk d'une polymérase et le Cleft de la seconde. (Figure 18.C.).

Les analyses des structures 3D à haute résolution de l'ARN pol I entière obtenues en 2013 (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013) font également mention de ces deux états oligomériques possibles, car les deux sont retrouvés dans les cristaux obtenus pour la cristallographie aux RX. L'ARN pol I sous forme monomérique en solution a été utilisée pour les reconstructions 3D de l'enzyme unique. Mais il est intéressant de noter que des structures précises du dimère dans les deux publications ont été obtenues, mettant en valeur deux ARN pol I tête-bêches, et les interactions sont effectivement retrouvées entre le Stalk (Rpa14/Rpa43) d'une ARN pol I et le Cleft de la deuxième. Notamment, les structures à haute résolution obtenues (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013) montrent qu'un élément spécifique du C-ter d'une sous-unité du Stalk (Rpa43), conservée évolutivement, vient s'insérer et envahir le Cleft de la polymérase adjacente, entre le Clamp et le domaine "protrusion" (Figure 18.D.E.) Cet élément est d'ailleurs nommé "connector" par Engel *et al* (2013). Par ailleurs, on notera qu'*in vivo*, un mutant de délétion de la partie Cterminale de Rpa43 incluant ce connector induit un défaut de croissance à 30°C, fortement accentué à 37°C (Peyroche *et al*, 2000) suggérant que cet élément joue un rôle important mais non-essentiel à la viabilité cellulaire (Figure 18.F.).



Etant donné que la dimérisation médiée par le Stalk avait déjà été observée en MET, cela suggérait que le dimère observé dans les cristaux correspondait à celui formé en solution.

Cette idée a été largement reprise par la suite, et notamment dans des études structurales plus récentes concernant l'ARN pol I en association avec le facteur d'initiation Rrn3 (Pils *et al*, 2016; Engel *et al*, 2016; Torreira *et al*, 2017). Il a ainsi été suggéré que l'ARN pol I existait dans un état d'équilibre entre dimères inactifs et monomères activables et actifs, et que le facteur Rrn3 pouvait s'associer à l'enzyme de façon stable, et empêcher sa dimérisation, rendant l'ARN pol I compétente pour l'initiation.

La présence de dimères d'ARN pol I, ainsi que leurs implications dans la régulation du pool d'enzyme disponible pour l'initiation de la transcription est maintenant établie *in vivo* (Torreira *et al*, 2017).

On notera également qu'une autre forme de dimère/multimère a été observée par mon équipe dans une étude alliant des techniques de "Miller spreads" en MET et de co-immunoprécipitation (Albert *et al*, 2011). Cette étude a montré que deux polymérases successives engagées en élongation étaient plus fréquemment en contact que ne le prédisait une distribution aléatoire des polymérases le long de l'ADNr. Pour rendre compte de la distribution des distances entre polymérases successives, près de 60% des ARN pol I en transcription sont associées. La fréquence de cette association dépend notamment de la sous-unité Rpa49. En effectuant la modélisation de deux ARN polymérases I en contact sur le même ADNr linéaire à transcrire, la position d'une polymérase par rapport à l'autre révélait que l'hétérodimère Rpa34-Rpa49 d'une ARN pol I pouvait être structuralement très proche de l'hétérodimère Rpa14-Rpa43 de la polymérase suivante (Albert *et al*, 2011). Ce modèle est en accord avec l'interaction observée des sous-unités Rpa43 et Rpa49 suggérée par leur co-expression hétérologue. La caractérisation détaillée de cette interaction faible n'a pas été poursuivie, et il reste possible que cette co-purification résulte de défauts de repliement de ces

sous-unités en l'absence de leurs partenaires respectifs (Rpa14 pour Rpa43 et Rpa34 pour Rpa49). Des études génétiques préalables suggéraient que les sous-unités Rpa49 et Rpa43 possédaient des liens fonctionnels (Beckouet *et al*, 2008), alors même que leurs positions respectives au sein d'une même ARN pol I étaient très éloignées. Des études sur le positionnement relatif de polymérase successives en transcription remettent en cause ce modèle (Neyer *et al*, 2016). En particulier, l'angle entre deux polymérase proposé dans Albert *et al* (2011) semble différent de l'angle mesuré sur un arbre de Miller observé dans des conditions natives (Cryo-fixation). De plus, l'interaction physique proposée à partir des observations des arbres de Miller semble peu compatible avec la distance physique séparant des complexes transcriptionnels (18 nm +/- 10 nm, Neyer *et al*, 2016). Ces deux données expérimentales (la distance entre polymérase successives et l'angle de sortie de l'ARN ) rendent peut probable le modèle proposé, au moins dans ces conditions d'étalement. L'organisation intra nucléaire des ADNr transcrits restant inconnue *in vivo*, des contacts entre ARN pol I successives le long de la fibre transcrite pourraient avoir lieu, mais cela reste à ce jour inaccessible expérimentalement.



## VI -1-b- Le Stalk de l'ARN pol I

Nous présenterons ici les caractéristiques structurales de ce module bien particulier qu'est le Stalk, facilement reconnaissable de par sa proéminence hors du cœur de l'enzyme. La structure à haute résolution du Stalk de l'ARN pol I a par ailleurs été obtenue par cristallographie (Kuhn *et al*, 2007) (Figure 19.A.). Cet hétérodimère est composé de deux sous-unités: Rpa14 et Rpa43 pour l'ARN pol I (similaires à Rpb4 e Rpb7 chez la pol II, Rpc17 et Rpc25 chez la pol III, et Rpo4 et Rpo7 chez l'Archée, respectivement), son ancrage au Core enzyme est comparable dans le cas des trois polymérases (Jasiak *et al*, 2006). Cependant, à la différence de l'ARN pol II ou de la polymérase archéenne, Rpa14 ne possède pas de domaine HRDC (Helicase RNaseD C-terminal domain), ce qui pourrait expliquer sa taille plus réduite. Or, comme la fonction *in vivo* ainsi que la position de ce domaine HRDC varient entre l'ARN pol II et la polymérase archéenne (Jasiak *et al*, 2006), l'absence de ce domaine dans le cas de l'ARN pol I pourrait avoir des conséquences fonctionnelles.

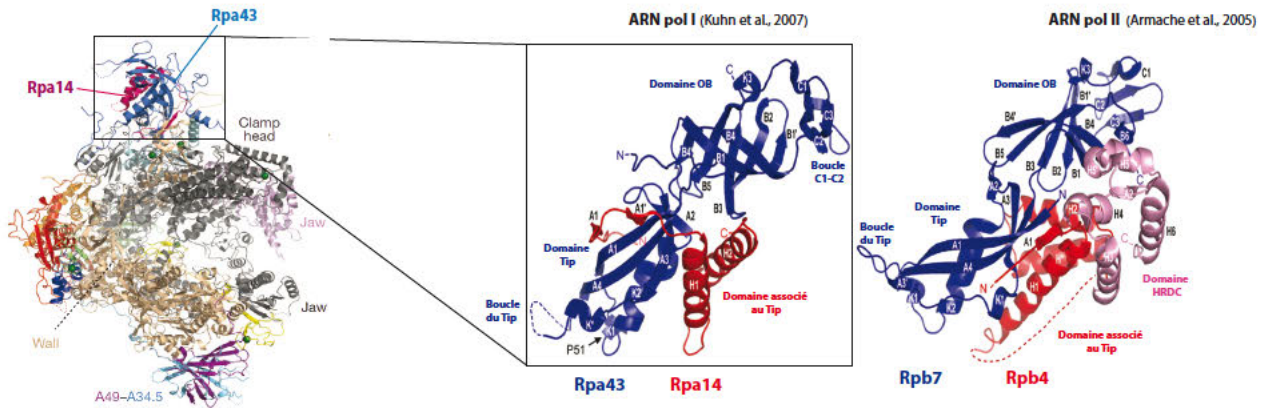
Une autre différence notable parmi les différentes ARN polymérases est que le Stalk de l'ARN pol I forme une interface plus large avec le cœur enzymatique (Figure 19.B.), notamment via la queue N-terminale de Rpa43, qui interagit avec une région du C-ter de la deuxième plus grande s.u. Rpa135. De plus, Rpa43 et Rpa14 interagissent également avec une hélice- $\alpha$  en N-ter de ABC23 (Rpb6) (Fernández-Tornero *et al*, 2013). Enfin, la plus grande sous-unité Rpa190, via ses régions C- et N-terminales, permet également la fixation du Stalk au cœur de l'enzyme.

Ainsi, le Clamp serait toujours en position fermée dans le cas de l'ARN pol I, ce qui serait en opposition au cas de l'ARN pol II dans lequel le Stalk est dissocié, et où le Clamp est ouvert (Cramer *et al*, 2001). En effet, le Stalk pol I se lie de façon étroite avec le cœur de l'enzyme, ce qui entraîne une légère inclinaison (environ 10°) du Stalk vers le Clamp de la pol I comparé à celui de l'ARN pol II. Cette inclinaison particulière permettrait notamment la

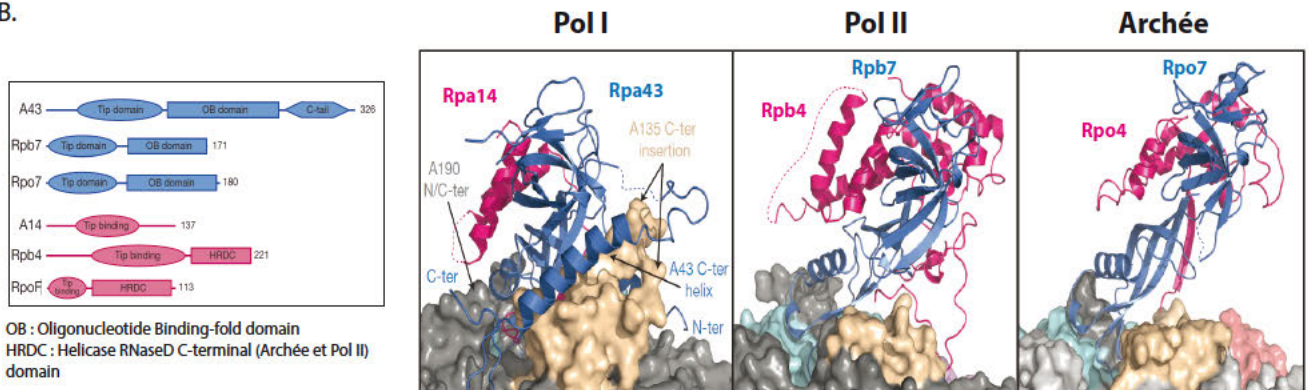
fixation du facteur spécifique d'initiation Rrn3. Le Stalk de l'Archée est lui replié dans le sens inverse.

Le Stalk de l'ARN pol I est notamment l'un des éléments principaux d'interaction avec le facteur d'initiation Rrn3.

A.



B.



**Figure 19. Structure du Stalk de l'ARN pol I, et comparaison avec d'autres ARN polymérases multimériques.**

A. Structures 3D de l'ARN pol I (Fernandez-Tornero et al., 2013) et des Stalks des ARN pol I et II. Les deux sous-unités de chaque Stalk sont indiquées en rouge (Rpa14/Rpb4) et bleu (Rpa43/Rpb7) respectivement. Le domaine HRDC (Helicase RNaseD C-terminal), absent pour l'ARN pol I, est indiqué en rose pour Pol II.

B. Structure et ancrage à la polymérase du Stalk dans le cas des ARN pol I, II et archéenne. A gauche est indiquée la schématisation des domaines de chaque Stalk. A droite, Les Stalks sont représentés en rubans, là où les polymérases respectives sont en représentation de surface.

### VI-1-c- Le facteur essentiel d'initiation Rrn3

Rrn3 est l'une des protéines initialement identifiée dans le crible génétique de Nogi *et al*, en 1991. Ce crible a permis l'identification de mutants *rrn* qui a permis l'identification de tous les gènes essentiels spécifiquement requis pour la transcription Pol I *in vivo* (détaillé dans le chapitre VI-2-1). Cette protéine est essentielle et conservée (Figure 20.A.): on la retrouve chez les mammifères, mais également chez différents types de levures, tels que *S.pombe* et bien sûr *S.cerevisiae*. L'expression de l'homologue humain Rrn3, aussi appelé TIF-IA, permet de restaurer la viabilité (Moorefield *et al*, 2000). Il a également été montré que Rrn3 s'associait à l'ARN polymérase I par le biais de la sous-unité Rpa43, qui fait partie du complexe du "Stalk" de la polymérase, constitué des sous-unités Rpa14 et Rpa43 (Peyroche *et al*, 2000). La structure 3D de Rrn3 pleine longueur a été obtenue à haute résolution par cristallographie (Blattner *et al*, 2011) (Figure 20.B.). Des structures de complexes Pol I-Rrn3 ont été obtenues récemment en Cryo-Microscopie Electronique à Transmission (Cryo-MET) (Engel *et al*, 2016; Pilsl *et al*, 2016; Torreira *et al*, 2017; Sadian *et al*, 2017) (Figure 20.C. et D.). Ces données structurales permettent de mieux comprendre la mécanistique de l'étape d'initiation de la transcription par l'ARN pol I. Rrn3 possède deux éléments principaux : une structure avec 10 répétitions HEAT (HEAT: structure formée par deux hélices- $\alpha$  reliées par un linker de petite taille) (Figure 20.A. et B.) et une région de surface riche en sérines, phosphorylables, appelée "serine patch".

La structure HEAT est composée de 20 hélices- $\alpha$  antiparallèles, notamment connues pour permettre les interactions protéine-protéine, additionnées de trois hélices en C-terminal regroupées contre les répétitions H6-H10. En solution, Rrn3 est retrouvée sous forme d'homodimère, mais ne se lie à l'ARN pol I que sous forme de monomère, et contient un ensemble de sérines nécessaires au recrutement de la pol I *in vivo*, au niveau des répétitions HEAT H2-H4 (Blattner *et al*, 2011). Plus particulièrement, il est maintenant connu que Rrn3

forme quatre interfaces de liaison avec l'ARN pol I: la première se fait *via* la sous-unité Rpa43 du stalk de la polymérase et les répétitions HEAT H2-H4 de Rrn3. La seconde est formée au niveau de la région du Clamp de Rpa135 de l'ARN pol I et la sixième hélice- $\alpha$  de la répétition H3. La troisième interaction s'effectue entre H5-H6 de Rrn3 et le domaine "dock" de Rpa190, la plus grande sous-unité de l'ARN pol I. A noter que le domaine "dock", dans le cas de l'ARN pol II, se lie avec les facteurs spécifiques d'initiation TFIIB, ainsi qu'au médiateur. Enfin, la dernière zone d'interaction met en relation la région C-terminale de la sous-unité AC40, certains résidus centraux de AC19 et une boucle de Rpb6 de l'ARN pol I, avec une boucle en C-ter de Rrn3 (Figure 20.C.).

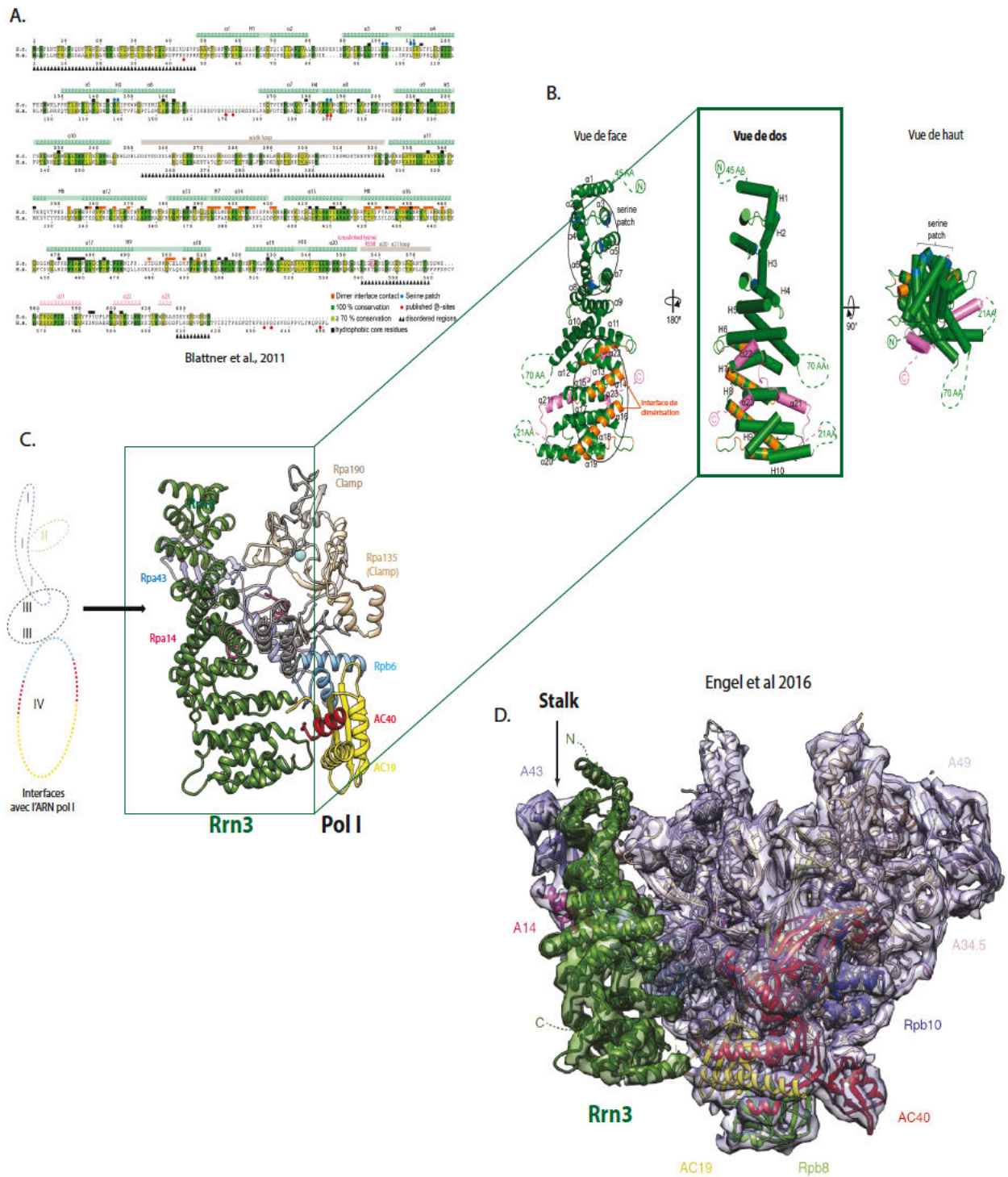
Le "serine patch" est une région de surface constituée de huit résidus sérines, arrangés en quatre paires (six de ces résidus sont conservés entre l'humain et la levure), et est requise *in vivo*. En effet, des mutations de ces sérines ont été effectuées: mutées en alanine, cela empêche la phosphorylation de ces résidus ; mutées en aspartate, cela mime un état phosphorylé de ces résidus (Blattner *et al*, 2011). Dans les deux cas, lorsque ces formes mutées de Rrn3 sont utilisées pour compléter une souche dans laquelle Rrn3 a été enlevée ( $\Delta$ rrn3), de sévères défauts de croissances sont observés. Les mêmes auteurs ont de plus montré que ce "serine patch" permettait également la fixation de Rrn3 à l'ARN pol I *in vitro*, et que sa phosphorylation nuisait au recrutement coopératif de l'ARN pol I et Rrn3 sur le promoteur d'un gène d'ADNr, *in vivo*.

Sachant qu'une minorité d'ARN polymérases I nucléaires est active en transcription (Milkereit *et al*, 1997), cela signifie intrinsèquement que différentes formes d'ARN pol I coexistent dans la cellule. Or, celles qui sont compétentes pour l'initiation sont associées au facteur d'initiation Rrn3. En effet, il a été montré que l'ARN pol I ne pouvait être recrutée au promoteur que si elle était associée à Rrn3, pour ensuite entamer son cycle de transcription, que ce soit *in vivo* ou *in vitro* (Yamamoto *et al*, 1996; Milkereit & Tschochner, 1998). Cependant, d'autres



données ont par la suite suggéré que le recrutement en lui-même était possible en l'absence de Rrn3, mais que le complexe ainsi formé n'était pas compétent pour la transcription (Aprikian *et al*, 2001). De fait, les études structurales récentes ont révélé que la fixation de Rrn3 sur la polymérase stabilisait un état monomérique de l'ARN pol I, compétent pour l'initiation de la transcription. Notamment, des détails sur la manière dont cette transition s'effectue ont été obtenus via ces études structurales sur des complexes d'ARN pol I monomériques et le complexe Pol I- Rrn3. Les principaux changements conformationnels sont associés au passage entre la forme dimérique et la forme monomérique (Pilsl *et al*, 2016). La formation de monomère induit la contraction partielle du Cleft, qui permet le relargage de "l'expand", et libère le site de liaison de l'ADN matrice, nécessaire à l'initiation de la transcription. Il est intéressant de noter que la position du domaine C-terminal de Rpa12 (TFIIS-like) est légèrement modifiée, et notamment que sa boucle catalytique devient mobile, ce qui empêche son activité stimulatrice du clivage d'être opérationnelle durant l'initiation (Engel *et al*, 2016). Dans une autre étude, Rpa12 semble même totalement désengagé du pore (Pilsl *et al*, 2016). Aucun changement structural significatif n'est détectable entre la forme monomérique et la forme monomérique associée à Rrn3. Cependant, le relâchement du "connector" rendu accessible au solvant dans la forme monomérique libère la surface de l'ARN pol requise pour la fixation de Rrn3. La fixation de Rrn3 à cette interface stabilise l'ARN pol I dans un état monomérique compétente pour l'initiation. En effet, une fois ce facteur fixé, la réassociation de l'élément connector sur la polymérase ne peut stériquement plus se faire.

Il a d'autre part été montré, sur la base d'une interaction double-hybride, que le recrutement du complexe Pol I-Rrn3 sur le complexe de pré-initiation se faisait via une interaction entre Rrn3 et la sous-unité Rrn6 du Core Factor (CF) (Peyroche *et al*, 2000).



**Figure 20. Le facteur d'initiation Rrn3 et son mode d'interaction avec l'ARN pol I.**

A. Alignement de séquences d'acides aminés de Rrn3 de *S.cerevisiae* et d'*Homo sapiens*. Les éléments secondaires de la structure sont indiqués au-dessus des séquences correspondantes. Notamment, les paires d'hélices formant les répétitions HEAT (H1-H10) et les hélices du Cter sont représentées au-dessus de la séquence en vert et rose respectivement. En-dessous des séquences sont indiquées par des triangles noirs les régions désordonnées. Surlignés en vert foncé et vert clair, on trouve les résidus conservés à 100% et à plus de 70% respectivement. La surface de dimérisation est indiquée en orange.

B. Structure 3D de Rrn3 en représentation en rubans (gauche) et en cylindres (milieu et droite). Les éléments de structure secondaires sont représentés tels qu'en A. Les hélices du C-ter de Rrn3 sont indiquées en rose.

C. Représentations des quatre interfaces d'interactions (I, II, III et IV) entre l'ARN pol I et Rrn3; schématisée à gauche, et en structure à droite, orientées comme la vue de dos en B.

D. Représentation 3D de l'enveloppe en Cryo-MET du complexe ARN pol I - Rrn3 dans laquelle ont été remplacées les structure à haute résolution des sous-unités de l'ARN pol I (différentes couleurs indiquées) et Rrn3 (vert).

#### VI -1-d- Les éléments caractéristiques du Core enzyme de l'ARN pol I

Globalement, on distingue quatre modules principaux, mobiles, façonnant l'architecture globale de l'ARN pol I autour du Cleft: "Le Core, le Jaw/Lobe, le Shelf et le Clamp" (Figure 21.A.). Le Cleft de la polymérase représente la fente entre les principaux modules de l'enzyme où s'introduira l'ADN, et est beaucoup plus large pour l'ARN pol I que dans le cas de l'ARN pol II (montré par superposition des structures à haute résolution des ARN pol I et II : environ 10Å plus large dans la structure obtenue par Fernandez-Tornero *et al.*, 2013 ; de 8 Å pour Engel *et al.*, 2013).

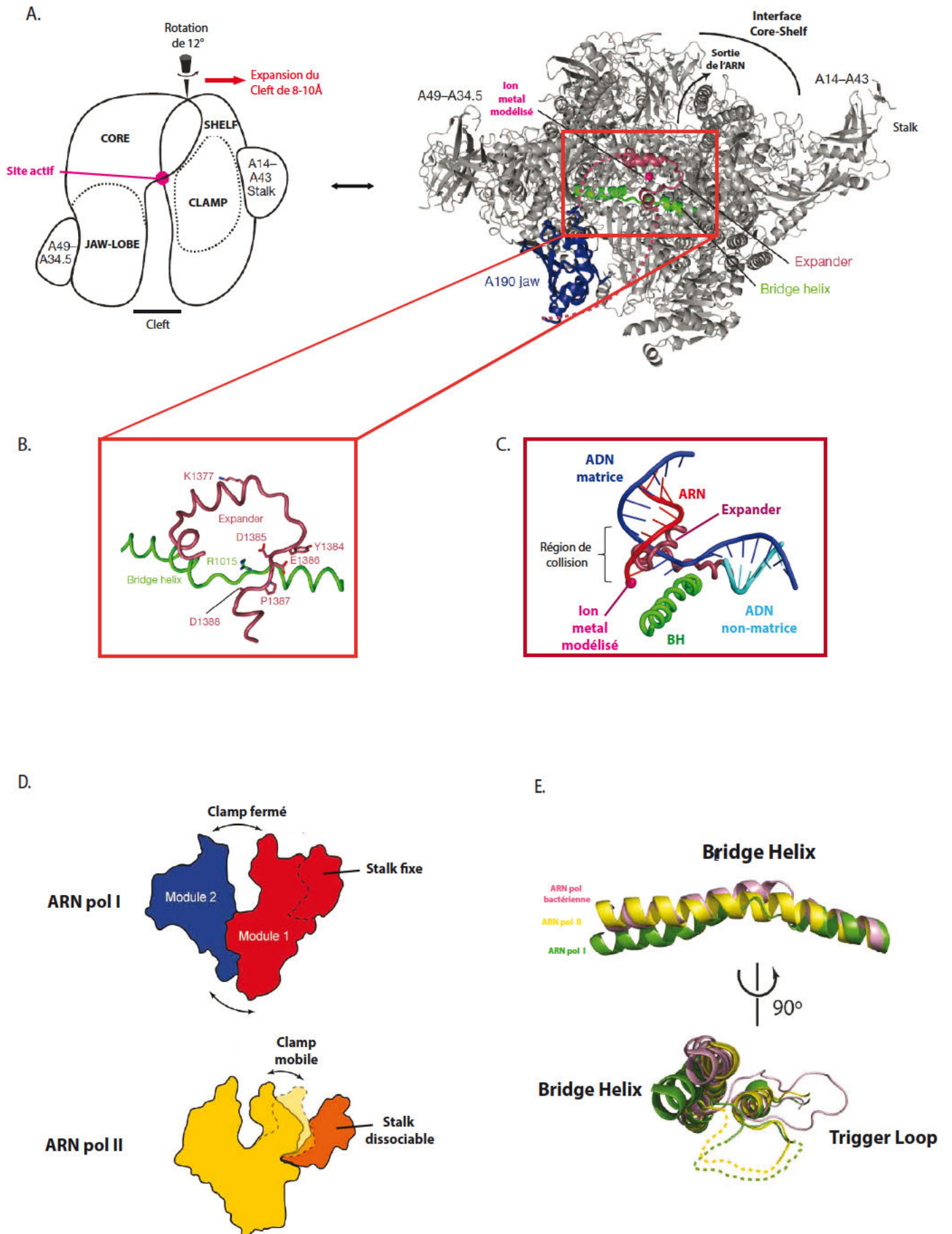
Un point important à noter ici est qu'on parle "d'expansion" du Cleft de l'ARN pol I, plutôt que "d'ouverture" du Cleft, comme c'est le cas pour l'ARN pol II par exemple. En effet, "l'expansion" diffère de "l'ouverture" par le mouvement de rotation des deux modules Core et Shelf de l'ARN pol I autour d'un axe passant par le site actif (voir schéma figure 21.A). Le mouvement de balancier du Clamp mobile caractérise également l'ARN pol II (Figure 21.D.), et n'est pas présent dans l'ARN pol I. Les modules "Core" et "Shelf" sont principalement représentés par les deux plus grandes sous-unités Rpa135 et Rpa190 respectivement. Le Cleft en expansion modifie le centre actif et est rendu stable par un élément structural, non-conservé dans le cas des ARN pol II et III, nommé "Expander" ou "DNA-mimicking loop" (Engel *et al.*, 2013 ; Fernandez-Tornero *et al.*, 2013 respectivement) (voir structure Figure 21.A.). En termes de séquence primaire, l'Expander est retrouvé entre deux régions d'acides aminés de Rpa190, qui font partie d'un autre module caractéristique, la "mâchoire" de Rpa190, ou "Jaw", élément majeur dans la structure de la pol I car retrouvé sur plusieurs sous-unités différentes, et qui forme comme son nom l'indique une mâchoire constituée de deux parties qui vont enserrer l'ADN entrant. Le module Jaw est présent dans ABC27, et dans deux régions de Rpa190 (aa 1260-1360 et 1400-1499, Fernandez-Tornero *et al.*, 2013). Ainsi, l'Expander est formé de deux hélices- $\alpha$  au niveau du site actif, et interagit avec la Bridge

Helix en la stabilisant en position dépliée (Figure 21.B.). Cet élément est donc structurellement situé exactement au niveau du site de liaison de l'ADN matrice (dans une pol I qui n'est pas en transcription) et stabilise le Cleft en expansion, via de nombreuses liaisons hydrogènes et interactions électrostatiques entre les résidus chargés négativement de cette boucle et les résidus chargés positivement à l'intérieur du Cleft. La position de l'Expander est donc incompatible avec une pol I en élongation. Il doit donc changer de conformation lors de l'activation de l'ARN pol I et l'initiation de la transcription, pour permettre au brin matrice d'entrer et d'atteindre le centre catalytique, ce qui est concomitant avec l'extension du Cleft. Ainsi, le "DNA-mimicking loop" ou "Expander" jouerait un rôle régulateur, empêchant la liaison de tout autre d'acide nucléique que l'ADNr (Figure 21.C.) ou déplaçant l'ADN transcrit durant la terminaison de la transcription. Il est intéressant de mentionner ici qu'une souche de levure dans laquelle on enlève la majeure partie de cet élément (délétion des résidus 1361-1390 de Rpa190) reste viable, mais affiche un retard de croissance à 37°C, et pas à 30°C (Fernandez-Tornero *et al*, 2013). On notera également que l'organisation spatiale de cet "Expander" coïncide avec la stabilisation de la partie C-terminale de la sous-unité Rpa12, suggérant que ces deux régions de la pol I se stabiliseraient entre elles.

De façon intéressante dans l'ARN pol I, ce Cleft largement en expansion va donc de paire avec une conformation intrinsèquement fermée de la pince ("Clamp"). Cet élément, dans le cas la pol II, est censé réguler l'ouverture/fermeture du site de liaison à l'ADN du Cleft par rotation des deux parties de cette pince. Or, un Clamp fermé n'est observé chez la pol II que lorsque tous les facteurs sont présents ou si l'enzyme est liée à la bulle de transcription (Kettenberger *et al*, 2004), ce qui n'est par exemple pas le cas si les éléments du Stalk de la pol II ne sont pas présents. Chez la pol I par contre, le fait que ce Clamp soit retrouvé fermé (Figure 21.D.) dans les structures cristallines provient certainement de la présence d'hélices qui interagissent avec ABC27 (Rpb5), et assurent également l'interaction entre le cœur

enzymatique et le Stalk de l'enzyme. L'idée que le Clamp adopterait une position toujours fermée pourrait expliquer le fait que l'enzyme puisse transcrire le long gène du précurseur d'ARNr (plus de 9kb) sans se dissocier de l'ADN matrice.

La Bridge Helix (BH) de l'ARN pol I est une région de la plus grande sous-unité Rpa190, est très conservée en termes de séquence et adopte une conformation spécifique (Engel *et al*, 2013 ; Fernandez-Tornero *et al*, 2013). De façon intéressante dans le cas de la pol I, et en parallèle avec l'expansion du Cleft, on peut noter que la BH se déroule d'un tour d'hélice complet (Rpa190 acides aminés 1012-1016) et permet l'apparition d'un nouveau repliement hélicoïdal de 10°, comparé aux autres ARN polymérases (exemples de l'ARN polymérase bactérienne et pol II eucaryote, voir superposition Figure 21.E.). La BH de l'ARN pol I serait flexible et pourrait notamment se dérouler selon les besoins lors de certaines étapes de la transcription. D'autre part, de même que dans le cas de l'ARN pol II, la "Trigger loop", juste adjacente à la BH n'aurait pas toujours la même conformation dans l'ARN pol I (Figure 21.E.). La Trigger Loop est l'un des éléments clefs dans le cadre du déclenchement du processus enzymatique à proprement parler, et se situe également dans la plus grande sous-unité Rpa190 (acides aminés 1186-1232) (Fernandez-Tornero *et al*, 2013).



**Figure 21. Architecture du coeur enzymatique de l'ARN pol I, et de ses modules spécifiques.**

A. Les panels de droite et de gauche représentent l'ARN pol I dans la même orientation. Panel de gauche : schématisation des quatre modules mobiles de l'enzyme et des hétérodimères Rpa14-Rpa43 et Rpa34-Rpa49, additionnés de la direction schématique d'expansion du Cleft (flèche rouge) et de la rotation (flèche noire) entre les modules Shelf et Core. Panel de droite : structure de l'enzyme entière, dans laquelle sont représentés en couleur : l' Expander (violet) et la Bridge Helix (BH) dépliée (vert), un ion metal catalytique modélisé (rose), et le domaine Jaw de Rpa190 (bleu foncé) (d'après Engel et al., 2013).

B. Zoom sur la région structurale montrant la stabilisation de la BH dépliée (en vert) par l'Expander (Violet) et les acides aminés importants pour cette stabilisation (d'après Engel et al., 2013).

C. Modélisation structurale de l'Expander (violet) de l'ARN pol I et de l'hybride ADN-ARN de l'ARN pol II (pdb : 1Y1W, en bleu-rouge respectivement) au niveau du centre actif, montrant les collisions stériques induites par cette superposition et l'impossibilité de maintenir les deux en même temps au centre actif. La BH (en vert), ainsi que l'ion catalytique modélisé (rose clair) sont également représentés (d'après Engel et al., 2013).

D. Comparaison schématique de l'ARN pol I et de l'ARN pol II, montrant la mobilité possible au niveau du Clamp et du Stalk. Le Clamp fermé dans le cas de l'ARN pol I est concomittant avec un Stalk indissociable (indiqué dans la zone rouge). Dans le cas de l'ARN pol II, le fait que le Stalk (orange foncé) soit dissociable permettrait au Clamp d'être plus mobile et d'avoir un mouvement de «balancier» (d'après Fernandez-Tornero et al., 2013).

E. Comparaison structurale des éléments BH et Trigger Loop entre trois ARN polymérase: ARN pol bactérienne (*Thermus thermophilus*, en rose), ARN pol II (jaune) et ARN pol I (vert). Concernant la Trigger Loop, celle de la bactérie est ordonnée dans le cristal, donc représentée en trait continu, alors que celles des ARN pol I et II ne le sont pas, et sont donc représentées en pointillés (d'après Fernandez-Tornero et al., 2013).

## VI-1-e- L'ARN polymérase I possède des sous-unités spécifiques : les "Built-in Transcription factors"

L'une des particularités de l'ARN pol I est la présence de sous-unités spécifiques, homologues à des facteurs de transcription de l'ARN pol II, et similaires à des modules de l'ARN pol III. On en compte trois, qui font partie intégrante de l'enzyme: les s.u. Rpa34 et Rpa49 qui forment un hétérodimère, et Rpa12. On peut ainsi les comparer à des facteurs de transcription faisant partie intégrante de l'enzyme et qui ne s'en dissocient pas, les "Built-in Transcription Factors" ou BTFs. Ces sous-unités ne sont pas essentielles chez la levure, à savoir que des mutants de délétion de ces sous-unités demeurent viables. Dans cette partie, nous détaillerons les rôles connus de chaque BTF, ainsi que leurs particularités structurales et fonctionnelles.

Les travaux de thèse ont débuté et se sont en grande partie consacrés à l'étude des caractéristiques de Rpa49, et à la caractérisation de mutants de délétion de cette sous-unité. Ici, je décrirais donc précisément ce qui est connu sur Rpa49 en termes de structures et de fonctions, afin d'apporter les bases nécessaires pour introduire les études génétiques et phénotypiques qui nous intéresseront dans la dernière partie de cette introduction (Chapitre VII), ainsi que dans les résultats et la discussion de ce manuscrit.

### *Rpa34- Rpa49*

Ces deux sous-unités sont chacune composées de deux modules, ou domaines structurés : un Nter et un Cter. Le Nter de l'une interagit avec le Nter de l'autre, formant un hétérodimère Rpa34-Rpa49. Le Cter de Rpa34 forme une longue queue chargée qui ancre la sous-unité au reste de la polymérase. Le Cter de Rpa49 est un domaine tWH ("tandem Winged Helix"), et est relié au Nter par un linker flexible de 83 acides aminés (Figure 22 A et B.).

L'hétérodimère des deux domaines Nter Rpa34(Nt)-Rpa49(Nt) est l'homologue structural de TFIIF de l'ARN pol II, alors que le C-ter de Rpa49 est similaire à TFIIE $\beta$  de l'ARN pol II



(voir ci-dessous). Lors de leur expression hétérologue, en plus de l'hétérodimère, deux homodimères Rpa34/Rpa34 et Rpa49/Rpa49 ont été observés lors de la purification des complexes formés. Une partie de l'hétérodimère Rpa49/Rpa34 a été identifiée dans la structure cristalline de l'ARN pol I. Tout comme TFIIF, Rpa49/Rpa34 s'associe au Lobe de l'ARN pol I via ses domaines de dimérisation qui forment une structure en triple tonneaux  $\beta$  (Figure 22 A et B.), position structurale cohérente avec les études antérieures de Cryo-MET (De Carlo *et al*, 2003) et de crosslink des sous-unités de l'ARN pol I (Jennebach *et al*, 2012). Il est également similaire aux sous-unités Rpc53-Rpc37 de l'ARN pol III (Figure 22.C.). Sa position en périphérie du core enzyme expliquerait pourquoi les deux sous-unités se dissocient ensemble de l'ARN pol I lors d'un traitement à l'urée (8M) (Huet *et al*, 1975), et pourquoi une préparation d'ARN pol I purifiée à partir d'une souche de levure dans laquelle le gène codant pour Rpa34 a été invalidé ne contient pas Rpa49 (Gadal *et al*, 1997). De plus, cela concorde également historiquement avec les densités électroniques obtenues initialement en Cryo-MET montrant deux densités électroniques séparées en-dehors de la région du Cleft et assignées à Rpa34 et Rpa49 (Bischler *et al*, 2002). Cet hétérodimère aurait un rôle en élongation, comme suggéré par des études de délétion de Rpa49 et de la dissociation de l'hétérodimère de Rpa49-Rpa34 montrant une modification de l'activité non-spécifique (indépendante du promoteur) de l'ARN pol I (Huet *et al*, 1975; Liljelund *et al*, 1992). Sa position permet également la stabilisation la sous-unité Rpa12, suggérant son implication "indirecte" dans la stimulation du clivage de l'ARN pol I, Rpa12-dépendante (Kuhn *et al*, 2007; Geiger *et al*, 2010). La déstabilisation de Rpa34/Rpa49 est aussi observée lorsque la délétion de Rpa12 est effectuée (Van Mullem *et al*, 2002). Une autre étude en 2008 suggère que Rpa49 pourrait également posséder un rôle dans l'association et le relargage du facteur d'initiation Rrn3 (Beckouet *et al*, 2008).

Rpa34 possède une extrémité C-terminale longue (90 acides aminés) et désordonnée, qui a un rôle essentiel d'ancrage de l'hétérodimère au reste de l'enzyme. Elle s'étend en effet sur plus de 70Å le long de la face externe de Rpa135, et atteint une cavité à la jonction entre AC40, ABC10β et Rpa135 (Figure 22.A. panel de droite). De plus, via l'utilisation de mutants de délétion de la partie C-terminale (acides aminés 199-338) de l'homologue murin PAF49, couplée à de la microscopie à fluorescence, il a été montré que cette longue queue Cterminale était importante pour la localisation nucléolaire de l'enzyme (Ushijima *et al*, 2008). Chez *S.cerevisiae*, ce domaine Cter se comporte comme un signal de localisation nucléolaire. Cet adressage pourrait s'effectuer via une interaction directe avec l'ARN pol I *in vivo*, indépendamment de Rpa49, mais cette hypothèse n'a pas été testée expérimentalement (Albert *et al*, 2011).

Le domaine C-terminal de Rpa49 a été cristallisé indépendamment de l'ARN Pol I (Geiger *et al*, 2010 ; Figure 22.C., à droite.). Il semble mobile, comme le suggère son absence des formes cristallines de l'ARN pol I entière (Engel *et al*, 2013 ; Fernandez-Tornero *et al*, 2013). La région Cterminale de Rpa49 forme un domaine "tWH" ("tandem Winged Helix"), à savoir un domaine constitué de deux motifs Winged Helix (WH) répétés. Plus globalement, il faut savoir que les domaines WH font partie de la famille des motifs Helix-Turn-Helix (HTH), comprennent généralement trois hélices-α et trois brins-β, dans l'ordre α1-β1-α2-β2-α3-β3, et forment des boucles en termes de structures secondaires (Teichmann *et al*, 2012). Les domaines WH sont connus pour pouvoir interagir avec l'ADN, mais aussi pour médier certaines interactions protéine-protéine. Ce domaine tWH de Rpa49 possède une surface chargée positivement, suggérant qu'elle pourrait notamment se lier à l'ADN chargé négativement. En effet, Geiger *et al*, ont montré en 2010 que le tWH de Rpa49 pouvait se lier à de l'ADN double brin via des résidus dans les hélices α2 et α3 de son domaine WH2, ainsi qu'à de l'ADN simple brin correspondant au brin non-matrice du promoteur de l'ARN pol I,

via sa queue C-terminale basique (voir positions des domaines Figure 22.B.). Les deux types d'activités de fixation à l'ADN ne s'excluent d'ailleurs pas mutuellement. Ces résultats suggèrent que le domaine tWH de Rpa49 serait impliqué dans les étapes précoces d'ouverture et/ou de stabilisation de la bulle de transcription, en se fixant à l'ADN simple brin non-matrice. On notera que ce domaine tWH possède de l'homologie avec TFIIE $\beta$  de l'ARN pol II, ainsi qu'avec Rpc34 et Rpc82 de l'ARN pol III, qui possèdent eux aussi des domaines WH (Figure 22.C.). On peut imaginer que la fusion de ce domaine Cter mobile de Rpa49 au reste de l'enzyme pourrait découler d'une adaptation moléculaire évolutive vis-à-vis du promoteur unique de l'ARN pol I, concordant avec une machinerie de transcription plus spécialisée (comparativement aux autres ARN polymérases qui possèdent différents types de promoteurs).



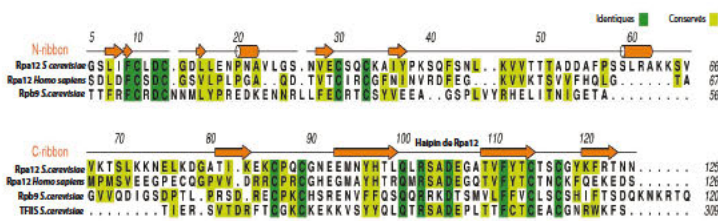
## Rpa12

Rpa12 est une petite sous-unité (125 acides aminés chez *S.cerevisiae*) composé de deux domaines de liaison au Zinc (Zn) ("Zn-binding ribbons") connectés par un Linker flexible, ayant de l'homologie pour la sous-unité Rpb9 et le facteur de transcription TFIIS de l'ARN pol II, et pour la sous-unité Rpc11 de l'ARN pol III (Figure 23.A.). Le domaine Nter est localisé entre le Jaw de Rpa190, le Lobe de Rpa135 et l'hétérodimère Rpa34-Rpa49, contactant ces trois éléments (Figure 23.B.). Ce positionnement est similaire à celui de Rpb9 de l'ARN pol II, même si on remarque une légère rotation et un déplacement vis-à-vis de ce dernier, sûrement due à la présence permanente de Rpa34-Rpa49. Le Nter de l'homologue Rpc11 de l'ARN pol III est également positionné selon le même schéma, au niveau des du Jaw et Lobe de Rpc160 et Rpc128 respectivement, et en contact avec Rpc37 (Figure 23.C.).

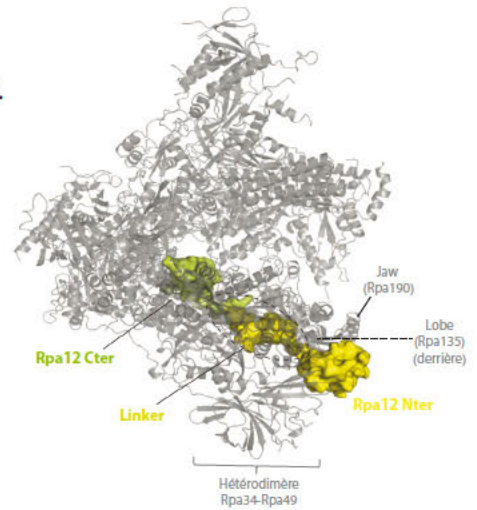
La région C-terminale de Rpa12 s'étend jusqu'au centre de l'enzyme, et atteint le site actif via une boucle possédant des résidus acides communs à ceux de TFIIS. Ce deuxième domaine de liaison au Zn vient se placer au niveau du pore d'entrée des nucléotides, positionnement qui coïncide très bien avec celui du facteur TFIIS dans le cas de l'ARN pol II, et la partie Cterminale de Rpc11 de l'ARN pol III (Figure 23.B.). En effet, cette position correspond à celle préalablement estimée par crosslink chimique (Jennebach *et al*, 2012), et va de pair avec la stimulation de l'activité intrinsèque de clivage des ARN polymérase eucaryotes, telle qu'énoncée dans le Chapitre-V-2-b, dans le cas de TFIIS. Ainsi, via son Cter, Rpa12 serait impliquée dans la stimulation de l'activité de clivage intrinsèque de l'ARN pol I. Cette activité *in vitro* peut être stimulée en augmentant le pH. Rpa12 permet d'augmenter l'activité basale de l'ARN pol I, ce qui a été suggéré par Kuhn et al. en 2007 par des expériences de clivage *in vitro* montrant qu'un mutant de délétion du Cter de Rpa12 de la polymérase I (Rpa12 $\Delta$ C(79-125)) était totalement inactif dans le clivage de l'ARN en comparaison avec une polymérase normale. On notera par ailleurs que le domaine Cter de Rpa12 n'est pas présent dans toutes les

structures en Cryo-MET de l'ARN pol I en complexe avec Rrn3 ou en élongation (PilsI *et al*, 2016; Neyer *et al*, 2016; Tafur *et al*, 2016). Cependant, Engel *et al* (2016) ont également obtenu une structure du complexe ARN pol I - Rrn3 dans laquelle cette partie C-terminale de Rpa12, qui va jusqu'au site actif dans la forme dimérique, a effectué une rotation et est légèrement retirée du site actif. Ce mouvement du C-ter de Rpa12 s'accompagnerait d'un repliement partiel de la partie centrale de la Bridge Helix, censée être déroulée. Toutes ces données suggèrent que cette région en C-ter, et au moins une partie du Linker, seraient flexibles, et devraient bouger hors de leur position initiale pour qu'une élongation processive s'effectue. Rpa12C-ter viendrait alors temporairement prendre sa place dans le pore lorsque le clivage de l'ARN est nécessaire (pause, backtrack...), comme le fait le facteur TFIIS qui ne vient que transitoirement pendant la transcription par l'ARN pol II. Ce phénomène se rapprocherait aussi grandement du cas de la sous-unité Rpc11, dont il a été décrit en 2015 qu'elle était mobile, et effectuait potentiellement un mouvement de va et vient dans et hors de l'ARN pol III durant la transcription (Hoffmann *et al*, 2015). Enfin, Rpa12 jouerait un rôle dans la terminaison de l'ARN pol I (Prescott *et al*, 2004; Kuhn *et al*, 2007).

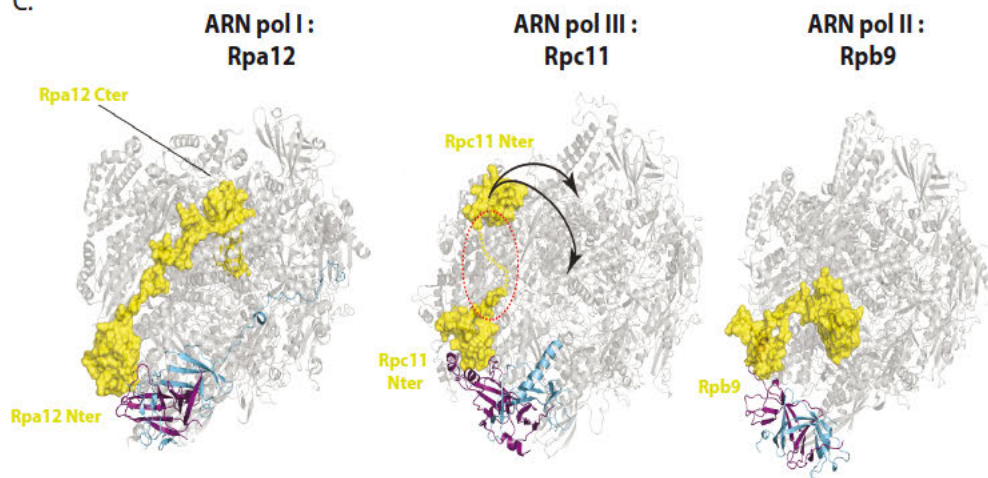
A.



B.



C.



### Figure 23. Etude de la sous-unité Rpa12 de l'ARN pol I de *S.cerevisiae*.

A. Alignement des séquences d'acides aminés de Rpa12 de *S.cerevisiae*, Rpa12 de *Homo sapiens*, Rpb9 et TFIIS de *S.cerevisiae*. Les résidus identiques sont surlignés en vert foncé, les résidus conservés en vert clair. Les flèches oranges représentent les brins- $\beta$ , et les cylindres oranges les hélices- $\alpha$ . La rangée du haut représente la partie Nter de chaque («N-ribbon»), et la rangée du bas la partie Cter («C-ribbon»). D'après Engel et al, 2013.

B. Structure de l'ARN polymérase I, dans laquelle la s.u. Rpa12 a été colorée. Sont indiquées les parties N- et Cterminales de Rpa12, ainsi que les régions de l'ARN pol I en contact avec Rpa12 (le Jaw de Rpa190, le Lobe de Rpa135, et l'hétérodimère Rpa34-Rpa49). D'après Fernandez-Tornero et al, 2013.

C. Comparaison structurale des sous-unités Rpa12 (Pol I), Rpb9 (Pol II) et Rpc11 (Pol III), indiquées en jaune. Les hétérodimères Rpa49-Rpa34, TFIIF- $\alpha/\beta$  et Rpc53-Rpc37 ont également été représentés (bleu-violet respectivement). Les flèches noires dans le cas de Rpc11 représentent la mobilité que possède sa partie Cterminale. D'après Hoffman et al, 2015.

## **VI-2- La transcription par l'ARN pol I**

En partant des bases structurales connues et décrites, il est maintenant plus aisé d'appréhender et de comprendre la plupart des étapes de la transcription par l'ARN pol I à proprement parler. Ainsi, les trois phases générales de la transcription des ARN polymérase ADN dépendantes multimériques (initiation, élongation et terminaison) sont reprises ici dans le cas de l'ARN pol I (Figure 24.B.), complétées à la lumière des données structurales obtenues ces dernières années.

Brièvement, chez *S.cerevisiae*, la formation du Complexe de Pré-Initiation (PIC) nécessite la présence du facteur d'initiation Rrn3, et d'un complexe protéique, le Core factor (CF), composé de trois protéines (Rrn6, Rrn11 et Rrn7). L'ARN pol I peut être purifiée sous une forme inactive, dimérique. Sa liaison au facteur Rrn3 ne change pas sa conformation, mais la stabilise sous forme monomérique ; elle peut alors se lier au CF via Rrn3, acquérant ainsi sa compétence pour l'initiation (ITC pour Initially Transcribing Complex) (Blattner *et al*, 2011; Engel *et al*, 2013; Fernández-Tornero *et al*, 2013; Neyer *et al*, 2016; Tafur *et al*, 2016).



## VI-2-a- La transcription par l'ARN pol I : régions promotrices et terminatrices de la transcription chez *S.cerevisiae*.

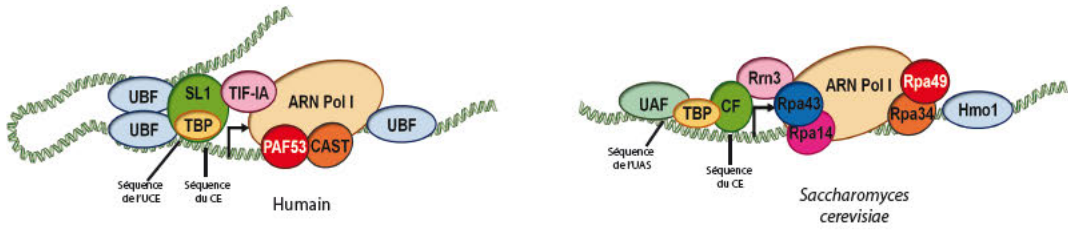
Chez *S.cerevisiae*., en effectuant des études de délétion de la région intergénique, les domaines nécessaires à la transcription par l'ARN polymérase (ARN pol I) ont été identifiés. En particulier, la région s'étendant du nucléotide (nt) -146 au nt +8 a été déterminée *in vivo* comme suffisante pour la transcription, et à ce titre définie comme la région promotrice de l'ARN pol I (Musters *et al*, 1989). Le domaine nommé CE pour "Core Element", ou "promoteur basal", est strictement essentiel pour l'initiation de la transcription (nt -28 à +8). Le domaine appelé UAS pour "Upstream Activating Sequence" (nt -148 à -50), possède une activité stimulatrice essentielle à la transcription *in vivo* (domaines indiqués préalablement Figure 7).

De façon similaire, des expériences ont également été menées chez les mammifères montrant que l'organisation du promoteur du gène d'ADNr chez les mammifères est bipartite. On retrouve le "Core Element" essentiel où le PIC est assemblé, en aval, allant de la position nt -45 à +18, ainsi que le domaine UCE ("Upstream Control Element") en amont comprenant la région des nucléotides -156 à -107 sans lequel le Core-Element n'est pas fonctionnel *in vivo* (Herdman *et al*, 2017). Bien qu'il n'y ait aucune conservation de séquence entre levure et humain, mise à part une forte proportion de déoxynucléotides A/T dans la région autour du +1 de la transcription, ces deux éléments ont des fonctions comparables dans ces deux organismes (Jones *et al*, 1988; Learned *et al*, 1986). On notera également que des travaux récents chez la souris ont montré que sans UBF, le PIC ne se forme pas (Herdman *et al*, 2017).

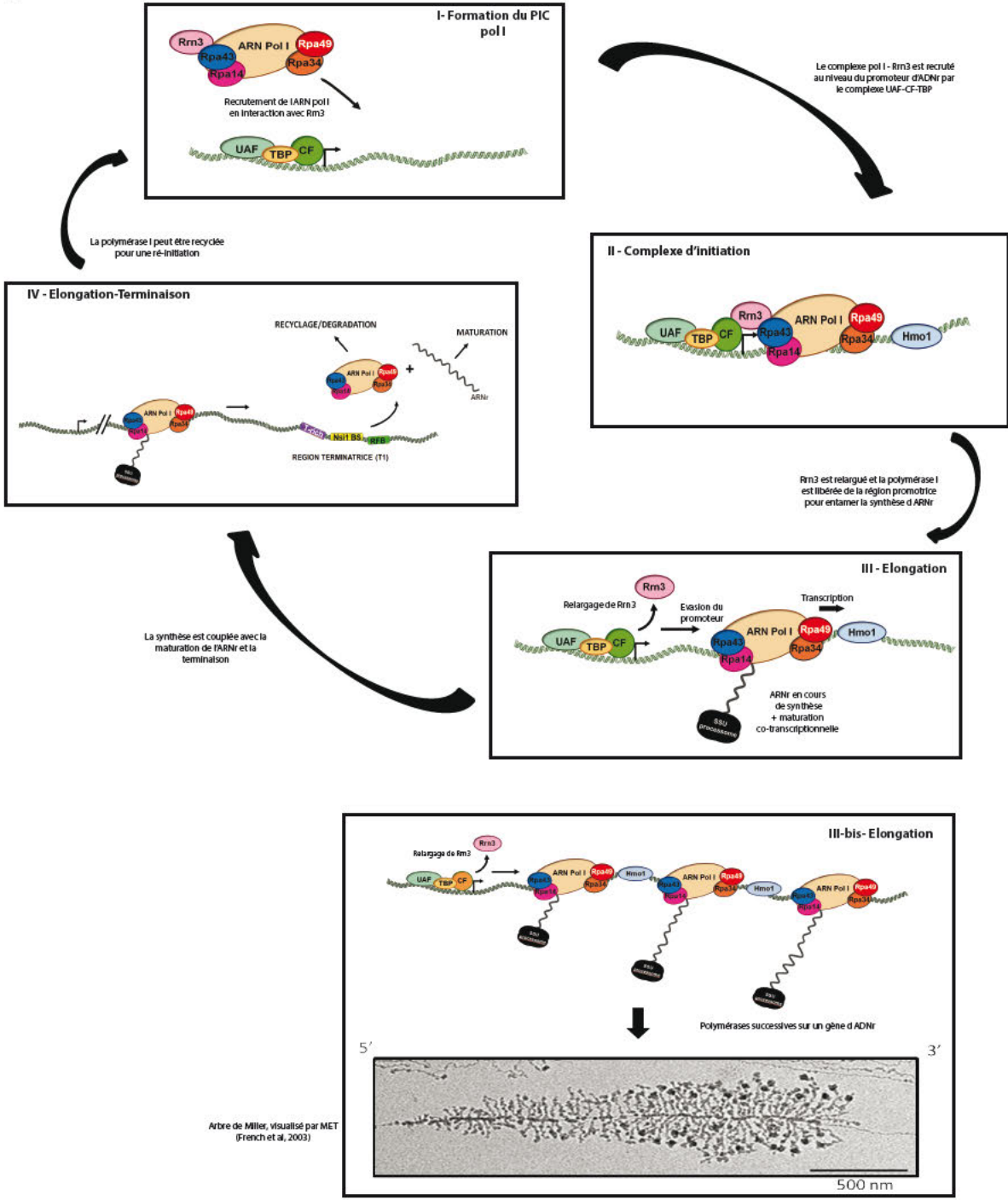
Le terminateur dans le cas de l'ARN pol I est défini comme une région nécessaire à la terminaison de la transcription, sur laquelle une protéine spécifique se fixe sur l'ADN pour

A.

Complexe de Pré-Initiation (PIC)



B.



**Figure 24. Cycle transcriptionnel de l'ARN pol I chez *S.cerevisiae*.**

A. Représentations schématiques des complexes de pré-initiation (PIC) sur l'ADNr, chez l'humain (à gauche) et chez *S.cerevisiae* (à droite). L'ARN pol I (beige), en interaction avec le facteur d'initiation Rrn3 (rose pâle) se lie à l'ADN via les facteurs d'initiation «Upstream Activating factor» (UAF, vert pâle) et «Core Factor» (CF, vert vif) en interaction avec la «TATA Binding Protein» (TBP, jaune). Les sous-unités des deux hétérodimères de l'ARN pol I ont été représentés : Rpa43-Rpa14 (bleu-rose vif) et Rpa34-Rpa49 (orange-rouge). Les facteurs/sous-unités correspondants chez l'humain sont représentés selon les mêmes codes couleurs que pour *S.cerevisiae*.

B. Schématisation du cycle transcriptionnel de l'ARN pol I en cinq encarts (I-, II-, III-, III-bis-, et IV-), reprenant les codes couleurs de l'encart A.

I- Formation du PIC : l'ARN pol I monomérique en complexe avec Rrn3 peut être recrutée sur l'ADNr.

II- Complexe de pré-initiation formé : l'ARN pol I est recrutée grâce aux complexes UAF et CF/TBP au niveau du promoteur d'ADNr.

III- Etape d'élongation processive de la transcription. Le passage du complexe d'initiation à celui d'élongation inclut la dissociation des facteurs d'initiation, et l'ARN pol I accrochée à l'ADNr sort de la région promotrice pour synthétiser le précurseur de l'ARNr 35S. L'ARN est représenté en noir. Au cours de cette activité de synthèse, une maturation intervient aussi, induisant l'apparition de boules terminales au bout des ARN synthétisés : les SSU processomes (complexes en noir).

III-bis- Même étape que III-, où sont représentés plusieurs ARN pol I successives en transcription sur un même gène d'ADNr, et leurs ARNr en cours de synthèse respectifs. Le panel du bas est un cliché d'un «arbre de Miller» en MET, où l'on visualise les mêmes éléments que sur le schéma du haut.

IV- Fin de la phase d'élongation et phase de terminaison de la transcription. L'ARN pol I arrivant à un terminateur, la synthèse de l'ARNr stoppe, le transcrit néo-synthétisé est relargué, et l'ARN pol I se décroche de l'ADNr, et peut être dégradée via le protéasome, ou recyclée pour un nouveau cycle de transcription.

agir principalement en tant que barrière capable de bloquer la translocation de l'ARN polymérase I : NsiI dans le cas de *S.cerevisiae* et TTF-I chez la souris. En termes de séquence, le terminateur, ou région terminatrice, est généralement noté T1, et est localisé 13 nucléotides en aval de la région transcrite par l'ARN pol I chez *S.cerevisiae* (Reeder & Lang, 1994). On constate une organisation similaire dans le cas des mammifères: le terminateur est composé d'une séquence répétée 8 fois (T1 à T8), située 21 nucléotides après la région transcrite (Grummt *et al*, 1985).

#### VI -2-b- La formation du complexe d'initiation de l'ARN pol I requiert des facteurs d'initiation qui lui sont spécifiques

De même que la polymérase archéenne ou les deux autres polymérases eucaryotes, il est acquis que l'ARN polymérase I de *S.cerevisiae* ne peut pas se fixer directement à l'ADN promoteur mais qu'elle doit reconnaître et se lier à un complexe multi-protéique assemblé sur le promoteur afin de former le Complexe de Pré-Initiation (PIC) (Figure 24.A. panel de droite). Il existe ainsi une séquence promotrice, où des facteurs d'initiation vont se fixer pour permettre l'assemblage du PIC et le recrutement efficace de l'ARN polymérase I, ainsi qu'une séquence activatrice de la transcription (UAS), requise pour stabiliser le PIC. Leur étude a permis une meilleure compréhension du mécanisme d'initiation pol I. Le "Core Element" est notamment la région de fixation d'un complexe protéique, le "Core Factor" (CF), qui permettra le recrutement efficace de l'ARN pol I. L' "Upstream Activating Sequence" permet le recrutement d'un deuxième complexe, l'Upstream Activating Factor (UAF).

#### Des éléments clés dans la formation du PIC de l'ARN polymérase I chez *S.cerevisiae* : "Core Factor" et "Upstream Activating Factor"

Chez *S.cerevisiae*, des études génétiques ont mis en évidence des mutants altérés pour la transcription de l'ADNr, mutants appelés *rrn* pour "rRNA synthesis defective" (Nogi *et al*,

1991). Cette méthode est basée sur un crible génétique de recherche de mutants ayant un défaut de synthèse d'ARNr. Elle utilise un système dans lequel le gène du pré-ARNr 35S est en fusion avec un promoteur GAL ("GAL7-35S rDNA") sur un plasmide multi-copies, et est transcrit par l'ARN pol II sous le contrôle du système GAL. Cette construction a par exemple permis, en présence de galactose, de permettre la survie de mutants où l'activité pol I est abolie, tel qu'un mutant de délétion du gène codant pour Rpa135, ou un mutant thermosensible de Rpa190 propagé à 37°C. Ainsi, via cette approche génétique, il a été possible d' identifier un grand nombre de gènes codant pour des protéines directement impliquées dans la synthèse de l'ARN 35S, telles que Rrn3, Rrn6, Rrn9, Rrn10 et Rrn11, mais aussi des sous-unités de l'ARN pol I : Rpa190, Rpa135, Rpa43, les sous-unités non-essentielles (Rpa49, Rpa34, Rpa14 ou Rpa12), où ayant une fonction essentielle en plus de l'activité de synthèse de l'ARNr comme les sous-unités communes (Rpb5, Rpb6, Rpb8, Rpb10 et Rpb12) ou partagées (Rpc40 et Rpc19), ou enfin le facteur non-essentiel Uaf30.

Ainsi, le recrutement de l'ARN pol I est directement dépendant de deux éléments protéiques, que sont le "Core Factor" (CF) et l' "Upstream Activating Factor" (UAF), en association avec la TBP (TATA Binding Protein) (Figure 24.A. à droite). Ces deux éléments cruciaux dans l'étape d'initiation contiennent les différentes protéines identifiées dans le crible génétique décrit ci-dessus.

Le CF est composé des trois protéines essentielles à la synthèse des ARNr par l'ARN pol I: Rrn6, Rrn7 et Rrn11. Des mutants de délétion (*rrn6Δ*, *rrn7Δ* ou *rrn11Δ*) de ces différentes protéines sont incapables d'initier correctement la synthèse de l'ARN 35S (Keys *et al*, 1994; Lalo *et al*, 1996). Le CF interagit avec l'UAF (Oakes *et al*, 1999), la TBP (Steffan *et al*, 1996), ainsi que le facteur Rrn3 (Peyroche *et al*, 2000; Pils1 *et al*, 2016; Engel *et al*, 2016; Torreira *et al*, 2017). La structure du CF de *S.cerevisiae* récemment obtenue par cristallographie aux RX (Engel *et al*, 2017) permet une étude plus approfondie de sa structure

et de ses interactions. Cette structure cristallographique révèle une organisation en deux modules (Modules I et II), reliés entre eux de façon flexible par une région charnière centrale (Figure 25.A.). Le Module I est formé d'un feuillet- $\beta$  à 7 brins  $\beta$  appartenant à Rrn6, ainsi que de Rrn11, qui forme une structure en répétition Tétratricopeptide (ou TPR, acides aminés qui constituent un motif structurel organisé en tandems d'hélices  $\alpha$  répétées, souvent impliqués dans les interactions protéine-protéine dans la formation de complexes protéiques (Blatch & Lässle, 1999)). Le Module II est constitué d'un domaine hélicoïdal de Rrn6 qui s'enroule autour de Rrn7. Le domaine C-ter de Rrn7 contient une insertion qui forme 6 hélices  $\alpha$  et permet de relier le Module I. Il est cependant à noter que des parties sont manquantes dans cette structure à haute résolution, telles que les parties mobiles N- et C-terminales de Rrn6, le N-ter de Rrn7, et une extension mobile du TPR de Rrn11. De plus, Rrn11 contient un domaine hélicoïdal d'association avec l'ADN ("Propeller/Promoter-Associated Domain", PAD). En effet, dans cette étude, ils ont pu déterminer les contacts entre l'ADN promoteur et les composants du PIC (CF et ARN pol I). Pour cela, ils ont utilisé un assemblage d'acides nucléiques basé sur la séquence naturelle du promoteur pol I de *S.cerevisiae*, composé de l'ADN de la position -80 à +23 (le +1 représentant le TSS), d'une bulle de transcription pré-ouverte (-8 à +8) et d'un produit ARN de 6 nucléotides. Ainsi, ils ont déterminé que la région d'ADN -35 à -25 interagissait avec les hélices  $\alpha_1$  et  $\alpha_2$  du domaine PAD de Rrn11, contactant respectivement le grand et le petit sillon de l'ADN, et que la région d'ADN proximale -20 à -12 était en interaction avec les domaines Protrusion et Wall de Rpa135 de l'ARN pol I elle-même.

De plus, TFIIB dans la pol II (et l'homologue archéen TFB) a été retrouvée comme ayant des similarités structurelles avec Rrn7 de l'ARN pol I, concordant avec les homologies de séquences préalablement prédites (Blattner *et al*, 2011; Naidu *et al*, 2011; Knutson & Hahn, 2011). Cependant, en modélisant le CF sur l'ADN sur la base d'une superposition de Rrn7 et

de TFIIB dans le complexe d'initiation pol II (Plaschka *et al*, 2016), il se trouve que Rrn7 ne peut pas se lier à l'ADN de la même manière que TFIIB, à cause de collisions stériques entre Rrn7 et l'ADN.

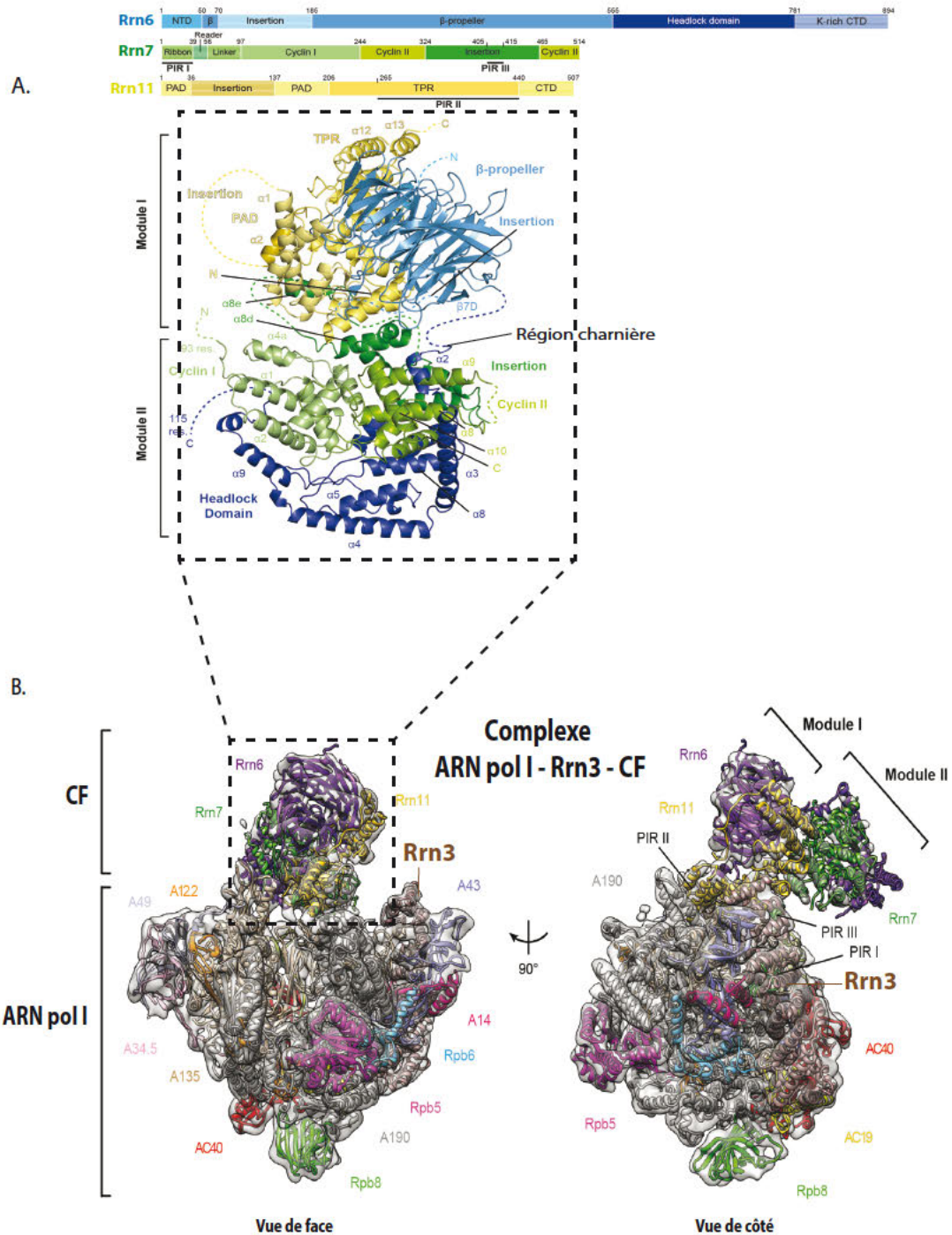
L'UAF, quant à lui est composé des facteurs essentiels Rrn5, Rrn9, Rrn10, des histones H3-H4 et de la sous-unité non-essentielle Uaf30. La déplétion des protéines Rrn5, Rrn9 et Rrn10 induit un très fort défaut de croissance des cellules (5-10 jours avant l'apparition de colonies) et une très forte diminution de la synthèse des ARNr 35S. *In vivo*, en l'absence de Rrn5, Rrn9 ou Rrn10, l'activité de synthèse d'ARNr par l'ARN pol I est abolie. Par contre, contrairement aux mutants de l'ARN pol I ou du Core facteur, une synthèse résiduelle d'ARNr est observée. C'est l'analyse de la sous-unité Uaf30 qui a permis de comprendre ce phénotype complexe. En l'absence de Uaf30, le complexe UAF est toujours capable de stimuler l'ARN pol I, mais dans ces conditions, près de 10% de la synthèse de l'ARNr est réalisée par l'ARN pol II (Siddiqi *et al*, 2001). La production par l'ARN pol II de pré-ARNr 35S, qui pourra être maturé et incorporé aux ribosomes, est donc possible à partir du locus endogène, mais exclusivement en l'absence de répression par l'UAF. Le complexe UAF a donc deux activités, la répression de l'ARN pol II, et la stimulation de l'initiation par l'ARN pol I. Contrairement à ce qui est observé *in vivo*, le complexe UAF n'est pas strictement essentiel pour la synthèse de l'ARNr par l'ARN pol I *in vitro* (Keener *et al*, 1998).

*In vitro*, l'initiation de la transcription est réduite dans des extraits cellulaires dépourvus de Rrn5, Rrn9 ou Rrn10, et elle est restaurée par addition du complexe UAF (Keys *et al*, 1996). Ceci n'est possible que si la région activatrice UAS est présente dans la région promotrice. L'effet stimulateur du facteur UAF est largement dépendant de la présence d'une autre protéine, la TBP ("TATA Binding Protein") (homologue TFA chez les archées). La TBP, n'est pas stricto sensu membre du complexe UAF ou CF, mais permet *in vitro* et *in vivo* d'agir telle une charnière entre ces deux facteurs (Lalo *et al*, 1996; Steffan *et al*, 1998). TBP se lierait à

l'ADN au niveau de la région -35 à -55 bp en amont du TSS. En effet, des expériences de biochimie et de biologie moléculaires ont montré l'interaction de la TBP avec l'UAF. En 1997, l'équipe d'Herbert Tschochner a observé la co-purification sur colonnes chromatographiques de Rrn9 de l'UAF avec la TBP (Milkereit *et al*, 1997). D'autres part, l'existence d'une interaction double-hybride entre Rrn9 et TBP (Steffan *et al*, 1998) corrobore cette observation

Le complexe massif comprenant la TBP, l'UAF et le CF, peut recruter l'ARN polymérase I, celle-ci étant déjà en interaction avec le facteur d'initiation Rrn3





**Figure 25. Complexe de Pré-Initiation (PIC) de l'ARN pol I chez *S.cerevisiae*. D'après Engel et al., 2017.**

A. Le panel du haut correspond à l'organisation schématique des domaines de chaque sous-unité du Core Factor (CF) : Rrn6 en bleu clair et foncé, Rrn7 en vert, et Rrn11 en jaune. Les sites d'interaction avec la polymérase (PIRs) sont indiqués en noir en-dessous du domaine correspondant. Le panel du bas représente la structure en cristallographie aux rayons X du Core Factor (CF) de *S.cerevisiae*, comprenant les sous-unités Rrn6, 7 et 11 selon le même code couleur, organisé en deux modules (modules I et II, en noir). Les régions mobiles et non résolues dans le cristal, tel qu'un domaine de Rrn11 (le PAD, pour «Promoter Associated Domain»), sont indiquées en pointillés selon la couleur de la sous-unité correspondante.

B. Deux orientations (tournées de 90°) de la reconstruction en Cryo-MET de la structure du complexe comprenant l'ARN pol I, Rrn3 et le CF. Les sous-unités du CF sont colorées telles qu'en A (sauf Rrn6 en violet). Rrn3 est représentée en marron, et les modules I et II du CF sont indiqués en noir.

### Données structurales sur le PIC de l'ARN pol I

Récemment, plusieurs études structurales en Cryo-MET ont permis d'énormes progrès dans la compréhension moléculaire des mécanismes de formation et de fonctionnement des complexes d'initiation de la transcription par l'ARN pol I (Engel *et al*, 2017; Han *et al*, 2017; Sadian *et al*, 2017). Notamment, la structure d'un complexe comprenant l'ARN pol I, Rrn3 et le CF en complexe avec l'ADN a été modélisée (Engel *et al*, 2017) (Figure 25.B.), permettant de proposer un modèle avec plusieurs états conformationnels successifs du complexe d'initiation :

1) La liaison du facteur Rrn3 stabilise l'ARN pol I sous une forme monomérique compétente pour l'initiation, caractérisée par un Cleft en expansion permettant l'accès au site actif (Blattner *et al*, 2011; Engel *et al*, 2016; Pilsel *et al*, 2016) .

2) Le Core Factor CF, grâce à son domaine PAD chargé positivement, reconnaît et se lie à un élément d'ADN en amont du TSS,

3) Via ses domaines de liaison à la polymérase (PIRs), le CF peut se lier à l'ARN pol I et charger l'ADN double brin fermé du promoteur dans le Cleft ouvert, formant le complexe fermé, ou "Close Complex, CC" (Figure 26.A.). Dans ce complexe, les domaines Protrusion et Wall de la seconde plus grande sous-unité Rpa135 sont en contact avec la région promotrice d'ADN.

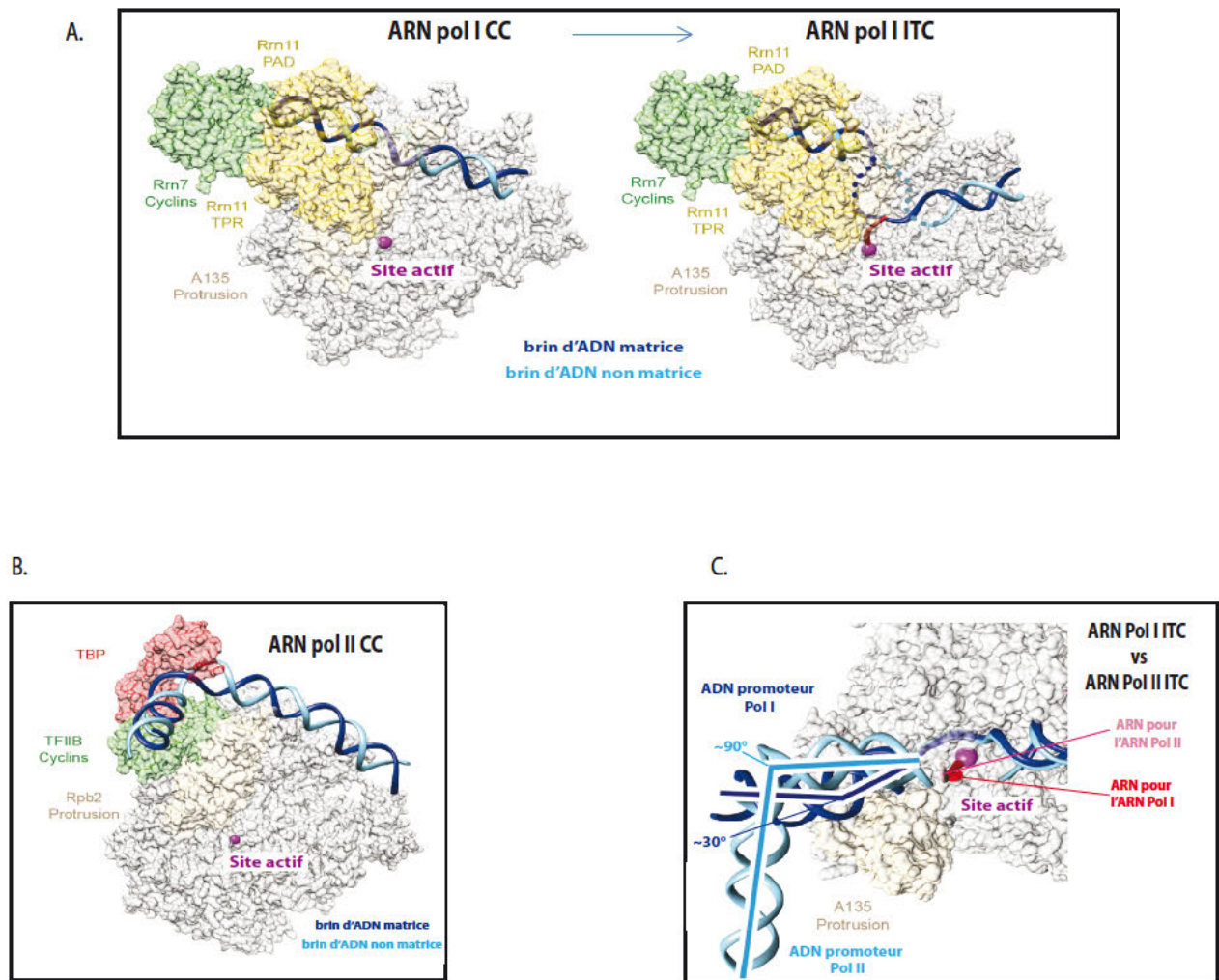
4) L'ADN double brin se déroule alors entre les domaines Protrusion et le Clamp de la polymérase, formant le complexe dit ouvert ("Open Complex, OC"). Ce complexe pourrait être stabilisé par la partie C-ter mobile de Rpa49 (domaine tWH) qui peut effectivement se lier à l'ADN (Geiger *et al*, 2010; Tafur *et al*, 2016).

5) La synthèse de l'ARN débute, amenant à la contraction du Cleft et à la formation du complexe de transcription initial ("ITC, pour Initially Transcribing Complex") (Figure 26.A.). Avec la longueur de l'ARN synthétisé augmentant, le domaine de Rrn7 qui occupait le site de

sortie de l'ARN de l'ARN pol I se déplace, induisant un changement de conformation du CF, et le passage vers un complexe en élongation ("EC, Elongating Complex") (voir aussi schéma récapitulatif Figure 28).

Ces études ont révélé des différences notables dans la structure du PIC de l'ARN pol I comparé à celle de l'ARN pol II. L'une des caractéristiques très intéressante est que, pour l'ARN pol I, l'ADN promoteur se déroule le long du Wall de Rpa135 et à travers le Cleft, alors qu'il se recourbe et est suspendu bien au-dessus du Wall dans le cas de l'ARN pol II, ce qui est dû à la liaison du domaine "Cyclin" de TFIIB entre l'ADN et le Wall (pour comparaison, voir les complexes "CC" Figure 26.A. (Pol I) et Figure 26.B. (Pol II)). De plus, les interactions entre Rrn7, Rrn11 et l'ADN induisent une courbure de l'ADN entrant dans la polymérase d'environ 30° seulement, comparé à l'ARN pol II où un angle de 90° est observé, induit par TBP-TFIIB (Figure 26.C.).

Il est important de mentionner que la TBP et surtout l'UAF essentiel *in vivo* pour l'activité de l'ARN pol I, ne sont pas présents dans ces reconstructions 3D, par manque de données structurales les concernant à l'heure actuelle. Leur présence au sein du complexe pourrait changer la conformation et la position de certains des éléments du PIC, ce qui reste encore à déterminer.



**Figure 26. Comparaison des structures des complexes fermés (CC) et initiaux (ITC) de la transcription par les ARN pol I et pol II chez *S.cerevisiae*.**

A. A gauche: modèle du complexe fermé («Close Complex, CC») de l'ARN pol I. A droite: modèle du complexe initial de transcription («Initially Transcribing Complex, ITC») de l'ARN pol I. Le site actif de chaque polymérase est indiqué en violet représentant l'ion métal catalytique. Les deux brins d'ADN (matrice et non-matrice) sont représentés en bleu foncé et clair respectivement.

B. Structure du complexe fermé (CC) de l'ARN pol II (pdb : 5F5Z). Sont représentés : la polymérase en gris en beige et le domaine Protrusion de la deuxième plus grande sous-unité Rpb2 ; le facteur TFIIB en vert, et son domaine «Cyclin» ; la TBP en rouge . Le site actif et les deux brins d'ADN (matrice et non-matrice) sont représentés tels qu'en A.

C. Comparaison de la position de l'ADN promoteur upstream des structures des complexes ITC («Initially Transcribing Complex») de l'ARN pol I et de l'ARN pol II (pdb : 4V1N), montrant une différences d'angle et donc de direction de l'ADN. La position de l'ARN est aussi représentée, en rouge et en rose pour l'ARN pol I et pol II respectivement, de même que le site actif pour les deux enzymes en violet.

Il est également à noter que même si la TBP est présente dans les préparations des différentes études, aucune densité n'a pu lui être attribuée en MET, suggérant que cet élément serait instable. Engel *et al* (2017) proposent entre autre que le positionnement de l'ADN au sein du PIC provoquerait l'apparition d'une conformation particulière du complexe, dans laquelle seuls les promoteurs capables de se déformer pourraient s'engager, impliquant plutôt la reconnaissance d'un état conformationnel particulier que la séquence du promoteur. D'autre part, l'une des différences les plus flagrantes entre ARN pol I et II concerne notamment le positionnement du CF lors de l'assemblage du PIC, malgré l'homologie entre Rrn7 et TFIIB. Les domaines "cyclin I et cyclin II" de Rrn7 ne contactent pas l'ARN pol I, contrairement à TFIIB où cette interaction est requise, ce qui en fait une différence majeure entre les deux enzymes pour la reconnaissance de l'ADN.

Il est également intéressant de noter que des différences de structure sont trouvées dans les trois modèles d'ARN pol I proposés récemment . Notamment, Engel *et al* (2017) et Sadian *et al* (2017) détectent la présence de Rrn3 dans leurs complexes OC et ITC, là où Han *et al* (2017) ne le détectent pas, suggérant sa dissociation précoce lors d'une étape de l'initiation (ou un problème de purification *in vitro* réalisée dans l'étude). D'autre part, même si nécessaire à l'assemblage du complexe, on remarquera que l'ARN produit n'est pas visible dans les reconstructions de Sadian *et al* (2017), suggérant que leur complexe serait un état pré-actif du complexe OC.

#### VI-2-c- L'équivalence de la formation du complexe d'initiation de l'ARN pol I chez les mammifères

L'organisation bipartite du promoteur de l'ADNr chez *S.cerevisiae* et les mammifères est relativement conservée, même si on ne retrouve pas réellement d'homologie de séquence au niveau des régions promotrices. De même que le facteur Rrn3, qui est très conservé, une

conservation fonctionnelle est probable pour les protéines impliquées dans la formation du complexe d'initiation chez les mammifères et *S.cerevisiae*. Chez l'Humain et la souris, des études biochimiques ont mis en évidence des protéines impliquées dans l'initiation, appartenant au complexe SL1 ("Selectivity Factor 1") chez l'Homme (nommé TIF-1B pour la souris) et à UBF (Figure 24.A. à gauche).

SL1 est un complexe de 300kDa, nécessaire à l'initiation de la transcription (Learned *et al*, 1985), structurellement conservé chez la souris, mais non-interchangeable *in vitro* avec son homologue murin. SL1 est composée de la TBP, elle-même associée à des facteurs particuliers, les TAF, pour "TBP associated Factors", tels que TAF<sub>110</sub>, TAF<sub>63</sub>, TAF<sub>48</sub> et TAF<sub>41</sub> chez l'Homme (Zomerdijk *et al*, 1994). En faisant le parallèle entre les systèmes de transcription humain et de *S.cerevisiae*, on peut comparer SL1 au Core Factor de levure, tous deux essentiels, même s'ils ne possèdent qu'une faible homologie de séquence. Rrn7 et TAF<sub>63</sub> partageraient cependant des homologies structurales et seraient fonctionnellement équivalentes au facteur de transcription pol II TFIIB (Knutson & Hahn, 2011). On retrouve ainsi l'homologue de Rrn6 de levure (TAF1C), de Rrn7 (TAF1B), de Rrn11 (TAF1D) (Russell and Zomerdijk, 2006), additionnées de deux sous-unités TAF1D et TAF12. SL1 interagit de plus avec l'homologue vertébré de Rrn3 (hsRrn3 pour l'humain, ou TIF-1A pour la souris),.

UBF ("Upstream Binding Factor") quant à elle est une protéine à cinq boîtes HMGB, constituée de deux polypeptides (UBF1 et UBF2). *In vitro*, ces deux peptides peuvent se fixer au promoteur des gènes d'ADNr sous forme d'hétéro- ou d'homodimères, bien que UBF1 soit un activateur transcriptionnel plus efficace qu'UBF2. On retrouve UBF au niveau des séquences promotrices des gènes d'ADNr, et elle coopère avec SL1 pour organiser le complexe d'initiation. Elle permet de stimuler la transcription *in vitro* (Bell *et al*, 1990), aurait de l'affinité pour la séquence activatrice (UCE) et permettrait le rapprochement de l'UCE avec

le CE . Notamment, en se fixant à l'ADN, elle entrainerait la création d'une boucle sur l'ADN, formant ainsi une structure favorable à l'assemblage du complexe de pré-initiation (structure appelée "enhanceosome").

#### VI-2-d- La transcription par l'ARN pol I : l'élongation de la transcription

Durant l'initiation, l'ADN double brin doit être ouvert, pour permettre l'accès au brin transcrit. Dans le cas de la pol II, ce processus requiert l'activité hélicase de TFIIH ATP dépendante, ce qui ne semble pas nécessaire ni pour l'ARN pol I ni l'ARN pol III. Le phénomène dit de "translocation" détermine la fin de la phase d'initiation de la transcription, et le début la transcription abortive pendant laquelle la synthèse de quelques nucléotides est effectuée. Une fois que l'hybride ADN-ARN est suffisamment long, le complexe passe dans la phase d'élongation de la transcription à proprement parler. Elle est définie par la séparation de l'ARN pol I du complexe d'initiation, et par la synthèse productive des premiers nucléotides. Comme pour toutes les ARN polymérase multimériques, l'élongation par l'ARN pol I n'est pas continue. En effet, dépendant de la topologie de l'ADN ou de la chromatine, la pol I ne se déplace pas à vitesse constante, et peut s'arrêter de façon transitoire dans sa progression. On peut alors mentionner deux paramètres de l'élongation de la pol I. Le premier est la vitesse d'élongation, à savoir le nombre de nucléotides polymérisés par unité de temps (souvent en secondes). Le second est la processivité, terme qui fait référence au nombre de nucléotides polymérisés par événement d'initiation, sans dissociation de l'enzyme. Tout au long de la phase d'élongation, l'ARN pol I interagit donc simultanément avec l'ADN qui lui sert de matrice, l'hybride ADN-ARN, et le transcrit ARN naissant.

Des données structurales récentes en Cryo-MET en molécules uniques ont par ailleurs permis de caractériser plus précisément les principaux remaniements structuraux au cours de l'élongation, entre l'ARN pol I et les acides nucléiques (ADN et ARN) (Tafur *et al*, 2016;

Neyer *et al*, 2016). Ainsi, Tafur *et al* (2016) ont obtenu quatre structures de complexes avec l'ARN pol I dans leurs conditions expérimentales de Cryo-MET, dont un complexe ouvert ("Open complex OC"), deux complexes en élongation de la transcription ("EC1 et EC2"), et un dernier complexe en élongation en présence du domaine tWH de Rpa49 ("EC\_tWH"), qui n'était pas présent dans les trois premiers, mais qui représente une plus faible proportion des molécules totales analysées. Pour obtenir ces complexes *in vitro*, ils ont utilisé deux différentes constructions d'acides nucléiques permettant de mimer la région d'ADN en transcription (la "bulle de transcription"). Un premier assemblage de 38 paires de bases (pb) contenant une bulle de transcription de 11 nucléotides (nt) et un oligonucléotide d'ARN de 20 nt, a permis la reconstruction d'un seul modèle, l'EC1. Un second assemblage plus long, de 70 pb, contient la séquence promotrice sauvage de l'ADNr, contenant une bulle de 15 nt, et un oligonucléotide d'ARN de 10 nt. Cet assemblage a permis d'obtenir les trois autres complexes (OC, EC2 et EC\_tWH) (les assemblages d'acides nucléiques sont indiqués Figure 27.A.). Dans les reconstructions des complexes en élongation (EC1, EC2 et tWH), plusieurs changements conformationnels peuvent être notés, notamment en comparant avec la structure du complexe ouvert (OC) (Figure 27.B.) : la région de Cleft qui va lier l'ADN se resserre progressivement jusqu'à adopter une position fermée en élongation ; la Bridge Helix (BH) change de conformation, se rapprochant plus de celle des autres ARN polymérases multimériques ; le domaine "Expander" ou DNA mimicking Loop" de la polymérase occupant le site de liaison à l'ADN se déplace ; le domaine Cterminal de Rpa12 est également déplacé, car aucune densité électronique ne peut lui être attribuée au niveau de sa position proche du site actif dans le Cleft (Figure 27.B.) ; et enfin le domaine tWH de Rpa49 contacte l'ADN entrant dans la polymérase, d'une manière similaire à TFIIE de l'ARN pol II (Figure 27.C.). Le complexe OC correspond à une conformation de la polymérase liée à l'ADN avant que la synthèse d'ARN n'ait débuté, ou, dans un autre cas, à une conformation dans laquelle l'ARN a



été clivé due à l'activité intrinsèque de clivage de l'ARN pol I. Les deux complexes EC 1 et EC2 sont très similaires conformationnellement, même si l'assemblage en acides nucléiques utilisé n'est pas le même. Dans le complexe EC\_tWH, une densité importante supplémentaire apparaît dans la reconstruction, entre le Stalk et l'ADN entrant (Figure 27.C.). Le remplacement moléculaire du domaine tWH de Rpa49 (pdb : 3NFI) obtenu en cristallographie aux RX, au sein de cette densité additionnelle, correspond. Notamment, son domaine WH2 et sa queue Cterminale, connus pour pouvoir lier l'ADN (Geiger *et al*, 2010), se placent juste à côté de ce brin d'ADN entrant dans la reconstruction. Concernant ce dernier point, il est intéressant de noter que cette position de Rpa49tWH dans un complexe EC corrèle avec la position trouvée par Engel *et al* (2017) dans leur étude sur la formation du PIC. En effet, dans cette étude, une densité correspondante à celle du tWH est retrouvée dans les classifications de Cryo-MET pour l'ITC, également contactant l'ADN entrant. Ceci suggère que le domaine flexible tWH de Rpa49 bouge avec l'ADN entrant durant la transition de l'ITC à l'EC. Dans ce complexe EC\_tWH, de même que dans les EC1 et EC2, on ne retrouve jamais de densité correspondant au Cter de Rpa12, alors que celles correspondantes au Nter et au linker sont toujours présentes. Au contraire, dans le complexe OC, toutes les densités de Rpa12 sont retrouvées. Ces données structurales suggèrent fortement que les domaines Cter des sous-unités Rpa49 et Rpa12 sont des éléments mobiles, pouvant adopter différentes conformations et jouer différents rôles durant la transcription par l'ARN pol I.



### Figure 27. Complexes d'élongation de l'ARN polymérase I.

A. Schématisation des deux assemblages ADN-ARN : de 38 paires de base (pb) (à gauche), utilisé pour l'Elongating Complex 1 (EC1), et de 70 pb (à droite) pour l'Open Complex (OC), l'EC2 et l'EC\_tWH. Les domaines d'interaction avec les sous-unités Rpa190 et Rpa135 de la polymérase sont indiqués en gris en beige respectivement. Les cercles pleins représentent les nucléotides modélisés, les cercles vides les non-modélisés.

B. Représentations schématiques de différents complexes de l'ARN pol I : l'ARN polymérase I Apo sans acide nucléique, sous forme dimérique (à gauche), l'ARN pol I en OC (au milieu) et celle en EC (à droite). La sous-unité Rpa12 est représentée en jaune : présente dans son intégralité pour les deux premiers complexes, et sans sa partie Cterminale dans le cas de l'EC. La Bridge Helix (BH) est également représentée en gris foncé, adoptant différentes conformations. Les flèches noires représentent la contraction du Cleft. L'élément Expand, ou DNA Mimicking Loop, n'est présent que dans la forme sans ADN (gauche).

C. Modélisation du complexe EC\_tWH (à gauche). La densité électronique correspondante au domaine tWH de Rpa49 est superposée en transparence avec la structure cristallographique de Rpa49-Cter (pdb : NFI). Les panels de droite zooment sur la région du tWH, montrant notamment sa queue C-terminale et ses domaines de liaison à l'ADN. D'après Tafur et al., 2016.

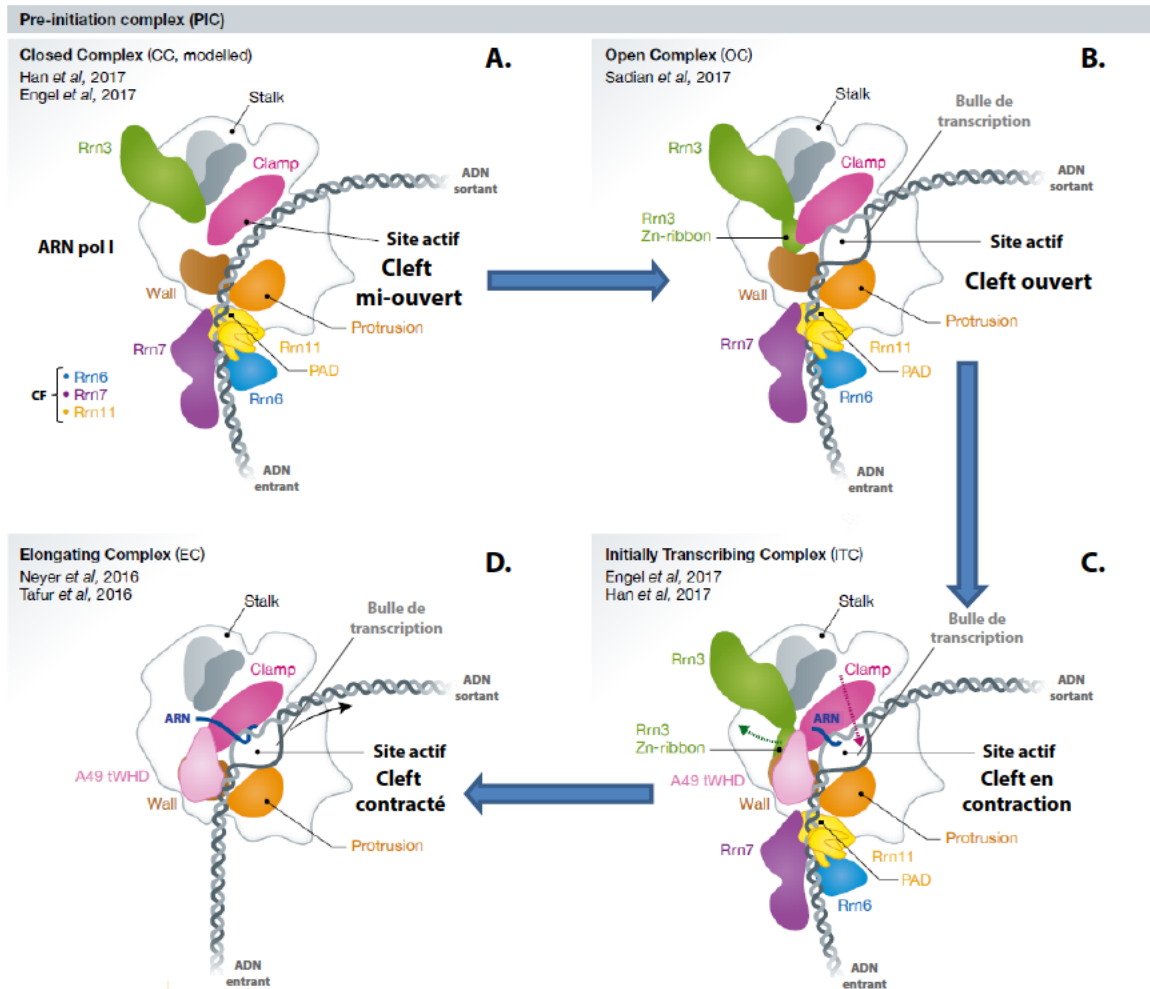
D. Le panel en haut à gauche représente la superposition moléculaire d'une ARN pol I en complexe d'élongation obtenue en Cryo-MET à 3,8Å en molécule unique (représentée sous forme de rubans) et de la reconstruction en Cryo-MET en tomographie de la densité d'une polymérase en transcription (surface en gris clair). Le panel en bas à gauche est une schématisation de trois polymérases successives en transcription sur l'ADNr, formant un angle d'environ 150°. La reconstruction en trois dimensions de ces ARN polymérases I successives a été effectuée (panel de droite), modélisant un arbre de Miller : les polymérases sont indiquées en orange, l'ADN en bleu foncé, et l'ARN en cours de synthèse en rouge. De plus, les densités observées au bout de ces ARNr (gris pâle) correspondent aux boules terminales, ou «SSU processomes». Enfin, les possibles complexes modificateurs de l'ARN naissant sont indiqués en bleu clair. Adapté de Neyer et al., 2016.

En parallèle, Neyer *et al* (2016) ont également obtenu des structures d'ARN pol I actives en élongation de la transcription, par deux méthodes de Cryo-MET. Premièrement, par Cryo-MET en molécules uniques, ils ont obtenu une structure à haute résolution (3,8Å) de l'ARN pol I en élongation en complexe avec un assemblage d'ADN et d'ARN, comparable aux complexes d'élongation décrits dans Tafur *et al*, 2016. Les mêmes caractéristiques sont ainsi retrouvées : une contraction du Cleft, via les modules Core et Shelf-Clamp de la polymérase ; la Bridge Helix change de conformation, et la densité électronique correspondante à la Trigger Loop devient plus faible, montrant que cet élément est également mobile ; l'élément "Expander" est déplacé, permettant à l'ARN pol I d'interagir avec l'hybride ADN-ARN ; enfin le domaine Cter de Rpa12 se déplace depuis son pore durant la contraction du Cleft. Cette modification de conformation est centrale : la modélisation de ce domaine dans le pore provoque une collision avec le module Shelf contracté. Par ailleurs, en comparant les structures d'élongation des ARN polymérases bactérienne, pol II et pol III, avec celle de l'ARN pol I obtenue, on remarque que toutes ces enzymes adoptent un état conformationnel de la transcription contracté, montrant que des bactéries aux eucaryotes, les mécanismes fondamentaux de la transcription ADN dépendante sont conservés.

Deuxièmement, en alliant des techniques de Miller's spreads et de tomographie en Cryo-MET, ils ont permis de caractériser structurellement plusieurs polymérases successives sur un même ADNr cellulaire, ajoutant un degré d'analyse structurale supplémentaire intéressant (Figure 27.D.). A partir de cellules de *S.cerevisiae* en phase exponentielle de croissance, l'ADN est étalé sur des grilles de MET, de façon similaire à la technique de "Miller spread" expliqué Chapitre II-1 de ce manuscrit. La différence majeure repose dans le fait qu'ils congèlent instantanément les grilles, entourant les échantillons de glace vitreuse, afin de s'affranchir des limitations de la technique conventionnelle, notamment la déshydratation et l'utilisation de contrastants. Cela permet d'observer les échantillons dans des conditions ex-

vivo, mais qui se rapprochent un peu plus des conditions natives. Les images et reconstructions obtenues permettent de visualiser et de révéler le positionnement des ARN pol I le long de l'ADNr, ainsi que les ARNs naissants qui en émergent, mais aussi la présence des boules terminales (SSU processomes) en 5' des ARNs synthétisés. Dans ces conditions expérimentales, environ 70 polymérases sont observées par gène d'ADNr, mais seulement 11 arbres de Miller entiers ont été obtenus, permettant une première approximation du nombre d'ARN pol I par gène. Toujours selon cette étude, la distance entre deux polymérases successives serait de 18 +/-10 nm. Il est notamment intéressant de noter que l'ADNr entrant et sortant de trois polymérases successives forme un angle ouvert de 150° environ.

Pour résumer, le PIC adopte différents états durant l'initiation de la transcription: premièrement, le PIC en complexe fermé ("CC, Close Complex") se lie au promoteur d'ADN double-brin non-déroulé. Puis la transition du CC au complexe ouvert ("OC, Open Complex") s'effectue via le déroulement de l'ADN double brin et l'insertion du brin d'ADN matrice au sein du site actif. La synthèse de l'ARN débute alors, formant le complexe initiateur de la transcription, encore lié aux facteurs d'initiation ("ITC, Initially Transcribing Complex"). La dissociation de ces facteurs d'initiation et la formation d'un hybride ADN-ARN assez long permettent enfin la formation du complexe d'élongation ("EC, Elongating Complex"). Un schéma global récapitulatif de ces étapes d'initiation et de passage à l'élongation par l'ARN pol I, et les complexes associés, est proposé Figure 28 (adapté de (Jochem *et al*, 2017)).



**Figure 28. Schéma global récapitulatif des différents états du Complexe de Pré-Initiation (PIC) formés par l'ARN pol I et ses facteurs associés (A, B et C.), jusqu'à la formation d'un complexe en élongation (D.) (d'après Jochem et al, 2017).**

Les facteurs d'initiation sont représentés: Rrn3 (en vert) ; les composants du Core Factor CF : Rrn6 en bleu, Rrn7 en violet, Rrn11 et son domaine acide PAD en jaune. Les éléments structuraux majeurs de l'ARN pol I participant à ces étapes d'initiation sont également représentés : le domaine Wall (marron) et le domaine Protrusion (orange) de Rpa135, le domaine Clamp (rose). Le Stalk de la polymérase est aussi indiqué. L'ADN entrant et sortant est représenté en gris, et l'ARN naissant en bleu foncé. La position du domaine tWH de Rpa49 est de plus représentée en rose clair dans les deux derniers complexes.

A. Complexe fermé («Close Complex, CC»). L'ADN double brin lié au CF, est inséré entre les domaines Wall et Protrusion de l'ARN pol I en interaction avec le facteur Rrn3, ce qui induit une courbure de l'ADN de 30°.

B. Complexe ouvert («Open Complex, OC»). L'ADN double brin est déroulé, et la bulle de transcription se forme au niveau du site actif de la polymérase, stabilisée par le domaine «Zn-ribbon» Rrn3.

C. Complexe initiateur de la transcription («Initially Transcribing Complex, ITC»). L'ARN pol I se détache de la région promotrice, ce qui est dû à la contraction du Cleft et la dissociation des facteurs d'initiation, toutes deux progressives, ainsi qu'au déplacement du domaine Zn-ribbon de Rrn3. Le Clamp se resserre autour de la bulle de transcription, et la synthèse de courts oligonucléotides débute.

D. Complexe en élongation («Elongating Complex, EC»). Le complexe en élongation se forme grâce à la dissociation complète des facteurs d'initiation et à la contraction totale du Cleft. La synthèse du précurseur de l'ARNr peut se faire de manière processive.

## VI-2-e- La transcription par l'ARN pol I : la terminaison de la transcription

Ce processus crucial comprend en réalité deux événements majeurs: la terminaison à proprement parler, à savoir la fin de la synthèse et le relargage du transcrit néo-synthétisé, mais également le recyclage efficace de l'ARN pol I. Cet événement est déclenché au niveau des terminateurs, des séquences d'ADN où peuvent se fixer des facteurs protéiques spécifiques tels que Nsi1 et Reb1 chez *S.cerevisiae*, ou TTF1 ("Transcription release Factor for Pol I") chez la souris. Ces protéines possèdent effectivement un domaine de liaison à l'ADN conservé, et sont interchangeable *in vitro* (Kuhn *et al*, 1990). Lorsque l'accrochage au site de liaison s'effectue, la polymérase en cours d'élongation doit s'arrêter et ainsi promouvoir la terminaison proche d'une région riche en pyrimidines. Chez *S.cerevisiae* ou chez la souris, la terminaison s'effectue respectivement 16 ou 13 nucléotides en amont de ce site de liaison. Il a été initialement suggéré qu'un défaut de terminaison pouvait être associé à des mutations dans le gène *REB1* (Reeder *et al*, 1999). De fait, chez la levure *S.cerevisiae*, il a été déterminé que l'élément d'ADN minimal permettant une terminaison de l'ARN pol I efficace *in vitro*, contenait un site de liaison proche du consensus reconnu par la protéine Reb1, long de 11 pb, au niveau du 3', ainsi qu'un élément de 10-15 pb riche en T, 12 à 20 pb en amont de ce site de liaison, nommé site T-rich 1 (Lang *et al*, 1994). Il est à noter que Reb1 étant un facteur de transcription essentiel de l'ARN Pol II (Ju *et al*, 1990), son implication directe dans la terminaison de l'ARN pol I a été expérimentalement difficile à établir *in vivo*. On sait maintenant que Nsi1, un paralogue de Reb1, est la protéine se liant au terminateur de l'ARN pol I. En effet, même si le site présent *in vivo* est proche de celui reconnu par Reb1, Reb1 n'est pas recrutée pour autant à ce terminateur. Seule la mutagenèse de ce site vers le consensus de Reb1 permet le recrutement cette protéine. Par contre, Nsi1 reconnaît ce site, et contribue directement à la terminaison *in vivo* (Reiter *et al*, 2012; Merkl *et al*, 2014). Par

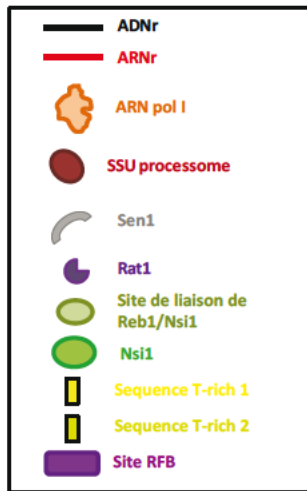
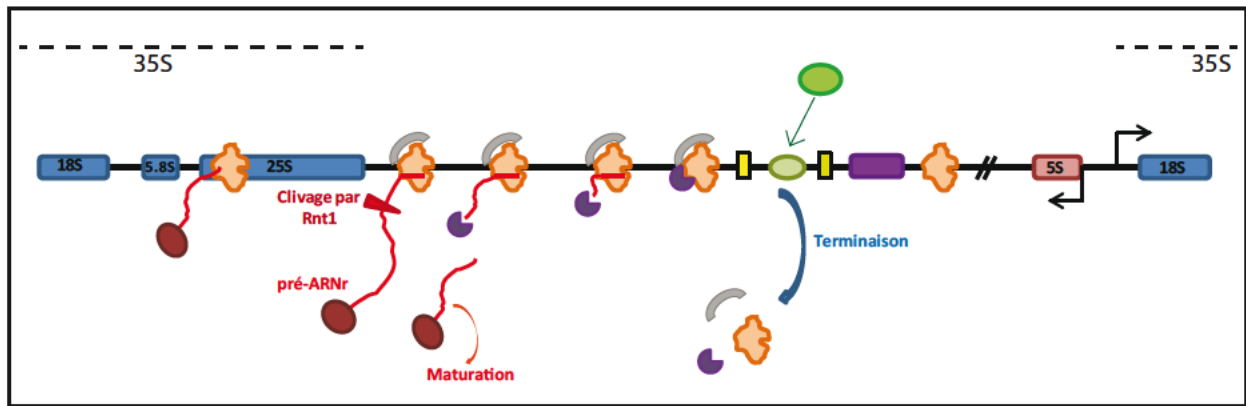


cohérence avec la littérature existante sur le site permettant la terminaison, (anciennement décrit comme le site de liaison de Reb1), je décrirai ici ce site comme Reb1/Nsi1.

Lors de la terminaison, la présence seule de ces protéines n'est pas suffisante pour la dissociation du complexe ternaire ADN-ARN-Pol I, comme montré sur des matrices immobilisées d'ADN (Lang *et al*, 1994). Des facteurs additionnels sont requis pour que la terminaison s'effectue efficacement *in vivo*. En effet, *in vitro*, seuls 20% des transcrits générés par la pol I subissent une terminaison et un relargage normaux, tandis qu' *in vivo*, 70% du nombre total des transcrits voient leur terminaison s'effectuer au niveau d'une région riche en T juste après le site de liaison de Reb1/Nsi1, le site T-rich 2 (Reeder *et al*, 1999). Des éléments d'ADN en cis supplémentaires ont ainsi été identifiés. Notamment, il a été montré que l'inactivation de l'endonucléase Rnt1 stabilisait les transcrits n'ayant pas effectué de terminaison, donc encore accroché à la polymérase après le site terminateur. Il a alors été suggéré que le clivage de l'ARN néo-synthétisé au site de Rnt1 (32-65 pb en amont du site riche en T, soit 59-92 pb en amont du site de liaison de Reb1) facilitait la terminaison. De plus, Fob1, une protéine se liant au RFB ("Replicating Fork Barrier") et bloquant les fourches de réplication de l'ADNr, a été suggérée comme un système alternatif impliquée dans la terminaison (El Hage *et al*, 2008). Un modèle sur le mécanisme de terminaison pol I a ainsi été proposé (El Hage *et al*, 2008; Kawauchi *et al*, 2008) (Figure 29), également inspiré de la terminaison de la transcription Pol II ("Torpedo-like mechanism"): la pol I en élongation effectuerait une pause au voisinage du site de liaison de Reb1/Nsi1, puis Rnt1 cliverait co-transcriptionnellement l'ARN naissant. L'exonucléase Rat1, en association avec l'hélicase Sen1, interviendrait alors pour dégrader la région 5' résultante du clivage par Rnt1, jusqu'à dégradation complète de l'ARN encore associé à la Pol I, entraînant par la suite la déstabilisation et le décrochage de l'ARN pol I et des facteurs associés. Il a également été suggéré qu'un dernier facteur, la poly-nucléotide kinase Gcr3, maintiendrait la

phosphorylation sur le produit de clivage de Rnt1 et permettrait l'action de Rat1 (une exonucléase 5'→3'), concluant efficacement la terminaison (Braglia *et al*, 2011).

Si les facteurs comme Nsi1, Fob1, Rat1 et Gcr3 sont requis pour la terminaison, le lien fonctionnel avec Rpa12, présumée faciliter la terminaison (Prescott *et al*, 2004), n'est à ce jour pas totalement établi.



Adapté de Kawauchi et al., 2008.

### Figure 29. L'étape de terminaison de l'ARN pol I.

Schématisation d'un modèle de terminaison de type «Torpedo», ressemblant à celui de l'ARN pol II. L'ARNr naissant (rouge vif) synthétisé par l'ARN pol I (orange) est clivé par Rnt1 (triangle rouge) co-transcriptionnellement. Le pré-ARNr est alors relargué pour être ensuite mûré. L'exonucléase 5'-3' Rat1 (en violet) est alors recrutée à l'extrémité 5' de l'ARN clivé et encore accroché à la polymérase, afin de le dégrader progressivement, aidée de l'activité hélicase de Sen1 (gris). La disparition totale de cet ARN en dégradation, et la fixation d'une protéine de type Nsi1 au site de liaison de la protéine Reb1 (en vert) permettraient de conclure efficacement la terminaison de la transcription. RFB = Replicating Fork Barrier.

## **Chapitre VII - Analyse fonctionnelle *in vivo* des modules structuraux de**

### **l'ARN pol I.**

Nos connaissances du système de transcription de l'ARN pol I en particulier grâce aux études structurales récentes constituent des avancées majeures dans la compréhension de la synthèse des ARNr *in vivo*. Pourtant, les observations *in vivo* sont parfois surprenantes, et peuvent différer fondamentalement des observations de phénotypes associés à des mutants dans les systèmes paralogues des ARN pol II et III. L'impact et le rôle de certaines sous-unités de l'ARN pol I a été étudié *in vivo* chez *S.cerevisiae*, et ce notamment via l'utilisation de mutants de ces sous-unités, conférant à certaines souches de levure des phénotypes variés et analysables. Nous parlerons plus concrètement dans ce chapitre des mutants de certaines sous-unités, dont les BTFs de l'ARN pol I, afin d'introduire plus en détails les résultats obtenus durant cette thèse et les perspectives possibles.

### **VII-1- Analyses des phénotypes associés à des mutants de certaines sous-unités de l'ARN**

#### **pol I.**

##### **VII-1-a- Le Stalk Rpa14-Rpa43 et Rrn3**

###### **Rpa14**

Rpa14 est une petite sous-unité non-essentielle, la seule sous-unité sans orthologue identifiée chez les métazoaires. Un mutant de délétion de Rpa14 n'est pas létal, mais a un défaut de croissance à 37°C (phénotype de thermosensibilité) (Smid *et al*, 1995). Lors de la purification de l'ARN Pol I, l'absence de Rpa14 induit la déstabilisation et la perte des sous-unités Rpa43 et ABC23, indiquant que cette sous-unité aurait un rôle non-essentiel dans la stabilité de ces sous-unités et de l'enzyme plus globalement. De plus, l'inactivation du gène codant pour

Rpa14 provoque une létalité synthétique (SL, "Synthetic Lethal") lorsque les gènes codant pour les sous-unités non-essentiels Rpa49, Rpa34 ou Rpa12 sont inactivés. Si le triple mutant portant les trois invalidations (*rpa49Δ* + *Rpa34Δ* + *Rpa12Δ*) est viable, les doubles mutations *rpa14Δ+rpa34Δ*, *rpa14Δ+rpa49Δ* et *rpa14Δ+rpa12Δ* sont létales. Le triple mutant *rpa34Δ+ rpa49Δ+ rpa12Δ* présente néanmoins un très fort défaut de croissance et une instabilité phénotypique importante.

De plus, il a récemment été montré que l'interaction entre l'ARN pol I et le facteur Rrn3 était diminuée de 75% dans des mutants *rpa14Δ*, ce qui n'est pas le cas dans une souche déletée uniquement de la queue C-terminale de Rpa14 (*rpa14ΔCter*,  $\Delta 101-137$ ) (Torreira *et al*, 2017). De façon surprenante, les mêmes auteurs ont également montré qu'une souche dans laquelle seule une boucle interne de Rpa14 ("TA-loop", désordonnée dans le cristal de l'ARN pol I dimérique) est déletée (*Rpa14ΔTAloop*;  $\Delta 53-77$ ), semble à l'inverse stimuler l'interaction Pol I - Rrn3 et leur association au promoteur de l'ARN pol I (observé par ChIP *in vivo*), suggérant que cette TA-loop de Rpa14 aurait pour rôle de limiter l'interaction Pol I - Rrn3.

### Rpa43

L'autre sous-unité du Stalk, Rpa43 est essentielle. Elle possède dans sa partie Cterminale une hélice (le "Connector") qui est nécessaire pour la dimérisation de deux ARN pol I. De plus, la sous-unité Rpa43 est également cruciale pour la formation du complexe ARN pol I-Rrn3. Des délétions de la partie C-terminale de Rpa43, comprenant cet élément Connector, induisent un phénotype de thermosensibilité, et empêchent l'interaction Pol I - Rrn3 (Peyroche *et al*, 2000). Les mêmes auteurs ont montré que les ARN pol I purifiées et mutées pour Rpa43 possédaient un fort défaut de transcription *in vitro*. De plus, la délétion du C-ter de Rpa43 réalisée par Torreira *et al* (2017) empêche également l'homodimérisation des ARN pol I (*A43DCt*,  $D307-326$ ).

### Rrn3

*In vivo*, il est possible d'interférer avec les mécanismes de contrôle de l'ARN pol I médiés par l'interaction Pol I - Rrn3, via la sous-unité Rpa43 du stalk. En effet, une souche de levure portant un gène codant pour une protéine de fusion fonctionnelle Rrn3-Rpa43, nommée CARA (Constitutive Association of Rrn3 and A43) a été créée (Laferté *et al*, 2006). Cette souche ne permet plus l'homodimérisation des ARN pol I via la surface d'interaction Rpa43-Cleft de deux polymérasés adjacentes, mais cette construction indissociable est suffisante pour compléter une souche de levure dans laquelle les gènes essentiels *RRN3* et *RPA43* ont été inactivés. Cette souche a été utilisée plus récemment dans une autre étude *in vivo*, dans laquelle la croissance cellulaire était stoppée par carence nutritionnelle (Torreira *et al*, 2017). Dans ce cas, les cellules induisent une dissociation rapide du complexe Pol I-Rrn3, ainsi que la formation d'homodimères d'ARN pol I inactifs en transcription, mécanismes réversibles lors de l'addition de nutriments dans le milieu, restaurant la croissance cellulaire. Les mêmes auteurs ont par ailleurs montré par CHIP que ce complexe chimérique Pol I-Rrn3 était normalement associé à l'ADNr dans des cellules en croissance normale, et était 6 à 7 fois plus associé dans des conditions de carence nutritionnelles. Ces données suggèrent que l'homodimérisation de l'ARN pol I est importante pour la dissociation complète du complexe Pol I-Rrn3 dans des conditions de carence nutritionnelle, et pour une inactivation complète de la transcription.

### VII-1-a- Les Built-in Transcriptions factors Rpa34-Rpa49 et Rpa12

Comme vu Chapitre VI-1-e, Rpa34 possède une longue queue Cterminale qui l'ancre à la polymérase, et les deux sous-unités Rpa49 et Rpa12 s'organisent en deux modules, N- et C-terminaux, reliés par un linker. Des phénotypes de croissance très différents, associés à des

modifications de production d'ARNr, peuvent être attribués à la délétion de toute ou une partie de ces sous-unités.

### Rpa34 et Rpa49

Rpa34 et Rpa49 forment un hétérodimère via leurs régions N-terminales, mais les approches structurales montrent clairement qu'ils représentent deux modules fonctionnels différents, reliés par un long linker nécessaire et suffisant pour l'association à l'ARN pol I. Si les mutants portant une délétion entière de Rpa34 (*rpa34Δ*) ou de Rpa49 (*rpa49Δ*) ont été initialement caractérisés, il est maintenant clair qu'il existe deux types d'allèles à considérer : 1) les mutants *rpa34Δ* / *rpa49ΔNter*, affectant donc le module TFIIIF-like, et 2) le mutant *Rpa49ΔtWH*, portant l'invalidation de la partie C-terminale de Rpa49. *In vivo*, les mutants de délétion *rpa49Δ* et *rpa34Δ* sont viables, mais conduisent à des défauts de synthèse du pré-ARNr 35S très différents. L'absence de Rpa49 conduit à une croissance très ralentie, et un défaut massif de synthèse des ARNr, alors que celle de Rpa34 n'induit pas de défaut de croissance (voir plus bas) (Liljelund *et al*, 1992; Gadai *et al*, 1997). Même si les phénotypes associés sont drastiquement différents, l'hétérodimère Rpa34-Rpa49 se dissocie facilement de la polymérase purifiée, constituant ainsi une nouvelle forme d'ARN pol I (historiquement notée Pol A\*) dont l'activité transcriptionnelle est altérée. Ce type d'altération, comme une sensibilité accrue à l' $\alpha$ -amanitine de pol A\* comparée à l'ARN pol I entière, suggère des problèmes lors de l'élongation (Liljelund *et al*, 1992). On sait maintenant que cette forme d'enzyme est inactive dans un test d'initiation de la transcription ("promoter-dependent assay") *in vitro* (Pilsel *et al*, 2016).

### - L'hétérodimère "TFIIIF like module "

Même si l'absence de Rpa34 ne provoque aucun défaut de croissance, la combinatoire de cette délétion avec celle d'autres protéines peut avoir des conséquences lourdes en termes de

croissance, pouvant provoquer la mort cellulaire (SL). Notamment, le mutant *rpa34Δ* de l'ARN pol I est SL avec *rpa14Δ* (voir plus haut), et *top1Δ* (Top1 = ADN Topoisomérase I) (Gadal *et al*, 1997). Dans mon équipe, Albert *et al* (2011) ont par ailleurs effectué une cartographie globale des interactions génétiques en utilisant la souche *rpa34Δ* comme cible, validant ces deux premiers mutants SL, et identifiant deux fonds génétiques additionnels dans lesquels la sous-unité Rpa34 devient totalement essentielle : *gcr2Δ* (un activateur de la transcription des gènes impliqués dans la glycolyse) et *stb5Δ* (un facteur de transcription impliqué dans la réponse au stress oxydatif et à la résistance à certaines drogues). Les mêmes auteurs ont également montré que la partie C-terminale de Rpa34 était impliquée dans la localisation nucléolaire de l'ARN pol I, et que la délétion de cette région de la sous-unité, comme celle de Rpa34 entière, provoquait une déstructuration détectable du nucléole (Teixeira *et al*, 2002; Albert *et al*, 2011) (Figure 30.A.). Il est à noter que la surexpression de Rpa49, en l'absence de Rpa34, corrige l'ensemble des phénotypes de *rpa34Δ*. *In vivo*, les défauts observés en l'absence de Rpa34 pourraient donc tous être indirects, et en réalité liés à une dissociation partielle de Rpa49 de l'ARN pol I (Beckouet *et al*, 2008; Albert *et al*, 2011). La fonction de Rpa34 se restreindrait à la stabilisation de Rpa49 (Beckouet *et al*, 2008), ce qui n'est pas le cas chez la souris par exemple, où l'orthologue de Rpa34 ("Paf49") possède des fonctions propres (Yamamoto *et al*, 2004).

La délétion de la partie N-terminale de Rpa49 n'entraîne pas de défaut de croissance, ni de sensibilité particulière à des drogues (Beckouet *et al*, 2008) (Figure 30.C.). Le complexe TFIIF est essentiel à la croissance cellulaire, de même que les sous-unités Pol III Rpc37-Rpc53, ce qui constitue donc une différence de phénotype marquante (Yan *et al*, 1999; Wu *et al*, 2011).



On mentionnera de plus que des altérations de la structure nucléolaire dans les mutants *rpa49Δ* et *rpa34Δ* ont été rapportées (Berger *et al*, 2008), en microscopie confocale à fluorescence sur un grand nombre de cellules de levures.

- Analyse du mutant *rpa49Δ* et de la partie C-terminale de Rpa49 : le domaine *tWH*

Le mutant *rpa49Δ* est viable, mais possède en revanche de forts défauts de croissance. A 25°C, cette délétion est létale, générant un mutant cryo-sensible (Cs) et à 30°C, la croissance du mutant est très ralentie (temps de doublement 2 à 3 fois plus long qu'un sauvage) (Liljelund *et al*, 1992) (Figure 30.B.). Des expériences de marquage métabolique de l'ARN naissant dans des souches sans Rpa49 ont montré que le taux de synthèse de l'ARNr 5.8S (l'un des produits de maturation du pré-ARNr 35S synthétisé par l'ARN pol I), comparé à celui du 5S (synthétisé par l'ARN pol III) était fortement réduit, montrant que bien que non-essentielle, Rpa49 est importante pour la synthèse d'ARNr (Gadal *et al*, 2002). Liljelund *et al*. (1992) ont également montré que l'insertion du marqueur d'auxotrophie *HIS3* au milieu de la phase codante de Rpa49 résulte en une protéine tronquée, mais qui peut s'incorporer dans l'ARN pol I. Dans une souche de levure hétérozygote diploïde, les versions tronquées et sauvages de Rpa49 sont en compétition pour l'assemblage dans la polymérase, même si en termes de phénotype, l'allèle sauvage est dominant. On notera dans le cas de ce mutant de troncation de Rpa49 un point important, car au cœur de la partie expérimentale de ce travail, qui sera largement repris dans les résultats et discussion de ce manuscrit : l'apparition spontanée de suppresseurs du défaut de croissance observé dans la souche de levure invalidée pour Rpa49. En effet, après un certain nombre de générations en milieu liquide riche, les cellules mutées peuvent "ré-adopter" un phénotype sauvage, et pousser normalement. De façon intéressante, la délétion de la région N-terminale de la sous-unité du Stalk Rpa43, de même que des troncations du facteur d'élongation de l'ARN pol II Spt5 (*spt5(1-930)*, *spt5(1-886)*, ou

*spt5(1-797)*), permettent aussi de restaurer la croissance d'une souche *rpa49Δ* (Beckouët *et al*, 2011; Viktorovskaya *et al*, 2011). On notera également qu'on observe une létalité synthétique entre *rpa49Δ* et *hmo1Δ* (Figure 30.C), et que la surexpression d'Hmo1, à partir d'un plasmide multi-copies (2 $\mu$ ) supprimer le phénotype de croissance de la souche *rpa49Δ* (Gadal *et al*, 2002) (Figure 30.D.). HmoI est l'une des sept HMG-box ("High Mobility Group") protéines, similaire à UBF chez l'Humain, et il a été montré par immunoprécipitation de la chromatine des régions transcrites de l'ADNr, que Hmo1 y était retrouvée, suggérant que cette protéine HMG-box peut structurer la chromatine à cet endroit (Berger *et al*, 2007).

Des mutants de délétion du Cter de Rpa49 possèdent le même phénotype de croissance que la délétion de Rpa49 entière (Figure 30.C.) (Beckouët *et al*, 2008). De plus, les mêmes auteurs ont montré que le mutant *rpa49Δ* et les mutants de délétion du C-ter de Rpa49, SL avec *rpa14Δ* et *hmo1Δ*, sont sensibles à certaines drogues comparés à des levures de référence (WT), telles que le Mycophénolate (MPA), le 6-Azauracil (6-AZA) ou la Caféine (CAF) (Figure 30.C.). Le MPA et le 6-AZA sont des inhibiteurs de l'élongation, réduisant le pool de dNTP dans la cellule. Le mutant *rpa49Δ* est aussi particulièrement sensible à la rapamycine. La rapamycine est un inhibiteur de la kinase TORC1. La voie TORC1 est un activateur essentiel de la formation du complexe ARN pol I - Rrn3. Cette sensibilité est en accord avec un défaut des mutants de délétion de Rpa49 dans la formation du PIC. Or, la caféine a deux effets concomitants, puisqu'elle réduit le pool de dNTP dans la cellule (comme le MPA ou 6-AZA), mais conduit également à une inhibition des PI3-kinases comme ATR ou TOR. Tout comme la rapamycine, la formation du PIC dans un mutant *rpa49Δ* est donc plus fortement inhibée par la caféine.

Dans les mutants *rpa49Δ*, le facteur essentiel d'initiation Rrn3 est recruté moins efficacement au niveau du promoteur d'ADNr, et sa dissociation de la polymérase en élongation est altérée ou affectée (Beckouët *et al*, 2008).

La souche *rpa49Δ* présente un nombre réduit d'ARN polymérase I par gène d'ADNr, ce que l'on peut observer sur des étalements de Miller en MET (Figure 30.E.) (Albert *et al*, 2011). On rappellera également qu'une désorganisation nucléolaire est observée dans une souche *rpa49Δ* (Figure 30.A.). La morphologie normale du nucléole dans cette souche peut être restaurée en diminuant le nombre de copies de gènes d'ADNr de 190 à 25 copies. De plus, mon groupe a montré que le nombre de ribosomes dans les cellules *rpa49Δ* était significativement réduit comparé à des cellules WT, ce qui n'est pas le cas pour des *rpa34Δ* par exemple. En effectuant des expériences de profils de polysomes après centrifugation sur gradient de sucrose, ils ont remarqué que des polysomes étaient détectés dans le mutant *rpa49Δ* comme dans le WT, mais que des "Halfmers" (petits épaulements pour chaque pic de polysome sur le profil) apparaissaient dans le cas de *rpa49Δ*, représentatifs d'un défaut de biogénèse de la sous-unité 60S du ribosome (Figure 30.F.).

L'une des délétions de la partie Cterminale de Rpa49 (*rpa49ΔCter*, acides aminés Δ367-415) (Beckouet *et al*, 2008) a été réitérée en 2010 (Geiger *et al*, 2010), induisant également un phénotype de cryo-sensibilité. Les mêmes auteurs ont également montré que le module de dimérisation seul ne pouvait pas se lier à l'enzyme, mais que son extension par le linker de Rpa49 ou la queue de Rpa34 le lui permettait.

Dans des tests d'initiation de la transcription *in vitro* sur des matrices de promoteurs d'ADNr, Pilsl *et al* ont utilisé l'ARN pol I dans laquelle il manque l'hétérodimère Rpa34-Rpa49, qui est presque totalement inactive en initiation de la transcription comparée à une souche sauvage (Pilsl *et al*, 2016). L'addition de l'hétérodimère recombinant entier Rpa34-Rpa49 dans la souche mutante stimule alors fortement l'activité de transcription. Il est alors intéressant de noter que l'addition du C-ter de Rpa49 (tWH) seul permet également cette même stimulation de la transcription, alors que l'addition du N-ter de Rpa49, en complexe avec le N-ter de Rpa34, ne le permet pas, suggérant que le tWH de Rpa49 aurait un rôle prédominant en

initiation. Enfin, le même type d'expérimentation *in vitro* a été réalisé, mais dans un test d'élongation de la transcription. Là, la polymérase délétée des sous-unités Rpa34 et Rpa49 est moins efficace en transcription qu'un sauvage. Ce défaut n'est toutefois pas comparable au défaut majeur d'initiation. De même, l'addition de Rpa34-Rpa49 recombinant, ou celle du tWH seul de Rpa49 stimule fortement l'élongation, là où l'addition des deux Nter de Rpa34 et Rpa49 ne le permet pas, suggérant là aussi un rôle du tWH de Rpa49 en élongation. Ces données seraient en accord les résultats d'études menées *in vitro* (Kuhn *et al*, 2007; Geiger *et al*, 2010) et *in vivo* (Beckouet *et al*, 2008) proposant un rôle de Rpa49 à la fois en initiation et élongation de la transcription.

Le fait que le domaine tWH de Rpa49 soit flexible empêche son positionnement structural exact dans les structures à haute résolution de l'ARN pol I, que ce soit sous forme de monomère ou de dimère, en complexe avec les facteurs d'initiation (Rrn3, PIC), ou en élongation (en complexe avec l'ADN et l'ARN). Cependant, dans certains cas, des densités électroniques peuvent lui être attribuées, révélant plusieurs localisations possibles sur la polymérase durant les différentes étapes de la transcription, ce que nous aborderons plus en détails dans la partie Discussion de ce manuscrit.

### La sous-unité Rpa12

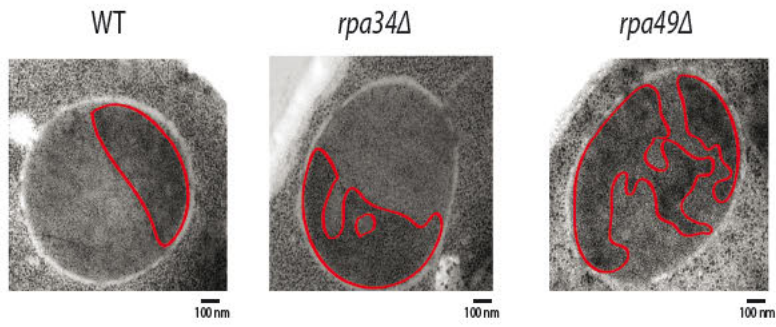
Rpa12 pourrait avoir un rôle dans l'assemblage et/ou la stabilité de l'ARN pol I, et plus précisément de sa plus grande sous-unité Rpa190. En effet, l'invalidation du gène codant pour la sous-unité Rpa12 entière conduit à un fort défaut de croissance à toutes les températures, mais est létale uniquement à haute température (>34°C). Ce phénotype pourrait être causé par l'instabilité de la polymérase. En l'absence de Rpa12, la concentration cellulaire en Rpa190 est bien plus basse que dans des cellules sauvages. De plus, ce phénotype de thermosensibilité peut être supprimé en surexprimant le gène codant pour Rpa190, suggérant que Rpa12

jouerait un rôle dans l'assemblage et la stabilité de Rpa190 dans la polymérase I (Nogi *et al*, 1993). La délétion de la partie N-terminale de Rpa12 (Rpb9-like) induit une dissociation de l'ARN pol I du peptide, et conduit comme la délétion totale de Rpa12 à un phénotype de thermosensibilité (Van Mullem *et al*, 2002). Par contre, ce n'est pas le cas de la délétion de la partie Cterminale (TFIIS-like), qui n'a aucun phénotype de croissance associé, et où le reste de Rpa12 demeure ancré à la polymérase. Ainsi, dans cette étude, il a été montré que le domaine N-ter de Rpa12, seul élément requis *in vivo*, était nécessaire à l'ancrage de la sous-unité au reste de la polymérase. L'absence de Rpa12 affecte également le recrutement de la s.u. Rpa49. Rpa12 $\Delta$ C ne conduit à aucune instabilité de Rpa49.

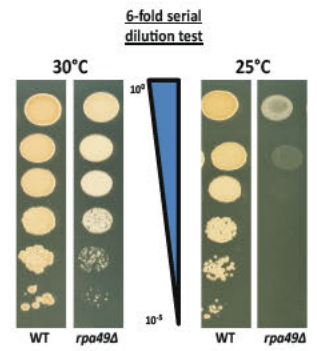
Il a été suggéré que délétion du gène *RPA12* entraîne un défaut de terminaison important, notamment révélé par microscopie électronique sur des arbres de Miller (Prescott *et al*, 2004; Kawauchi *et al*, 2008), où l'on observe sur un gène d'ADNr que des ARN pol I sont détectables au niveau des espaces intergéniques, région d'accumulation inhabituelle comparée à un sauvage (Figure 30.G.).

Dans le système de l'ARN Pol III, Rpc11 (Chédin *et al*, 1998; Hoffmann *et al*, 2015), paralogue de Rpa12 impliqué dans le clivage, est essentiel à la croissance. Par contre, les phénotypes observés en absence de Rpa12 sont comparables à ceux associés à l'absence des paralogues du système de l'ARN pol II: Rpb9 et TFIIS. Le phénotype de croissance associé à l'invalidation de Rpb9 est comparable à celui de Rpa12. De la même façon, l'absence d'un phénotype de croissance qui pourrait être lié à la stimulation du clivage (*rpa12 $\Delta$ c*) est à rapprocher du phénotype lié à la délétion du gène codant pour TFIIS (*DST1*) (Malagon *et al*, 2004). De façon intéressante, les mêmes auteurs ont suggéré un rôle de TFIIS dans la transition initiation-élongation de la transcription par l'ARN pol II.

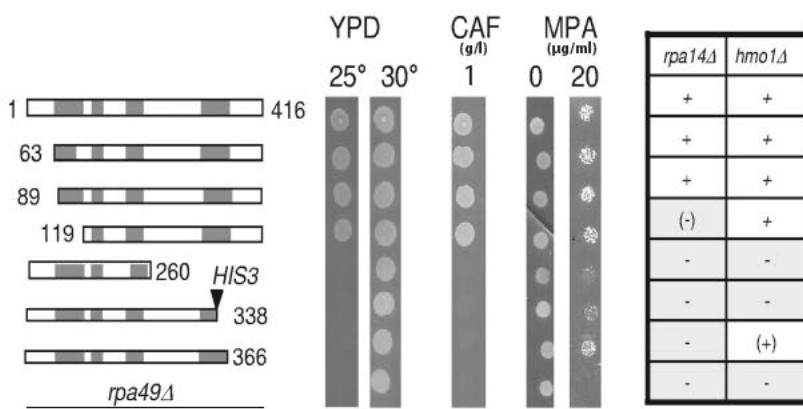
A.



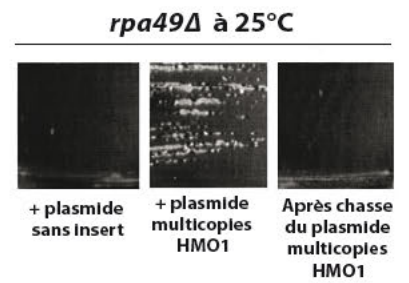
B.



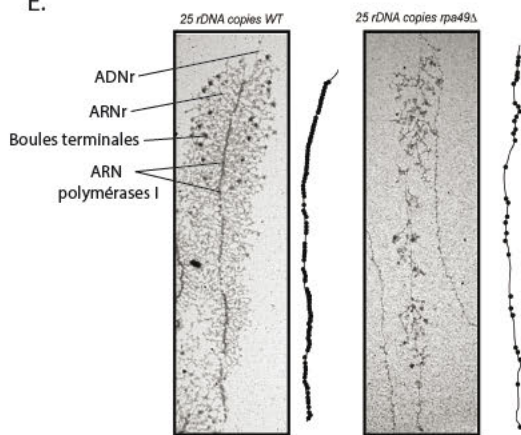
C.



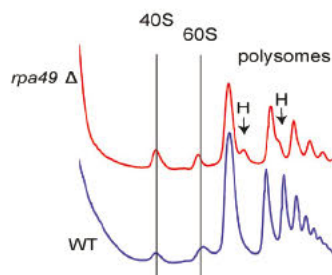
D.



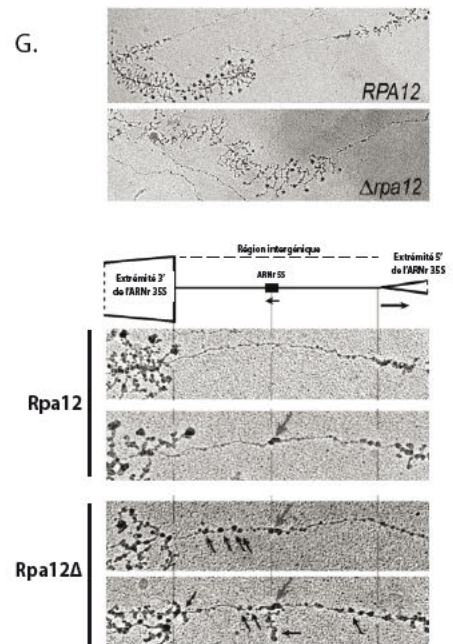
E.



F.



G.



**Figure 30. Phénotypes associés à des mutants des sous-unités Rpa34, Rpa49 et Rpa12 de l'ARN pol I de *S.cerevisiae*.**

A. Cellules de *S.cerevisiae* cryo-fixées et cryo-substituées de souches WT, *rpa34Δ* et *rpa49Δ*, montrant respectivement : un nucléole normal, une destructuration mineure et une destructuration majeure du nucléole. Une segmentation manuelle a été effectuée pour délimiter le nucléole pour chaque cas (lignes rouges). D'après Albert et al., 2011.

B. Test de croissance en dilution en séries d'une souche sauvage (WT) et d'une souche mutante *rpa49Δ*, à 25°C et 30°C, montrant le phénotype de cryosensibilité (25°C) et le défaut de croissance (30°C) de la souche mutante. Six dilutions sont représentées de haut en bas (du moins au plus dilué).

C. Mutagenèse de la sous-unité Rpa49 de l'ARN pol I et analyse de la croissance des souches de levures ainsi générées. Plusieurs délétions sont effectuées, comme indiquées sur les schémas de Rpa49 à gauche. La délétion pleine longueur est représentée en bas. Au milieu: des tests en goutte de croissance ont été réalisés sur trois milieux différents pour tester la sensibilité à deux drogues des différentes souches: un milieu riche (YPD), et deux milieux avec drogues (YPD + CAF (Caféine) et YPD + MPA (Mycophénolate). Les concentrations pour chaque drogue sont indiquées. A droite un tableau représente la croissance/létalité de chaque mutant quand associé à la délétion soit de Rpa14, soit de Hmo1. Un + ou (+) signifie la croissance de la souche double mutante, le (-) une croissance très ralentie, et le - la létalité (d'après Beckouet et al., 2008).

D. Test de croissance en stries sur boîtes de pétri d'une souche de levure *rpa49Δ* à 25°C, transformées par différents plasmides: à gauche, un plasmide sans insert ; au milieu, un plasmide multicopies surexprimant le gène codant pour la protéine Hmo1, et supprimant le défaut de croissance de la souche ; à droite, le même plasmide que précédemment, chassé sur un milieu sélectif, ne permettant plus la suppression du phénotype *rpa49Δ*. D'après Gadal et al., 2002.

E. Expérience de Miller's spread dans deux souches de levure avec 25 copies d'ADNr : sauvage (WT) et *rpa49Δ*. On visualise pour chaque arbre de Miller les ARN pol I en transcription le long d'un gène d'ADNr, et les ARNr en cours de synthèse, ainsi que des boules terminales au bout de ces ARN naissants. L'ADNr et les ARN pol I sont tracés à droite de chaque arbre en noir. D'après Albert et al., 2011.

F. Profil de polysomes (DO260nm) après centrifugation sur gradient de sucrose des extraits cellulaires des souches WT (ligne bleue) et *rpa49Δ* (ligne rouge). Les pics correspondants aux sous-unités 40S et 60S, ainsi que les pics des polysomes, sont indiqués pour chaque souche. Dans le cas *rpa49Δ*, on observe des épaulements à droite de chaque pic de polysome, correspondant aux «Halfmers» (H, en noir), représentatifs d'un défaut de biogénèse de la 80S.

G. Expériences de «Miller spreads» et visualisation en MET en coloration négative de gènes d'ADNr de souches sauvages pour Rpa12 (Rpa12) ou délétées de la sous-unité (Rpa12Δ), montrant un défaut de terminaison dans le cas Rpa12Δ. Le panel du haut illustre les deux types «d'arbres de Miller» représentant les gènes d'ADNr en transcription, ainsi que des régions intergéniques les séparant. Le panel du bas propose deux clichés de chaque souche. Les flèches noires sur les clichés indiquent la présence inhabituelle d'ARN pol I au sein des régions intergéniques dans le cas Rpa12Δ, et pas dans le cas Rpa12 sauvage. La région correspondante au gène de l'ADNr 5S transcrit par l'ARN pol III est également visible. D'après Prescott et al., 2004.

## PARTIE EXPERIMENTALE

La partie expérimentale de ce manuscrit traitera des études génétiques, biochimiques et structurales sur des mutants de l'ARN pol I, dans le but de caractériser les rôles possibles des sous-unités spécifiques de cette enzyme. Notamment, le mutant de délétion de la sous-unité Rpa49 (*rpa49Δ*), l'un des "Built-in Transcription Factors" dont la caractérisation a débuté dès les années 1990 (Liljelund *et al*, 1992), nous a servi de point de départ pour nos travaux. Les résultats obtenus durant cette thèse sont présentés sous la forme d'un article en cours/fin d'écriture pour soumission, et seront donc rédigés en anglais. Dans ce travail, j'ai repris et continué les études antérieures sur les supresseurs du défaut de croissance de la souche *rpa49Δ* commencées par Christophe Normand, Adrien Chauvier, Titouan Genty et Sylvain Audibert. J'ai ensuite cherché à caractériser le mécanisme de suppression en me focalisant sur un supresseur spécifique représentatif de la région où l'on trouve la majorité de nos mutations supresseurs : *RPA135-F30IS*. Les études de transcription *in vitro* sur les différentes souches mutantes ont été réalisées en collaboration avec Michael Pils, doctorant dans l'équipe de Herbert Tschochner à l'Université de Regensburg, en Allemagne. Christophe Dez et Isabelle Leger-Silvestre m'ont montré et aidé à réaliser les études de marquage métaboliques des ARNs néo-synthétisés et les expériences de Miller's spreads, respectivement. Tout ces travaux ont été réalisés sous la supervision d'Olivier Gadal, directeur de mon équipe.



## Genetic analysis of RNA polymerase I unveils roles of Rpa12 subunit during transcription initiation

Tommy Darrière, Michael Pils, Adrien Chauvier, Titouan Genty, Sylvain Audibert, Christophe Dez, Isabelle Leger-Silvestre, Christophe Normand, Olga Calvo, Carlos Fernández-Tornero, Herbert Tschochner and Olivier Gadal

### **Summary : (205 words)**

The majority of the transcriptional activity of exponential growing cells is ensured by RNA Polymerase I (Pol I) which produces a large rRNA precursor. Pol I transcription cycle is achieved through complex structural rearrangements of the enzyme unveiled by recent structural studies. *In vivo* Rpa49 subunit, and specifically its C-terminal tandem Winged Helix (tWH) domain is required for both initiation and elongation of the transcription cycle. Here, we characterized novel extragenic suppressors of the growth defect caused by the absence of Rpa49. Suppressor mutations were identified on the two largest subunits of the Pol I: Rpa190 Rpa135 and Rpa12. Suppressor mutant Rpa135-F301S and Rpa12-S6L restored normal rRNA synthesis and increased Pol I density on rDNA genes in absence of Rpa49-tWH. Mapped on the Pol I structure, we observe that they are clustering at an interface between the jaw and lobe modules of the two largest subunits and the subunit Rpa12. Our genetic data in *S. cerevisiae* point toward a specific role of Rpa12 at the interface between the jaw and lobe modules during transcriptional initiation.

## **Introduction :**

The nuclear genome of eukaryotic cells is transcribed by three RNA polymerases (Chambon, 1975). RNA polymerase II (Pol II) transcribes most of the genome, and is responsible for all messenger RNAs production. RNA polymerase III and I are specialized in the synthesis of a limited number of transcripts. RNA polymerase III (Pol III) produces small structured RNAs including tRNAs and the 5S ribosomal RNA. Finally, RNA polymerase I (Pol I) produces a single transcript, the large polycistronic precursor (the 35S pre-rRNA in yeast) processed by multiple successive steps into the mature rRNAs (in yeast, 25S, 18S, and 5.8S rRNAs). Despite producing a single transcript, Pol I is by far the most active eukaryotic RNA polymerase contributing up to 60% of the total transcriptional activity in exponential growing cells (Warner, 1999). Pol I rRNA synthesis constitutes the first step in ribosome biogenesis, and is a rate limiting process for cell growth. Such massive transcriptional output can be best visualized by using the DNA spreads method developed by Miller *et al*, 1969, where the 35S rRNA genes (rDNA) adopt a “Christmas tree” conformation, with up to 120 polymerases per transcribed gene (Miller & Beatty, 1969). Altered cell proliferation, often associated with a modified rate of rRNA synthesis via a Pol I activity deregulation, has been associated with various types of cancer (Drygin *et al*, 2010).

In the budding yeast *Saccharomyces cerevisiae*, full subunits composition and structural informations are now available for the three nuclear RNA polymerases (Fernández-Tornero *et al*, 2013; Engel *et al*, 2013)(Hoffmann *et al*, 2015; Cramer *et al*, 2001). Like the other two nuclear RNA polymerases, Pol I contains a core of shared or homologous subunits that are largely conserved up to Archea (Werner, 2008). The two largest subunits (Rpa190 and Rpa135) bear the catalytic site. Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are shared with Pol II and Pol III, while Rpc40 and Rpc19 are also present in Pol III. This 9-subunits core is

associated with the stalk, a structure formed in Pol I by the heterodimeric complex Rpa43/Rpa14 docking the essential Rrn3 initiation transcription factor to the enzyme (Peyroche *et al*, 2000; Yamamoto *et al*, 1996; Blattner *et al*, 2011). Pol I-Rrn3 complex is interacting with promoter bound factors, the Core Factor (CF), forming the initially transcribing complex (ITC) (Engel *et al*, 2017). Additionally, Pol I and Pol III have subunits functionally and structurally related to Pol II specific basal transcription factors, called the "Built-in Transcription Factors" (Werner, 2008). In Pol I and Pol III, their presence results in a higher number of subunits, from 12 subunits in Pol II, to 14 and 17 for Pol I and Pol III respectively, and correlates with a massive transcript's production from few genes (Werner, 2008). Rpa34 and N-terminal domain of Rpa49 (Rpa49Nt) in Pol I (i.e. Rpe53 and Rpc37 in Pol III) are related to specific basal transcription factor TFIIF (Geiger *et al*, 2010; Landrieux *et al*, 2006; Wu *et al*, 2011). Rpe34 in Pol III and Rpa49 C-terminal domain (Rpa49Ct) bear a tandem Wing Helix (tWH) motif similar to TFIIE (Geiger *et al*, 2010; Landrieux *et al*, 2006). Finally, Rpa12 (Pol I) and Rpe11 (Pol III) C-terminal domains are both implicated in stimulating endogenous transcript cleavage activity like TFIIS in the case of Pol II (Van Mullem *et al*, 2002; Kuhn *et al*, 2007).

Despite clear similarities, yeast genetic studies of Pol III and Pol I Built-in Transcription Factors reveal striking differences. If each of the Pol III subunits is strictly essential for cell growth, none of the Pol I built-in transcription Factors is needed for cell survival. Rpa34 deletion, or invalidation of Rpa49 N-terminal domain, removing TFIIF like heterodimer, has a very slight effect *in vivo* (Gadal *et al*, 1997; Beckouet *et al*, 2008; Geiger *et al*, 2010). Full or C-terminal (A49Ct) deletion of Rpa49 leads to a similar phenotype, with a strong growth defect at all temperatures, exacerbated below 25°C (Liljelund *et al*, 1992; Beckouet *et al*, 2008; Geiger *et al*, 2010). Full deletion of Rpa12 leads to a strong growth defect at 25°C and 30°C, and is lethal at high temperature (Nogi *et al*, 1993). Deletion of the C-terminal

extension of Rpa12 abolishes stimulation of intrinsic cleavage without any detectable growth defect (Van Mullem *et al*, 2002; Kuhn *et al*, 2007). Finally, the triple deletion (simultaneous deletion of *RPA49*, *RPA34* and *RPA12*) is viable, cumulating growth defects associated with each of the single mutant (Gadal *et al*, 1997).

In absence of Rpa49, Pol I is functional but has well documented defects, both *in vivo* and *in vitro*, in initiation and elongation (Liljelund *et al*, 1992; Beckouet *et al*, 2008; Albert *et al*, 2011; Pilsl *et al*, 2016; Gadal *et al*, 2002; Huet *et al*, 1975). Restoration of active rRNA synthesis in absence of Rpa49 has been used to identify regulatable factors, like Hmo1 and Spt5, or other Pol I subunits such as Rpa43 (Gadal *et al*, 2002; Beckouet *et al*, 2008; Viktorovskaya *et al*, 2011). We decided to make use of the spontaneous occurrence of extragenic suppressors of the growth defect of *rpa49* null mutant (Liljelund *et al*, 1992) to better understand the role of this subunit, or reveal new kinds of interactions between the different subunits of the RNA pol I during transcription. We could show that the suppressing phenotype was caused by specific point mutations in the two largest Pol I subunits, Rpa190 and Rpa135. Mapping mutations onto the Pol I structure identified a small area at the interface between the jaw and the lobe modules of the two largest subunits, in which Rpa12 linker domain is inserted. The subunit Rpa12 turns out to be essential for the suppression to occur. Moreover, Pol I with some point mutations in Rpa12 is able to correct the growth defect and initiate transcription in absence of Rpa49. Altogether, our genetic data gathered in absence of A49Ct domain, suggest that efficient RNA polymerase I initiation involves the jaw/lobe module interface and specifically Rpa12.

## **Results :**

### **Isolation of extragenic suppressor mutant of the growth defect in absence of *rpa49*.**

To better understand how cell growth is achieved in absence of Rpa49, we decided to characterize extragenic suppressors of *rpa49* deletion. *RPA49* full deletion mutants have a strong growth defect at 30°C and are unable to grow at 25°C. However, despite a deletion in the genome, spontaneous suppressors were previously observed (Liljelund *et al*, 1992). We reproduced this observation, and quantified the frequency of occurrence of individual clones that were able to grow at 25°C. We observed a low frequency of colony occurrence, comparable with the spontaneous mutation rate of a single control gene (*CANI* ;  $< 5.10^{-6}$ ). To isolate a large number of suppressors, we then used UV irradiation conditions. UV irradiation, with a survival rate of about 50% (5s, 5W/m<sup>2</sup>), increases about 10 fold the rate of suppressor mutation occurrence. After three days, we could identify clones growing at 25°C, and we selected individual colonies, called SGR for Suppressor of Growth defect of *RPA49* deletion, restoring growth at 25°C with various growth rates. We ranked SGR from 1 to 186 based on their growth rates at 25°C, *SGR1* restoring growth to a level comparable to wild-type condition (Figure 1A). By crossing those 186 SGR clones with a strain of the opposite mating type bearing only the deletion of *rpa49* we could obtain diploid cells homozygote for *RPA49* deletion and heterozygote for each suppressor. Growth restoration at 25°C of the diploids showed that all obtained suppressor phenotypes were fully or partially dominant. We focused on the best suppressor clones SGR1 and SGR2, and performed tetrad analysis to follow segregation of observed suppression phenotype. We could show that each suppressor phenotype was linked to a single locus in the genome, and that both SGR mutants had no associated growth defect, as shown for SGR1 in Figure 1A. Using global genomic mapping of SGR1 and SGR2, derived from "Genetic Interaction Mapping" (GIM) methods (Decourty *et*

al, 2008) (material and methods; Suppl. Figure 1), we found a genomic linkage close to genes encoding the two largest Pol I subunits : *RPA135* for SGR1 and *RPA190* for SGR2 (Suppl. Figure 1). Sequencing genomic DNA revealed that SGR1 bears a double mutation while SGR2 bears a single one (*RPA135-(I218T/R379K)* and *RPA190-A1557V* alleles respectively). By genetic analysis and genomic sequencing, we could identify an additional mutant SGR3 in *RPA135*: *RPA135-R305L*. Note that heterogeneity of growth induced by strong UV mutagenesis prevented suppressors cloning in the 183 remaining SGR clones.

We next made use of the dominant phenotype of these suppressors to isolate more alleles that can suppress in *RPA190* and *RPA135*. By propagating plasmids bearing wild-type *RPA135* or *RPA190* in a mutagenic XL1-red *E.coli* strain, we constructed a library of randomly generated mutants (see material and methods). After phenotypic selection of *rpa49Δ* mutants bearing a mutagenized subunit Rpa190 or Rpa135 at 25°C, each plasmid bearing a suppressor allele was extracted, sequenced and re-transformed into yeast to confirm the suppressor phenotype. From this, we could isolate 9 novel alleles of Rpa190 and 13 of Rpa135 which were able to restore growth of *rpa49* deletion mutant at 25°C (supplementary Table 1). Like for original SGR strains, the suppression strength was evaluated based on growth restoration at 25°C, relative to wild-type (WT) growth. Note that using mutagenesis of Rpa190 or Rpa135, we could identify strong suppressors alleles, more effective than SGR1, 2 or 3 (which have a medium suppressor strength). In conclusion, we identified novel alleles of the two largest subunits as extragenic suppressors of *rpa49Δ*-associated growth defect.

### **Rpa190 and Rpa135 alleles can bypass the need of Rpa49 for optimal growth.**

The growth of the strains bearing one of the 6 following alleles (*RPA190-E1274K*, *RPA190-C1493R*, *RPA190-L1262P*, *RPA135-R379G*, *RPA135-Y252H* and *RPA135-F301S*) has been evaluated by 10-fold dilution test (Figure 1B), showing a good suppression by all of them,

stronger in the case of Rpa135 alleles. Previous genetic studies have isolated other genetic backgrounds alleviating the growth defect of *rpa49Δ* at 25°C: *rpa43-35,326* (Beckouet *et al*, 2008) ; rDNA copy number (Albert *et al*, 2011), Hmo1 over-expression (Gadal *et al*, 2002) or Spt5 truncations (Viktorovskaya *et al*, 2011). In all cases, rRNA synthesis was only partially restored in absence of Rpa49, and transcription defect was still detected. Here, we focused on Rpa135-F301S allele, the most effective suppressor of the growth defect: *rpa49Δ rpa135-F301S* double mutant is growing at 25°C like a WT strain (Figure 1B). In absence of Rpa49, both TFIIF and TFIIE like modules are absent of Pol I. *rpa49ΔC* allele removes A49Ct domains, the TFIIE like module. In strain bearing the *rpa49ΔC* allele, the TFIIF module (Rpa34 and Rpa49-N) remains associated to the polymerase (Liljelund *et al*, 1992; Beckouet *et al*, 2008). However, *rpa49ΔC* or *rpa49* full deletion have very similar growth defect (Liljelund *et al*, 1992; Beckouet *et al*, 2008). To restrict our genetic analysis to the absence of only A49Ct module, we decided to focus on *rpa49ΔC* mutant. We integrated the *RPA135-F301S* point mutation in the genome at its native Rpa135 locus in three genetic backgrounds: Wild-type, *rpa49Δ* (full deletion), *rpa49ΔC*. Growth rate was determined in each of these yeast strains, in presence or in absence of *RPA135-F301S* at 30°C. In WT strain, the suppressor allele *RPA135-F301S* has no effect on growth (doubling time of 102 min). In *rpa49Δ* strain, doubling time was 180 min and Rpa135-F301S suppression restored growth up to 135 min. Similar suppression was observed when using *rpa49ΔC* background.

To relate Rpa49 phenotype to nascent rRNA synthesis and Pol I density on transcribed genes, we next used mutant cells with a decreased and stable number of rDNA repeats, all being actively transcribed with a very high loading rate of Pol I (Cioci *et al*, 2003; Albert *et al*, 2011). Thus, in this background, there is almost no variation in the number of polymerases per rDNA gene due to the absence of variation in the ratio of active/inactive copies. We generated three strains in this background bearing *rpa49ΔC*, *RPA135-F301S* and *rpa49ΔC*

*RPA135-F301S* alleles. Doubling time (Figure 2A) and *de novo* synthesis of rRNA, were determined (Figure 2 B). Labeling of nascent rRNA was performed using a 2 min pulse with <sup>3</sup>H adenine. Due to experimental heterogeneity in cell's culture of *rpa49ΔC* mutant, we performed labeling in three independent cultures. Even under permissive conditions (30 °C), *rpa49ΔC* had a *circa* fivefold defect on synthesis of rRNA (compare lane 1 to 2-5). In presence of *rpa135-F301S*, the absence of A49tWH had no consequence on rRNA synthesis (compare line 5 and 6).

To evaluate Pol I density on transcribed genes, we performed Miller's spreads, the only technique allowing the counting of individual Pol I on a single rRNA gene (Miller & Beatty, 1969; Albert *et al*, 2011). We had previously showed that in the same genetic background, full deletion of *rpa49* resulted in a three-fold decrease of Pol I density per gene (Miller & Beatty, 1969; Albert *et al*, 2011). In the strain expressing the *rpa49ΔC* allele, we observed a four-fold decrease of pol I density per gene (Figure 2C). Rpa135-F301S allele alone had no influence on Pol I density. Interestingly, despite a full restoration of rRNA production, we only recovered about half of the WT Pol I density per gene in the double mutant *rpa49ΔC rpa135-F301S* (about 46 RNA Pol I per gene).

Taken together, these results show that a strain containing Rpa135-F301S allele is fully competent for rRNA synthesis, even in absence of the A49Ct. However, the Pol I density on rRNA gene is only partially restored.

### **In vitro, Pol I bearing Rpa135-F301S point mutation supports initiation of transcription in absence of A49tWH domain**

A49tWH is essential in promoter-dependent transcription assay (Pilsel *et al*, 2016). Our *in vivo* analysis showed that in Rpa135-F301S mutant background, A49tWH is not required for rRNA synthesis. We use the promoter-dependent *in vitro* transcription system to confirm this



hypothesis (Figure 3). After depletion of Rpa49, purified Pol I lacks both Rpa34 and Rpa49 subunits and is the so-called Pol A\* complex (Huet et al, 1975; Gadad et al, 1997; Pilsel et al, 2016). Pol I lacking subunits Rpa49/Rpa34 (Pol A\*) was almost inactive when compared with wild-type Pol I (WT) or Pol I bearing Rpa135-F301S (compare lanes 1 and 5 with lanes 13 in Figure 3). As a control, addition of recombinant Rpa34/Rpa49, Rpa49-tWH alone or Rpa34/Rpa49N-ter stimulated transcription of Pol A\* (lane 14, 15 and 16). Interestingly, Pol I bearing Rpa135-F301S did not require Rpa49/Rpa34 for promoter dependent activity (compare lane 5 to 9). As expected, addition of recombinant Rpa34/Rpa49, or Rpa49-tWH alone, to WT or Pol I bearing Rpa135-F301S did not further stimulate activity.

In conclusion, Pol I bearing Rpa135-F301S is fully active in absence of Rpa49-tWH in a reconstituted assay in vitro.

### **Most suppressors are clustered in a hotspot of Pol I.**

Pol I structural data is now available, in an inactive form (Fernández-Tornero *et al*, 2013; Engel *et al*, 2013), in complex with Rrn3 (Pilsel *et al*, 2016; Engel *et al*, 2016; Torreira *et al*, 2017), in complex with other initiation factors (Engel *et al*, 2017; Han *et al*, 2017; Sadian *et al*, 2017) or in elongating forms (Neyer *et al*, 2016; Tafur *et al*, 2016). Rpa135 and Rpa190 residues suppressing *rpa49ΔC* growth defect have been mapped onto the structure of the RNA Pol I in which Rpa49 structure have been fully determined (Han *et al*, 2017; Figure 4A). Most of the suppressor mutations, and all those providing full restoration of growth (see supplementary table 1), appeared to be clustered at a specific interface between the two largest subunits Rpa190 and Rpa135 (figure 4B) between the lobe (Rpa135 - salmon) and the jaw (Rpa190 - blue). This small region is characterized by the presence of a β-strand in the structure of Rpa12 (Rpa12 –yellow - residues 46-51; Figure 4B), following four β-strands of Rpa190, thus forming a shared five-strands anti-parallel β-Sheet. This Rpa12 β-strand is also

facing the Rpa135 lobe domain (residue 252 to 315 of Rpa135) in which six independent mutations are found, including Rpa135-F301S. Notably, Rpa135 contains a pattern of conserved amino acids "DSF" (D299, S300, F301), which are very conserved among eukaryotic species. Interestingly, we found suppressor mutations for each of these three amino acids (see supplementary table 1). Lateral chain of the mutated residues are involved in electrostatic interaction stabilizing this jaw/lobe interface (Figure 4C). Substituted residues resulted in a destabilization of this interface (see supplementary table 1), suggesting a specific rearrangement of the interface Lobe/jaw in each of those mutants.

### **Point mutations in Rpa12 can also suppress the growth defect due to the absence of Rpa49.**

We then tested if alleles of Rpa12 itself could behave as suppressors. We generated a library of randomly mutagenized Rpa12 (see material and methods), and could identify two dominant alleles (Rpa12-S6L and Rpa12-T49A), which suppressed the growth defect of *rpa49 $\Delta$ C* (Figure 5A). However, Rpa12 suppressor alleles did not restore WT growth when expressed in *rpa49 $\Delta$*  mutant (compare the size of the colonies Figure 5.A.). We integrated the *RPA12-S6L* point mutation in the genome at its native locus in a low rDNA copies number genetic background, with or without *rpa49 $\Delta$ C* deletion. We analyzed the restoration of rRNA synthesis by pulse labeling (Figure 5B), and Pol I loading rate by Miller's spread (Figure 5C). Like Rpa135-F301S, rRNA synthesis and Pol I loading rate were largely restored when combining *rpa49 $\Delta$ C* with *RPA12-S6L* background.

Interestingly, when focusing on the Pol I 3D structure, that Rpa12-S6L and Rpa12-T49A obtained by random mutagenesis were specifically localized in the "hotspot". The threonine 49 of Rpa12 is located on the  $\beta$ -strand (Rpa12 aa 46-51) previously described (Figure 4B), facing the residues D299, S300, F301 of Rpa135 and the Rpa190-E1274 (Figure 4B). The

other mutant Rpa12-S6L is in the N-terminal domain of Rpa12. In conclusion, point mutations in Rpa190, Rpa135 or Rpa12 within hotspot domains are all bypassing the requirement for A49-Ct *in vivo*.

### **Rpa190 Expander/DNA mimicking loop, Rpa14 and Rpa34 are not involved in the suppression phenotype**

Pol I structural determination reveals the presence of an extended loop inside the DNA-binding cleft folded in a expander/DNA mimicking loop conformation when Pol I is in an inactive, dimeric form (Engel *et al*, 2013; Fernández-Tornero *et al*, 2013). This element is not found in Pol II and III. Those residues are inserted between the four  $\beta$ -strands of Rpa190 corresponding to mutation's hotspot. Short deletion of this domains (1361-1390) of Rpa190 resulted in slow growth phenotype (Fernández-Tornero *et al*, 2013). We postulate that expander/DNA mimicking loop could be responsible of the suppression phenotype. We generated a novel allele *rpa190 $\Delta$ loop* (deletion of residues 1342-1411), which has no associated growth defect (Figure 6A). When combined with *rpa49* full deletion, we could not generate a viable double mutant. Therefore, DNA mimicking loop presence is required for Pol I activity in absence of Rpa49. Next, we tested the involvement of this loop in the suppression activity of Rpa135-F301S allele. When comparing *rpa49 $\Delta$  rpa135-F301S* double mutant with triple mutant *rpa49 $\Delta$  rpa135-F301S rpa190 $\Delta$ loop*, we observed no growth difference (Figure 6A); Therefore, we conclude that expander/DNA mimicking loop of Rpa190 is not required for suppression, but is required for the viability of *rpa49* deletion mutant.

We also introduced *RPA135-F301S* in yeast strains lacking other subunits of the Pol I, Rpa34 and Rpa14 (Figure 6B and C). In the viable double mutant lacking both Rpa49 and Rpa34, the heterodimer partner of Rpa49 (*rpa34 $\Delta$  rpa49 $\Delta$* ), the suppression by Rpa135-affected (Suppl. Figure 2A). The double deletion mutant lacking both Rpa49 and Rpa14, just like

*rpa190Δloop rpa49Δ*, is not viable (Gadal *et al*, 1997). In this case, if we introduced the suppressor *RPA135-F301S* by genetic crosses, we observed that the resulting triple mutant (*rpa14Δ rpa49Δ rpa135-F301S*) could grow (Suppl. Figure 2B).

We conclude that Rpa14, Rpa34 and DNA mimicking loop in Rpa190 are not required for suppression

### **Rpa12 is required for suppression of Rpa49 growth defect.**

We now want to evaluate if specific region of Rpa12 are involved in the suppression of *rpa49Δ* growth defect (Figure 7A). Rpa12 C-terminal region (TFIIS like) is inserted toward active center to stimulate intrinsic cleavage activity, but is displaced during productive initiation and elongation step. Rpa12 C-terminal deletion resulted in normal growth (Van Mullem *et al*, 2002) (Figure 7A- lane 2), even if Pol I bearing Rpa12ΔC is unable to stimulate cleavage activity (Kuhn *et al*, 2007). Full deletion of *RPA12* results in slight growth defect at 24°C, exacerbated at 30°C (Nogi *et al*, 1993)(Figure 7A-lane 3). Counter-wise, *RPA49* deletion results in growth defect at 30°C, exacerbated at 24°C (Liljelund *et al*, 1992)(Figure 7A-lane 4). Due to the absence of growth defect of the *rpa12ΔC* mutant, we expected a growth defect of the double mutant *rpa49Δ rpa12ΔC* similar to *rpa49Δ*. This is not the case, and we observe a synergistic phenotype, with an aggravated growth defect detected at both 24°C and 30°C.(Figure 7A – lane 5). The double mutants lacking simultaneously Rpa12 and Rpa49 is viable, but has major growth defects (Gadal *et al*, 1997)(Figure 7A – lane 6).

We then introduced the *RPA135-F301S* allele and assessed the suppression phenotype of the strain lacking Rpa49 at 25°C, with or without Rpa12 subunit (Figure 7B). Note that double mutant lacking both Rpa49 and Rpa12 grow very poorly at 30°C and can only be documented after 10 days of growth (Figure 7B, right panel). We observed that the suppression activity

was completely abolished in absence of Rpa12 (Figure 7B, compare 2<sup>nd</sup> to 4<sup>th</sup> lanes), strongly suggesting that Rpa12 presence within the hotspot is necessary for the suppression to occur.

Following random mutagenesis screen of Rpa12, we have not isolated Rpa12 alleles in C-terminal domain. To explore if Rpa12 C-terminal extension was necessary for the *rpa49Δ* phenotype suppression, we introduced the truncation of Rpa12 C-terminus in alleles Rpa12-S6L and Rpa12-T49A. As shown in figure 7C, *RPA12-S6L* and *RPA12-S6L-ΔC* resulted in similar suppression of *rpa49Δ* growth defect. *RPA12-T49A* behaved as Rpa12-S6L (data not shown). Furthermore, *rpa49ΔC* or *rpa49Δ* acted similarly in all genetic interaction tests with Rpa12 alleles.

In conclusion, we could show that Rpa12, but not its C-terminal extension, is absolutely required for suppression of the growth defect of *rpa49* deletion mutant.

## **Discussion**

In this work, we investigated how cells can improve growth in absence of Rpa49. We could show that modulating a very specific area of Pol I resulted in a functional Pol I lacking this subunit. We propose that our mutation point toward a very specific conformational change during initiation event.

### ***In vivo* Rpa49 modulate Rrn3-Pol I interaction**

Our previous studies point toward a specific involvement of Rpa49 in association and dissociation of Rrn3 from Pol I. In absence of Rpa49, Rrn3 is poorly recruited to pre-initiation complex and is not efficiently dissociated during elongation (Beckouet *et al*, 2008). We proposed that polymerases I transcribing rDNA are in close proximity, which could result in dissociation of Rrn3 by interaction with neighbouring polymerases (Albert *et al*, 2011). However, a recent structural study by chromatin spread (Neyer *et al*, 2016) established that physical distance separating Pol I is not compatible with such model. Later, Pilsl *et al* showed that Pol I-Rrn3, in which Rrn3 is over-expressed, but lacking Rpa49, is unable to initiate in an *in vitro* transcription assay (Pilsl *et al*, 2016). Furthermore, elongating polymerase structure show that the tWH domain of Rpa49 is associated with a position very close to the stalk, likely creating a sterical clash with the position occupied by Rrn3 (Tafur *et al*, 2016).

In this work, we show that genetically modified polymerases lacking Rpa49 or Rpa49-tWH, with a point mutation in Rpa190, Rpa135 or Rpa12 at a position diametrically opposed to Rrn3-bound position can initiate and grow properly. At this stage, the relationship between F301S residue of Rpa135 and Rrn3 appears indirect, but it remains to be experimentally investigated.

### **What is the role of Rpa12 in suppression?**

Rpa12 subunit, through a TFIIIS like domains at its C-terminus, is involved in stimulating the intrinsic cleavage activity of Pol I. Purified Pol I lacking Rpa12 C-terminal domain is inactive in cleavage stimulation. Furthermore, Rpa12-C-terminal domain can contact the active site of the polymerase in an inactive conformation, and is retrieved in both initiation competent form and elongating polymerases. The direct evidence that cleavage is not involved came from the presence of an fully functional suppressor phenotype of Rpa12-S6L- $\Delta$ C thus lacking the domain required for stimulating the cleavage.

The Rpa12 N-terminal domain at the surface of Pol I is required for docking subunit to the enzyme, and affect the accumulation of the largest subunit Rpa190 at 37°C, and is therefore indirectly required for cleavage. A linker region of Rpa12 is connecting a N-terminal module (Rpb9-like) at the surface of the Pol I to the C-terminal mobile region (TFIIIS-like).

We could show here Rpa12 linker position has an impact on initiation. Furthermore, in vitro, purified Pol I lacking Rpa12 is inactive in an in vitro promoter dependent assay (data not shown).

### **Major conformational changes during DNA-binding cleft implicates Rpa12 at the lobe/jaw interface**

Along the transcription cycle, RNA polymerase I undergoes major conformational changes, mainly affecting the width of the DNA-binding cleft (reviewed in Fernandez-Tornero, 2018). During transcription initiation, the cleft aperture narrows from a semi-open configuration, as seen in the cryo-EM structures of the enzyme bound to Rrn3 (Engel et al., 2016; Pilsl et al., 2016; Torreira et al., 2017), to a fully closed conformation observed in initially-transcribing complexes (Engel et al., 2017; Han et al., 2017). This allows gripping of the transcription bubble inside the cleft (Figure 7A). Following Rrn3 release, DNA binding is further secured by the Rpa49 linker, which crosses the cleft from the lobe to the clamp over downstream

DNA, and by the Rpa49 tWH domain, which binds upstream DNA in the vicinity of the clamp (Han et al., 2017; Tafur et al., 2016).

Cleft closure is achieved by relative movement of two structural units, located on opposite sides of the cleft, that pivot using five hinges (Fernandez-Tornero et al., 2013). The unit constituted by the shelf and clamp modules is essentially rigid. On the contrary, the unit constituted by the core and lobe modules undergoes very interesting internal rearrangements, which lies were our mutated residues are clustered (Movie 1). The most prominent reorganization within this latter unit affects the Rpa190 jaw domain, the outer rim of which shifts away from DNA by about 3.7 Å using the lobe/jaw interface as hinge. This movement also involves the linker region in subunit Rpa12, which contains a beta-strand (residues 46-50) that completes a four-stranded beta-sheet in the Rpa190 jaw domain. As a result, a short alpha-helix within the Rpa12 linker region shifts its position by about 3.0 Å.

Rearrangements in the jaw and Rpa12 linker regions are likely essential to allow pivoting of the shelf-clamp unit against the core-lobe unit. Without such motion, cleft closure would be impossible (Movie 1). As mentioned, structural results suggest that the Rpa49-tWH and its linker domains are involved in securing the closed cleft conformation (Han et al., 2017; Tafur et al., 2016). In the Rpa49ΔC, cleft closure is likely disfavored (Figure 7B). We hypothesize that the suppressor mutants favor the closed cleft conformation by likely increasing the flexibility of the lobe/jaw/Rpa12 interface and, thus, promoting the pivoting between the two major structural units that is required for cleft closure.

We conclude that Rpa12 linker domain, by participating in a hinge in the lobe/jaw interface is involved in cleft-closure necessary to insert DNA toward the active center. There are indirect evidences that similar domains are acting in initiation, directly or indirectly, in Pol II and Pol III. Rpb9, which is in Pol II similar to Rpa12-Nterminal module, but lacking TFIIIS domain, was required for a proper start site selection (Hull *et al*, 1995). In *rpb9Δ*, +1 start site was



affected. For Pol III, TFIIIS was shown in vivo and in vitro to stimulate Pol III transcription initiation (Ghavi-Helm *et al*, 2008). However, motion at the lobe/jaw interface during cleft closure is only detectable in Pol I (see movie 2 and 3 for Pol II and Pol III respectively).

## **Figure legends**

**Figure 1. Alleles of Rpa190 and Rpa135 suppress growth defect of *rpa49Δ* mutant at various level.** (A) SGR1 mutant restore growth of *rpa49Δ* mutant. 10-fold serial dilutions of wild-type (WT), *rpa49Δ* strain, SGR1/*rpa49Δ* double mutant, and SGR1 single mutant were spotted on rich media to check for growth at 30°C and 25°C for three days. (B) 10-fold dilution of wild-type (WT) and *RPA49* deletion (*rpa49Δ*) is compared to *rpa49Δ* bearing various plasmids: pGL190\_3 (Rpa190-E1274K), pGL190\_11 (Rpa190-C1493R), pGL190\_23 (Rpa190-L1262P), pGL135\_6prim (Rpa135-R379G), pGL135\_54 (Rpa135-Y252H) and pGL135\_33 (Rpa135-F301S). Growth was evaluated after 3 days at 25°C. Strains and plasmids used are listed in supplementary table 2 and 3 respectively.

**Figure 2. Rpa135-F301S allele in absence of Rpa49-tWH allows restoration of growth, of rRNA synthesis and modulates Pol I occupancy on rDNA genes.** (A) Doubling times of WT, *rpa49ΔC*, RPA135-F301S and double mutant *rpa49ΔC/RPA135-F301S* in low copy number background (see supplementary table 2). (B) *In vivo* labeling of newly synthesized RNAs. wild-type (lane 1), *rpa49ΔC* (lanes 2-4) and *RPA135-F301S* (Lane 5) and double mutant *rpa49ΔC/RPA135-F301S* (Lane 6) were grown to an OD600 of 0.8. Cells were then pulse-labeled with [<sup>8-3</sup>H] adenine for 2 min. Samples were collected and total RNAs were extracted, separated by gel electrophoresis. (C) Representative Miller's spread from WT, *rpa49ΔC*, *RPA135-F301S* and double mutant *rpa49ΔC/RPA135-F301S*. Panels on the right of each micrograph show interpretive tracing of the genes. Polymerases that appear on the gene are shown on the tracing by black dots. Number of polymerases counted on genes is indicated below. N represents the number of individual spread genes used for the quantification (see material and methods).

**Figure 3. Pol I bearing Rpa135-F301S does not require A49tWH for promoter dependent activity *in vitro*.** Promoter dependent transcription assay using extracts from wild-type (lane 1-4), *RPA135-F301S* (lane 5-8), double mutant *RPA135-F301S* / PolA\*(*rpa49Δ*) (lane 9-12) and single mutant PolA\*(*rpa49Δ*) (lane 13 to 16). Pol I purified after depletion of Rpa49 (noted here PolA\*, see material and methods) results in Pol I lacking both Rpa34 and Rpa49. Each reactions contain same amounts of Pol I (4 nM end concentration), recombinant CF and Rrn3; recombinant Rpa49/Rpa34, Rpa49 residues 110 to 415 (Rpa49tWH), Rpa49 residues 1 to 186 /Rpa34 (Rpa34/Rpa49N-ter) were added as shown in the table and preincubated with DNA-template. Reaction was started by addition of transcription-buffer containing NTPs and proceed for one hour. The expected run-off transcript (300 nt) is marked with black arrow.

**Figure 4. Mapping of the modified residues in Rpa190, Rpa135 and Rpa12 alleles on Pol I structure.** (A) Two different views of the original initial transcribing complex model and its 14 different subunits (PDB 5W66; (Han *et al*, 2017). (B) Most mutated suppressor residues are clustered at the interface between Jaw (Rpa190, blue) and lobe (Rpa135, salmon) modules of the Pol I. Note that Rpa12 (residue 46 to 51; yellow) is part of this interface. (C) Enlargement of areas point toward the modified residues (Rpa190-N863, -S1259, -L1262, -E1274, -C1493, Rpa135-Y252, D299, S300, F301, R305 and Rpa12-S6, T49) shown in panel B.

**Figure 5. Rpa12-S6L allele in absence of Rpa49-tWH allows restoration of growth and rRNA synthesis, and modulates Pol I occupancy on rDNA genes.** (A) 10-fold dilution of *rpa49Δ* mutant bearing various plasmids: a pRS316 empty plasmid (*rpa49Δ*), YCp50-26 bearing *RPA49* (WT), pRS316-A12-S6L (*RPA12-S6L*) or pRS316-A12-T49A (*RPA12-T49A*). Growth was evaluated after 3 days at 25°C or 2 days at 30°C. (B) *In vivo* labeling of newly

synthesized RNAs. For comparison, wild-type (lane 1), *rpa49ΔC* (lanes 2) depicted in figure 2B are shown with *RPA12-S6L* single mutant and *rpa49ΔC/RPA12-S6L* double mutant. (C) Representative Miller's spread from WT, *rpa49ΔC* (others than in figure 2C), *RPA12-S6L* and double mutant *rpa49ΔC/RPA12-S6L*.

**Figure 6. DNA mimicking loop of Rpa190, Rpa14 and Rpa34 are not required for suppression.** Deletion of DNA mimicking loop (A) or deletion of RPA34 (B) does not modulate the suppression activity of *RPA135-F301S*. (C) *RPA135-F301S* suppresses the synthetic lethality between *rpa14Δ* and *rpa49Δ*. 10-fold serial dilutions was performed on rich media, growth was evaluated after 3 days at 24°C

**Figure 7. Rpa12 alleles can modulate rpa49 growth defect (A)** Growth of double mutants *rpa49Δ rpa12ΔC*, or *rpa49Δ* combined with depletion of *rpa12*. Depletion of Rpa12 was achieved using *pGAL-RPA12* construct on glucose containing medium (strain OGT30-1c). 10-fold serial dilutions of OGT30-1c bearing pRS316-A12 (WT), pRS316-A12-DCter expressing Rpa12 bearing a C-terminal deletion of residues 65-125 (*rpa12ΔC*) or an empty plasmid pRS316 (*-rpa12*). Growth phenotype of *rpa49Δ* combined with *RPA12* depletion was tested using strain OGT30-3c bearing respectively pRS316-A12 (*rpa49Δ*), pRS316-A12-DCter (*rpa12ΔC rpa49Δ*) or an empty plasmid pRS316 (*-rpa12 rpa49Δ*). Growth was assessed after 4 days at 24°C or 30°C. **(B)** Suppression activity of *RPA135-F301S* is abolished in absence of Rpa12. Rpa12 under regulatable promoter (pGAL) was either expressed (+*RPA12*; left panel) on galactose containing medium, or repressed (*-RPA12*; right panel) on glucose containing medium. 10-fold serial dilutions of OGT30-1c (WT), OGT30-3c (*rpa49Δ*), yTD42-1a (*RPA135-F301S*) and yTD43-1a (*rpa49Δ RPA135-F301S*). Note that upon *RPA12* over-expression (left panel), *rpa49Δ* growth is barely observed (compare to

panel A, right panel – 4<sup>th</sup> lane). In this genetic background, complete suppression of *rpa49Δ* growth defect by Rpa135-F301S is detected. Depletion of *RPA12* abolishes the suppression activity of Rpa135-F301S (compare left and right panel). Extended incubation (10 days) was used on the right panel, instrumental to detect growth of the double mutant *rpa49Δ rpa12Δ* on plate. (C) C terminal extension of Rpa12 is not required for suppression. 10-fold serial dilutions of OGT30-1c (RPA49-WT) bearing pRS316-A12 (WT) or pRS316-A12-DCter (*rpa12ΔC*) and OGT30-3c (*rpa49Δ*) bearing pRS316-A12 (WT), pRS316-A12-DCter (*rpa12ΔC*), pRS316-A12-S6L (RPA12-S6L) and pTD10 (RPA12-S6L-ΔC). Growth was assessed after 4 days at 24°C.

**Figure 8. Schematic representation of Pol I lacking Rpa49 structural** (A) Free monomeric Pol I, with mobile Rpa49Ct and linker. (B) Initially transcribing complex (ITC) upon DNA insertion in presence of Rpa49 (purple). Rpa49Ct is interacting with upstream DNA and Rpa49-linker is folded, closing the cleft. (C) Pol I lacking Rpa49 is defective in DNA insertion and cleft closure. Pol I is unable to efficiently close the cleft and Jaw/lobe module position is clashing with downstream DNA (red asterisks). (D) Suppressor mutations (green) facilitate jaw/lobe module displacement and DNA insertion toward the active center in absence of Rpa49.

## **MATERIALS & METHODS**

### *Yeast strains, tetrad analysis and plasmid constructions*

The yeast strains, plasmids and oligonucleotides used in this study are listed in supplementary tables 2, 3 and 4 respectively. Yeast media and genetic techniques were described previously (Sherman *et al*, 1986). Yeast strains were constructed by meiotic crossing and plasmid DNA transformation according to standard protocols, (Schiestl & Gietz, 1989)(Sambrook *et al*, 1989). Yeast tetrad dissections were performed on rich media plates, and then evaluated on SD medium plates corresponding to the yeast strain auxotrophies (Singer instrument). Genetic interaction mapping (GIM) analysis of RPA49 deletion mutant was performed as described previously (Decourty *et al*, 2008). Microarray data were normalized using MATLAB (MathWorks, Inc., Natick, MA) as previously described (Albert *et al*, 2011).

### *In vivo labeling, RNA extractions and analysis*

Metabolic labeling of pre-rRNAs was performed as previously described (Hermann-Le Denmat *et al*, 1994) with the following modifications. Strains were pre-grown in synthetic glucose containing medium lacking adenine to an optical density at 600nm of 0.8 at 30°C. 1ml cultures were labeled with 50 $\mu$ Ci of [8-<sup>3</sup>H] adenine (NET06300 PerkinElmer) for 2 min. Cells were collected by centrifugation and pellets were frozen in liquid nitrogen. RNAs were then extracted as previously described (Beltrame & Tollervey, 1992) and precipitated with ethanol. For high molecular weight RNAs analysis, 1/5th of the total of RNAs were glyoxal denatured and resolved on a 1.2% agarose gel. Low molecular weight RNAs were resolved on 8% polyacrylamide/8.3M urea gels.

### *Miller spreads experiments and analysis*

Chromatin spreading was mainly performed as described previously with minor modifications (Osheim *et al.*, 2009). Carbon-coated grids are made hydrophilic by glow discharge instead of ethanol treatment. Negatively stained chromatin was obtained by short incubation with heavy metal followed by quick drying of the sample. Images were obtained on a JEOL JEM-1400 HC electron microscope (40 to 120 kV) with an Orius camera (11Mpixels). The position of the RNA polymerases I, and the rDNA fiber were determined by visual inspection of micrographs using Image J (<http://rsb.info.nih.gov/ij/>). Digital pictures were processed by software programs Image J and Adobe Photoshop® (v. CS6).

#### *In vitro transcription promoter dependent assays*

Promoter-dependent *in vitro* transcription reactions were performed as previously described (Pilsel *et al.*, 2016; Tschochner, 1996) with some modifications. Briefly, 1.5 ml reaction tubes (Sarstedt safety seal) were placed on ice. 0.5–1 µl template (50–100 ng DNA) was added, which corresponds to a final concentration of 5–10 nM per transcription reaction (25 µl reaction volume). 1–2 µl CF (0.5 to 1 pmol/µl; final concentration 20–40 nM) and 1–3 µl Pol I-Rrn3 (final concentration 4–12 nM) were added to each tube. 20 mM HEPES/KOH pH 7.8 were added to a final volume of 12.5 µl. Transcription was started adding 12.5 µl transcription buffer 2 X. The samples were incubated at 24°C for 30 min at 400 r.p.m. in a thermomixer. 200 µl Proteinase K buffer (0.5 mg/ml Proteinase K in 0.3M NaCl, 10mM Tris/HCl pH 7.5, 5mM EDTA and 0.6% SDS) was added to the supernatant to stop transcription. The samples were incubated at 30°C for 15 min at 400 r.p.m. in a thermomixer. Ethanol (700 µl) p.a. were added and mixed. Nucleic acids were precipitated at -20°C overnight or for 30 min at -80°C. The samples were centrifuged for 10 min at 12,000g and the supernatant was removed. The precipitate was washed with 0.15 ml 70% ethanol. After centrifugation, the supernatant was removed and the pellets were dried at 95°C for 2min.

RNA in the pellet was dissolved in 12  $\mu$ l 80% formamide, 0.1M TRIS-Borate-EDTA (TBE), 0.02% bromophenol blue and 0.02% xylene cyanol. Samples were heated for 2 min under vigorous shaking at 95°C and briefly centrifuged. After loading on a 6% polyacrylamide gel containing 7M urea and 1 X TBE RNAs were separated applying 25 watts for 30–40 min. The gel was dried after 10 min rinsing in water for 30 min at 80°C using a vacuum dryer. Radiolabelled transcripts are visualised using a PhosphoImager.



## Supplementary tables

| Sub.  | Alleles                   | Strenght | Hotspot    | Location                 | Mutant effect                   |
|-------|---------------------------|----------|------------|--------------------------|---------------------------------|
| A190  | N863T                     | Weak     | Funnel     | Funnel/A12-linker        | Destabilization of Funnel/Rpa12 |
| A190  | S1259L                    | Medium   | Jaw        | Jaw/Shelf hinge          | Hinge conformation              |
| A190  | L1262P                    | Weak     | Jaw        | Jaw/Shelf hinge          | Hinge conformation              |
| A190  | E1274K                    | Medium   | Jaw        | Jaw/A12-linker interface | Destabilization of Jaw/Rpa12    |
| A190  | C1493R                    | Strong   | Jaw        | Jaw/A12-linker interface | Destabilization of Jaw/Rpa12    |
| A135  | Y252H                     | Medium   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A135  | D299G                     | Medium   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A135  | S300F                     | Medium   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A135  | F301S                     | Strong   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A135  | F301L                     | Strong   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A135  | <b>SGR3</b><br>R305L      | Medium   | Lobe       | Jaw/Lobe interface       | Destabilization of Jaw/Lobe     |
| A12.2 | S6L                       | Medium   | N-terminal | Jaw/A12-linker interface | Destabilization of Jaw/Rpa12    |
| A12.2 | T49A                      | Medium   | Linker     | Jaw/A12-linker interface | Destabilization of Jaw/Rpa12    |
| A190  | L608S                     | Medium   | -          |                          |                                 |
| A190  | E611K                     | Medium   | -          |                          |                                 |
| A190  | S936A                     | Weak     | -          |                          |                                 |
| A190  | <b>SGR2</b><br>A1557V     | Weak     | -          |                          |                                 |
| A135  | D157G                     | Medium   | -          |                          |                                 |
| A135  | D157N                     | Medium   | -          |                          |                                 |
| A135  | <b>SGR1</b><br>I218/R379K | Medium   | -          |                          |                                 |
| A135  | R379G                     | Weak     | -          |                          |                                 |
| A135  | G580D                     | Medium   | -          |                          |                                 |
| A135  | C584Y                     | Weak     | -          |                          |                                 |
| A135  | I913V                     | Medium   | -          |                          |                                 |

**Supplementary table 1.** List of 24 individual suppressor mutations in *RPA190*, *RPA135* and *RPA12* subunit (Sub.) of the growth defect of *rpa49Δ* strain. The suppressors have been classified according to growth rate when combine with *rpa49Δ* mutant : weak, medium or strong. SGR1, 2 and 3 depict alleles originally isolated (see text). Of those 24 mutants, 13 mutations are found in a specific hot-spot shown in Figure 4.

**Supplementary table 2: Yeast strains used in this study**

| Referred as   | strain                   | Genotype  | source                      |
|---|--------------------------|---|-----------------------------|
| <i>WT</i> figure 1, 6                               | BY4741                   | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>  | Euroscarf                   |
| <i>rpa49Δ</i> in figure 1                           | OGT9-6a                  | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpa49Δ::alphaNATMX6</i>  | This study                  |
| <i>rpa49Δ - KAN</i> in figure 5A                    | OGT8-11a                 | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ::KANMX6</i>   | This study                  |
| <i>SGR1</i> in figure 1                             | OGT15-7b                 | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 HIS3::RPA135(I218T, R379K)</i>   | This study                  |
| <i>SGR1 rpa49Δ</i> in figure 1                      | LH514D                   | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpa49Δ-alphaNAT RPA135(I218T, R379K)</i>   | This study                  |
| <i>SGR2</i>   | AH29R                    | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ rpa49Δ- kanmx4 RPA190-1557V</i>   | This study                  |
| <i>SGR3</i>   | LH11D                    | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpa49Δ-alphaNAT RPA135-R305L</i>   | This study                  |
| <i>WT</i> in figure 2,5B,5C                         | yTD27                    | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135-TAP-HIS3, rDNA copy number ~25 + pCJF4-LEU-GAL49</i>                              | This study                  |
| <i>RPA135-F301S</i> in figure 2                     | yTD28                    | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135-F301S-TAP-HIS3, rDNA copy number ~25, + pCJF4-LEU-GAL49</i>                       | This study                  |
| <i>rpa49ΔC</i> in figure 2,5B,5C                    | yTD29                    | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135 -TAP-HIS3 rpa49ΔC(186-416)::HPH, rDNA copy number ~25, + pCJF4-LEU-GAL49</i>      | This study                  |
| <i>rpa49ΔC RPA135-F301S</i> in figure 2             | yTD30                    | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135-F301S-TAP-HIS3 rpa49ΔC(186-416)::HPH, rDNA copy number ~25, + pCJF4-LEU-GAL49</i> | This study                  |
| <i>WT</i> in figure 3                               | BY4741<br>A135-TEV-ProtA | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPA135-TAP-HIS3</i>  | (Pils1 <i>et al</i> , 2016) |
| <i>Pol A*</i> in figure 3                           | yJPF162-1a               | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPA135-TAP-HIS3 KANMX-HIS3::pGAL::HA-RPA49</i>   | (Pils1 <i>et al</i> , 2016) |
| <i>Pol A*</i> <i>RPA135-F301S</i> in figure 3       | yTD6-6b                  | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ::HPH RPA135-F301S-TAP-HIS3</i>  | This study                  |
| <i>RPA135-F301S</i> in figure 3                     | yTD6-6c                  | <i>MATalpha his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 RPA135-F301S-TAP-HIS3</i>  | This study                  |
| <i>BY4741 rpa49ΔC</i>                               | yTD25                    | <i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpa49ΔC(186-416)::KAN RPA135-F301S-TAP-HIS3</i>  | This study                  |
| <i>RPA12-S6L</i> in figure 5B,5C                    | yTD31                    | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135- TAP-HIS3 RPA12-S6L-KAN-MX, rDNA copy number ~25, + pCJF4-LEU-GAL49</i>           | This study                  |
| <i>rpa49ΔC RPA12-S6L</i> in figure 5B,5C            | yTD23_1a                 | <i>MATa ade2-1 ura3-1 his3-11,15 trp1-1 leu-3,112 can1-100 job1Δ::NAT-MX RPA135- TAP-HIS3 RPA12-S6L-KAN-MX rpa49ΔC(186-416)::HPHMX, rDNA copy number ~25</i>      | This study                  |
| <i>rpa49Δ</i> in figure 6, 7                        | OGT15-9d                 | <i>MATalpha his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ::HPHMX</i>  | This study                  |
| <i>rpa190Δloop</i>                                  | SCOC2260                 | <i>Mata; ade2, arg4, leu2-3,112, trp1-289,ura3-52, RPB6::TAP-K.L.URA3, Rpa190-ΔLOOP-KANMX</i>   | This study                  |
| <i>rpa190Δloop</i> in figure 6A                     | yTD48-1a                 | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa190Δloop::KANMX</i>   | This study                  |
| <i>rpa190Δloop RPA135-F301S</i> in figure 6A        | yTD51-2c                 | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 rpa190Δloop::KANMX RPA135-F301S-URA3K1</i>  | This study                  |
| <i>rpa190Δloop RPA135-F301S rpa49Δ</i> in figure 6A | yTD51-8a                 | <i>MATa his3Δ1 leu2Δ0 lys2Δ0 met15Δ0 ura3Δ0 RPA135-F301S-URA3K1 rpa49Δ::HPHMX</i>   | This study                  |
| <i>RPA135-F301S rpa49Δ</i> in figure 6A             | yTD51-5a                 | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ::HPHMX</i>  | This study                  |

**Supplementary table 2 (continued)**

|  |           |  |            |
|--|-----------|--|------------|
| <i>rpa49Δ</i><br>Figure 6B, 6C                           | yCN223-2a | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 rpa49Δ::KANMX</i>  | This study |
| <i>rpa49Δ</i><br><i>RPA135-F301S</i><br>Figure 6B, 6C    | yTD2-3d   | <i>MATa his3Δ1 leu2Δ ura3Δ0 lys2Δ0 rpa135Δ::KANMX</i><br><i>rpa49Δ::HPHMX</i> + pGL135_33 (ARS/CEN <i>RPA135-F301S</i> ) | This study |
| <i>rpa34Δ</i><br>Figure 6B                               | yCN224-1a | <i>MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 rpa34Δ::KANMX</i>  | This study |
| <i>rpa34Δ</i><br><i>RPA135-F301S</i><br>Figure 6B        | yTD36-2b  | <i>MATalpha his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 rpa34Δ::NATMX</i><br><i>RPA135-F301S::URA3</i>                                   | This study |
| <i>rpa34Δ rpa49Δ</i><br>Figure 6B                        | yTD37-7d  | <i>MATalpha his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 rpa34Δ::NATMX</i><br><i>rpa49Δ::HPHMX</i>  | This study |
| <i>rpa34Δ rpa49Δ</i><br><i>RPA135-F301S</i><br>Figure 6B | yTD37-3d  | <i>MATalpha his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 rpa34Δ::NATMX</i><br><i>rpa49Δ::HPHMX RPA135-F301S::URA3</i>                     | This study |
| <i>rpa14Δ</i><br>Figure 6C                               | yCN225-1a | <i>MATalpha his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa14Δ:: KANMX</i>  | This study |
| <i>rpa14Δ</i><br><i>RPA135-F301S</i><br>Figure 6C        | yTD38-3d  | <i>MATa his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 rpa14Δ::NATMX</i><br><i>RPA135-F301S::URA3</i>                                       | This study |
| <i>rpa14Δ rpa49Δ</i><br><i>RPA135-F301S</i><br>Figure 6C | yTD39-8a  | <i>MATa his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0 rpa14Δ::NATMX</i><br><i>rpa49Δ::HPHMX RPA135-F301S::URA3</i>                         | This study |
| figure 7<br>with plasmids as<br>indicated in legends     | OGT30-1c  | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 KANMX-pGAL::RPA12</i>   | This study |
| figure 7 with<br>plasmids as indicated                   | OGT30-3c  | <i>MATalpha his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ:: HPHMX</i><br><i>KANMX-pGAL::RPA12</i>                                    | This study |
| <i>RPA135-F301S</i> in<br>figure 7B                      | yTD42-1a  | <i>MATa his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 RPA135-F301S::URA3 KANMX-</i><br><i>pGAL::RPA12</i>                                   | This study |
| <i>rpa49Δ</i><br><i>RPA135-F301S</i><br>in figure 7B     | yTD43-1a  | <i>MATapha his3Δ1 leu2Δ0 lysΔ0 ura3Δ0 rpa49Δ:: HPHMX RPA135-</i><br><i>F301S::URA3 KANMX-pGAL::RPA12</i>                 | This study |

**Supplementary table 3: Plasmids used in this study**

| Name               | Description   | Source                           |
|--------------------|---|----------------------------------|
| pVV200             | <i>Plasmid 2<math>\mu</math> with bla, TRP1</i>   | (Van Mullem <i>et al</i> , 2003) |
| pNOY80             | <i>Plasmid CEN6 ARS4, with marker URA3 and RPA135 cloned bay Gap repair with flanking regions (EcoRI-XhoI, NsiI-EcoRV) from SRP3</i>              | (Yano <i>et al</i> , 1992)       |
| pGL190_3           | <i>Plasmid pVV200 (2<math>\mu</math> URA3 RPA190) bearing the mutation RPA190-E1274K suppressor of the rpa49<math>\Delta</math> growth defect</i> | This study                       |
| pGL190_11          | <i>Plasmid pVV200 (2<math>\mu</math> URA3 RPA190) bearing the mutation RPA190-C1493R suppressor of the rpa49<math>\Delta</math> growth defect</i> | This study                       |
| pGL190_23          | <i>Plasmid pVV200 (2<math>\mu</math> URA3 RPA190) bearing the mutation RPA190-L1262P suppressor of the rpa49<math>\Delta</math> growth defect</i> | This study                       |
| pGL135_6prim       | <i>Plasmid pNOY80 (CEN4 URA3 RPA135) bearing the mutation RPA35-R379G suppressor of the rpa49<math>\Delta</math> growth defect</i>                | This study                       |
| pGL135_54          | <i>Plasmid pNOY80 (CEN4 URA3 RPA135) bearing the mutation RPA35-Y252H suppressor of the rpa49<math>\Delta</math> growth defect</i>                | This study                       |
| pGL135_33          | <i>Plasmid pNOY80 (CEN4 URA3 RPA135) bearing the mutation RPA35-F301S suppressor of the rpa49<math>\Delta</math> growth defect</i>                | This study                       |
| pTD1 3b 135TAP     | <i>Plasmid pNOY80 bearing RPA135-TAP-HIS3</i>   | This study                       |
| pTD2 6c 135TAP     | <i>Plasmid pNOY80 bearing RPA135-F301S-TAP-HIS3</i>   | This study                       |
| pTD5               | <i>Plasmid pTD1_3b_135TAP cut by NsiI + SdaI from position 8106 to 10430, deleted for URA3 marker</i>   | This study                       |
| pTD6               | <i>Plasmid pTD2_6c_135TAP cut by NsiI + SdaI from position 8106 to 10430, deleted for URA3 marker</i>   | This study                       |
| pRS316             | <i>Original plasmid CEN4, with marker URA3</i>  | (Sikorski & Hieter, 1989)        |
| Ycp50-26           | <i>ARS/CEN RPA49</i>  | P. Thuriaux, unpublished         |
| pRS316-A12         | <i>Plasmid pRS316 KAN containing RPA12 coding sequence</i>  | This study                       |
| pRS316-A12-S6L     | <i>Plasmid pRS316 containing RPA12 coding sequence mutated for C2384T bp (amino acid S6L)</i>   | This study                       |
| pRS316-A12-S6L-KAN | <i>Plasmid pRS316 KAN containing RPA12 coding sequence mutated for C2384T bp (amino acid S6L)</i>   | This study                       |
| pRS316-A12-T49A    | <i>Plasmid pRS316 containing RPA12 coding sequence mutated for A2512G bp (amino acid T49A)</i>  | This study                       |
| pRS316-A12-DCter   | <i>Plasmid pRS316 KAN containing RPA12 coding sequence truncated for its Cterminal part amino acids (Rpa12<math>\Delta</math>65-125)</i>          | This study                       |
| pTD9               | <i>Plasmid pRS316-A12-DCter mutated for A2512G bp (amino acid T49A)</i>   | This study                       |
| pTD10              | <i>Plasmid pRS316-A12-DCter mutated for C2384T bp (amino acid S6L)</i>  | This study                       |
| pCJPF4             | <i>pFL36cII with LEU2 marker, CEN4, containing RPA49 coding region</i>  | This study                       |
| pCJPF4-GAL49-1     | <i>Plasmid pCJPF4 KAN with RPA49 coding region under GAL promoter (#624/625 oligos used for PCR insert recombination in pCJPF4)</i>               | This study                       |

**Supplementary table 4: Oligonucleotides used in this study**

| Name            | Sequence  | N°   |
|-----------------|---|------|
| F3-structure    | AAGTCACAATTCTCCAATTTAAAAGTCGTCACCACGACGGCAGACGATGCGTTTCCATCTTCTCTTAGAGCCAAGAAATGAGGCGCGCCACTTCTAAA  | 1371 |
| o-pGAL_RPA12_R3 | CTAATACGGCATTAGGATTTTCCAGGAGATCACCACAGTCCAAGCAAAAAATTAACGATCCTACAACAGACATgcactgagcagcgaatctg        | 1633 |
| o-pGAL_RPA12_R2 | CTAATACGGCATTAGGATTTTCCAGGAGATCACCACAGTCCAAGCAAAAAATTAACGATCCTACAACAGAcattttgagatccgggttt           | 1634 |
| o-pGAL_RPA12_F4 | TTTATGTCAAGTGAAGGGCATTGCATGGCAAGTGTAAGAATCCAAGAAAGgaattcgagctcgtttaaac                              | 1635 |
| F3-RPA12-CCT5ok | agttaagatctcagatgaaggtgctactgtcttctatacatgcacttctgtggttacaaagttcogtaccacaattgaTGAGGCGCGCCACTTCTAAA  | 1682 |
| R1-A12-OK       | tctatagatgtcacatgatgaaagcggggatgatattaatgtacaaattgtaatatgtgcgaacacaaccaatcaGAATTCGAGCTCGTTTAAAC     | 1559 |
| o-checkA2-K-up  | cactgaatgtcacgatagattatcgtgt  | 1556 |
| o-checkA2-K-do  | tttcaaatattgctataaaaatggatgata  | 1557 |
| RPA49F4         | CCACTATTGCAGTCGTTATCAACCTTTTGCACCTTATCTAGTAgaattcgagctcgtttaaac                                     | 624  |
| RPA49R2         | CACCTTCAATTTTCGATTTTCAGAAAACAGACCTTTTTCACGGACATtttgagatccgggttt                                     | 625  |
| o-F2-186-RPA49  | CTACGTTTCGACTTATACAAAAAAGAAGTCTGAAAAGGACGAATTCTGATTACACGGTGAA AACGAGAGACTAGAATACCGGATCCCCGGGTAAATTA | 1515 |
| o-R1_RPA49      | GCTGCTTTATAGAACCATCTAGCAAAAAAAAAAACCAGAGCAAACCAATGCTCAGTATATA CCTCTAGGTAACtgaattcgagctcgtttaaac     | 208  |

## **Acknowledgements**

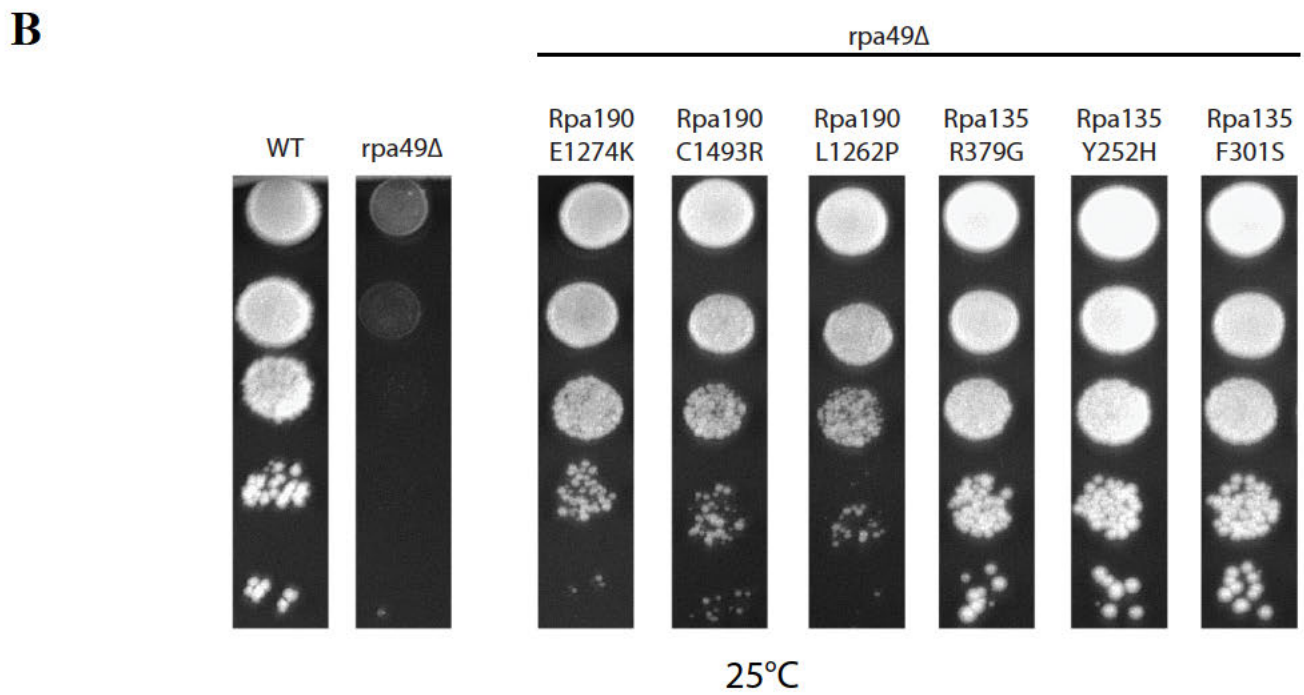
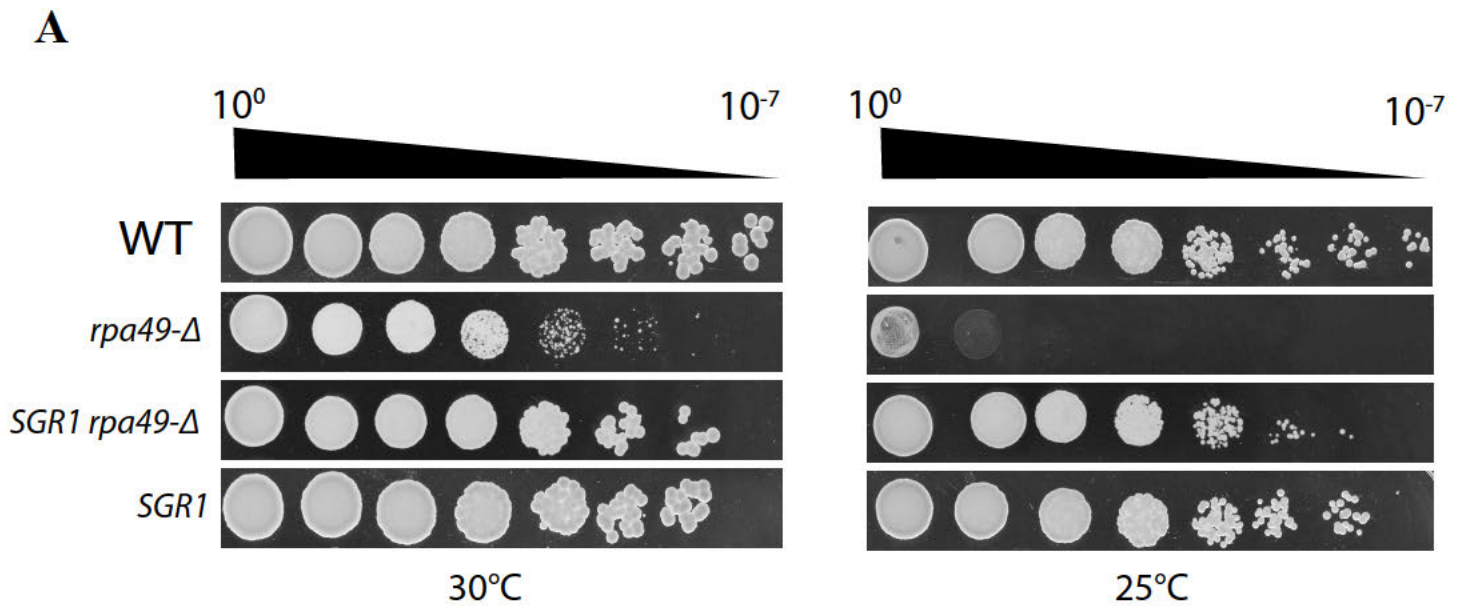
This work benefited from the assistance of the imaging platform of Toulouse TRI. This work was supported by the *Agence Nationale de la Recherche* (ANDY) and the IDEX of Toulouse University (Clemgene and Nudgene). We thank Margaux Cescato, Lilian Guillot and Jorge Perez-Fernandez for strain and plasmid constructions. YCp50-26 carrying *RPA49* was given by P. Thuriaux. We also thank Melanie Panarotto for Miller spreads analysis. This work benefited from exchange with all unit members including Anthony Henras, Frederic Beckouët, Lise Dauban and Marta Kwapisz. We thank Stephanie Balor and Vanessa Soldan from METI platform for the TEM acquisitions. This work benefited from the assistance of the imaging platform of Toulouse TRI.

## Bibliography

- Albert B, Léger-Silvestre I, Normand C, Ostermaier MK, Pérez-Fernández J, Panov KI, Zomerdijk JCBM, Schultz P & Gadal O (2011) RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *J Cell Biol* **192**: 277–293
- Beckouet F, Labarre-Mariotte S, Albert B, Imazawa Y, Werner M, Gadal O, Nogi Y & Thuriaux P (2008) Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol* **28**: 1596–1605
- Beltrame M & Tollervey D (1992) Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J* **11**: 1531–1542
- Blattner C, Jennebach S, Herzog F, Mayer A, Cheung ACM, Witte G, Lorenzen K, Hopfner K-P, Heck AJR, Aebersold R & Cramer P (2011) Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev* **25**: 2093–2105
- Chambon P (1975) Eukaryotic nuclear RNA polymerases. *Annu Rev Biochem* **44**: 613–638
- Cioci F, Vu L, Eliason K, Oakes M, Siddiqi IN & Nomura M (2003) Silencing in yeast rDNA chromatin: reciprocal relationship in gene expression between RNA polymerase I and II. *Mol Cell* **12**: 135–145
- Cramer P, Bushnell DA & Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**: 1863–1876
- Decourty L, Saveanu C, Zemam K, Hantraye F, Frachon E, Rousselle J-C, Fromont-Racine M & Jacquier A (2008) Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. *Proc Natl Acad Sci U S A* **105**: 5821–5826
- Drygin D, Rice WG & Grummt I (2010) The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu Rev Pharmacol Toxicol* **50**: 131–156
- Engel C, Gubbey T, Neyer S, Sainsbury S, Oberthuer C, Baejen C, Bernecky C & Cramer P (2017) Structural basis of RNA polymerase I transcription initiation. *Cell* **169**: 120–131.e22
- Engel C, Plitzko J & Cramer P (2016) RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat Commun* **7**: 12129
- Engel C, Sainsbury S, Cheung AC, Kostrewa D & Cramer P (2013) RNA polymerase I structure and transcription regulation. *Nature* **502**: 650–655
- Fernández-Tornero C, Moreno-Morcillo M, Rashid UJ, Taylor NMI, Ruiz FM, Gruene T, Legrand P, Steuerwald U & Müller CW (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**: 644–649
- Gadal O, Labarre S, Boschiero C & Thuriaux P (2002) Hmo1, an HMG-box protein, belongs to the yeast ribosomal DNA transcription system. *EMBO J* **21**: 5498–5507
- Gadal O, Mariotte-Labarre S, Chedin S, Quemeneur E, Carles C, Sentenac A & Thuriaux P (1997) A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* **17**: 1787–1795
- Geiger SR, Lorenzen K, Schrieck A, Hanecker P, Kostrewa D, Heck AJR & Cramer P (2010) RNA polymerase I contains a TFIIIF-related DNA-binding subcomplex. *Mol Cell* **39**: 583–594
- Ghavi-Helm Y, Michaut M, Acker J, Aude J-C, Thuriaux P, Werner M & Soutourina J (2008) Genome-wide location analysis reveals a role of TFIIIS in RNA polymerase III transcription. *Genes Dev* **22**: 1934–1947
- Han Y, Yan C, Nguyen THD, Jackobel AJ, Ivanov I, Knutson BA & He Y (2017) Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *elife* **6**:
- Hermann-Le Denmat S, Werner M, Sentenac A & Thuriaux P (1994) Suppression of yeast RNA polymerase III mutations by FHL1, a gene coding for a fork head protein involved in rRNA processing. *Mol Cell Biol* **14**: 2905–2913
- Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJH, Sachse C & Müller CW (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**: 231–236
- Huet J, Buhler JM, Sentenac A & Fromageot P (1975) Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc Natl Acad Sci U S A* **72**: 3034–3038
- Hull MW, McKune K & Woychik NA (1995) RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev* **9**: 481–490
- Kuhn C-D, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R & Cramer P (2007) Functional architecture of RNA polymerase I. *Cell* **131**: 1260–1272
- Laferté A, Favry E, Sentenac A, Riva M, Carles C & Chédin S (2006) The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev* **20**: 2030–2040
- Landrieux E, Alic N, Ducrot C, Acker J, Riva M & Carles C (2006) A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* **25**: 118–128
- Liljelund P, Mariotte S, Buhler JM & Sentenac A (1992) Characterization and mutagenesis of the gene encoding

- the A49 subunit of RNA polymerase A in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **89**: 9302–9305
- Miller OL & Beatty BR (1969) Visualization of nucleolar genes. *Science* **164**: 955–7
- Van Mullem V, Landrieux E, Vandenhaute J & Thuriaux P (2002) Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Mol Microbiol* **43**: 1105–1113
- Neyer S, Kunz M, Geiss C, Hantsche M, Hodirnau V-V, Seybert A, Engel C, Scheffer MP, Cramer P & Frangakis AS (2016) Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature* **540**: 607–610
- Nogi Y, Yano R, Dodd J, Carles C & Nomura M (1993) Gene RRN4 in *Saccharomyces cerevisiae* encodes the A12.2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol Cell Biol* **13**: 114–122
- Osheim YN, French SL, Sikes ML & Beyer AL (2009) Electron microscope visualization of RNA transcription and processing in *Saccharomyces cerevisiae* by Miller chromatin spreading. *Methods Mol Biol* **464**: 55–69
- Peyroche G, Milkereit P, Bischler N, Tschochner H, Schultz P, Sentenac A, Carles C & Riva M (2000) The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* **19**: 5473–5482
- Pilsel M, Crucifix C, Papai G, Krupp F, Steinbauer R, Griesenbeck J, Milkereit P, Tschochner H & Schultz P (2016) Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat Commun* **7**: 12126
- Sadian Y, Tafur L, Kosinski J, Jakobi AJ, Wetzel R, Buczak K, Hagen WJ, Beck M, Sachse C & Müller CW (2017) Structural insights into transcription initiation by yeast RNA polymerase I. *EMBO J*
- Sambrook J, Fritsch EF & Maniatis T (1989) Molecular cloning: a laboratory manual. *Molecular cloning: a laboratory manual*. Available at: <https://www.cabdirect.org/cabdirect/abstract/19901616061>
- Schiestl RH & Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* **16**: 339–346
- Sherman F, Fink GR & Hicks JBN (1986) *Methods in Yeast Genetics. A Laboratory Course Manual*. New York: Cold Spring Harbor Laboratory
- Tafur L, Sadian Y, Hoffmann NA, Jakobi AJ, Wetzel R, Hagen WJH, Sachse C & Müller CW (2016) Molecular structures of transcribing RNA polymerase I. *Mol Cell* **64**: 1135–1143
- Torreira E, Louro JA, Pazos I, González-Polo N, Gil-Carton D, Duran AG, Tosi S, Gallego O, Calvo O & Fernández-Tornero C (2017) The dynamic assembly of distinct RNA polymerase I complexes modulates rDNA transcription. *elife* **6**:
- Tschochner H (1996) A novel RNA polymerase I-dependent RNase activity that shortens nascent transcripts from the 3' end. *Proc Natl Acad Sci U S A* **93**: 12914–12919
- Viktorovskaya OV, Appling FD & Schneider DA (2011) Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *J Biol Chem* **286**: 18825–18833
- Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**: 437–440
- Werner F (2008) Structural evolution of multisubunit RNA polymerases. *Trends Microbiol* **16**: 247–250
- Wu C-C, Lin Y-C & Chen H-T (2011) The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIC, Bdp1, and the RNA polymerase III active center. *Mol Cell Biol* **31**: 2715–2728
- Yamamoto RT, Nogi Y, Dodd JA & Nomura M (1996) RRN3 gene of *Saccharomyces cerevisiae* encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. *EMBO J* **15**: 3964–3973



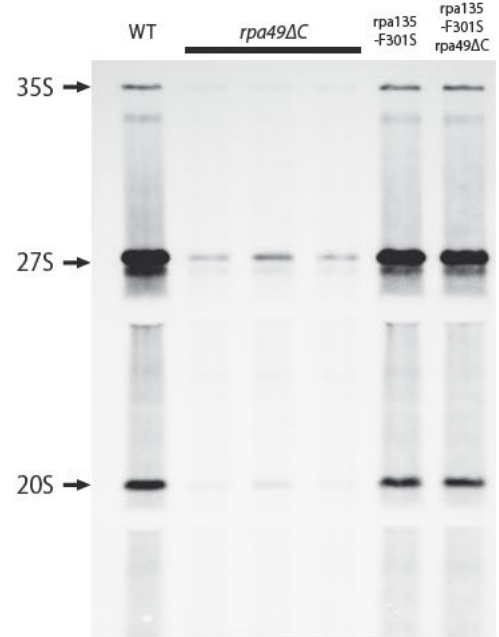


**Figure 1**

A.

|                         | Doubling time<br>(min) |
|-------------------------|------------------------|
| WT                      | 104 +/- 2              |
| Rpa49ΔC                 | 292 +/- 11             |
| RPA135-F301S            | 98 +/- 2               |
| Rpa49ΔC<br>RPA135-F301S | 118 +/- 3              |

B.



C.

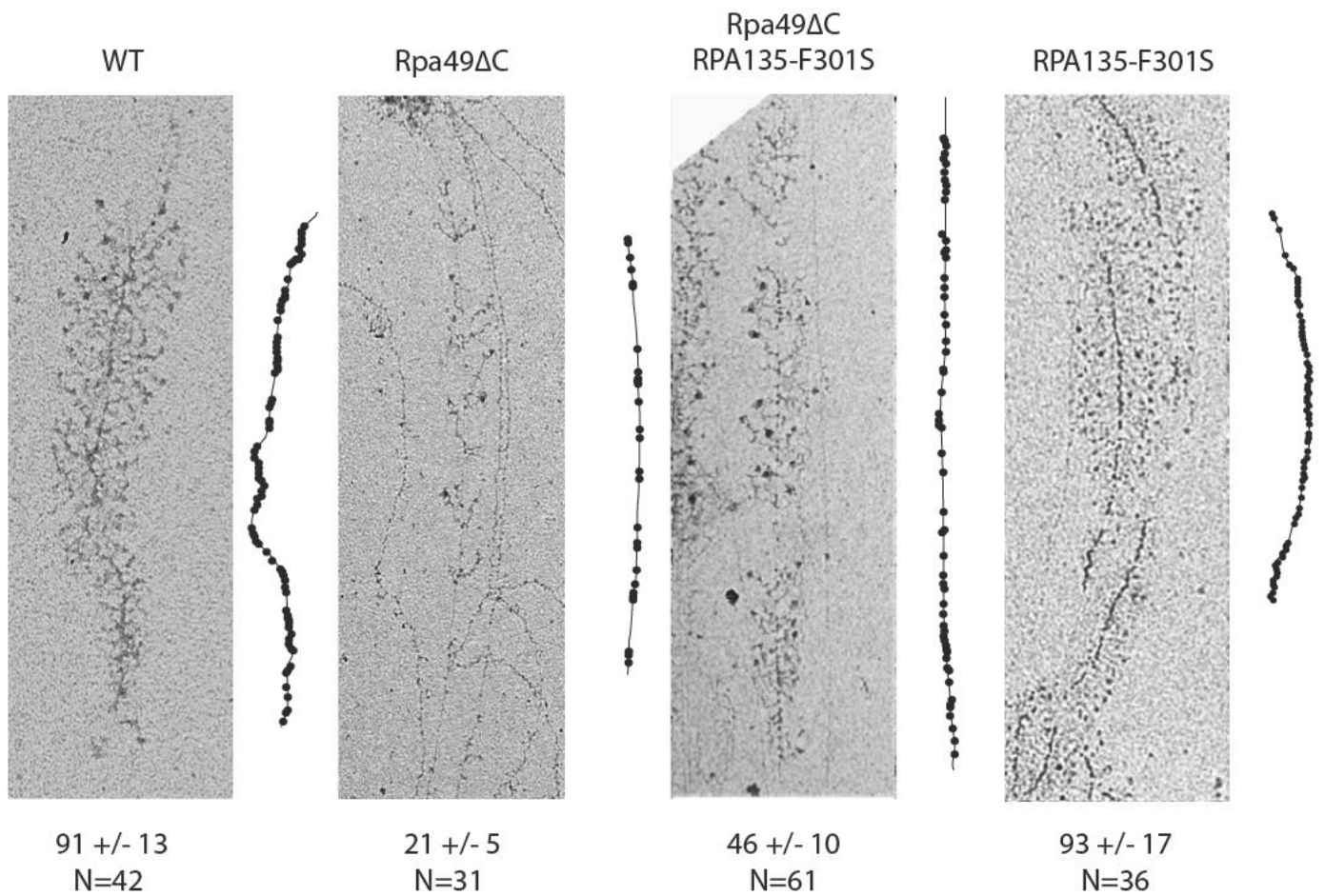
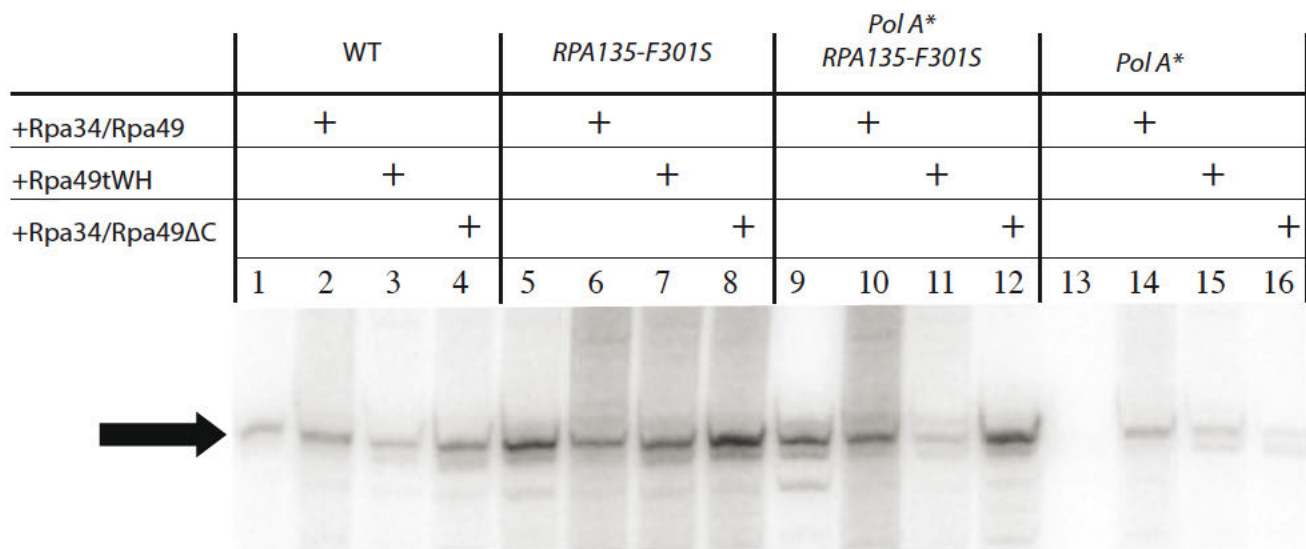
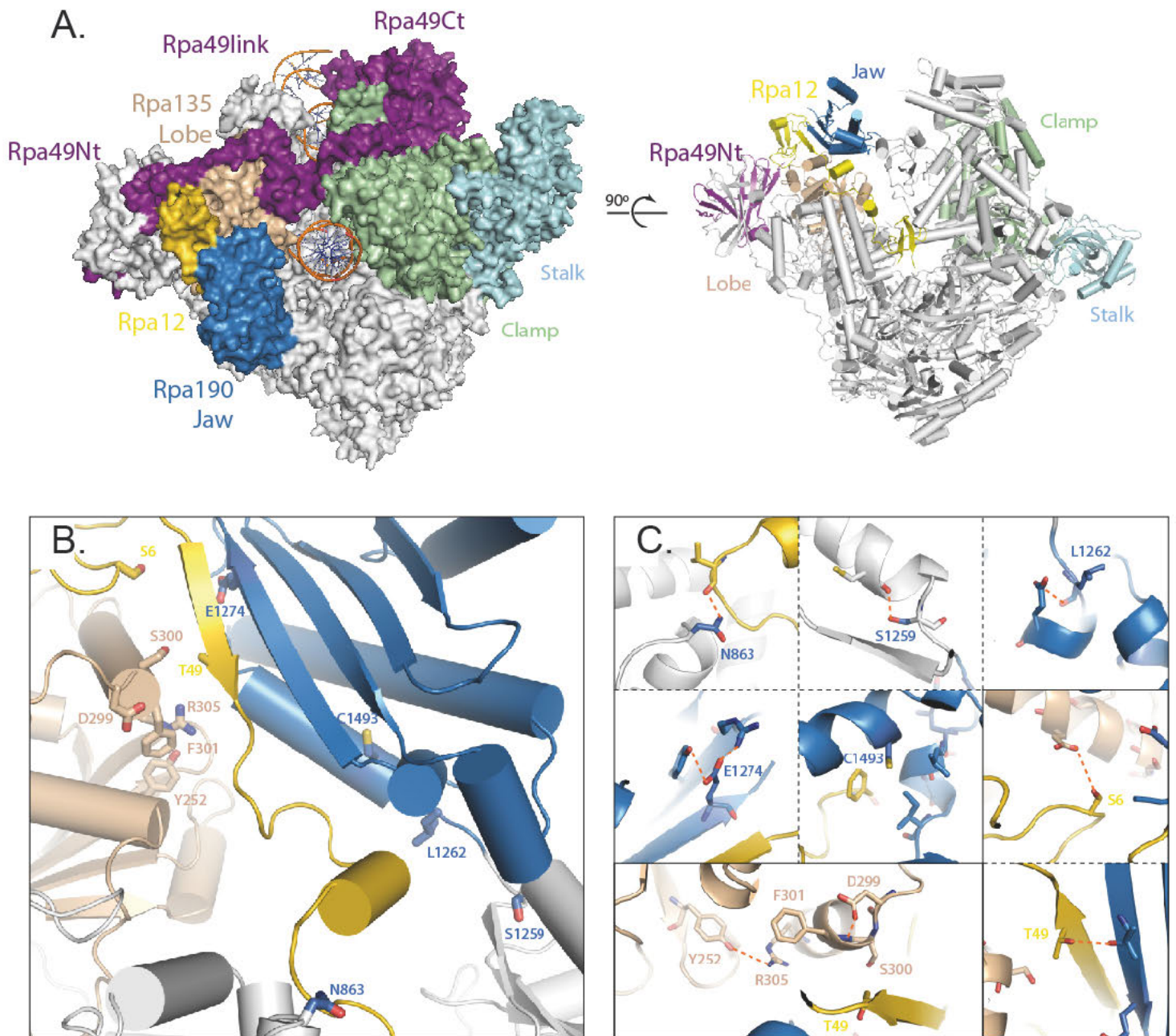


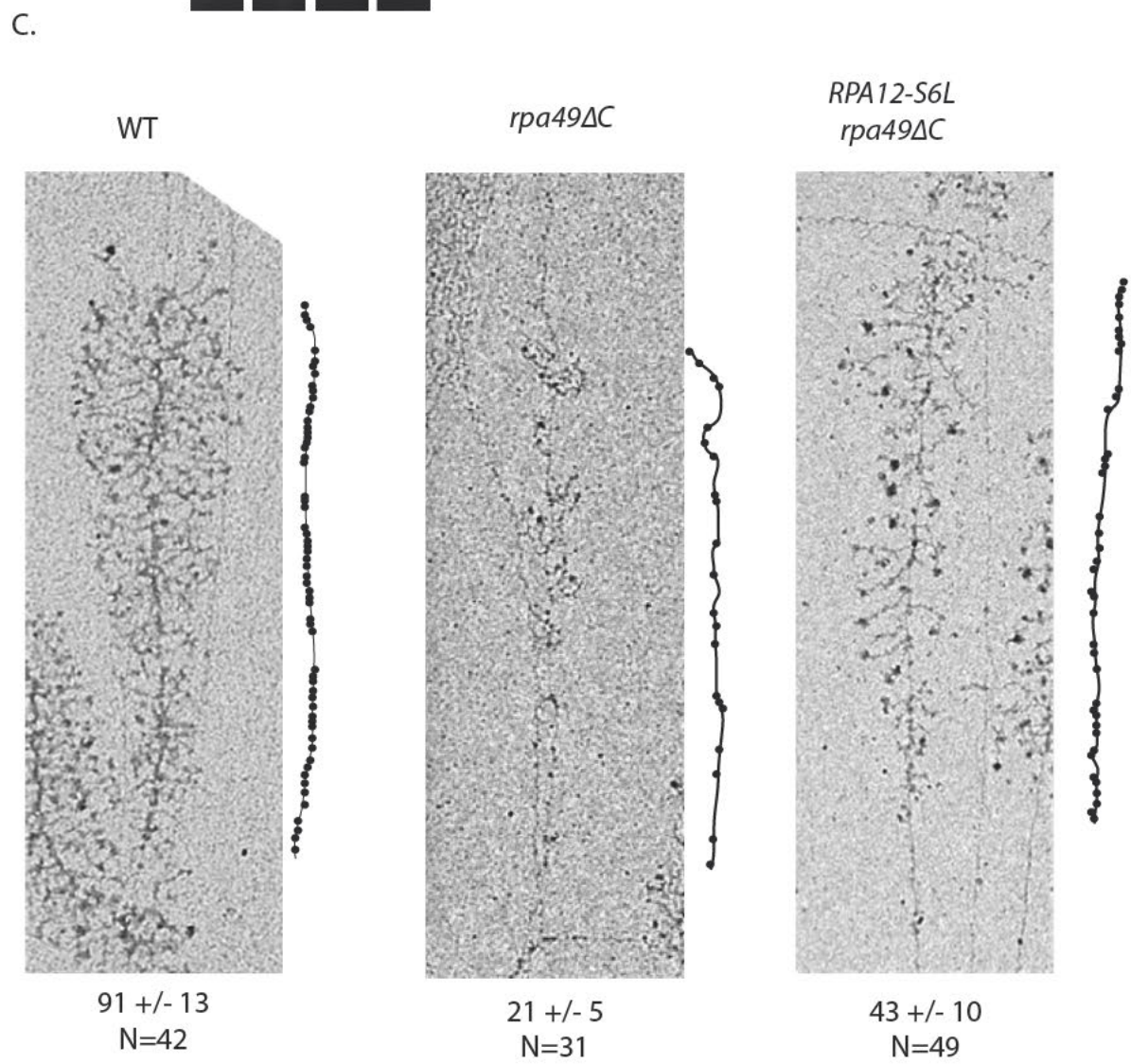
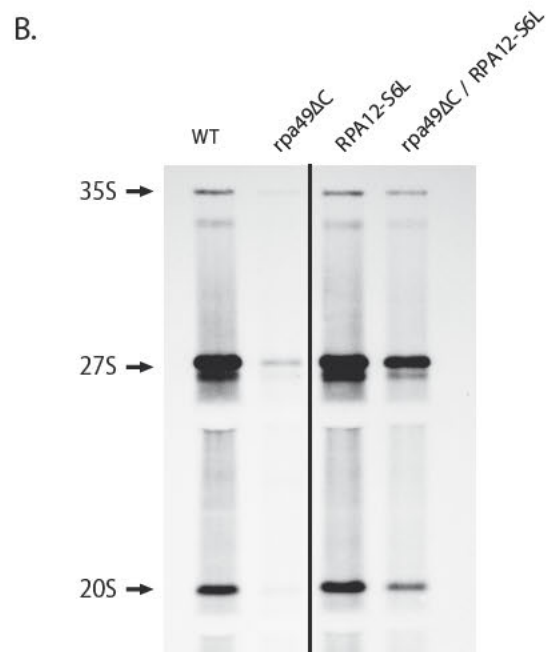
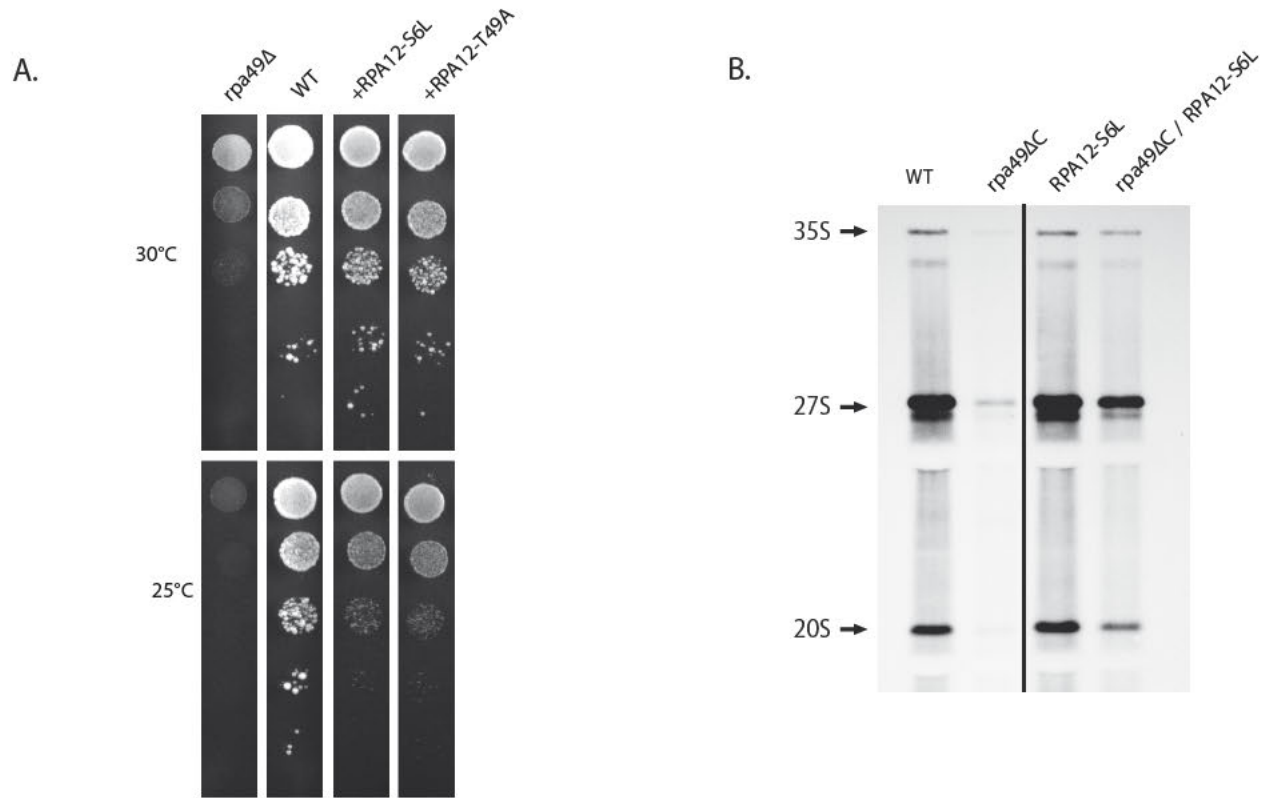
Figure 2.



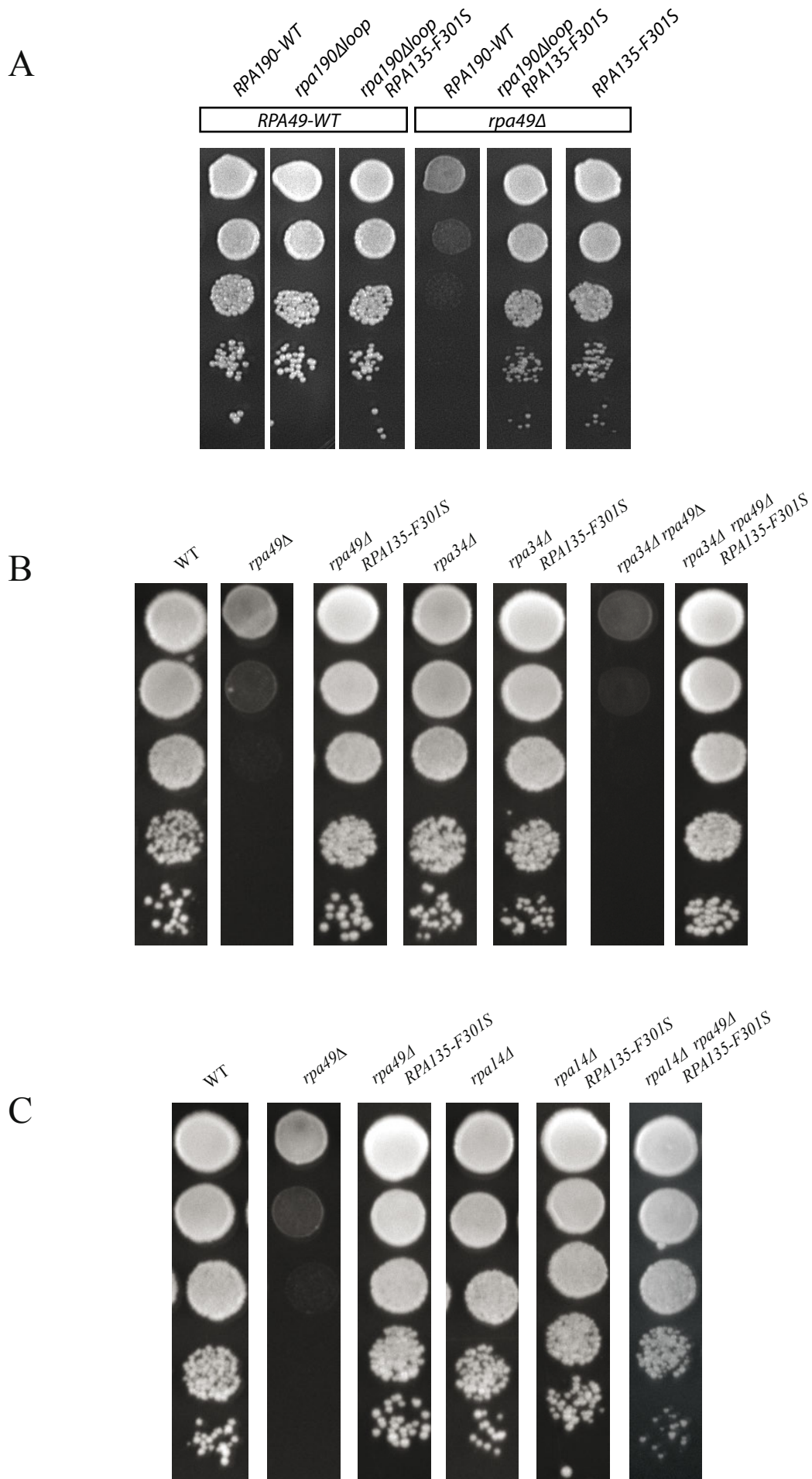
**Figure 3.**



**Figure 4.**

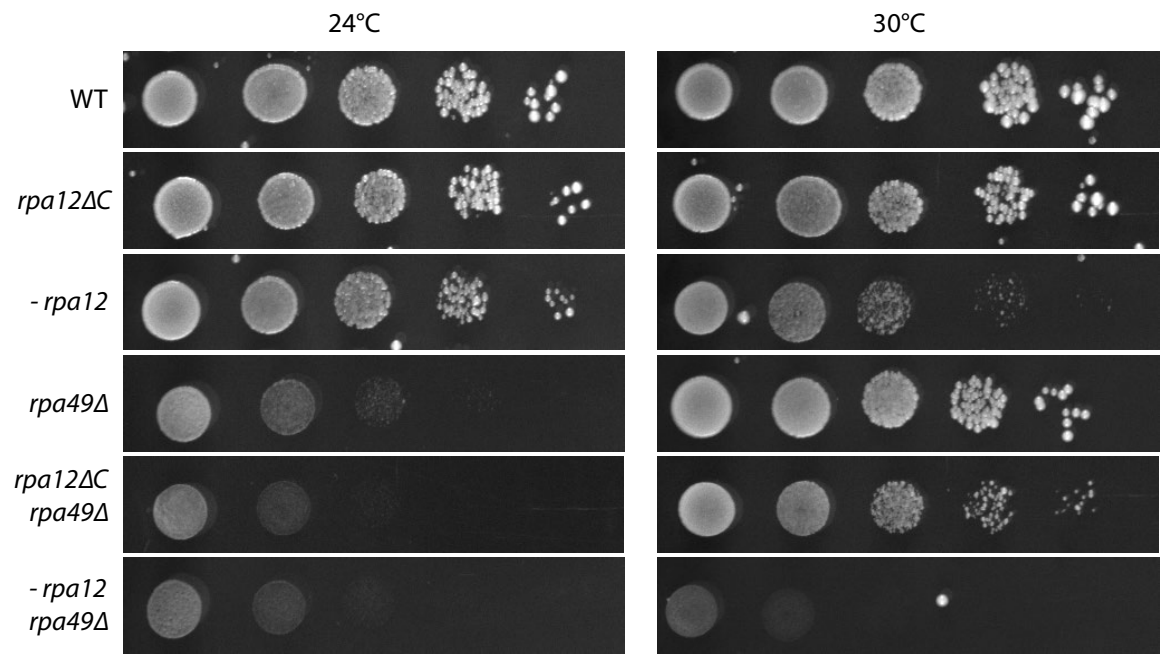


**Figure 5.**

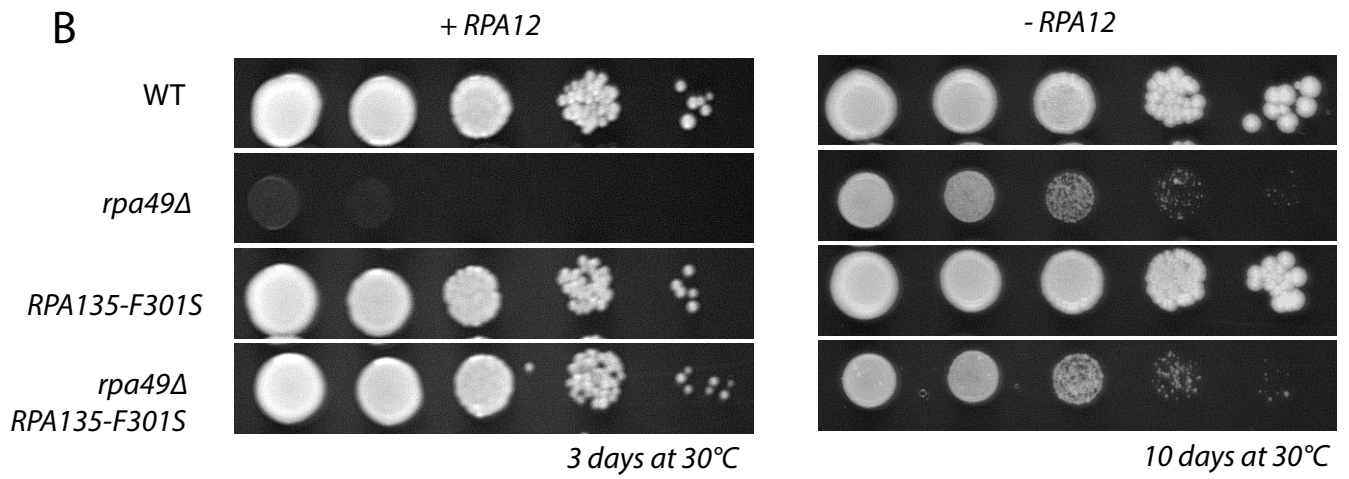


**Figure 6.**

A



B



C

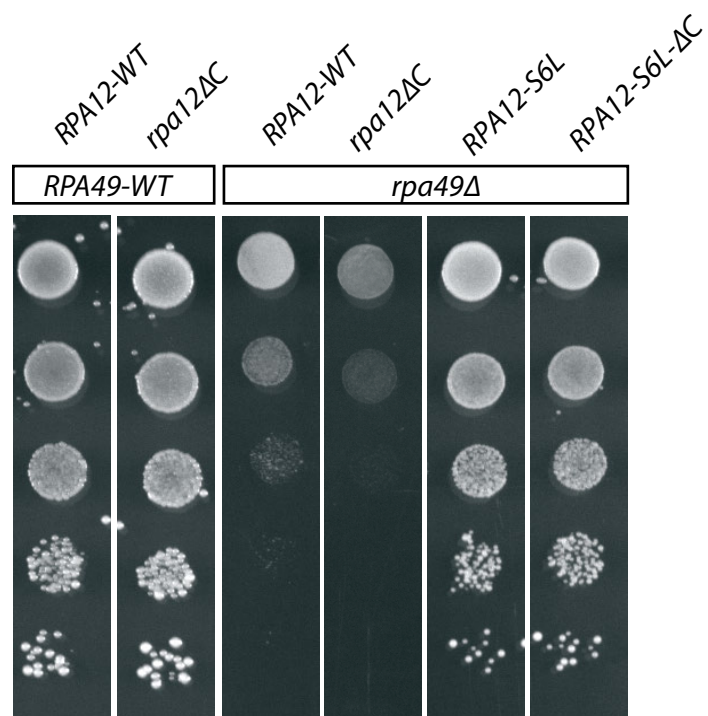
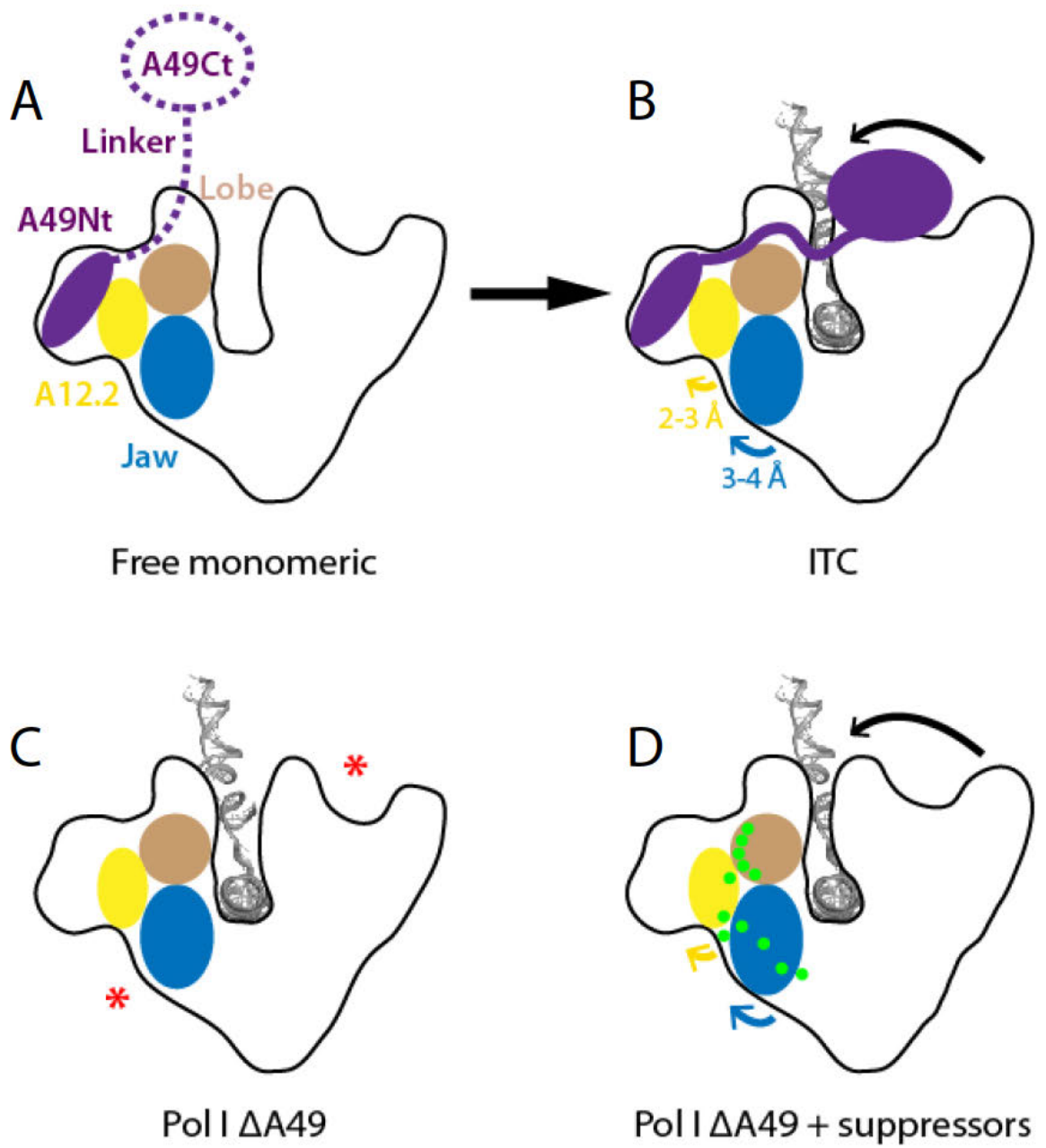


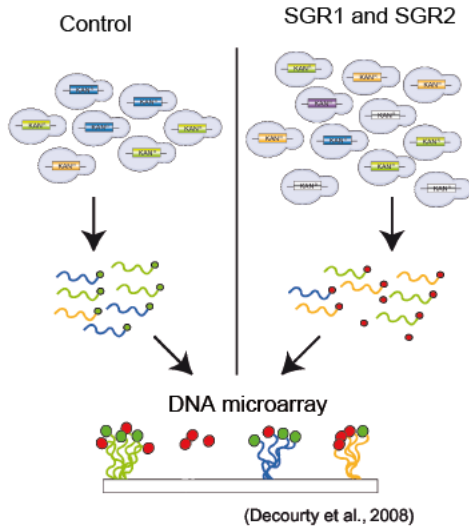
Figure 7.



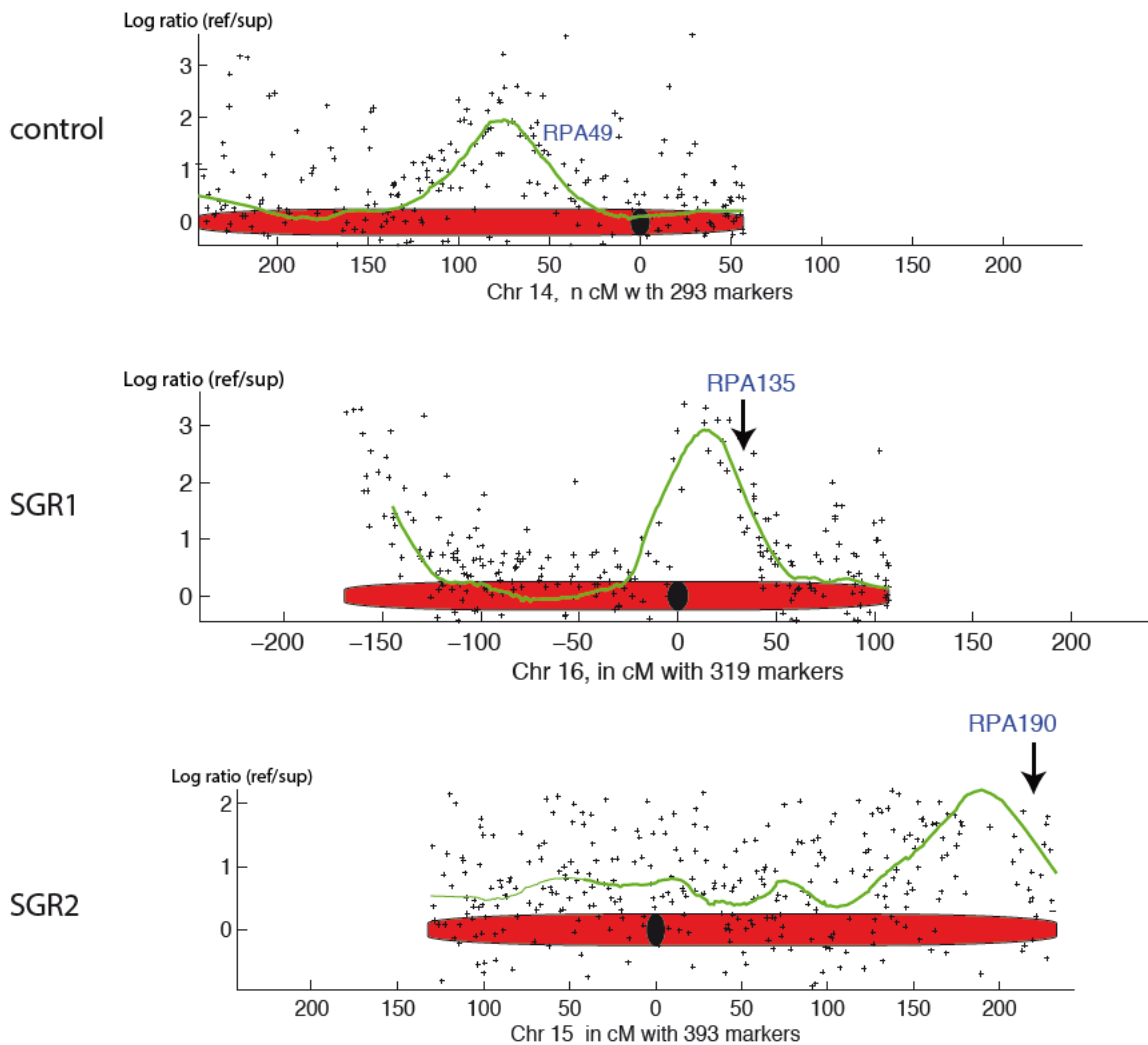
**Figure 8.**



A



B



Suppl. Figure 1 : Genetic interaction mapping (GIM) to identify suppressor mutations in SGR1 and SGR2. (A) Schematic representation of GIM interaction assay. Enrichment ratio between control and SGR are used as read-out to map genetic interactions. (B) Relative enrichment of each barcode (black cross) along chromosome (in red) are used to map genetic linkage. Green curve represents mean in a sliding windows of 20 barcodes. Local maximum in such curve was used to identify locus of interest: *RPA49* as positive control (upper panel, chr 14); *RPA135* in SGR1 (middle panel, chr 16) ; *RPA190* in SGR2 (lower panel, chr 15).

## **DISCUSSIONS/PERSPECTIVES**

Dans l'objectif de mieux caractériser et comprendre le fonctionnement des sous-unités spécifiques de l'ARN pol I, et donc plus globalement les mécanismes qui régissent l'activité de transcription de cette enzyme, nous avons réalisé des études génétiques sur des mutants de délétion de Rpa49, et sur des suppresseurs du défaut de croissance associé. Les résultats obtenus révèlent une région de la polymérase proche de Rpa12, au niveau d'une interface entre Rpa135 et Rpa190, qui aurait un rôle lors de l'initiation de la transcription. Cette région contient notamment l'élément "Expander" ou "DNA mimicking loop", décrit chapitre VI-1-d-, dont nous reparlerons ci-après. Ces données *in vivo* sont à relier aux structures récentes de l'ARN pol I en complexe avec Rn3 (Pisl *et al*, 2016 ; Engel *et al*, 2016 ; Torreira *et al*, 2017), en complexes d'initiation (Engel *et al*, 2017 ; Han *et al*, 2017 ; Sadian *et al*, 2017), et en élongation avec hybride ADN-ARN (Tafur *et al*, 2016 ; Neyer *et al*, 2016). Réunis, tous ces aspects permettent de proposer des hypothèses quant aux modes de fonctionnement moléculaires de l'enzyme et de ses facteurs, autant *in vitro* qu'au sein de l'organisme modèle *S.cerevisiae* dans notre cas, ainsi que d'imaginer de futures expériences pour compléter cette compréhension.

### **I – De la génétique aux modules structuraux sur l'ARN polymérase I chez *S.cerevisiae*.**

#### **I - 1 - Analyse d'une polymérase sans la sous-unité Rpa49 ou sans le C-ter de Rpa49.**

Rpa49 fait partie des "Built-in Transcription Factors" de l'ARN pol I, ces sous-unités homologues à des facteurs de transcription de l'ARN pol II, mais qui restent associées à

l'ARN pol I lors de la transcription. Afin de mieux caractériser le rôle de cette sous-unité, et en particulier de sa partie C-terminale flexible (domaine Rpa49-tWH), nous avons utilisé des mutants de délétion de Rpa49 ou de ce domaine tWH (*rpa49Δ / rpa49ΔtWH*). Nos résultats (tests de croissance, arbres de Miller, et *in vitro* dans des "promoter dependant assays"), montrent un fort défaut de croissance et d'initiation de la transcription dans une souche de levure qui n'exprime pas Rpa49 ou Rpa49-tWH, défauts qui peuvent être corrigés par les suppresseurs extragéniques isolés dans ce travail (mutations ponctuelles dans Rpa135, Rpa190 ou Rpa12). On notera un point important concernant le nombre d'ARN polymérase I par gène d'ARNr, étudiable dans les expériences de Miller's spread : ce nombre dans les conditions WT est d'environ 100 ARN pol I/gène dans nos conditions d'étude, alors qu'il est de seulement 20 en moyenne dans le cas *rpa49ΔtWH*, et de 45-50 environ dans le double mutant *rpa49ΔtWH* + RPA135-F301S, donc en présence de la mutation ponctuelle suppresseur. Ainsi, l'ajout de cette mutation dans la souche *rpa49ΔtWH* ne restaure pas totalement le nombre de polymérase I par gène d'ARNr, bien que la croissance de cette souche soit quasiment sauvage. Cela sous-entendrait que malgré un nombre moins élevé de polymérase par gène, la synthèse d'ARNr serait finalement aussi efficace à l'échelle cellulaire, ce qui est notamment cohérent avec les données de marquage métabolique des ARN néo-synthétisés effectués dans ces différentes souches (voir partie Résultats). Ainsi, en présence de la mutation suppresseur, l'efficacité des ARN pol I serait augmentée lorsque la sous-unité Rpa49 est absente. Ceci pourrait être dû au fait que les polymérase initient mieux / plus facilement au niveau du promoteur, ou a une meilleure efficacité en élongation/processivité des ARN pol I, ce qui reste méconnu. Pour comparaison, les arbres obtenus dans la condition RPA135-F301S en présence de Rpa49 sont exactement les mêmes que pour un WT, avec un nombre de polymérase par gène identique, de même que la synthèse d'ARNr évaluée par marquage métabolique qui est identique.

On rappelle que d'autres suppresseurs extragéniques existent, tels que la délétion du N-ter de Rpa43, la surexpression de HmoI, ou des mutants de troncation de Spt5. On notera que l'analyse ultrastructurale du nucléole de levure en microscopie électronique telle que réalisée par Albert *et al*, 2011, n'a pas été réalisée dans nos travaux en présence de nos suppresseurs. Cette étude montrait une altération forte de la structure nucléolaire dans un *rpa49Δ* comparé à un sauvage, révélateur de l'impact de cette sous-unité spécifique dans la structuration du nucléole. Il serait ainsi intéressant de regarder si le nucléole se restructure, et si oui à quel niveau (partiellement, totalement?) dans les souches avec suppresseurs.

Un crible génétique ("Genetic Interaction Mapping" (GIM, voir partie Résultats) a permis d'identifier dans les souches suppresseurs des mutations positionnées dans les deux plus grandes sous-unité de la polymérase, qui représentent à elles seules plus de 50% de la masse moléculaire de l'enzyme. En observant la localisation sur la structure de l'ARN pol I de la grande majorité de ces mutations suppresseurs, il est apparu clair que la région juste à côté de Rpa12 était impliquée dans le mécanisme de suppression, et c'est donc logiquement que nous avons cherché et trouvé de nouveaux suppresseurs dans Rpa12 même.

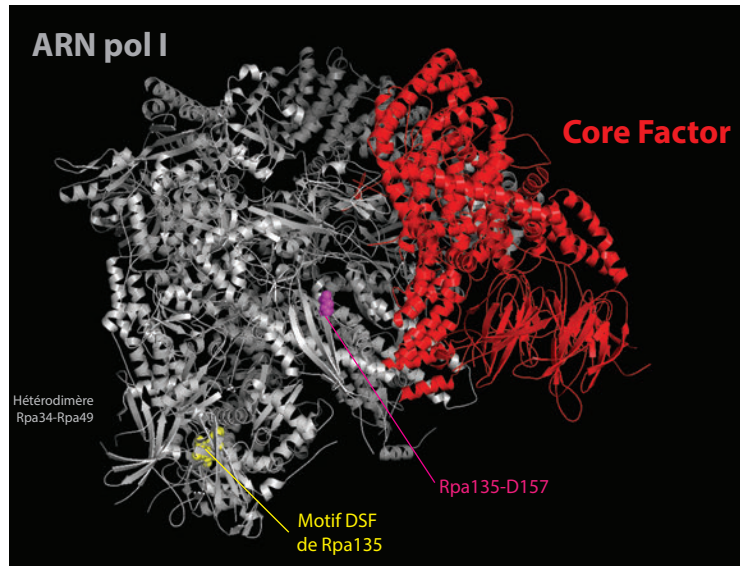
Il est bien sûr important de noter que le crible n'est pas saturé, et que des allèles suppresseurs pourraient être trouvés dans d'autres sous-unités de l'ARN pol I, notamment autour de cette région très localisée, mais pas uniquement. Il est intéressant de noter que quatre des mutations ponctuelles totales identifiées ne sont cependant pas retrouvées dans ce "Hot-spot" proche de Rpa12 N-ter. Or, on note tout de même que trois d'entre elles sont retrouvées proches du centre catalytique, et notamment de la position qu'adopte le C-ter de Rpa12 inséré dans le canal allant au site actif, ce qui n'est pas sans lien avec les résultats obtenus concernant cette sous-unité Rpa12. La force de suppression en croissance de ces mutants portant ces 3 mutations est classée comme "intermédiaire" comparée à l'allèle Rpa135-F301S par exemple. Enfin, une dernière mutation (RPA135-D157N), est placée sur une face externe de Rpa135 et

est isolée du reste des allèles. Cet allèle suppresseur possède une force de suppression en croissance assez faible, et pourrait au vu de sa localisation jouer un rôle dans l'interaction de l'ARN pol I avec des facteurs extérieurs lors de certaines étapes de la transcription (formation du PIC, transition ITC-EC, terminaison, etc...) (Figure 31A). Notamment, en observant la structure de l'ARN pol I en complexe de PIC (Engel *et al*, 2017), on remarque que la position de cette mutation fait face au Core Factor, et plus précisément au domaine TPR de Rrn11 (voir Chapitre VI -2-b-A/). Pour rappel, le domaine TPR est connu pour promouvoir les interactions protéine-protéine, souvent dans l'assemblage de complexes multi-protéiques (Blatch & Lässle, 1999). En observant la structure, on remarque qu'un écart assez faible entre ces deux positions demeure, mais il est alors important de noter que cette structure à haute résolution obtenue en 2017 n'inclut ni l'UAF ("Upstream Activating Factor") ni la TBP ("TATA Binding Protein"), qui pourraient peut être s'intercaler à ce niveau de l'enzyme lors de la formation du PIC. Ces deux facteurs n'ont par ailleurs jamais été obtenus lors d'études structurales de complexes de l'ARN pol I, ou d'études génétiques (suppression, co-létalité...), ce qui pourrait être utile à la prochaine étape dans la compréhension du fonctionnement de l'ARN pol I en initiation.

### **I - 2 - Les positions possibles du domaine C-terminal de Rpa49.**

Plus particulièrement, nous parlerons ici de la partie C-terminale de Rpa49, possédant une région tandem Winged Helix (tWH), qui n'a pas pu être cristallisée avec l'ARN pol I entière lors des études à haute résolution de 2013 (Engel *et al*, 2013 ; Fernandez-Tornero *et al*, 2013) du fait de sa flexibilité. Cependant, des études structurales plus récentes ont proposé différentes positions du domaine tWH de Rpa49, en observant des densités électroniques en Cryo-MET pouvant correspondre à ce domaine, et en remplaçant la structure à haute résolution de Rpa49-tWH cristallisée à part au sein de ces densités. Or, dépendant de l'étude, et donc de la phase du cycle transcriptionnel étudiée (formation du PIC, élongation...), la position des

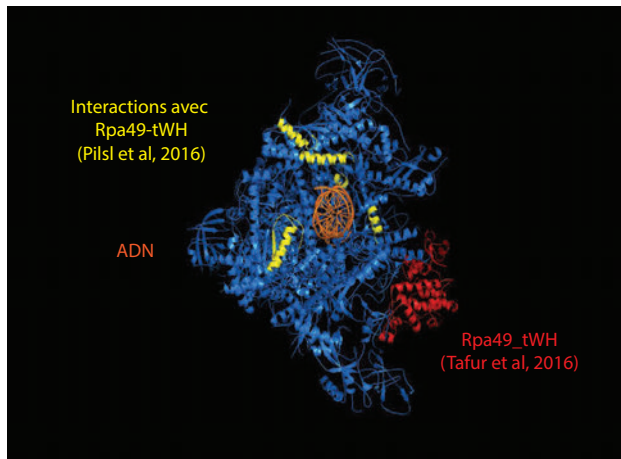
A.



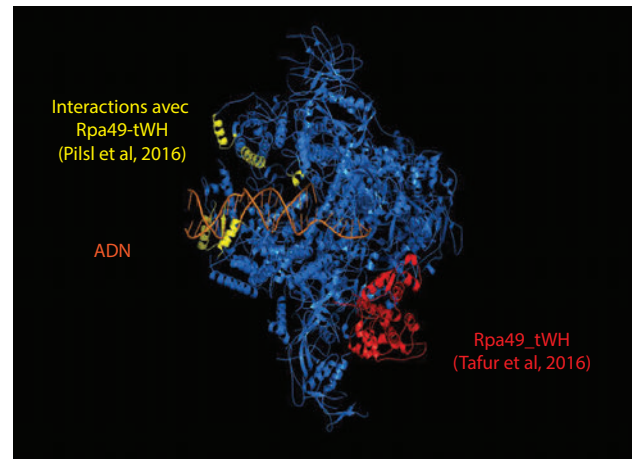
Engel *et al*, 2017

B.

### ARN pol I en complexe d'élongation

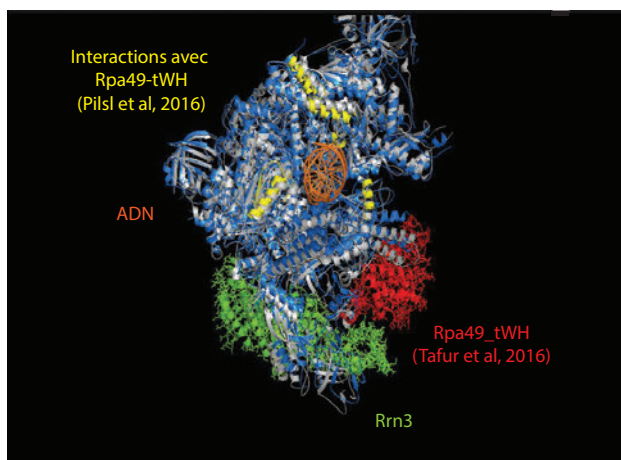


90°

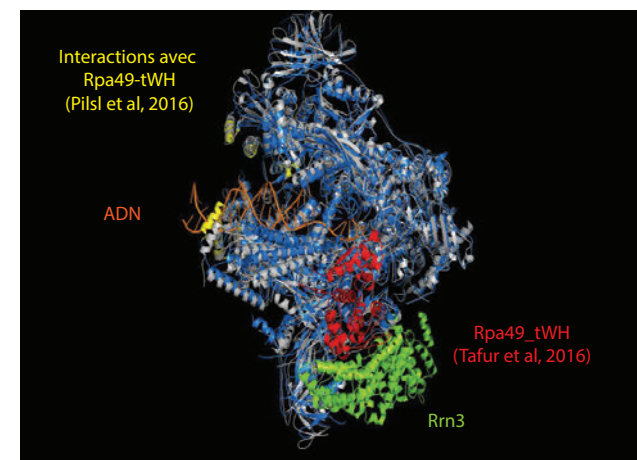


↓ Superposition avec ARN pol I en complexe avec Rrn3 ↓

C.



90°



PilsI *et al*, 2016 ; Engel *et al*, 2016 ; Tafur *et al*, 2016

**Figure 31. Structures de l'ARN pol I et position du domaine C-terminal de Rpa49.**

A. Structure de l'ARN pol I (en gris) en complexe avec le Core Factor (CF en rouge) (Engel et al, 2016 ; pdb : 5N5Z). La mutation ponctuelle RPA135-D157N suppresseur du défaut de croissance de la souche *rpa49Δ* est représentée par des sphères (en rose). Pour information, les acides aminés du motif DSF (299-301) de Rpa135, qui peuvent également être mutés et font partie du «Hotspot» de suppresseurs, sont indiqués en jaune.

B. Structure de l'ARN pol I (en bleu) en complexe d'élongation avec l'ADN (orange), avec le domaine tWH de Rpa49 (en rouge) (Tafur et al, 2016 ; «EC\_tWH», pdb : 5M64). Pour comparaison, les résidus de l'ARN pol I (Clamp de Rpa190(256–303) ; les régions entourant l'élément Expander Rpa190(1,320–1,337 et 1,440–1,456) ; la région du Lobe de Rpa135(222-226) et ABC23(87-123)) interagissant avec ce domaine tWH dans l'étude de Pils et al, 2016, sont indiqués en jaune, suggérant une différence de positionnement de Rpa49-tWH dans les deux structures (élongation et initiation respectivement). Deux orientations de la polymérase sont représentées.

C. Superposition de la structure du B. (bleue), avec celle de l'ARN pol I (en gris) en complexe avec Rrn3 (vert). Cette superposition montre une proximité immédiate, voire collisionnelle, entre ce facteur Rrn3 et le domaine Rpa49-tWH (rouge) de la structure en élongation (Tafur et al, 2016). Les deux orientations de la polymérase sont représentées comme en B.

Le logiciel Pymol est utilisé pour les visualisations et modélisations 3D.

densités attribuées à ce tWH semble varier (Figure 31.B et C). Ainsi, sur la base des structures disponibles dans la RCSB-PDB (<http://www.rcsb.org/pdb/home/home.do>), et en modélisant *in silico* ces positions, on réalise que Rpa49-tWH peut adopter au moins deux placements différents sur la structure :

1) Pilsl *et al* (2016) proposent, à partir d'images obtenues en Cryo-MET d'une sous-population de molécules d'ARN pol I compétentes pour l'initiation, une modélisation de la position du domaine tWH au-dessus du Cleft de l'ARN pol I, où peut s'insérer l'ADN (Figure 31.B.). On notera que plusieurs positions possibles du tWH ont été trouvées dans cette étude, mais que celle-ci représente la très grande majorité des molécules analysées. Le domaine tWH de Rpa49 ferait un "pont" entre les domaines Clamp et Protrusion de l'ARN pol I, position qui pourrait correspondre à celle déterminée par immuno-EM (Bischler *et al*, 2002) et cohérente avec les données de crosslink chimiques et de spectrométrie de masse obtenues précédemment (Jennebach *et al*, 2012). Dans la figure 31.B, les positions des résidus d'interaction entre Rpa49-tWH et le reste de la polymérase, indiqués par Pilsl *et al*, 2016, ont été surlignés en jaune. Pour plus de clarté dans les superpositions nous intéressant *in silico* ensuite, on note que la structure utilisée pour cette modélisation est celle de l'ARN pol I en élongation (EC\_tWH) ( Tafur *et al*, 2016 ; Pdb : 5M64). Dans cette structure, l'ADN est également modélisé.

2) Tafur *et al*, 2016, ont obtenu plusieurs complexes de l'ARN pol I en élongation (EC). Or, un d'entre eux nous intéresse tout particulièrement : EC\_tWH, contenant une densité attribuée au C-ter de Rpa49. Dans ce complexe en élongation, la position de Rpa49-tWH (en rouge Figure 31.C.) change comparé à 1), et est proche du Stalk (Rpa14-Rpa43) de la polymérase. On peut ainsi superposer cette structure avec une de l'ARN pol I en complexe avec le facteur Rrn3 (ici tirée de l'étude de Engel *et al*, 2016). Si l'on modélise Rrn3 (en vert sur la Figure



31.C.), on observe clairement qu'un clash stérique pourrait avoir lieu entre les deux protéines (Rpa49-Cter et Rrn3).

2)-bis- Enfin, il est intéressant de mentionner que l'étude récente du PIC (Engel *et al*, 2017) de l'ARN pol I fait également mention d'une densité électronique attribuable au domaine tWH de Rpa49, à un endroit similaire à celui trouvé par Tafur *et al* (2016), en élongation. Globalement, la région de l'ARN pol I où l'on retrouve cette densité dans les deux cas est la même, mais la position et l'orientation exacte de ce domaine ne sont pas totalement identiques, suggérant plusieurs conformations possibles. Cela signifierait que le C-ter flexible de Rpa49 pourrait changer de conformation et de localisation avant la formation totale du PIC, à un état conformationnel de transition vers l'état d'élongation.

Ces différentes positions suggèrent que le C-ter de Rpa49, mobile durant la transcription, pourrait par exemple accompagner l'ADN au début de la formation du PIC, et jouer un rôle dans le relargage de Rrn3 lors de la transition ITC - EC, comme proposé préalablement (Beckouet *et al*, 2008). Toutes ces transitions, et leurs rôles *in vivo*, restent méconnues, et de nouvelles études génétiques, biochimiques et structurales ciblées sont nécessaires afin de connaître et caractériser les différentes transitions conformationnelles de l'ARN pol I lors de la transcription.

Enfin, on notera que des structures à haute résolution de complexes de l'ARN pol I en état de pause/Backtrack lors de l'élongation, ou même lors de la terminaison de la transcription, sont également manquantes pour parfaire la compréhension moléculaire de cette enzyme.

## II - Un lien avec le facteur d'initiation Rrn3 ?

La structure du facteur d'initiation Rrn3 en complexe avec l'ARN pol I a été obtenue, et sa localisation sur l'enzyme, au niveau des sous-unités Rpa14 et Rpa43 du Stalk, correspond à un état "compétent pour initiation" (PilsI *et al*, 2016 ; Engel *et al*, 2016 ; Torreira *et al*, 2017). Or, il a été suggéré que le relargage de Rrn3 en élongation, ou à un état transitionnel d'initiation à élongation, dépendrait des sous-unités Rpa34 et Rpa49 (Beckouet *et al*, 2008). Notamment, il a été montré dans cette étude que le recrutement et le relargage de Rrn3 sont altérés dans un mutant *rpa49Δ* (montré par ImmunoPrécipitation de la Chromatine, ou "ChIP"). En effet, comme nous l'avons vu précédemment, la position de Rrn3 dans les structures n'est pas compatible avec celle de Rpa49-tWH en élongation (clash stérique probable entre les deux protéines). Rpa49-tWH pourrait jouer un rôle dans le déplacement de Rrn3 à l'une des étapes cruciales pour passer d'une polymérase en initiation à une polymérase en élongation processive, via des interactions directes. En comparant les structures des complexes ARN pol I - Rrn3 avec les structures en élongation comprenant le tWH de Rpa49, on peut visualiser *in silico* les régions et les résidus précis qui pourraient entrer en collision entre les deux protéines. Notamment, des expériences de mutagenèse dirigée sur les résidus de l'hélice- $\alpha$  de Rpa49(308-315), et en particulier Q309 et K314, pourraient être envisagées pour abolir ce type de collision avec Rrn3. Il serait alors intéressant de regarder *in vivo* la croissance de ces souches, et la capacité de recrutement et relargage du facteur Rrn3, et de comparer si des effets similaires à ceux observés dans une souche *rpa49Δ* sont observés. Si c'était le cas, un rôle précis de Rpa49-tWH pourrait être proposé dans la dissociation de Rrn3 lors de la transcription. Pour étayer cette hypothèse, des expériences de Chip sur Rrn3 dans les souches *rpa49Δ* contenant nos supresseurs sont judicieuses et complémentaires, et en cours au laboratoire. Le recrutement de Rrn3 au promoteur de l'ARN pol I, et son relargage par la suite, pourraient être évalués et comparés à une souche dans laquelle seule la mutation

*rpa49Δ* est présente. Pour rappel, toutes les mutations ponctuelles suppresseurs des phénotypes associés au mutant *rpa49Δ* sont structurellement très éloignées du Stalk de la polymérase, et donc de la région d'interaction avec Rrn3, ainsi que de la position de Rpa49-tWH en élongation (Tafur *et al*, 2016), suggérant au premier abord qu'aucun lien direct ne pourrait être établi. Cependant, les réarrangements allostériques, et le positionnement exact du long linker reliant les domaines N- et C-ter (tWH) de Rpa49 ne sont pas connus précisément, et pourraient faire intervenir cette région du Hot-spot de suppresseurs, à l'interface entre Rpa190, Rpa135 et Rpa12 (voir partie Résultats).

Enfin, un dernier point intéressant concerne la construction de la fusion Rrn3-Rpa43, nommée CARA. Cette construction est viable chez la levure *S.cerevisiae*, ne provoque qu'un léger défaut de croissance, mais ne peut plus former d'homodimères d'ARN pol I (Laferté *et al*, 2006). La rapamycine est une drogue ciblant la voie TORC1 qui inhibe l'activité transcriptionnelle. Suite à l'inhibition de la voie TORC1 par la rapamycine, la forme protéique CARA devient résistante à l'inhibition transcriptionnelle de l'ARN pol I (Laferté *et al*, 2006). Il est à noter que *in vivo*, la croissance de la souche CARA, conservant une transcription de l'ARNr en condition de carence nutritionnelle, est hyper-sensible aux traitements par la rapamycine (Laferté *et al*, 2007). On note que la fusion CARA n'est fonctionnelle que si largement surexprimée. Des études ont été effectuées dans mon équipe sur des souches de levure portant cette fusion de protéines. Notamment, on notera que la combinaison CARA + *rpa49Δ* dans une souche de *S.cerevisiae* est létale (Beckouët *et al*, 2008). De plus, un fort défaut de croissance est également observé dans une souche combinant CARA + *rpa34Δ*. Or, l'expression de la sous-unité du Stalk Rpa43 seule dans ce dernier contexte permet de restaurer totalement la croissance de la souche double mutante (F.Beckouët, communication personnelle, résultats non-publiés). Il serait intéressant de tester les effets de la surexpression

de Rpa43 dans le contexte *CARA*+ *rpa49Δ*, avec par exemple un promoteur inductible pour l'expression de RPA49, afin de voir si la croissance peut aussi être restaurée.

La viabilité de la souche *CARA*, dans laquelle la construction constitutive *Rrn3*-*Rpa43* fait que cette fusion de protéines est indissociable de l'enzyme, atteste d'un déroulement du cycle de transcription compatible avec la croissance (formation du PIC, transitions en initiation et initiation-élongation, terminaison...). On peut donc émettre l'hypothèse dans cette souche que la partie C-terminale de *Rpa49* serait requise et forcerait le déplacement de *Rrn3*, permettant de passer correctement la/les transition(s) vers une élongation processive, rendant la souche viable (sans forcément dissociation complète de la fusion *Rpa43*-*Rrn3*). Dans ce même contexte *CARA*, si *Rpa49* n'est plus présente (*CARA* + *rpa49Δ*), le domaine tWH ne pourrait plus jouer son rôle crucial de déplacement de *Rrn3*. Ainsi, si le facteur *Rrn3*, indissociable de la polymérase, ne pourrait plus être déplacé, la transition vers l'élongation ne pourrait plus se faire correctement, conduisant à la létalité. De plus, seule la délétion totale de *Rpa49* a été évaluée avec *CARA* : la combinaison avec la délétion du C-ter de *Rpa49* uniquement n'a pas été testée, ce qu'il faudrait également caractériser. Le phénotype attendu serait similaire à celui de *CARA* + *rpa49Δ*, si *Rpa49*-tWH possédait un tel rôle crucial de déplacement, voire de dissociation/relargage de *Rrn3*. Il serait alors intéressant de travailler sur la souche portant la construction *CARA* dans un contexte *rpa49Δ*/*rpa49ΔtWH*, et d'y ajouter nos mutations suppresseurs, afin de voir si la croissance peut être restaurée, et avec quelle force. Dans cette optique, l'ajout de nos mutations suppresseurs situées dans la région autour de *Rpa12*-Nter, donc éloignées de cette région *Stalk*-*Rrn3*, ne devrait pas changer le phénotype directement, sauf si des réarrangements allostériques de grande envergure au sein de la polymérase entière s'effectuaient du fait de leur ajout, ce qu'il serait intéressant de tester. Si nos suppresseurs du Hot-spot identifiés précédemment supprimaient efficacement la létalité *CARA* + *rpa49Δ*, on pourrait alors imaginer qu'ils seraient impliqués dans le déplacement du C-ter de *Rpa49* (et de

son linker associé) vers la position de Rrn3. Ceci reste imaginable vu la taille assez élevée (83 résidus) du linker flexible de Rpa49, qui pourrait passer au-dessus ou même au sein de la région du Hot-spot de mutations.

### **III - La région à l'interface entre Rpa190, Rpa135 et Rpa12 aurait un rôle dans l'initiation de la transcription**

Nos travaux révèlent qu'en l'absence de Rpa49, ou du C-ter de Rpa49, ces souches mutantes possèdent un défaut d'initiation important, associé à un fort défaut de croissance. Nos données expérimentales obtenues sur les supresseurs montrent que nos mutations permettent à la polymérase de se passer du domaine tWH de Rpa49 (C-terminal) pour effectuer une initiation efficace.

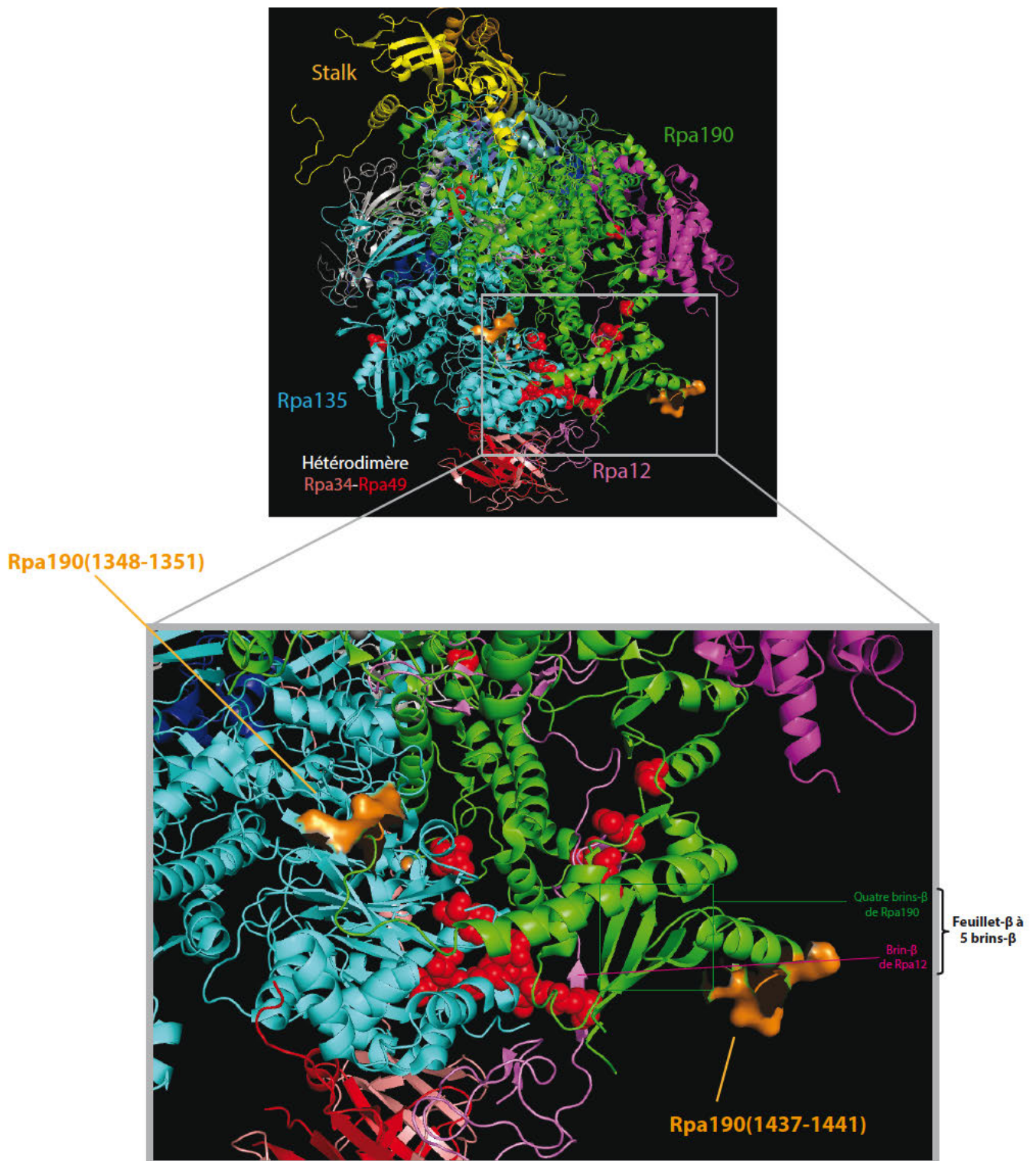
L'hypothèse de travail que nous formulons à ce stade est qu'un élément de l'ARN pol I inhiberait l'initiation de la transcription, par exemple pour éviter une fixation non-spécifique à de l'ADN non ribosomique, ou lorsque le complexe multi protéique nécessaire à l'initiation au niveau du promoteur de l'ARN pol I n'est pas complet/prêt pour l'initiation. Nous pensons ainsi que l'un des rôles du domaine Rpa49-tWH serait de lever cette inhibition, en interagissant avec cet élément même ou sa région proche, lorsque les conditions requises à l'initiation sont réunies. En l'absence de Rpa49-tWH, nos mutations ponctuelles supresseurs seraient capable de faciliter un/des changement(s) conformationnel(s) au sein de cette région de l'ARN pol I, permettant de mimer un rôle similaire à celui de Rpa49-tWH. On se demande ainsi quel peut être cet élément potentiellement inhibiteur impliqué?

Pour répondre à cette question, il faut avant tout rappeler que les mutations supresseurs extragéniques du défaut de croissance lié à *rpa49Δ* étudiées dans nos travaux sont très

majoritairement retrouvées dans une région ciblée de la polymérase, un "Hot-spot" à l'interface entre Rpa190, Rpa135, et Rpa12. Nous avons donc examiné *in silico* les éléments structuraux présents à proximité de ce Hot-spot, impliquant ces trois sous-unités de la polymérase. On note la présence d'un "canal" où s'insérerait Rpa12, entouré par des motifs structuraux précis de Rpa190 et Rpa135 (modules "Jaw" et "lobe" respectivement), dans lesquels on retrouve notre Hot-spot de mutations supprimeurs. Le domaine impliqué est extrêmement localisé sur la surface de la Pol I. En effet, nous avons montré que la partie C-ter de Rpa12 n'était pas impliquée dans ce mécanisme (aucune mutation supprimeur n'y est retrouvée, et un mutant de délétion de la partie C-ter de Rpa12 n'affecte pas la suppression). En revanche, en enlevant la totalité de Rpa12, la suppression n'est plus possible. La région N-ter et la région du linker proche du N-ter de Rpa12 sont donc cruciaux pour ce phénomène de suppression. En étudiant *in silico* la structure 3D de la polymérase, et plus précisément de cette région, on observe la présence d'un brin- $\beta$ , Rpa12(46-50), qui est phase avec une série de quatre brins- $\beta$  de Rpa190 formant une structure secondaire en feuillet- $\beta$  à cinq brins- $\beta$  (Figure 32, panel du bas). Concernant les quatre brins de Rpa190, dans la séquence primaire, on remarque que les deux premiers brins (en position 1266-1293) sont distants de près de 163 résidus des deux autres brins terminant ce motif structural (position 1456-1475). Au sein de ses 163 résidus, on trouve une région (position 1351-1437, Figure 32) dont la structure n'est pas résolue, et ce dans la plupart des études structurales analysées (Engel *et al*, 2013 ; Fernandez-Tornero *et al*, 2013 ; Tafur *et al*, 2016, Engel *et al*, 2017). Cette région est probablement cruciale pour l'inhibition de l'initiation de l'ARN pol I. En effet, c'est au sein de ce domaine que l'on trouve la boucle de 40 résidus (position 1360-1400), correspondant à l'élément caractéristique et spécifique de l'ARN pol I nommé "Expander" (Engel *et al*, 2013) ou "DNA-Mimicking Loop" (Fernandez-Tornero *et al*, 2013). La DNA-Mimicking Loop occupe une position précise dans le dimère d'ARN pol I de Engel *et al*, 2013 (pdb : 4C2M),

et est aussi positionnée dans deux des trois structures de Fernandez-Tornero *et al*, 2013. Elle est cependant absente dans la structure résolue à plus haute résolution dans cette dernière étude. Mais même lorsque la DNA-Mimicking Loop occupe une position précise, on note que les régions de part et d'autre de cet élément (positions 1351-1360 et 1400-1437) ne sont pas structurellement résolues.

Cette région de la polymérase contient donc des éléments nécessairement mobiles lors du cycle transcriptionnel. La région (1351-1437) de rpa190 serait ainsi capable d'inhiber l'initiation, en occupant le site d'insertion de l'ADN. Nous proposons que les mutations suppresses permettent de faciliter le déplacement de la DNA-mimicking loop pour permettre une initiation efficace. Cette hypothèse implique qu'en présence de Rpa49, la DNA-mimicking loop n'inhibe pas l'initiation, car déplacée par le domaine Rpa49-tWH. On notera également que les positions 1337 et 1440 de Rpa190 sont en interaction directe avec Rpa49-tWH dans une sous-classe de molécules ARN pol I-Rrn3 (voir Discussion I-2-, Pilsel *et al*, 2016). Cette région de contact, de part et d'autre de la DNA mimicking loop, renforce notre hypothèse : Rpa49-tWH stimulerait l'initiation en déplaçant la DNA mimicking loop, ce qui pourrait s'effectuer soit en facilitant des changements conformationnels dans cette région de l'enzyme, soit en interagissant physiquement avec la DNA mimicking loop. En l'absence de ce Rpa49-tWH, l'initiation serait donc fortement inhibée. Nos mutations suppresses joueraient ainsi un rôle similaire à celui de Rpa49-tWH en conditions WT. Pour tester directement cette hypothèse, il faudrait montrer que la double invalidation de Rpa49-tWH et de la DNA mimicking loop est capable d'effectuer une initiation efficace.



**Figure 32. Existe-t-il un lien entre l'élément «Expand/DNA mimicking loop», Rpa12 et Rpa49 en initiation de la transcription par l'ARN pol I?**

On retrouve ici sur la structure de l'ARN pol I (Pdb : 4C3I, Fernandez-Tornero et al, 2013) la position des mutations (sphères rouges) supprimeuses du défaut de croissance du mutant *rpa49Δ* ou *rpa49ΔC* dans Rpa190 (en vert) et Rpa135 (en cyan), autour de Rpa12 (en rose clair). Les deux positions correspondantes aux régions des résidus Rpa190(1348-1351) et Rpa190(1437-1441) ont été représentées en surfaces orange. La région de Rpa190 entre ces deux positions (d'environ 90 résidus) est manquante dans la structure, mais le placement de la grande majorité des mutations supprimeuses dans Rpa135 et dans le reste de Rpa190 coïncide avec cette position, suggérant une possible interaction entre cet élément de Rpa190, Rpa12 et le domaine C-ter de Rpa49 (ou son linker) durant la transcription.



#### **IV – Rpa12, un élément mobile de l'ARN pol I**

Nos travaux suggèrent donc que Rpa12 n'aurait pas seulement un rôle dans l'activité de stimulation du clivage (Backtrack/pauses/terminaison) ou d'assemblage de la polymérase (Nogi *et al*, 1993 ; Van Mullem *et al*, 2002 ; Kuhn *et al*, 2007), mais serait aussi requis pour la mobilité de l'élément "DNA mimicking loop". On rappelle qu'en l'absence de Rpa12, les suppresseurs ne sont plus capables de supprimer le défaut de croissance de *rpa49Δ* ou de moduler l'efficacité de l'initiation de la transcription. Plus précisément, la région proche de Rpa12, comprenant des régions de Rpa135 et Rpa190, et notamment l'élément "Expander" ou "DNA Mimicking Loop" de Rpa190, joueraient un rôle crucial dans les changements de conformation nécessaires pour passer certaines transitions entre les différents complexes d'initiation, vers un état d'élongation processive en complexe avec l'ADN.

Enfin, un dernier point concerne le rôle de Rpa12 dans la stabilité de l'ARN pol I, et plus précisément de sa plus grande sous-unité Rpa190. En effet, l'invalidation du gène codant pour Rpa12 chez *S.cerevisiae* provoque de légers défauts de croissance à 24/30°C, et n'est létale qu'à haute température (>37°C) , ce qui serait dû à l'instabilité de la polymérase. Dans de tels mutants, la concentration cellulaire en Rpa190 est bien plus basse que dans des cellules sauvages. De plus, ce phénotype de thermosensibilité peut être supprimé en surexprimant le gène codant pour Rpa190, suggérant que Rpa12 posséderait un rôle dans l'assemblage et la stabilité de Rpa190 dans l'ARN pol I (Nogi *et al.*, 1993).

Chez l'Homme, il est connu que des inhibiteurs spécifiques de l'ARN pol I, tel que le BMH-21, un intercalant de l'ADN, est capable de provoquer la mort cellulaire de lignées cellulaires tumorales, sans affecter la croissance des cellules contrôles (Peltonen *et al*, 2014 ; Colis *et al*, 2014). Concrètement, le BMH-21 se lie à l'ADN ribosomique, ce qui empêche l'ARN pol I de progresser sur l'ADNr et induit la dégradation de sa plus grande sous-unité Rpa194

(homologue humain de Rpa190) par le protéasome. Cela montre que les ARN pol I bloquées en cours d'élongation peuvent être ciblées par le protéasome, révélant des opportunités thérapeutiques nouvelles dans la lutte contre le Cancer. Il existe une analogie troublante entre le mécanisme d'action du BMH21 et le phénotype observé en l'absence de Rpa12. Nogi (1993), puis Van Mullen 2002), postulaient que l'absence de Rpa12 conduisait à un défaut d'assemblage de l'ARN Pol I, avec un défaut de stabilité de la plus grande sous-unité Rpa190. On pourrait proposer une hypothèse alternative : en l'absence de Rpa12, un mécanisme similaire de dégradation des ARN pol I engagées en élongation pourrait être imaginé chez *S.cerevisiae*. Cette hypothèse est directement testable. Seules les ARN pol I engagées en transcription sont dégradées par le protéasome. En l'absence de Rpa12, mais sans activité de synthèse d'ARNr, on prédit que rpa190 est stable. L'inhibition du protéasome devrait stabiliser Rpa190 en l'absence de Rpa12. Enfin, l'absence de Rpa12 devrait phénocopier le traitement par le BMH-21 dans des lignées tumorales humaines.

Cette hypothèse de travail est donc testable, et permettrait l'étude de la dégradation des ARN pol I en pause lors de la phase d'élongation. La dégradation de l'ARN pol II lors de pauses dites "terminales" est maintenant bien connu. Un mécanisme similaire pour l'ARN pol I, identifié par le traitement BMH21, impliquerait alors Rpa12, qui participerait donc à deux types de contrôle : la stimulation du clivage des ARN en "Backtrack", et le ciblage vers la machinerie de dégradation de l'ARN pol I en pause "terminale". Cette perspective est clairement un sujet qu'il serait important de poursuivre.

## REFERENCES

- Aitchison JD & Rout MP (2012) The yeast nuclear pore complex and transport through it. *Genetics* **190**: 855–883
- Albert B, Léger-Silvestre I, Normand C & Gadal O (2012) Nuclear organization and chromatin dynamics in yeast: biophysical models or biologically driven interactions? *Biochim Biophys Acta* **1819**: 468–481
- Albert B, Léger-Silvestre I, Normand C, Ostermaier MK, Pérez-Fernández J, Panov KI, Zomerdijk JCBM, Schultz P & Gadal O (2011) RNA polymerase I-specific subunits promote polymerase clustering to enhance the rRNA gene transcription cycle. *J Cell Biol* **192**: 277–293
- Andersen JS, Lam YW, Leung AKL, Ong S-E, Lyon CE, Lamond AI & Mann M (2005) Nucleolar proteome dynamics. *Nature* **433**: 77–83
- Aprikian P, Moorefield B & Reeder RH (2001) New model for the yeast RNA polymerase I transcription cycle. *Mol Cell Biol* **21**: 4847–4855
- Armache K-J, Mitterweger S, Meinhart A & Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**: 7131–7134
- Arnold JJ, Smidansky ED, Moustafa IM & Cameron CE (2012) Human mitochondrial RNA polymerase: structure-function, mechanism and inhibition. *Biochim Biophys Acta* **1819**: 948–960
- Arnone JT, Walters AD & Cohen-Fix O (2013) The dynamic nature of the nuclear envelope: lessons from closed mitosis. *Nucleus* **4**: 261–266
- Bar-Nahum G, Epshtein V, Ruckenstein AE, Rafikov R, Mustaev A & Nudler E (2005) A ratchet mechanism of transcription elongation and its control. *Cell* **120**: 183–193
- Basehoar AD, Zanton SJ & Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709
- Beck M, Förster F, Ecke M, Plitzko JM, Melchior F, Gerisch G, Baumeister W & Medalia O (2004) Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**: 1387–1390
- Beckouet F, Labarre-Mariotte S, Albert B, Imazawa Y, Werner M, Gadal O, Nogi Y & Thuriaux P (2008) Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol* **28**: 1596–1605

- Beckouët F, Mariotte-Labarre S, Peyroche G, Nogi Y & Thuriaux P (2011) Rpa43 and its partners in the yeast RNA polymerase I transcription complex. *FEBS Lett* **585**: 3355–3359
- Bell SD, Magill CP & Jackson SP (2001) Basal and regulated transcription in Archaea. *Biochem Soc Trans* **29**: 392–395
- Bell SP, Jantzen HM & Tjian R (1990) Assembly of alternative multiprotein complexes directs rRNA promoter selectivity. *Genes Dev* **4**: 943–954
- Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G & Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**: 1524–1529
- Berger AB, Cabal GG, Fabre E, Duong T, Buc H, Nehrbass U, Olivo-Marin J-C, Gadal O & Zimmer C (2008) High-resolution statistical mapping reveals gene territories in live yeast. *Nat Methods* **5**: 1031–1037
- Berger AB, Decourty L, Badis G, Nehrbass U, Jacquier A & Gadal O (2007) Hmo1 is required for TOR-dependent regulation of ribosomal protein gene transcription. *Mol Cell Biol* **27**: 8015–8026
- Bischler N, Brino L, Carles C, Riva M, Tschochner H, Mallouh V & Schultz P (2002) Localization of the yeast RNA polymerase I-specific subunits. *EMBO J* **21**: 4136–4144
- Blatch GL & Lässle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**: 932–939
- Blattner C, Jennebach S, Herzog F, Mayer A, Cheung ACM, Witte G, Lorenzen K, Hopfner K-P, Heck AJR, Aebersold R & Cramer P (2011) Molecular basis of Rrn3-regulated RNA polymerase I initiation and cell growth. *Genes Dev* **25**: 2093–2105
- Blombach F, Smollett KL, Grohmann D & Werner F (2016) Molecular Mechanisms of Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors and Open Complex Formation. *J Mol Biol* **428**: 2592–2606
- Bouteille M & Hernandez-Verdun D (1979) Localization of a gene: the nucleolar organizer. *Biomedicine* **30**: 282–287
- Braglia P, Kawauchi J & Proudfoot NJ (2011) Co-transcriptional RNA cleavage provides a failsafe termination mechanism for yeast RNA polymerase I. *Nucleic Acids Res* **39**: 1439–1448
- Brun I, Sentenac A & Werner M (1997) Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* **16**: 5730–5741
- Buratowski S, Hahn S, Guarente L & Sharp PA (1989) Five intermediate complexes in

- transcription initiation by RNA polymerase II. *Cell* **56**: 549–561
- Byrne R, Levin JG, Bladen HA & Nirenberg MW (1964) The in vitro formation of a dna-ribosome complex. *Proc Natl Acad Sci U S A* **52**: 140–148
- Bystricky K, Heun P, Gehlen L, Langowski J & Gasser SM (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc Natl Acad Sci U S A* **101**: 16495–16500
- De Carlo S, Carles C, Riva M & Schultz P (2003) Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. *J Mol Biol* **329**: 891–902
- Cermakian N, Ikeda TM, Miramontes P, Lang BF, Gray MW & Cedergren R (1997) On the evolution of the single-subunit RNA polymerases. *J Mol Evol* **45**: 671–681
- Chamberlin M & Berg P (1962) Deoxyribo nucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci U S A* **48**: 81–94
- Chédin S, Riva M, Schultz P, Sentenac A & Carles C (1998) The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIIS-like subunit and is important for transcription termination. *Genes Dev* **12**: 3857–3871
- Cheung ACM & Cramer P (2012) A movie of RNA polymerase II transcription. *Cell* **149**: 1431–1437
- Cmarko D, Verschure PJ, Rothblum LI, Hernandez-Verdun D, Amalric F, van Driel R & Fakan S (2000) Ultrastructural analysis of nucleolar transcription in cells microinjected with 5-bromo-UTP. *Histochem Cell Biol* **113**: 181–187
- Colis L, Ernst G, Sanders S, Liu H, Sirajuddin P, Peltonen K, DePasquale M, Barrow JC & Laiho M (2014) Design, synthesis, and structure-activity relationships of pyridoquinazolinecarboxamides as RNA polymerase I inhibitors. *J Med Chem* **57**: 4950–4961
- Compe E & Egly J-M (2012) TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol* **13**: 343–354
- Cramer P, Bushnell DA & Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**: 1863–1876
- Davey CA, Sargent DF, Luger K, Maeder AW & Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* **319**: 1097–1113
- Dez C, Dlakić M & Tollervey D (2007) Roles of the HEAT repeat proteins Utp10 and Utp20 in 40S ribosome maturation. *RNA* **13**: 1516–1527

- Dieci G, Fiorino G, Castelnuovo M, Teichmann M & Pagano A (2007) The expanding RNA polymerase III transcriptome. *Trends Genet* **23**: 614–622
- Dilworth DJ, Suprpto A, Padovan JC, Chait BT, Wozniak RW, Rout MP & Aitchison JD (2001) Nup2p dynamically associates with the distal regions of the yeast nuclear pore complex. *J Cell Biol* **153**: 1465–1478
- Dragon F, Gallagher JEG, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlege RE, Shabanowitz J, Osheim Y, Beyer AL, Hunt DF & Baserga SJ (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**: 967–970
- Drygin D, Rice WG & Grummt I (2010) The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu Rev Pharmacol Toxicol* **50**: 131–156
- Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA & Noble WS (2010) A three-dimensional model of the yeast genome. *Nature* **465**: 363–367
- Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D & Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* **318**: 1777–1779
- Engel C, Gubbey T, Neyer S, Sainsbury S, Oberthuer C, Baejen C, Bernecky C & Cramer P (2017) Structural basis of RNA polymerase I transcription initiation. *Cell* **169**: 120–131.e22
- Engel C, Plitzko J & Cramer P (2016) RNA polymerase I-Rrn3 complex at 4.8 Å resolution. *Nat Commun* **7**: 12129
- Engel C, Sainsbury S, Cheung AC, Kostrewa D & Cramer P (2013) RNA polymerase I structure and transcription regulation. *Nature* **502**: 650–655
- Fazal FM, Meng CA, Murakami K, Kornberg RD & Block SM (2015) Real-time observation of the initiation of RNA polymerase II transcription. *Nature* **525**: 274–277
- Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, Richardson TM, Kriwacki RW, Pappu RV & Brangwynne CP (2016) Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**: 1686–1697
- Fernández-Tornero C, Moreno-Morcillo M, Rashid UJ, Taylor NMI, Ruiz FM, Gruene T, Legrand P, Steuerwald U & Müller CW (2013) Crystal structure of the 14-subunit RNA polymerase I. *Nature* **502**: 644–649
- Fishburn J, Tomko E, Galburt E & Hahn S (2015) Double-stranded DNA translocase activity of

- transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. *Proc Natl Acad Sci U S A* **112**: 3961–3966
- Gadal O, Labarre S, Boschiero C & Thuriaux P (2002) Hmo1, an HMG-box protein, belongs to the yeast ribosomal DNA transcription system. *EMBO J* **21**: 5498–5507
- Gadal O, Mariotte-Labarre S, Chedin S, Quemeneur E, Carles C, Sentenac A & Thuriaux P (1997) A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* **17**: 1787–1795
- Gallagher JEG, Dunbar DA, Granneman S, Mitchell BM, Osheim Y, Beyer AL & Baserga SJ (2004) RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. *Genes Dev* **18**: 2506–2517
- Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A & Nehrbass U (2004) Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**: 63–73
- Geiger SR, Lorenzen K, Schrieck A, Hanecker P, Kostrewa D, Heck AJR & Cramer P (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell* **39**: 583–594
- Gnatt AL, Cramer P, Fu J, Bushnell DA & Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**: 1876–1882
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG (1996) Life with 6000 genes. *Science* **274**: 546, 563–7
- Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, Michaelis J & Werner F (2011) The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* **43**: 263–274
- Grummt I, Maier U, Ohrlein A, Hassouna N & Bachellerie JP (1985) Transcription of mouse rDNA terminates downstream of the 3' end of 28S RNA and involves interaction of factors with repeated sequences in the 3' spacer. *Cell* **43**: 801–810
- Haeusler RA & Engelke DR (2006) Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Res* **34**: 4826–4836
- Hage A El, Koper M, Kufel J & Tollervey D (2008) Efficient termination of transcription by RNA polymerase I requires the 5' exonuclease Rat1 in yeast. *Genes Dev* **22**: 1069–1081
- Han Y, Yan C, Nguyen THD, Jackobel AJ, Ivanov I, Knutson BA & He Y (2017) Structural

- mechanism of ATP-independent transcription initiation by RNA polymerase I. *elife* **6**:
- Helmann JD & Chamberlin MJ (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839–872
- Henras AK, Plisson-Chastang C, O’Donohue M-F, Chakraborty A & Gleizes P-E (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* **6**: 225–242
- Herdman C, Mars J-C, Stefanovsky VY, Tremblay MG, Sabourin-Felix M, Lindsay H, Robinson MD & Moss T (2017) A unique enhancer boundary complex on the mouse ribosomal RNA genes persists after loss of Rrn3 or UBF and the inactivation of RNA polymerase I transcription. *PLoS Genet* **13**: e1006899
- Herskowitz I (1988) Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol Rev* **52**: 536–553
- Hoelz A, Glavy JS & Beck M (2016) Toward the atomic structure of the nuclear pore complex: when top down meets bottom up. *Nat Struct Mol Biol* **23**: 624–630
- Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJH, Sachse C & Müller CW (2015) Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**: 231–236
- Hoffmann NA, Sadian Y, Tafur L, Kosinski J & Müller CW (2016) Specialization versus conservation: How Pol I and Pol III use the conserved architecture of the pre-initiation complex for specialized transcription. *Transcription* **7**: 127–132
- Huet J, Buhler JM, Sentenac A & Fromageot P (1975) Dissociation of two polypeptide chains from yeast RNA polymerase A. *Proc Natl Acad Sci U S A* **72**: 3034–3038
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hermsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC & Banfield JF (2016) A new view of the tree of life. *Nature microbiology* **1**: 16048
- Jasiak AJ, Armache K-J, Martens B, Jansen R-P & Cramer P (2006) Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* **23**: 71–81
- Jennebach S, Herzog F, Aebersold R & Cramer P (2012) Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Res* **40**: 5591–5601
- Jin Q, Trelles-Sticken E, Scherthan H & Loidl J (1998) Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *J*



*Cell Biol* **141**: 21–29

Jochem L, Ramsay EP & Vannini A (2017) RNA polymerase I, bending the rules? *EMBO J* **36**: 2664–2666

Jones MH, Learned RM & Tjian R (1988) Analysis of clustered point mutations in the human ribosomal RNA gene promoter by transient expression in vivo. *Proc Natl Acad Sci U S A* **85**: 669–673

Ju QD, Morrow BE & Warner JR (1990) REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. *Mol Cell Biol* **10**: 5226–5234

Juven-Gershon T & Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol* **339**: 225–229

Kawauchi J, Mischo H, Braglia P, Rondon A & Proudfoot NJ (2008) Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev* **22**: 1082–1092

Keener J, Josaitis CA, Dodd JA & Nomura M (1998) Reconstitution of yeast RNA polymerase I transcription in vitro from purified components. TATA-binding protein is not required for basal transcription. *J Biol Chem* **273**: 33795–33802

Kettenberger H, Armache K-J & Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**: 347–357

Kettenberger H, Armache K-J & Cramer P (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **16**: 955–965

Keys DA, Lee BS, Dodd JA, Nguyen TT, Vu L, Fantino E, Burson LM, Nogi Y & Nomura M (1996) Multiprotein transcription factor UAF interacts with the upstream element of the yeast RNA polymerase I promoter and forms a stable preinitiation complex. *Genes Dev* **10**: 887–903

Keys DA, Vu L, Steffan JS, Dodd JA, Yamamoto RT, Nogi Y & Nomura M (1994) RRN6 and RRN7 encode subunits of a multiprotein complex essential for the initiation of rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*. *Genes Dev* **8**: 2349–2362

Knutson BA & Hahn S (2011) Yeast Rrn7 and human TAF1B are TFIIB-related RNA polymerase I general transcription factors. *Science* **333**: 1637–1640

Kohler R, Mooney RA, Mills DJ, Landick R & Cramer P (2017) Architecture of a transcribing-translating expressome. *Science* **356**: 194–197

- Komissarova N & Kashlev M (1997) RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *J Biol Chem* **272**: 15329–15338
- Kornberg RD (2005) Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**: 235–239
- Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A & Darst SA (2000) A structural model of transcription elongation. *Science* **289**: 619–625
- Kuhn A, Bartsch I & Grummt I (1990) Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. *Nature* **344**: 559–562
- Kuhn C-D, Geiger SR, Baumli S, Gartmann M, Gerber J, Jennebach S, Mielke T, Tschochner H, Beckmann R & Cramer P (2007) Functional architecture of RNA polymerase I. *Cell* **131**: 1260–1272
- Laferté A, Favry E, Sentenac A, Riva M, Carles C & Chédin S (2006) The transcriptional activity of RNA polymerase I is a key determinant for the level of all ribosome components. *Genes Dev* **20**: 2030–2040
- Lafontaine DL & Tollervey D (1998) Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem Sci* **23**: 383–388
- Lalo D, Steffan JS, Dodd JA & Nomura M (1996) RRN11 encodes the third subunit of the complex containing Rrn6p and Rrn7p that is essential for the initiation of rDNA transcription by yeast RNA polymerase I. *J Biol Chem* **271**: 21062–21067
- Landrieux E, Alic N, Ducrot C, Acker J, Riva M & Carles C (2006) A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* **25**: 118–128
- Lang WH, Morrow BE, Ju Q, Warner JR & Reeder RH (1994) A model for transcription termination by RNA polymerase I. *Cell* **79**: 527–534
- Learned RM, Cordes S & Tjian R (1985) Purification and characterization of a transcription factor that confers promoter specificity to human RNA polymerase I. *Mol Cell Biol* **5**: 1358–1369
- Learned RM, Learned TK, Haltiner MM & Tjian RT (1986) Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* **45**: 847–857
- Léger-Silvestre I, Trumtel S, Noaillac-Depeyre J & Gas N (1999) Functional compartmentalization of the nucleus in the budding yeast *Saccharomyces cerevisiae*. *Chromosoma* **108**: 103–113.

- Liljelund P, Mariotte S, Buhler JM & Sentenac A (1992) Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **89**: 9302–9305
- Lisica A, Engel C, Jahnel M, Roldán É, Galburt EA, Cramer P & Grill SW (2016) Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc Natl Acad Sci U S A* **113**: 2946–2951
- Louder RK, He Y, López-Blanco JR, Fang J, Chacón P & Nogales E (2016) Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**: 604–609
- Malagon F, Tong AH, Shafer BK & Strathern JN (2004) Genetic interactions of DST1 in *Saccharomyces cerevisiae* suggest a role of TFIS in the initiation-elongation transition. *Genetics* **166**: 1215–1227
- Merkl P, Perez-Fernandez J, Pilsil M, Reiter A, Williams L, Gerber J, Böhm M, Deutzmann R, Griesenbeck J, Milkereit P & Tschochner H (2014) Binding of the termination factor Nsi1 to its cognate DNA site is sufficient to terminate RNA polymerase I transcription in vitro and to induce termination in vivo. *Mol Cell Biol* **34**: 3817–3827
- Milkereit P, Schultz P & Tschochner H (1997) Resolution of RNA polymerase I into dimers and monomers and their function in transcription. *Biol Chem* **378**: 1433–1443
- Milkereit P & Tschochner H (1998) A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J* **17**: 3692–3703
- Millhouse S & Manley JL (2005) The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol Cell Biol* **25**: 533–544
- Miwa K, Kojima R, Obita T, Ohkuma Y, Tamura Y & Mizuguchi M (2016) Crystal structure of human general transcription factor TFIIIE at atomic resolution. *J Mol Biol* **428**: 4258–4266
- Moorefield B, Greene EA & Reeder RH (2000) RNA polymerase I transcription factor Rrn3 is functionally conserved between yeast and human. *Proc Natl Acad Sci U S A* **97**: 4724–4729
- Mougey EB, O'Reilly M, Osheim Y, Miller OL, Beyer A & Sollner-Webb B (1993) The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes Dev* **7**: 1609–1619
- Van Mullem V, Landrieux E, Vandehaute J & Thuriaux P (2002) Rpa12p, a conserved RNA polymerase I subunit with two functional domains. *Mol Microbiol* **43**: 1105–1113

- Musters W, Knol J, Maas P, Dekker AF, van Heerikhuizen H & Planta RJ (1989) Linker scanning of the yeast RNA polymerase I promoter. *Nucleic Acids Res* **17**: 9661–9678
- Naidu S, Friedrich JK, Russell J & Zomerdijk JCBM (2011) TAF1B is a TFIIIB-like component of the basal transcription machinery for RNA polymerase I. *Science* **333**: 1640–1642
- Neyer S, Kunz M, Geiss C, Hantsche M, Hodirnau V-V, Seybert A, Engel C, Scheffer MP, Cramer P & Frangakis AS (2016) Structure of RNA polymerase I transcribing ribosomal DNA genes. *Nature* **540**: 607–610
- Nogi Y, Vu L & Nomura M (1991) An approach for isolation of mutants defective in 35S ribosomal RNA synthesis in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **88**: 7026–7030
- Nogi Y, Yano R, Dodd J, Carles C & Nomura M (1993) Gene RRN4 in *Saccharomyces cerevisiae* encodes the A12.2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol Cell Biol* **13**: 114–122
- Nudler E, Mustaev A, Lukhtanov E & Goldfarb A (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**: 33–41
- Oakes M, Siddiqi I, Vu L, Aris J & Nomura M (1999) Transcription factor UAF, expansion and contraction of ribosomal DNA (rDNA) repeats, and RNA polymerase switch in transcription of yeast rDNA. *Mol Cell Biol* **19**: 8559–8569
- Olins AL & Olins DE (1974) Spheroid chromatin units (v bodies). *Science* **183**: 330–332
- Osheim YN, French SL, Sikes ML & Beyer AL (2009) Electron microscope visualization of RNA transcription and processing in *Saccharomyces cerevisiae* by Miller chromatin spreading. *Methods Mol Biol* **464**: 55–69
- Oudet P, Gross-Bellard M & Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**: 281–300
- Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH & O’Shea CC (2017) ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**: 113–118
- Paget MS (2015) Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* **5**: 1245–1265
- Pasero P & Marilley M (1993) Size variation of rDNA clusters in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Mol Gen Genet* **236**: 448–52.
- Peltonen K, Colis L, Liu H, Trivedi R, Moubarek MS, Moore HM, Bai B, Rudek MA, Bieberich CJ & Laiho M (2014) A targeting modality for destruction of RNA polymerase I that

- possesses anticancer activity. *Cancer Cell* **25**: 77–90
- Peyroche G, Milkereit P, Bischler N, Tschochner H, Schultz P, Sentenac A, Carles C & Riva M (2000) The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* **19**: 5473–5482
- Pilsl M, Crucifix C, Papai G, Krupp F, Steinbauer R, Griesenbeck J, Milkereit P, Tschochner H & Schultz P (2016) Structure of the initiation-competent RNA polymerase I and its implication for transcription. *Nat Commun* **7**: 12126
- Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J & Cramer P (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**: 353–358
- Ponting CP (2002) Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res* **30**: 3643–3652
- Prescott EM, Osheim YN, Jones HS, Alen CM, Roan JG, Reeder RH, Beyer AL & Proudfoot NJ (2004) Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc Natl Acad Sci U S A* **101**: 6068–6073
- Reeder RH, Guevara P & Roan JG (1999) *Saccharomyces cerevisiae* RNA polymerase I terminates transcription at the Reb1 terminator in vivo. *Mol Cell Biol* **19**: 7369–7376
- Reeder RH & Lang W (1994) The mechanism of transcription termination by RNA polymerase I. *Mol Microbiol* **12**: 11–15
- Reiter A, Hamperl S, Seitz H, Merkl P, Perez-Fernandez J, Williams L, Gerber J, Németh A, Léger I, Gadal O, Milkereit P, Griesenbeck J & Tschochner H (2012) The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *EMBO J* **31**: 3480–3493
- Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei P-J, Burlingame AL & Kornberg RD (2016) Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* **166**: 1411–1422.e16
- Sadian Y, Tafur L, Kosinski J, Jakobi AJ, Wetzell R, Buczak K, Hagen WJ, Beck M, Sachse C & Müller CW (2017) Structural insights into transcription initiation by yeast RNA polymerase I. *EMBO J*
- Saeki H & Svejstrup JQ (2009) Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. *Mol Cell* **35**: 191–205
- Sayre MH, Tschochner H & Kornberg RD (1992) Purification and properties of *Saccharomyces cerevisiae* RNA polymerase II general initiation factor a. *J Biol Chem* **267**: 23383–23387

- Schultz LD & Hall BD (1976) Transcription in yeast: alpha-amanitin sensitivity and other properties which distinguish between RNA polymerases I and III. *Proc Natl Acad Sci U S A* **73**: 1029–1033
- Schultz P, Célia H, Riva M, Sentenac A & Oudet P (1993) Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals. *EMBO J* **12**: 2601–2607
- Schwarzacher HG & Wachtler F (1991) The functional significance of nucleolar structures. *Ann Genet* **34**: 151–60.
- Schwarzacher HG & Wachtler F (1993) The nucleolus. *Anat Embryol* **188**: 515–536
- Schweizer E, MacKechnie C & Halvorson HO (1969) The redundancy of ribosomal and transfer RNA genes in *Saccharomyces cerevisiae*. *J Mol Biol* **40**: 261–277
- Sekine S, Murayama Y, Svetlov V, Nudler E & Yokoyama S (2015) The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol Cell* **57**: 408–421
- Siddiqi IN, Dodd JA, Vu L, Eliason K, Oakes ML, Keener J, Moore R, Young MK & Nomura M (2001) Transcription of chromosomal rRNA genes by both RNA polymerase I and II in yeast uaf30 mutants lacking the 30 kDa subunit of transcription factor UAF. *EMBO J* **20**: 4512–4521
- Smid A, Riva M, Bouet F, Sentenac A & Carles C (1995) The association of three subunits with yeast RNA polymerase is stabilized by A14. *J Biol Chem* **270**: 13534–13540
- Steffan JS, Keys DA, Dodd JA & Nomura M (1996) The role of TBP in rDNA transcription by RNA polymerase I in *Saccharomyces cerevisiae*: TBP is required for upstream activation factor-dependent recruitment of core factor. *Genes Dev* **10**: 2551–2563
- Steffan JS, Keys DA, Vu L & Nomura M (1998) Interaction of TATA-binding protein with upstream activation factor is required for activated transcription of ribosomal DNA by RNA polymerase I in *Saccharomyces cerevisiae* in vivo. *Mol Cell Biol* **18**: 3752–3761
- Steitz T (1998) A mechanism for all polymerases. *Nature* **391**: 231–232
- Taddei A & Gasser SM (2012) Structure and function in the budding yeast nucleus. *Genetics* **192**: 107–129
- Tafur L, Sadian Y, Hoffmann NA, Jakobi AJ, Wetzel R, Hagen WJH, Sachse C & Müller CW (2016) Molecular structures of transcribing RNA polymerase I. *Mol Cell* **64**: 1135–1143
- Teichmann M, Dumay-Odelot H & Fribourg S (2012) Structural and functional aspects of

- winged-helix domains at the core of transcription initiation complexes. *Transcription* **3**: 2–7
- Teixeira MT, Dujon B & Fabre E (2002) Genome-wide nuclear morphology screen identifies novel genes involved in nuclear architecture and gene-silencing in *Saccharomyces cerevisiae*. *J Mol Biol* **321**: 551–561
- Thiry M & Lafontaine DLJ (2005) Birth of a nucleolus: the evolution of nucleolar compartments. *Trends Cell Biol* **15**: 194–199
- Thompson CM, Koleske AJ, Chao DM & Young RA (1993) A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**: 1361–1375
- Tisseur M, Kwapisz M & Morillon A (2011) Pervasive transcription - Lessons from yeast. *Biochimie* **93**: 1889–1896
- Torreira E, Louro JA, Pazos I, González-Polo N, Gil-Carton D, Duran AG, Tosi S, Gallego O, Calvo O & Fernández-Tornero C (2017) The dynamic assembly of distinct RNA polymerase I complexes modulates rDNA transcription. *elife* **6**:
- Trumtel S, Leger-Silvestre I, Gleizes PE, Teulieres F & Gas N (2000) Assembly and functional organization of the nucleolus: ultrastructural analysis of *Saccharomyces cerevisiae* mutants. *Mol Biol Cell* **11**: 2175–89.
- Tsang CK, Bertram PG, Ai W, Drenan R & Zheng XFS (2003) Chromatin-mediated regulation of nucleolar structure and RNA Pol I localization by TOR. *EMBO J* **22**: 6045–6056
- Ushijima R, Matsuyama T, Nagata I & Yamamoto K (2008) Nucleolar targeting of proteins by the tandem array of basic amino acid stretches identified in the RNA polymerase I-associated factor PAF49. *Biochem Biophys Res Commun* **369**: 1017–1021
- Vannini A & Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* **45**: 439–446
- Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH & Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**: 157–162
- Viktorovskaya OV, Appling FD & Schneider DA (2011) Yeast transcription elongation factor Spt5 associates with RNA polymerase I and RNA polymerase II directly. *J Biol Chem* **286**: 18825–18833
- Walmacq C, Kireeva ML, Irvin J, Nedialkov Y, Lubkowska L, Malagon F, Strathern JN & Kashlev M (2009) Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J Biol Chem* **284**: 19601–19612

- Wang W, Li G-W, Chen C, Xie XS & Zhuang X (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333**: 1445–1449
- Weinzierl ROJ (2013) The RNA polymerase factory and archaeal transcription. *Chem Rev* **113**: 8350–8376
- Wendte JM, Haag JR, Singh J, McKinlay A, Pontes OM & Pikaard CS (2017) Functional Dissection of the Pol V Largest Subunit CTD in RNA-Directed DNA Methylation. *Cell Rep* **19**: 2796–2808
- Wendte JM & Pikaard CS (2016) Targeting Argonaute to chromatin. *Genes Dev* **30**: 2649–2650
- Werner F (2013) Molecular mechanisms of transcription elongation in archaea. *Chem Rev* **113**: 8331–8349
- Werner F & Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* **9**: 85–98
- Wery M, Ruidant S, Schillewaert S, Leporé N & Lafontaine DJ (2009) The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance. *RNA* **15**: 406–419
- White CL, Suto RK & Luger K (2001) Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J* **20**: 5207–5218
- Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088–5090
- Wu C-C, Lin Y-C & Chen H-T (2011) The TFIIF-like Rpc37/53 dimer lies at the center of a protein network to connect TFIIIC, Bdp1, and the RNA polymerase III active center. *Mol Cell Biol* **31**: 2715–2728
- Yamamoto K, Yamamoto M, Hanada K, Nogi Y, Matsuyama T & Muramatsu M (2004) Multiple protein-protein interactions by RNA polymerase I-associated factor PAF49 and role of PAF49 in rRNA transcription. *Mol Cell Biol* **24**: 6338–6349
- Yamamoto RT, Nogi Y, Dodd JA & Nomura M (1996) RRN3 gene of *Saccharomyces cerevisiae* encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. *EMBO J* **15**: 3964–3973
- Yan Q, Moreland RJ, Conaway JW & Conaway RC (1999) Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem* **274**: 35668–35675
- Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF,



- Schramm A, Baker BJ, Spang A & Ettema TJG (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**: 353–358
- Zenke M, Grundström T, Matthes H, Wintzerith M, Schatz C, Wildeman A & Chambon P (1986) Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J* **5**: 387–397
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K & Darst SA (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**: 811–824
- Zomerdijk JC, Beckmann H, Comai L & Tjian R (1994) Assembly of transcriptionally active RNA polymerase I initiation factor SL1 from recombinant subunits. *Science* **266**: 2015–2018