Exhaustive Identification of the retroelement Ty1 Integrase partners in yeast Saccharomyces cerevisiae : characterization of the role of Casein kinase II in Ty1 retrotransposition in vivo

Indranil Adhya

université PARIS-SACLAY

NNT : 2018SACLS589

Thèse de doctorat

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Exhaustive identification of the retroelement Ty1 Integrase partners in yeast *Saccharomyces cerevisiae*

Characterization of the role of Casein kinase II in Ty1 retrotransposition in vivo

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-sud

École doctorale n°577 : Structure et dynamique des Systèmes vivants (SDSV)

Spécialité de doctorat : Sciences de la vie et de la santé

Thèse présentée et soutenue à Saclay, 14 Décembre 2018

Indranil ADHYA

Composition du Jury :

Dr. Pierre CAPY Professeur, CNRS	Président
Dr. Martin TEICHMANN Professeur, Université de Bordeaux	Rapporteur
Dr. Marc LAVIGNE Directeur de Recherche, Institut Cochin, Paris	Rapporteur
Dr. Julie SOUTOURINA Directrice de Recherche, CEA, Saclay	Examinateur
Dr. Anna BABOUR Chargée de Recherche (CR1), INSERM	Examinateur
Dr. Joël ACKER Directeur de Recherche, CEA, Saclay	Directeur de thèse
Dr. Pascale LESAGE Directrice de Recherche, CNRS	Invité

ABSTRACT

LTR-retrotransposons are widespread transposable elements in eukaryotes. Like retroviruses, they replicate by reverse transcription of their RNA into cDNA, which is integrated into the host genome by their own integrase (IN). High-throughput sequencing studies clearly established that integration does not occur randomly throughout the host-cell genome. Deep insights on retroviral biology have been gained by their study in yeast using the Ty1 LTR-retrotransposon as a working model. The Ty1 retrotransposon of the yeast Saccharomyces cerevisiae integrates upstream of class III genes, the genes transcribed by RNA polymerase III (Pol III). Recent data revealed the importance of AC40, a Pol III subunit in this targeting. An interaction between the Ty1 IN and AC40 is necessary for integration site choice at the Pol III genes. Nevertheless, the molecular mechanism remains largely unknown.

To obtain a global view of the entire phenomenon that occurs on the integration site we would like to exhaustively determine the proteins that interact with Ty1 IN and analyze their role in both Ty1 integration and RNA Pol III transcription. To achieve this goal, we have developed proteomic approaches to identify new Ty1 integrase cellular partners.

Firstly, we have performed a tandem chromatin affinity purification using the HBH (histidine and biotin) tag and the integrase as bait. Such procedure was recently set up in the lab in order to identify RNA pol III machinery partners. After *in vivo* crosslinking of yeast cells, a chromatin extract solubilized by sonication was sequentially purified by Nickel and streptavidin affinity chromatography. Integrase and the co-purified proteins were identified by mass spectrometry. We have identified several novel Ty1 IN partners that seem interesting and their molecular role in Ty1 retrotransposition will be studied. However, in the tenure of my PhD, I have particularly worked to decipher the molecular role of the casein kinase II protein in Ty1 retrotransposition *in vivo*. CK2 is a serine/threonine kinase involved in a variety of cellular processes such as cell differentiation, protein stability and proliferation. We have identified the four subunits of the CK2 holoenzyme complex in our TChAP experiments and have validated that it is a *bonafide* Ty1 IN partner *in vivo*. In good agreement with previous studies we have shown that CK2 holoenzyme negatively regulates Ty1 retrotransposition *in vivo.* Interestingly, we demonstrated that CK2 phosphorylates Ty1 IN *in vitro* and the identified phosphorylation sites are correlated to be phosphorylated *in vivo.* Interestingly, we have observed that the expression of both endogenous Ty1 IN and Ty1 IN expressed ectopically on a plasmid shows a slight high

abundance in the absence of CK2 *in vivo*. This phenomenon encouraged us to study the regulation of Ty1 IN expression by CK2. We have preliminary data, showing that CK2 phosphorylation directly affects Ty1 IN expression *in vivo.* Ty1 IN phosphomutant was developed that show no phosphorylation and such mutant has shown slightly more abundance compared to WT Ty1 IN *in vivo.* We are currently analyzing more precisely the amino acids phosphorylated by CK2 and investigating their importance in the expression of Ty1 IN. On the other hand, we have also demonstrated that Ty1 IN is a substrate of the proteasome. As indicated in the literature, that Ty1 proteins including Ty1 IN is ubiquitinylated and the fact that there exists a cross talk between phosphorylation and ubiquitinylation, we have hypothesized that may be CK2 phosphorylation leads to ubiquitinylation of Ty1 IN and further degradation by the 26S proteasome under normal growth conditions. Currently, we are performing experiments to test our hypothesis.

ACKNOWLEDGEMENTS

Firstly, I would take this opportunity to thank and express my utmost gratitude to my entire team for guiding me throughout the project right from the first day and teaching me the realms of scientific research. **Dr. Joel ACKER** has been instrumental in shaping my knowledge and scientific skills in all these years. A big thank you to accept me in your laboratory and teaching me everything from scratch. It has been enlightening to work closely with you and discuss scientific ideas on a daily basis. Apart from the research, you have been an inspiration to me both as a scientist and as a human being. I have always received fair criticism from you that has helped me excel and improved me both professionally and personally. I have always felt and realized that your criticisms were only meant for my good and betterment.

Dr. Christine CONESA has always been there as a constant support. With her expertise, she has always guided me and helped me analyze my results. You have always been there helping me with any kind of experiments I have asked for. On a personal basis, you have been a very gentle and kind person to me. Thank you very much for telling me when I was wrong and when I didn't work enough and appreciating me when I had good results. I owe you a great deal of respect and gratitude for performing all the experiments when I am not able to do them and letting me put your results in my thesis.

Even after shouldering immense amount of responsibilities as the head of the service **Dr. Olivier LEFEBVRE,** always made sure that he discussed the progress of my work and appreciated me when things were moving forward. Thank you very much for giving me your time and patience.

Carine CHAGNEAU has been instrumental in performing most of the experiments in the project. Though we haven't been able to interact much, I remain grateful and owe you respect for your picturesque results and allowing me to put your results in my thesis. Your contribution to the project and publication will always be remembered with a sense of gratitude. I would also like to thank, **Celia GOUZERH,** who joined our lab this year and made very significant contribution to the project and for being a kind person to me.

I owe my thanks and gratitude to the entire team of **Dr. Pascale LESAGE** for all their help and support in making the project what it is today. I have had very enlightening and stimulating

discussions with her during the meetings and conferences. I wish the very best to **Anastasia BARKOVA** and thank you to her to corroborate so strongly with me in the project. I would like to thank **Dr. Amna Asif-LAIDEN** for her advice in improving my thesis and her expertise in bioinformatics that helped our project reach greater heights of success.

I would like to thank the entire service for providing me with a positive and friendly working environment. I am grateful to all the respected jury members, **Dr. Pierre CAPY**, **Dr. Martin TEICHMANN**, **Dr. Anna BABOUR, Dr. Julie SOUTOURINA** and **Dr. Marc LAVIGNE** for taking the time and agreeing to review my work.

Last but not the least, I owe everything to my wife, Kuheli BOSE and my family for being a constant support and helping me through all the difficult times. Thank you to all once again.

Rapport du travail de thèse

1. Éléments transposable

Les éléments transposables sont des éléments mobiles largement répandus dans les génomes eucaryotes et procaryotes. Ils ont la capacité de se “transposer“, c’est à dire de se déplacer d'un site chromosomique ~~e~~ à un autre, d'où leur nom de transposons. Les TE jouent un rôle important dans l'évolution et le maintien de la diversité génétique.

La première classification des éléments transposables, proposée par Finnegan en 1989, était basée sur le mécanisme de transposition ~~.~~Les éléments transposables peuvent être divisés en deux grandes classes : (I) les rétrotransposons et (II) les transposons d'ADN. Les rétrotransposons ont un mécanisme de transposition réplicative, avec l’utilisation d’un ARN intermédiaire obtenu par transcription inverse de l'ARNm en ADNc. Les transposons d'ADN ont un mécanisme de transposition conservatrice, la séquence d’ADN est transférée d’un endroit du génome à un autre. Ces derniers codent une enzyme connue sous le nom de transposase qui assure cette transposition. Les éléments de classe I sont en outre divisés en deux sous-classes basées sur les caractéristiques structurales a) les rétrotransposons à répétition terminale longue (LTR) qui ressemblent à des rétrovirus et b) les rétrotransposons non-LTR. Les premiers sont flanqués de longues répétitions terminales aux deux extrémités du génome et codent les protéines Gag et Pol comme les rétrovirus. Les rétrotransposons LTR constituent environ 8% du génome humain. La plupart des informations concernant le mécanisme de rétrotransposition des rétrotransposons LTR provient de l'étude approfondie des éléments Ty dans la levure (Voytas et al 2002, Sandmeyer et al 2015), mais on suppose que le mécanisme est similaire parmi les rétrotranspositions LTR dans divers hôtes. Ce sont des éléments autonomes car ils codent pour toutes les enzymes catalytiques nécessaires à leur rétrotransposition telles que la protéase (PR), la transcriptase inverse (RT) et Integrase (IN). Les éléments non-LTR ne contiennent pas de longues répétitions à leurs extrémités génétiques. Ils encodent également les protéines Gag et Pol avec une queue poly-A à l'extrémité 3'. On les trouve en abondance chez l'homme (33 % du génome) et chez d'autres organismes eucaryotes. Ils sont composés d'éléments nucléaires intercalés (LINE) et de petits éléments nucléaires intercalés (SINE).

1.1 L'élément Ty de la levure *Saccharomyces cerevisiae*

Les seuls éléments mobiles présents chez la levure Saccharomyces cerevisiae sont des rétrotransposons à LTR. Ils sont répartis en cinq familles : Ty1, Ty2, Ty3, Ty4 et Ty5 (Clare et Farabaugh 1985 ; Warmington et al. 1985 ; Hansen et al. 1988 ; Stucka et al. 1992 ; Voytas et Boeke 1992). Le séquençage du génome de S. cerevisiae a permis de déterminer leur localisation dans le génome. Ils présentent une certaine homologie entre eux. Par ailleurs, les rétroélements sont également apparentés aux rétrovirus par leur cycle de réplication et leur structure génomique, ce qui en fait de bons modèles pour les étudier. Les éléments Ty1, Ty2, Ty4 et Ty5 appartiennent à la superfamille Ty1-copia tandis que l'élément Ty3 appartient à la famille Ty3-gypsy. La principale différence se situe au niveau de l'organisation génétique du gène POL. Au cours de mon travail de thèse, j'ai plus particulièrement travaillé sur le rétroélément Ty1 de la levure Saccharomyces cerevisiae. **C'**est le rétroélement le plus abondant. On dénombre plus de 32 copies actives dans son génome et son ARN représente environ 1% de l'ARNm total présent dans la cellule (Curcio et al 2015). Cependant, la rétrotransposition est un événement extrêmement rare elle est estimée à environ de 10^{-7}-10^{-8} transpositions /élément /génération. Le génome de Ty1 est constitué de 5918 pb et est flanqué de séquences LTR de chaque côtés (334 pb). Il est composé de 2 gènes qui se chevauchent partiellement : TyA, qui codent pour la protéine de la nucléocapside (GAG), et TyB qui codent pour les fonctions catalytiques nécessaires à la rétrotransposition, à savoir la protéase (PR), l'intégrase (IN) et la transcriptase/RNase H inverse (RT/RH). Ces 2 gènes sont co-transcrits et un décalage de lecture situé à leur jonction va permettre l'expression de la protéine GAG et d'une protéine de fusion Gag-Pol.

Le cycle de réplication est similaire à celui des rétrovirus, , mais contrairement aux rétrovirus, il est exclusivement intracellulaire car il ne possède pas de gêne enveloppe. L'élément Ty1 est transcrit dans le noyau et son ARNm est transporté dans le cytoplasme et va conduire à la synthèse de la protéine Gag et d'une protéine de fusion Gag-Pol. Toutes ces protéines vont être incorporées avec de l'ARN de Ty1 dans des VLP (virus-like particules) où les protéines seront maturées par sa propre protéase pour produire les différents polypeptides nécessaires à la rétrotransposition (PR, IN, et RT) Le cDNA obtenu par reverse transcription de l'ARN de Ty1 présent dans les VLP va ensuite se lier à l'intégrase (IN) pour former un complexe de pré-intégration (PIC), qui sera importé dans le noyau où le cDNA sera intégré dans le génome (Curcio et al 2015). Des cribles à haut débit ont montré que l'intégration des éléments Ty n'est

pas aléatoire dans le génome les insertions sont localisées principalement en amont des gènes transcrits par l'ARN Pol III (Pol III), à l'exception de Ty5 qui a une préférence plus spécifique pour l'hétérochromatine située dans les régions subtélomériques. Par la suite, des études ont montré que la présence d'un promoteur fonctionnel de l'ARN Pol III était nécessaire pour l'intégration de Ty1 au niveau des gènes Pol III (Devine et al 1996). Cependant les mécanismes moléculaires impliqués sont encore largement inconnus. Récemment, Bridier Nahmias et ses collaborateurs ont démontré qu'une association entre la sous-unité AC40 de la Pol III et l'intégrase de Ty1 est nécessaire à la spécificité d'intégration en amont des gènes Pol III. L'utilisation d'un mutant de perte d'interaction avec l'intégrase amène une redistribution des insertions dans les régions sub-télomériques sans affecter la fréquence de rétromobilité de Ty1 in vivo. Des mécanismes très similaires ont été décrits pour d'autres rétroéléments et rétrovirus et un modèle dit de « tethering » ou d'ancrage a été proposé dans lequel une protéine cellulaire interagit avec une intégrase et dirige le PIC vers des régions spécifiques du génome. C'est le cas par exemple de l'intégrase de Ty5 qui interagit avec la protéine Sir4 pour s'intégrer dans des régions sub-télomèriques ou encore de l'intégrase du VIH qui s'associe avec la protéine LEDGF / p75 pour s'intégrer aux promoteurs Pol II. Ces résultats suggèrent donc que ce modèle de "tethering » moléculaire pourrait également s'appliquer dans le cas de Ty1, avec l'ARN Pol III comme facteur d'ancrage.

L'ARN pol III est une enzyme complexe composée de 17 sous-unités responsables principalement de la transcription des ARNt, de l'ARNr 5S et d'autres petits ARN non codants. Les études de séquençages à haut débit des intégrations de Ty1 ont révélé l'existence d'insertions situées également en dehors des gènes transcrits par l'ARN Pol III, en particulier dans des régions sub-télomériques dans lesquelles l'ARN Pol III n'est pas présente. Ces données suggèrent donc soit l'existence d'autres facteurs d'ancrage responsables de cette spécificité soit l'existence d'autres mécanismes moléculaires.

Par ailleurs, de nombreux criblages génétiques ont été réalisés et ont permis d'identifier des facteurs cellulaires qui affectent la rétrotransposition in vivo de Ty1 (Scholes et al 2001, Griffith et al 2003, Nyswaner et al 2008 et Risler et al 2012). Cependant, les rôles de ces protéines n'ont pas été clairement définis et ne permettent pas de proposer d'autres facteurs d'ancrage potentiels de l'intégrase. Par conséquent, pour identifier des nouveaux partenaires qui interagissent avec l'intégrase de Ty1, des approches protéomiques nous ont semblé plus adaptées. A ce jour, peu d'études protéomiques ont été réalisées, excepté à ma connaissance

Celle de (Cheung et al 2016). Cependant, leurs travaux n'ont pas réussi à fournir une liste importante de partenaires potentiels.

Le but de de mon projet de thèse était de développer une approche protéomique robuste pour identifier de manière exhaustive de nouveaux partenaires de l'intégrase de Ty1 *in vivo* Ty1 et de caractériser leur rôle fonctionnel dans la rétrotransposition.

2. Identification exhaustive des partenaires Ty1 IN

Pour réaliser cet objectif, nous avons développé et adapté une méthode de purification en tandem de fractions chromatiniennes (TChAP), qui a été mise au point dans notre laboratoire pour identifier de nouveaux partenaires de l'ARN Pol III in vivo (Nguyen et al 2015). Le principe consiste à purifier l'intégrase de Ty1 (IN) à partir de cellules de levure crosslinkées *in vivo* et à identifier les protéines co-purifiées par spectrométrie de masse. Notre objectif était de purifier les partenaires Ty1 IN dans des conditions de croissance différentes qui stimulent ou entravent la rétrotransposition, en espérant pouvoir identifier des partenaires spécifiques dans ces conditions. La rétrotransposition est optimale à basse température (15°C) et est très fortement inhibée à 30°C. Nous avons donc décidé de réaliser des TChAP sur des cellules cultivées à 20°C (la température classique de test de rétrotransposition) et à 30°C. Cependant, nous avons dû optimiser la procédure du TChAP pour nous adapter à notre projet pour deux raisons. Premièrement, comme le Ty1 IN endogène est à peine détecté dans des conditions de croissance normale, l'intégrase a été exprimée à partir d'un vecteur d'expression plasmidique. Pour éviter tout effet délétère qui pourrait être causé par la surexpression et l'expression constitutive de Ty1 IN, nous avons choisi un plasmide centromérique avec une expression régulée par un promoteur Tet-Off. Enfin, la production d'une protéine de fusion HBH (Histidine-Biotine-Histidine) permet une purification de la protéine en conditions totalement dénaturantes qui solubilisent notre intégrase. Cependant, nous avons dû optimiser les conditions de sonication pour la préparation de la chromatine dans un tampon contenant de l'urée parce que les conditions expérimentales initialement développées dans le laboratoire n'étaient pas adaptées à l'utilisation de tampons chaotropiques.

Lors de ma première tentative de TChAP à partir de cellules cultivées à 20°C, nous avons observé un rendement très faible de l'intégrase contrairement à celui obtenu à partir de cellules cultivées à 30°C. Des études ayant montré que l'intégrase et la reverse transcriptase de Ty1

forment un complexe in vivo (Wilhelm et al 2006), nous avons testé leur co-expression ce qui a permis d'améliorer très nettement l'expression de IN à 20°C et d'obtenir des quantités de protéine suffisantes pour des analyses par spectrométrie de masse (MS).

En raison de la grande sensibilité de la technique de spectrométrie de masse, un grand nombre de protéines est détectée. Une analyse importante des données de MS basée sur plusieurs critères (background, reproductibilité…) est absolument essentielle pour sélectionner des partenaires potentiels. Le partenaire le plus important de l'intégrase que nous avons identifiée est l'ARN Pol III, en accord avec la littérature, suggérant que notre TChAP est une bonne approche protéomique. Nous avons également identifié plusieurs autres partenaires intéressants qui ont fait l'objet d'une analyse plus approfondie. Pour n'en nommer que quelques-unes, nous avons identifié les quatre sous-unités du complexe caséine kinase II qui régule positivement la transcription de Pol III et qui sont connues pour réguler négativement la rétrotransposition. Parmi les autres candidats potentiels qu'il pourrait être intéressant d'étudier on peut citer le facteur de transcription Dst1, Rap1, Reb1, le complexe Tup1-Cyc8, voir encore la protéine Bre1, une E3 ubiquitine ligase qui a déjà été identifiée comme un régulateur négatif de la rétromobilité Ty1 in vivo.

L'étape suivante consistait à vérifier si ces protéines étaient de véritables partenaires du Ty1 IN. Nous avons démontré que la plupart de ces protéines étaient capables de co-immunoprécipiter avec l'intégrase in vivo, confirmant la robustesse de nos expériences TChAP et analyses MS.

Comme plusieurs études ont montré que CK2 pouvait jouer un rôle dans la régulation de la transcription par la Pol III et que d'autres suggéraient que CK2 pouvaient également jouer un rôle dans la rétrotransposition de Ty1 (Nyswaner et al 2008), cette kinase nous a paru le candidat idéal à analyser. J'ai donc décidé de poursuivre ce travail en me focalisant plus particulièrement sur CK2 et son rôle potentiel sur la rétrotransposition.

3. La caséine kinase 2, un régulateur de la rétrotransposition de Ty1 ?

La levure CK2 est une sérine-thréonine-kinase constituée de 2 sous-unités catalytiques (Cka1, Cka2) et de 2 sous-unités régulatrices (Ckb1, Ckb2) ayant de nombreux substrats dont plusieurs composants de la machinerie de transcription Pol III. Des études (Acker et al 2013)) ont montré que CK2 régule positivement la transcription par l'ARN Pol III chez la levure . De plus, cette

kinase est présente sur les gènes transcrits par la Pol III, comme les ARNt (Graczyk et al 2011). CKB2 a également été identifié dans un crible génétique réalisé pour isoler des facteurs qui répriment la rétromobilité du Ty1 (Nyswaner et al 2008). Leurs données suggèrent que CK2 pourrait agir au niveau post-transcriptionnel pour réguler la rétrotransposition de Ty1 car la délétion de CKB2 n'a eu aucun effet sur le niveau d'ARNm Ty1. Récemment, dans un crible protéomique de Cheung et al., 2015, Ckb2 a également été identifié comme un partenaire potentiel l'intégrase de Ty1. Néanmoins, aucune étude n'a été réalisée pour analyser le le rôle de CK2 dans la régulation de la rétrotransposition de Ty1. Enfin, Irwin et al en 2005 ont également montré une augmentation de la fréquence de rétrotransposition de Ty3 en absence de CKB2. Comme les intégrases de ces 2 rétroéléments Ty1 et Ty3 partagent une certaine homologie et surtout possèdent ont une spécificité d'intégration au niveau des gènes transcrits par la Pol III. Une régulation globale de la rétrotransposition par CK2 peut être envisagée.

3. 1 Rôle de CK2 sur la mobilité de Ty1

Nous avons identifié des peptides des quatre sous-unités du CK2 dans nos données de spectrométrie de masse. Pour valider ces données, des expériences de co-immunoprécipitation (CoIP) ont montré que CK2 est associée à IN. Enfin une interaction entre la sous-unité Ckb2 et IN1 a été détectée par la technique du double hybride chez la levure.

En parallèle à nos études, A. Barkova de l'équipe de P. Lesage a analysé le rôle de différentes sous-unités CK2 sur la mobilité d'un élément spécifique Ty1 in vivo, en utilisant divers mutants de délétion. La présence d'un gène rapporteur (his3AI) au niveau d'un élément Ty1 endogène (Ty1-his3AI) permet d'analyser spécifiquement sa fréquence de transposition (Curcio et al. 1991). L'absence de l'une ou l'autre des sous-unités catalytiques n'a aucun effet sur la mobilité de Ty1 alors que la délétion de ckb2 entraine déjà une augmentation significative de celle-ci, en accord avec les études de Nyswaner et al 2008, qui avaient identifié *CKB2* comme un répresseur de la mobilité Ty1. Par ailleurs, l'absence conjuguée de CKA2 avec l'une ou l'autre des sous-unités régulatrices (Δcka2Δckb2 et Δcka1Δckb2) augmente de manière très importante la rétrotransposition de Ty, suggérant que le partenaire préférentiel de IN serait l'holoenzyme CK2.

3. 2 L'effet du CK2 sur l'expression de Ty1 *in vivo*

Pour déterminer si l'augmentation de la mobilité de Ty1 en l'absence de CK2 est liée à une augmentation de l'expression de Ty1 in vivo, nous avons analysé l'expression des éléments Ty1. Dans un premier temps, des tests de qRT-PCR ont été réalisés dans le laboratoire partenaire (A. Barkova) pour déterminer le niveau d'ARNm. Dans le cas de l'ARNm spécifique du Ty1-His3AI, aucune augmentation significative n'a été détectée en présence ou non de sous-unités de CK2. Par contre, la quantité d'ARNm globale de l'ensemble des éléments Ty1 endogènes est plus important dans les souches Δcka2Δckb2 et Δcka2Δckb1 suggérant un rôle de CK2 sur la transcription de ces gènes.

Par ailleurs des analyses de western blot ont également montré une augmentation de la quantité d'intégrase endogène présente dans les cellules en l'absence de sous-unités de CK2. Des résulats similaires ont été obtenus lorsque l'intégrase est exprimée de manière ectopique à partir d'un vecteur plasmidique suggérant donc un effet post-traductionnel.

Il est difficile de conclure qu'une légère augmentation de la quantité de la quantité d'intégrase puisse expliquer la forte augmentation de la mobilité de Ty1 en absence de CK2. Cependant, les résultats présentés suggèrent un problème de stabilité de l'intégrase qui pourrait être régulée par CK2. Cette hypothèse est actuellement testée dans le laboratoire.

3. 3 Ty1 IN phosphorylation par CK2 *in vitro*

Après avoir démontré que CK2 interagit avec Ty1 IN (IN1), nous nous sommes ensuite demandés si cette intégrase était bien un substrat de cette kinase. Nous avons démontré que CK2 de levure était capable de phosphoryler l'intégrase IN1 in vitro et 11 sites de phosphorylation ont été identifiés par spectrométrie de masse. Par la suite, nous avons montré que ces sites étaient également phosphorylés in vivo chez la levure, ce qui suggère que IN1 est bien un substrat de cette kinase. Pour déterminer si Ck2 a un effet direct sur l'expression de Ty1 IN, nous avons construit des phosphomutants mimant une absence de phosphorylation. Nos données préliminaires réalisés avec certains de ces mutants suggèrent qu'effectivement l'état de phosphorylation de ces acides aminés pourraient influencer le niveau d'expression de l'intégrase.

3. 4 Ty1 IN est dégradé par le protéasome

Nos données suggèrent que le Ty1 IN pourrait être une protéine à courte durée de vie dans des conditions normales de croissance. Diverses études ont montré que ce type de protéine était dégradée par le protéasome (Lee et al 1996, Goldberg et al 1997). Nous nous sommes demandé si c'était le cas pour IN1, si IN1 était effectivement un substrat du protéasome. Pour tester cette hypothèses, l'expression d'intégrase ectopique, a été analysée par western blot en présence de MG132, un inhibiteur du protéasome. Ces expériences préliminaires montrent effectivement une accumulation d'intégrase dans ces conditions, suggérant que l'intégrase est bien un substrat du protéasome. En absence de CK2, une légère augmentation du niveau d'intégrase peut également être détectée en présence de MG132, suggérant que d'autres voies métaboliques pourraient également réguler cette dégradation.

Parmi les protéines associées à IN1 identifiés au cours de nos approches protéomiques (TCHAP), de nombreuses sous-unités du protéasome ont été identifiées en bonne corrélation avec cette hypothèse. Enfin, des analyses protéomiques réalisées à grande échelle ont également identifié les rétroéléments Ty1 comme substrat du protéasome (**Kaake et al 2010)**

Projet 2

1. l'interaction entre ARN polymérase et Ty1 IN *in vivo*

Le groupe du Dr Pascale LESAGE a démontré que l'intégrase de Ty1 interagit avec la sous-unité AC40 de l'ARN polymérase III et que cette interaction est très importante pour la sélectivité d'intégration des éléments Ty1 en amont aux des gènes d'ARNt (Bridier Nahmias et al 2015). Des expériences de double hybride, réalisées chez la levure, ont permis d'identifier le domaine d'interaction avec AC40 qui est situé dans la partie C-terminale de l'intégrase (AA 578-635). La comparaison des séquences peptidiques des régions C-terminales des rétroéléments homologues Ty1, Ty2 et Ty4 a révélé une forte homologie dans cette région. En particulier, un motif de 6 acides aminés, KNMRSLE où les acides aminés K617, S621 et L622 sont parfaitement conservés dans les trois intégrases. Des expériences de double hybride chez la levure ont ensuite montré que la mutation de chacun de ces 3 acides aminés (K617A, S621A ou L622A) entraine une perte d'interaction avec AC40, suggérant qu'ils sont nécessaires pour l'association avec AC40. Notre groupe a démarré une collaboration sur ce projet et j'ai été plus

particulièrement impliqué dans la caractérisation de l'interaction entre les machinerie Pol I et Pol III avec Ty1 IN par des test biochimiques *in vivo* et *in vitro*.

1.1 Identification des sous-unités d'ARN polymérase qui interagissent et avec Ty1 IN

Nos expériences de TChAP ont révélé que l'ARN Pol I (Pol I) est l'un des partenaires potentiels de Ty1 IN. Comme plusieurs sous-unités sont communes aux Pol I et Pol III, en particulier AC40 la cible majeure de Ty1 IN, il n'est pas surprenant que la Pol I puisse s'associer avec Ty1 IN. Comme pour la Pol III, nous avons également confirmé par des expériences de co-IP que Ty1 IN est associée à la Pol I *in vivo*. Cependant, pour identifier avec précision toutes les sous-unités impliquées, nous avons utilisé la technique du double hybride levures et testé les interactions entre chacune des sous-unités des Pol III ou Pol I avec Ty1 IN. C'est ainsi que nous avons montré que plusieurs sous-unités spécifiques (C31, C11, C25 Pol III et A190, A14 Pol I) interagissent avec Ty1 IN. Cependant, pour confirmer une interaction directe, des analyses supplémentaires comme par exemple des essais de GST pull-down sont absolument nécessaires.

4.2 Ty1 IN est associé aux gènes cibles in vivo

D'après le modèle d'intégration, l'intégrase se lie à l'ADN cible pour initier le processus d'intégration. Cependant, une telle interaction entre Ty1 IN et son ADN cible n'a jamais été démontrée auparavant. Nous nous sommes donc demandé si une liaison entre les gènes Pol III et Ty1 IN pouvait être détectée par des expériences de ChIP. Nous avons réalisé ce type d'analyse en utilisant une intégrase exprimée de manière ectopique chez la levure et révélé une forte occupation sur tous les gènes Pol III que nous avons testés ainsi que sur le gène transcrit par la Pol I.

4.3 Caractérisation de la "targeting sequence" (TS) de Ty1 IN

L'étape suivante a consisté à caractériser les séquences importantes pour la spécificité d'intégration. Des études réalisées conjointement dans nos 2 laboratoires ont permis de mettre en évidence que la séquence KNMRSL, fortement conservée entre les rétroéléments Ty1, Ty2 et Ty4, semble être une séquence TS.

Dans un premier temps, j'ai pu montrer que l'interaction entre Ty1 IN et AC40 était directe en réalisant des expériences de co-immunoprécipitations à partir d'extrait bactériens dans lesquels AC40 et Ty1 IN étaient co-exprimées. J'ai également confirmé une perte de cette interaction lorsque les acides aminés K617, S621 ou L622 situés dans le TS étaient mutés. Des expériences de ChiP ont par ailleurs montré que l'intégrase ainsi mutée n'était plus associée ni aux gènes Pol III ni au gène Pol I in vivo. L'ensemble de ces données confirment l'importance de la séquence TS que nous avons définie dans la spécificité d'intégration.

5. Conclusion

Au cours de mon projet de thèse, nous avons adapté la technique du TChAP pour l'identification des partenaires de Ty1 IN et une liste assez conséquente de nouveaux interactant potentiels a été obtenue. La validation de ces cibles potentielles par CoIP prouve que le TChAP est une bonne approche protéomique. Cependant, les données brutes de spectrométrie de masse sont encore analysées avec soin pour les affiner, de manière à ne négliger aucune cible intéressante. En plus de la Pol III, nous avons trouvé d'autres partenaires potentiels qu'il serait intéressant d'étudier par la suite. Certains d'entre eux sont présents sur les gènes d'ARNt et sont impliqués dans la régulation de la transcription par la Pol III. Par conséquent, ils peuvent être un lien important entre les machineries Pol III et l'intégration de Ty1 in vivo. Parmi les autres partenaires, certains sont associés à d'autres endroits du génome et pourraient éventuellement être des facteurs d'ancrage secondaires qui pourraient expliquer les insertions de Ty1 identifiés en dehors des gènes Pol III.

Il n'est pas possible d'étudier simultanément le rôle de plusieurs ces partenaires, il nous a donc fallu faire des choix et nous avons donc décidé d'aller plus loin dans l'étude du rôle moléculaire du CK2 dans la rétrotransposition de Ty1. En accord avec la littérature, nous avons observé que CK2 réprime la rétrotransposition de Ty1 in vivo. Nous avons également observé que Ty1 IN qui est à peine détecté dans des conditions de croissance normales montre une abondance légèrement supérieure en l'absence de CK2 suggérant qu'il pourrait y avoir un problème de stabilité de la protéine dans des conditions physiologiques. Nous avons obtenu des résultats préliminaires très prometteurs montrant que le Ty1 IN est phosphorylée par CK2 et que cette phosphorylation interfère avec l'expression du Ty1 IN in vivo. De plus, nous avons également démontré que Ty1 IN est un substrat du protéasome. Par ailleurs, nous effectuons actuellement des analyses complémentaires pour étudier le lien possible entre la phosphorylation de Ty1 IN

par CK2 et sa dégradation par le protéasome. Nous avons émis l'hypothèse d'un modèle de dégradation de Ty1 IN par le protéasome, qui dépend de la phosphorylation. Actuellement, nous testons notre hypothèse.

De plus, nous avons été étroitement impliqués dans la caractérisation de la séquence de ciblage (TS) du Ty1 IN avec l'équipe du Dr Pascale LESAGE. Nous avons fourni des preuves convaincantes de l'importance de cette séquence TS sur l'interaction avec l'ARN Pol III et avec les gènes transcrits par la Pol III, en bon accord avec un modèle de "tethering » dans lequel l'ARN Pol III serait le facteur principal d'ancrage de l'u-intégrase du Ty1 chez la levure Saccharomyces cerevisiae.

Actuellement, d'autres expériences sont réalisées en laboratoire pour compléter cette analyse, et la rédaction de deux articles de recherche portant sur le rôle du CK2 dans la rétrotransposition in vivo de Ty1 et la caractérisation de la séquence de ciblage du Ty1 IN est en préparation. Je serai premier co-auteur avec Anastasia BARKOVA (Laboratoire de Pascale LESAGE) sur l'étude concernant CK2, qui devrait être soumise en 2019, tandis que je serai également co-auteur e de l'étude sur la caractérisation du TS.

TABLE OF CONTENTS

ABBREVIATIONS

TE	**Transposable element**
LTR	**Long terminal repeat**
RT	**Reverse transcriptase**
IN	**Integrase**
SSC	**Self-splicing elements**
LINEs	**Long interspersed nuclear elements**
SINEs	**Short interspersed nuclear elements**
ORF	**Open reading frame**
EN	**Endonuclease**
IS	**Insertion sequence**
TIR	**Terminal inverted repeats**
TSD	**Target site duplication**
TPRT	**Target primed reverse transcription**
VLP	**Virus like particle**
PR	**Protease**
NAC	**Nucleic acid chaperon**
RNP	**Ribonucleoprotein**
CUT	**Cryptic unstable transcript**
ASRNA	**Anti-sense RNA**
MMS	**Methyl methane sulphonate**
4 NQO	**4-nitroqinoline-1-oxide**
UV	**Ultraviolet**
ESR	**Environment stress response**
SRE	**Stress response elements**
RIG	**Retrotranscript indicator gene**
SGA	**Synthetic genetic array**
TF IIIA	**Transcription factor IIIA**
TFIIIB	**Transcription factor IIIB**
TFIIIC	**Transcription factor IIIC**
snRNP	**Small nuclear ribonucleoprotein**
SRP	**Signal recognition particle**

ICS	**Internal control sequence**
TBP	**TATA binding protein**
TChAP	**Tandem chromatin affinity purification**
HBH	**Histidine Biotin Histidine**
GST	**Glutathione S- transferase**
CK2	**Casein kinase II**
3 AT	**3-amino-1,2,4-triazole**
SDS PAGE	**Sodium dodecyl sulphate polyacrylamide gel electrophoresis**
TS	**Targeting sequence**
CNC	**Copy number control**
DUB	**deubiquitinating enzymes**

Chapter 1

INTRODUCTION

1. Transposable elements (TEs)

1.1 What are transposable elements?

Transposable elements are mobile elements widespread in the eukaryotic and prokaryotic genome. They have the ability to transpose from one site of the genome to the other which is why they are known as transposons. Dr. Barbara McClintock was the first scientist to report in the 1940s about TEs when she was working on unstable alleles in maize and discovered that genomic instability in maize is caused by DNA elements that jump from one place to another. Almost, 20 years later she was honored with the Nobel Prize in 1983 for her discovery of TEs. TEs occupy a substantial fraction of the eukaryotic genome and play a significant role in evolution and in maintaining genetic diversity. The percentage composition of TEs in various species is shown in Figure 1.

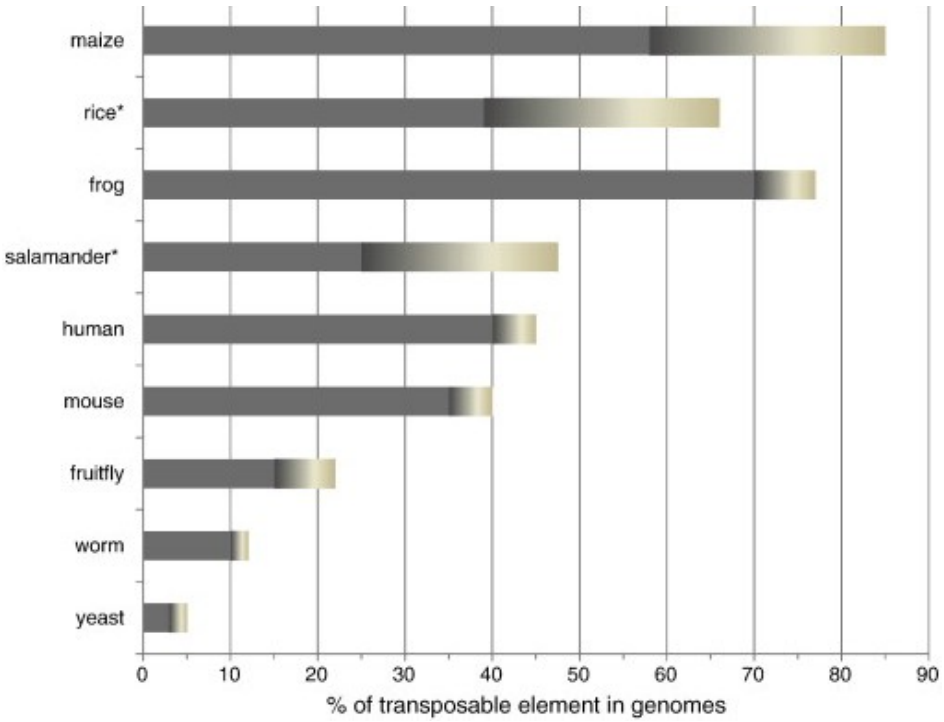


Figure 1: **Distribution of TEs in different species**. The colour gradient in the bar are the differences in the reported TE% in these species in various research studies (Adpated from **Chénais et al 2012**)

1.2 Classes of Transposable elements

Transposable elements are normally classified depending on their mechanism of transposition. **Finnegan 1989** was the pioneer in transposable element systemic classification based on what was reported in human, drosophila and yeast models (**Singer et al 1982, Finnegan & Fawcet 1986, Williamson 1983**). He proposed that transposable elements can be divided into two major classes (I) Retrotransposons and (II) DNA transposons. Retrotransposons replicate via an RNA intermediate through reverse transcription of the mRNA to cDNA and DNA transposons directly transpose from DNA to DNA. The DNA transposons encode an enzyme known as transposase that mediate their transposition from one site of the genome to the other. The class I elements are further divided into two subclasses based on their structural features. The first subclass is the Long terminal repeat (LTR) retrotransposons resembling retroviruses. They are flanked by long terminal repeats on either ends of the genome and encode the Gag and Pol proteins like the retroviruses. The other subclass includes the non-LTR elements which do not contain long repeats at their genetic ends, though later research showed the presence of short repeated sequences. They also encode the Gag and Pol proteins with poly-A tail at the 3' end. This classification was referred to as the Finnegan proposal. The simplified classification has been represented in Figure 2 from the original publication.

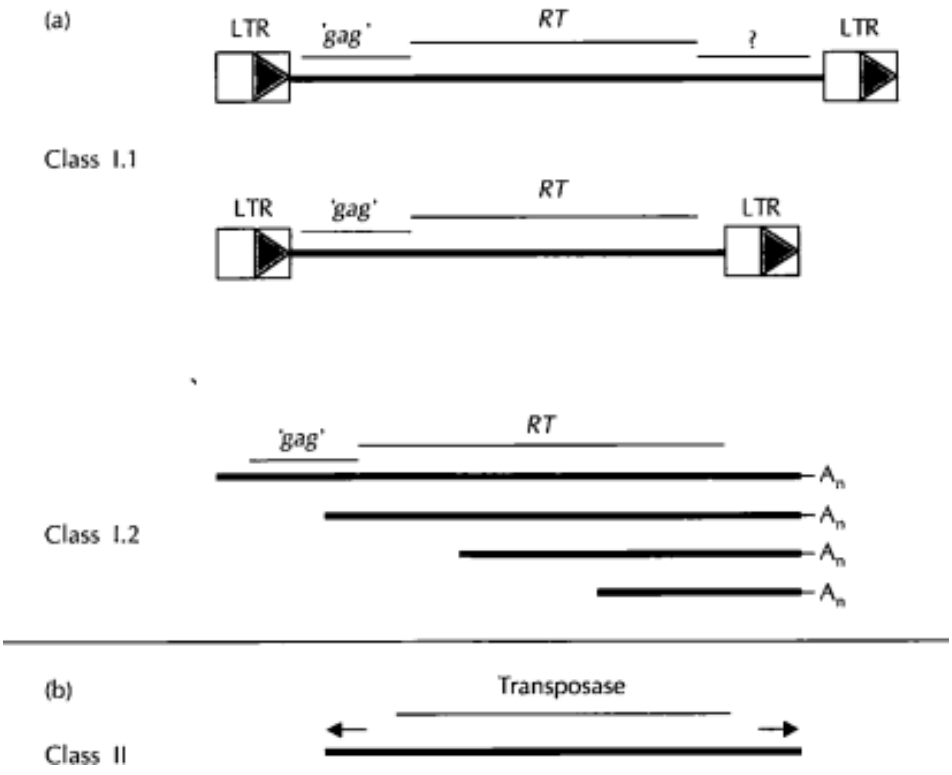


Figure 2**. Classification of transposable element by David J. Finnegan.** Classification of the transposable elements as Class I elements (retroelement) and Class II elements (DNA transposons). The Class I is subdivided into LTR and non-LTR elements. LTR elements are flanked by long terminal repeats shown by arrows and non LTRs have poly-A tail at the 3' end. The Class II transposons have inverted terminal repeats shown by opposite facing arrows and encode the transposase for mobility (Adapted from **Finnegan 1989**).

Later, classifications were refined and there have been major updates in the Finnegan proposal, known as the Wicker and Repbase propositions (**Wicker et al 2007, Kapitonovo and Jurka 2008)**. The Wicker proposition is in alignment with the basic structure provided by Finnegan. The basic criterion divides the TEs into two classes depending on the presence or absence of an RNA intermediate (Class I as retrotransposons and Class II as DNA transpososons) but each class is further divided into subfamilies based on the genetic organization and reverse transcriptase (RT) phylogeny. The Wicker proposition is represented in figure 3.

The Repbase proposal is very similar to the Wicker proposal where the TEs are basically divided into two categories, retrotransposons and DNA transposons and they refer to them as *types* rather than *classes.* The subfamilies in each type are divided based on the enzyme used for their transposition and are similar in both the cases. Though people over the years, have critically analyzed these classification systems and found weaknesses, the basic criterion to classify TEs according to their mechanism of mobility remains the same. One shortcoming is that these classifications considered only the eukaryotic TEs and not the prokaryotic ones. The elements in each of the subfamilies are discussed later. **Derbyshire and Curcio, 2003** divided TEs according to the enzymes/integrase involved in integrating the DNA to the host genome and the transposition intermediate. They proposed four classes (1) the DDE-transposases, (2) rolling-circle (RC) or Y2-transposases, (3) tyrosine (Y1)-transposases and (4) the serine (S)-transposases and each class was further divided taking the transposition mechanism into account. There always remains a debate as to which classification should be considered in the current scenario. In the recent times, **Piégu et al 2015,** supports the classification method of **Derbyshire and Curcio 2003** and have also proposed a new class to the existing classification. They considered the Self splicing elements (SSC) that didn't gain proper attention as TEs for a long time.

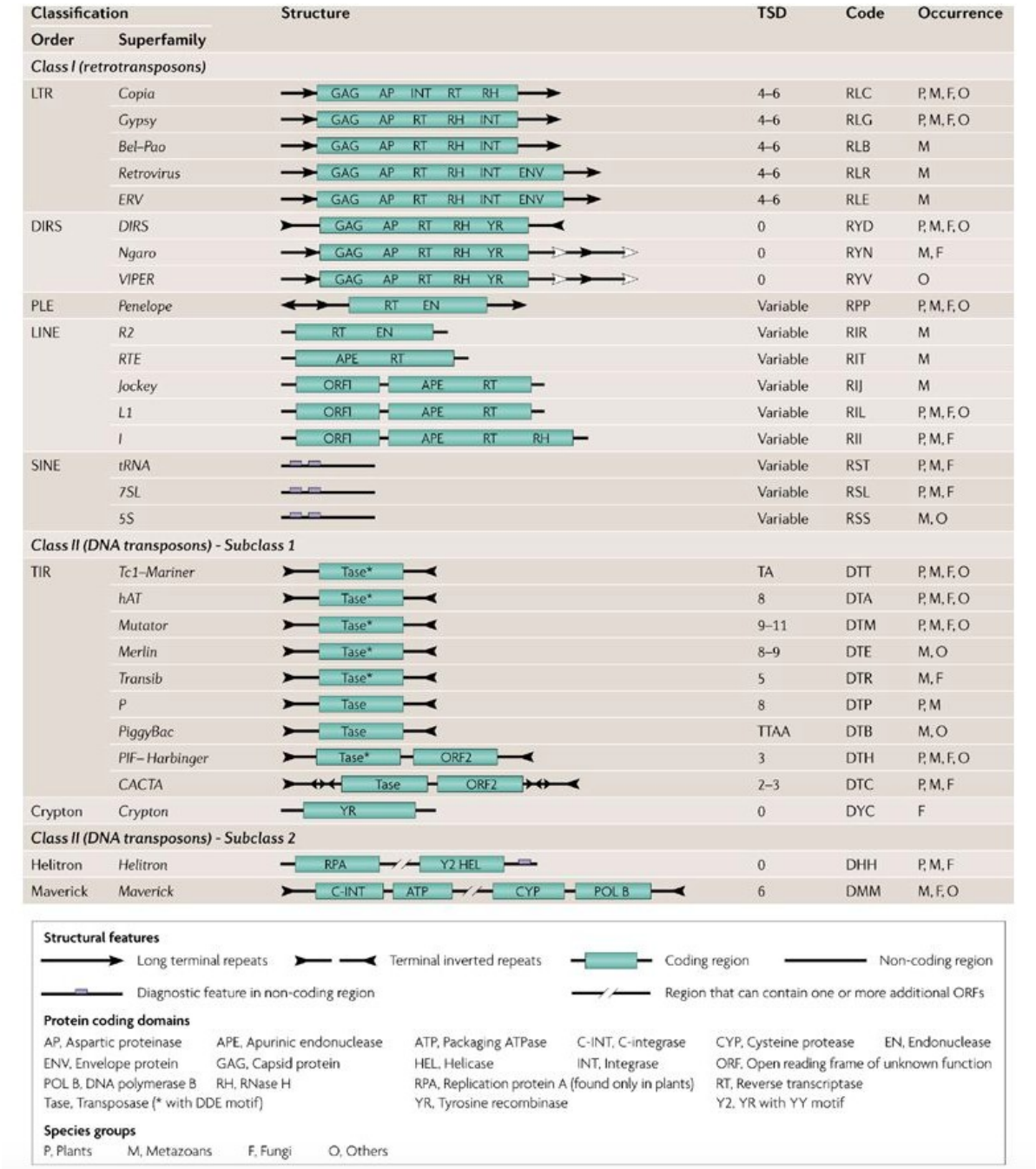


Figure 3. **Classification of transposable element by Wicker et al 2007.** They classified the transposable elements as Class I elements (retroelement) and Class II elements (DNA transposons) like Finnigan but divided each class as many subfamilies based on the genetic organization, mechanical features and RT phylogeny. (Adapted from **Wicker et al 2007)**

1.2.1 Retrotransposons (Class I)

Retrotransposons constitute approximately 45% of the human genome (**Bannert et al 2004**) and replicate via an RNA intermediate. They resemble retroviruses and are found in all eukaryotes. As mentioned before, based on the genetic structure they are divided into LTR and non-LTR retroelements. More recently DIRS and Penelope have been added to this category (Figure 3).

1.2.1a LTR retrotransposons

LTR retrotransposons constitute approximately 8% of the human genome. It includes the endogenous retroviruses in humans. The LTRs are the direct repeat sequences flanking the ORFs of the retroelements and the ORF contains the GAG and POL coding regions. Most of the information regarding the mechanism of LTR retrotransposons come from the extensive study of Ty elements in yeast **(Voytas et al 2002, Sandmeyer et al 2015)**, but it is assumed that the mechanism is similar among LTR retroelements in varied hosts. With the availability of genome sequence data from varied organisms, it has been well established that most organisms contain LTR retrotransposons from multiple distinct lineages. Although, they contain long terminal repeats, they differ in their genomic organization of the Pol genes and in the DNA sequence. The International Committee on Taxonomy on Viruses has provided a taxonomic framework (Figure 4a) of the large pool of retrotransposons based on the relationship among the amino acid sequences of the reverse transcriptase protein (RT), the highest conserved Pol protein. Two large families of LTR retrotransposon families found in eukaryotes are Pseudoviridae and Metaviridae. They are also differentiated on the basis of the order of Pol encoding genes (Figure 4b). They are autonomous elements as they encode for all the catalytic enzymes required for their retrotransposition such as the protease, reverse transcriptase (RT) and Integrase (IN). The replication life cycle will be discussed later.

This taxonomic framework went through revision over the years as diverse elements have been discovered. Three genera have been described under *Pseudoviridae*, namely pseudoviruses, hemiviruses and sireviruses. Sireviruses are derived from plants and constitute a distinct lineage according to their RT amino acide sequences. Pseudoviruses and Hemiviruses are differentiated by the primer they use for reverse transcription. The former uses a full tRNA primer whereas the latter uses a half tRNA primer. Likewise, the *Metaviridae* is further divided into three genera

namely, metaviruses, errantiviruses and semotiviruses. They are differentiated on the basis of phylogenetic analysis of their RT amino acid sequences.

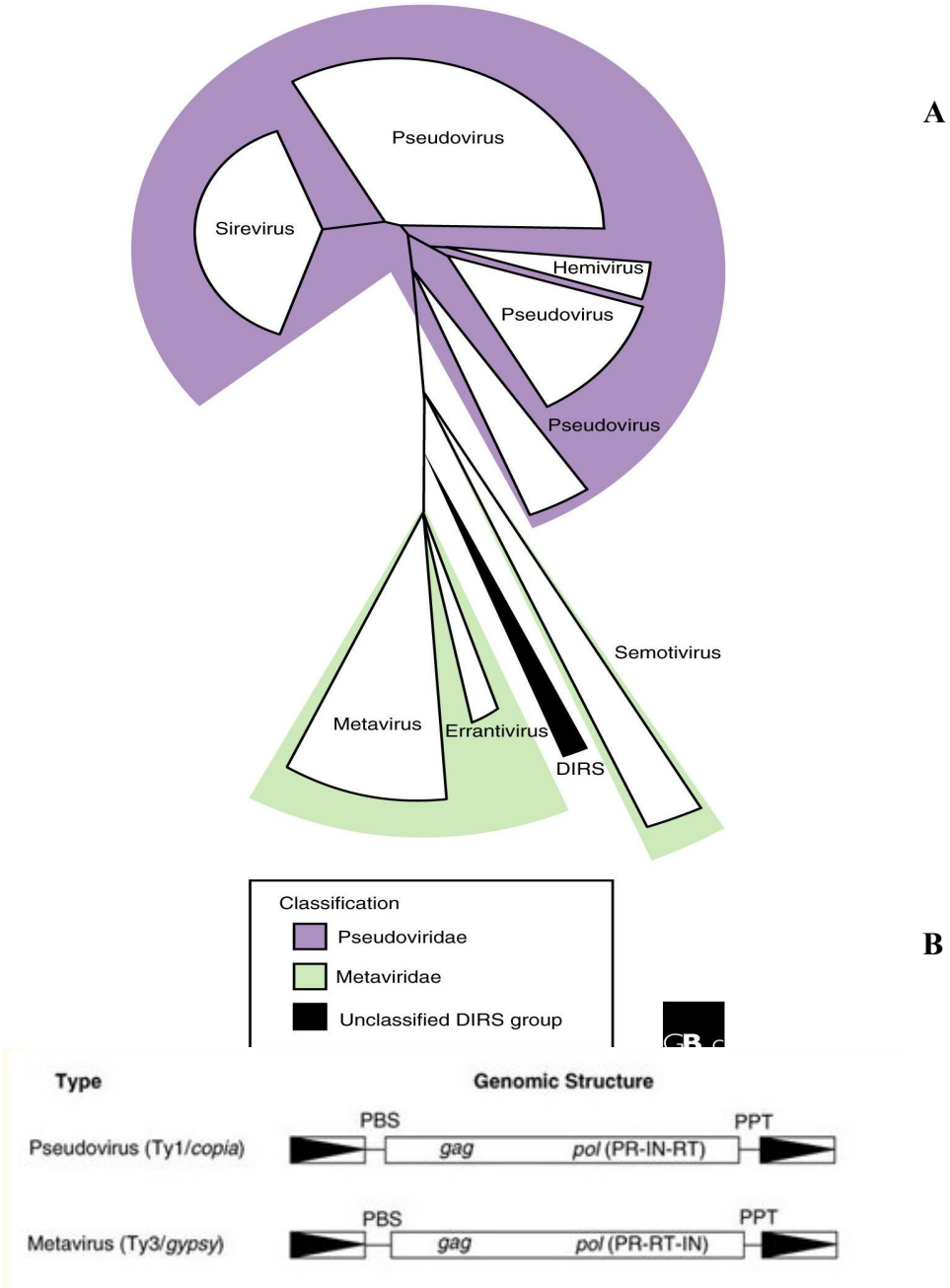
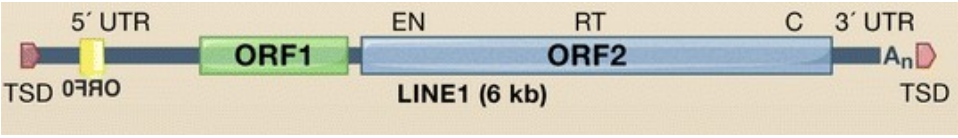


Figure 4. **Schematic diagram of the classification of LTR retroelements.** (A) The sectors represent the elements that make up the distinct lineages. The DIRS are retroelements that do not encode an Integrase or Protease but instead encodes a Tyrosine recombinase (YR). (B) **The genomic structure of the two families.** The orientation of the IN and RT in the *Pol* gene is reversed. (Image adapted from **Haveker *et al* 2004**)

1.2.1b non- LTR retrotransposons

The non-LTR retroelements are abundantly found in humans and other eukaryotic organisms. It accounts for 33% of transposable elements in humans and have played a significant role in genetic diversity and are linked to various diseases. Non-LTR retrotransposons are as old as the earliest multi-cellular organisms (**Malik et al 1999**). They are composed of the autonomous Long Interspersed Nuclear Elements (LINEs) and the non-autonomous Short Interspersed Nuclear elements (SINEs). The LINE 1s (L1s) are the only active autonomous elements in the human genome today and have been evolving for the past 150 million years. They are extensively studied for their potential role in various cancers. L1s comprise of almost 17% of the human genome and the L2s and L3s comprise an additional 4% (**Goodier et al 2016**).

The difference with the LTR elements being that these elements have a different genetic organization and replication mechanism. They do not contain any terminal repeats. The LINEs consist of a 5'UTR region functioning as the promoter and the 3'UTR ending in a poly-A tail. There are generally two ORFs. ORF1 encodes a 40 kDa RNA binding protein required for retrotransposition and the ORF2 encodes a 150 KDa endonuclease and reverse transcriptase (RT) (**Han et al 2010**). The mechanism of reverse transcription of the RNA copy to DNA differs from that of the LTR elements because non LTR elements do not possess a tRNA primer binding site, instead the LINE RNA associated with the RT and endonuclease enter the nucleus to the integration site and the 3'- OH primes to the single strand break on the target DNA made by its own nuclease after which the LINE RNA is reverse transcribed. This mechanism is known as 'target primed reverse transcription' (**Finnigan 2012**). The SINEs depend on the transposition machinery of the LINEs to propagate in cells, due to which they are called non- autonomous elements. In humans SINEs are known as ALU elements because they have a restriction site for the enzyme *Alu*I. They are 300 bp in size and are mostly hybrid elements. The 5' is derived from a tRNA, containing a Pol III promoter and the 3' containing poly-A tails like LINEs. The LINE related sequence helps them to be identified by LINE proteins to drive its retrotransposition (**Dewannieux et al 2003**). The schematic representation of the genetic structure of the non-LTRs are represented in Figure 5.



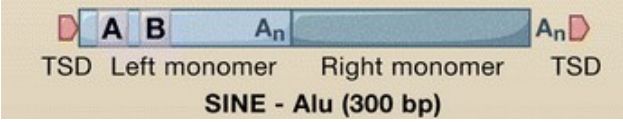


Figure 5. **Schematic diagram of the genetic structure of non-LTR elements.** Upper Panel shows the structure of LINE1 in humans containing a 5'UTR as the promoter and two ORFS encoding the RNA binding protein, the endonuclease (EN) and the reverse transcriptase (RT). The lower panel shows the structure of SINE-Alu elements in human which is a hybrid composed of a 5' derived from a tRNA gene containing the Box A and Box B of the Pol III promoter sequence and the 3' end derived from the LINE sequence ending with poly-A tails. (Image adapted from **Goodier et al 2016**)

1.3 DNA transposons (Class II)

Class II elements are present in almost all eukaryotes. In prokaryotes, a simplified form of Class II elements is known as insertion sequences (IS). They are divided into two big subclasses depending on the number of DNA strands that are incised during the transposition process. None of the class II elements move via an RNA intermediate.

The subclass I comprises of the terminal inverted repeats (TIRs) which vary in length and size. Under this subclass, there exists nine super-families which differ in their TIR sequence and target site duplication (TSD) size (**Wicker et al 2007**). The *Tc1-mariner* contains two TIR and a transposase where the catalytic DDE motif is conserved (**Shao et al 2001**). The hAT superfamily derives its name from three TE familes; hobo in drosophila, Ac-dc in maize, Tam3 in snapdragons. They are approximately 4 kb in size with a short 5-27 bp TIR and possess TSDs of 8 bp (**Kempken et al 2001**). The *mutator* superfamily occurring in eukaryotes can have up to 100 bp of TIR but sometimes have shorter TIRs. They produce around 9-11 TSDs and were identified in Entamoeba protozoan lately (**Pritham et al 2005**). *Merlin* superfamily of elements are present only in animals. A fully functional element encodes the DDE transposase and can be larger than 10 kb. The transposase of the superfamily *Transib* contains the conserved DDE motif and is found in Drosophila and mosquitoes (**Kapitonov 2003**). The RAG1 protein required for the VdJ recombination of DNA repair is derived from *Transib* transposase (**Kapitonov 2005**). The P superfamily produces 8 bp TSDs and was first identified in insects but are also found in metazoans and algae (**Hammer et al 2005**). The *piggybac* transposon family are widely found in animals and prefer TTAA chromosomal sites for integration (**Sarkar et al 2003**). They contain a DDD transposase motif comparable to the DDE motif of the other transposase and retroviral integrase. The piggyback transposon system has been deployed as

tools in genetic engineering. The *PIF-Harbinger* integrates at TAA chromosomal sites and contains 25 bp TIR (**Jurka et al 2001**). It contains two ORFs, one encoding a DDE transposase and the other encoding a DNA binding protein. *Cryptons* are not well characterized and have been found only in fungi (**Goodwin et al 2003**). They encode a tyrosine recombinase like other retroelements such as DIRS. They do not possess the RT indicating that they transpose via a DNA intermediate.

The subclass 2 is different in the replication process compared to the subclass 1 members. During replication, these members do not undergo a double stranded cleavage. Instead, they transpose by displacement of only one strand. The *Helitrons* transpose by a mechanism known as rolling circles without generating any TSDs (**Kapitonov 2001**). The ends of the elements are marked by a TC or CTRR motif (R is any purine). They encode a single-strand DNA binding protein and a tyrosine recombinase. They have been found mostly in plants but also constitute approximately 2% of *C. elegans* genome and 3% of the genome of bats, *Myotis lucifugus*. *Mavericks* have been found irregularly in eukaryotes but not in plants (**Pritham et al 2007**). They are long is size ranging from 10-20 kb with long TIRs flanking its ends. The coding region of these elements are relatively long, encoding up to 11 proteins but vary normally. They encode a DNA polymerase B and an integrase lacking the RT.

1.4 Significance of transposable elements

There are many studies delineating the importance of transposable elements as a key player in evolution. Initially, these transposable elements were regarded as "Junk DNAs" and it took quite some time for the scientific community to realize the significance of these elements. Normally, the insertion of TEs into genes or gene regulatory sequences can disrupt the function of the gene leading to fatal consequence to the genome. But transposition can also have a positive effect on the host leading to genetic diversity and innovation. **Chenais 2012** reviews the impact of transposable elements on the genome. The fact that these transposable elements occupy such massive portion of host genomes such as almost 85% of the maize genome, their role in genome expansion can't be ignored. Many studies report the variation of the genome size of different species of the same genus due to TE amplification (**Biémont et al 2015**). For instance, **Zuccolo et al 2007** showed that the genome size variation in the *Oryza* genus is due to both polyploidization and LTR retrotransposon proliferation. **Sun et al 2012** showed that the salamander genome ranging from 14 gb to 120 gb consists abundantly of

Ty3/gypsy LTR retroelements. This large variation of the genome size has been attributed to LTR proliferation. Chromosomal rearrangements are yet another role that these TEs play in shaping the host genome. Chromosomal rearrangements were first observed by Dr. Barbara McClintock were the TEs induced deletions and insertions in the genome. Since then, chromosomal rearrangements have been observed in various organisms including plants, invertebrates and vertebrates (**Bennetzen 2005**, **Konkel and Batzer 2010**). In humans LINEs and Alu elements have been shown to cause genetic rearrangements having implications in various types of cancers. These non-LTR elements are often found located near chromosomal rearrangement locus. They contribute to deletions, duplications, inversions and translocations in the human genome.

Transposable elements have also been shown to regulate gene expression by different mechanisms. It has been observed that many human exons have been derived from TEs (**Zhang and Chasin 2006**). For example, the C-terminal exon 219.1 of the human leptin receptor has been derived from TEs and possesses a functional role. Another role of TEs in gene regulation, is insertions at promoter and enhancer regions of genes (**Teng et al 2011**). Approximately 25% of the human promoter region has been composed of TE sequences. These TE insertions can upregulate a gene expression leading to cancers and other genetic diseases. All the active non-LTR retroelements in the human genome such as the L1, Alu and SVA are involved in causing genetic disorder. The following table 1 shows the contribution of TEs to human diseases.

Another important role of TEs is its ability to confer adaptability to the host under environmental stress. This idea was suggested by Dr. McClintock 1984 and indicates that under stress the TE proliferation increases to such an extent that it induces genetic variability and this variability is genetically transferred to the progeny. Several studies in plants have shown that TE expression is induced under stress such as low temperature, nitrate limitation and wound. **Fujino et al 2011** showed that the Tam3 transposition is induced when *Anthirhinum* plants are exposed to low temperature. Even transcription of Tnt1 transposon is increased in tobacco during fungal attack (**Grandbastien et al 2005**). Studies have demonstrated that TE insertions upregulate the nearby genes (**Guio et al 2014**). They have shown that Bari-Jheh transposition upregulates Jheh1 and Jheh2 gene to confer adaptability to Drosophila during oxidative stress response. Thus, we see both the beneficial and negative effects of the TE insertions. Nevertheless, in yeast Ty elements have been shown to upregulate under several stress conditions which is discussed in a later section.

Gene	Type of TE	Disease/genetic disorder
APC	*Alu*	Colon cancer [63]
BRCA1	*Alu*	Breast cancer/ovarian cancer [48]
BRCA2	*Alu*	Breast cancer/ovarian cancer [17]
MLVI2	*Alu*	Leukemia [64]
NF1	*Alu*	Neurofibromatosis type I [65]
APC	*Alu*	Cancer of the colon [8, 20]
BAAT	LTR	Hypercholanemia [66]
MSLN	LTR	Cancer [67]
ADH1C	LTR	Role in alcoholism [68]
HSD17B1	LTR	Breast cancer [69]
FKTN	L1	Fukuyama-type congenital muscular dystrophy [70, 71]
DMD	L1	Duchenne muscular dystrophy [71]
CYBB	L1	Chronic granulomatous disease [72]
RP2	L1	X-linked retinitis pigmentosa [73]
F8	*Alu*	Hemophilia A [74]
CYBB	L1	Chronic granulomatous disease [72]
U2AF65	*Alu*	Loss of hnRNP C binding, leading to aberrant exonization [75]
OAT	*Alu*	OAT deficiency [76]
COL4A3	*Alu*	Alport syndrome [77]
GUSB	*Alu*	Sly syndrome [78]
PDHX	L1	Pyruvate dehydrogenase complex deficiency [79]
RPS6KA3	L1	Coffin-Lowry syndrome [80]

Table 1. **Human diseases caused by TE insertions**. TE insertions can be through the classical target primed reverse transcription (TPRT) process leading to insertion into the locus of genes enhancing the expression or through insertion mediated deletion at the gene loci disrupting the gene function. (Adapted from **Ayarpadikannan et al 2014**)

2. The various Ty elements in the yeast *Saccharomyces cerevisiae*

The Ty elements, also known as retrotransposons are ubiquitous elements in the yeast genome. "Ty" stands for "Transposons of yeast". With the availability of the *S.cerevisiae* genome sequence the Ty elements in the genome were mapped on all the chromosomes and gave useful information as to how these retroelements shape the genome architecture (**Goffeau et al 1996**). Generally, they possess a 330 bp long termini called the delta sequence and are present in about 100 copies in the genome. These delta sequences are present in direct orientation unlike the prokaryotic transposable elements that carry inverted repeat sequences. They are classified under the LTR retroelements and replicate in the genome via an RNA intermediate that is reverse transcribed back to its cDNA (**Boeke *et al*, 1985)**. Five classes of Ty elements are known to be present in the yeast *Saccharomyces cerevisiae* namely, Ty1, Ty2, Ty3, Ty4 and Ty5 **(Clare and Farabaugh 1985; Warmington *et al*. 1985; Hansen *et al* 1988; Stucka et al. 1992; Voytas and Boeke 1992)**. They bear homology to each other and have high nucleotide and amino acid sequence identity. These Ty elements are often compared to the retroviruses because of their life cycle and the Gag/Pol proteins encoded by their genome. The Ty genome structure is similar to retroviruses but differ in their genome size, number in the haploid genome and integration specificity. Moreover, the Ty elements do not encode the Env protein that makes the Ty cycle exclusively intracellular (**Curcio *et al,* 2015**). Each step of the replication cycle of the Ty elements are similar to retroviral life cycle. Thus, they have been extensively studied as an important model to study how retroviruses infect their hosts and how they can influence the genome organization. Two distinct groups of the LTR retroelements are found in the eukaryotic genome and are distinguished by their organization of the *POL* gene and similarities among their reverse transcriptase protein (**Xiong and Eickbush 1990**). The Ty1, Ty2, Ty4 and Ty5 elements belong to the Ty1-*copia* superfamily whereas the Ty3 falls under the Ty3-*gypsy* family. The genomes of most organisms have multiple copies of distinct *copia* and *gypsy* retrotransposons. Each retrotransposon family can vary in their copy number among individuals which in a way reflects the age of such elements over the course of genome evolution.

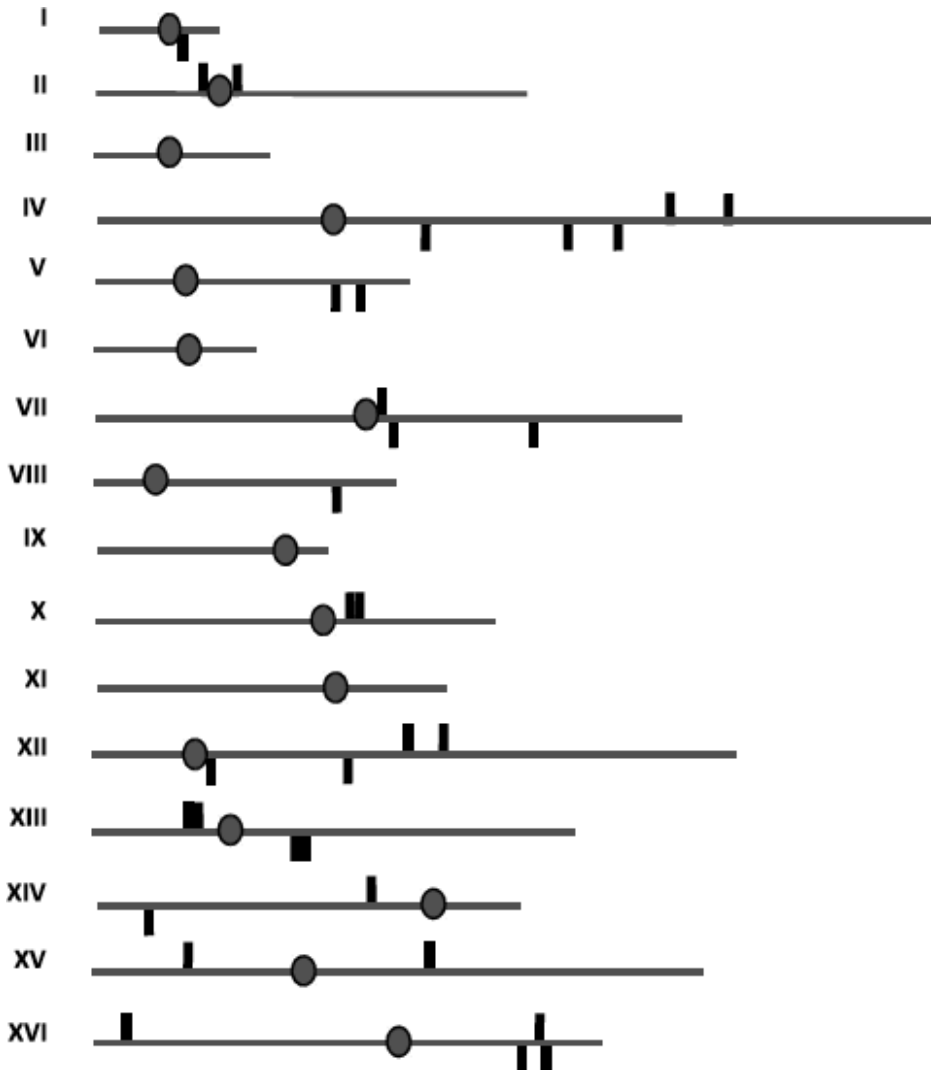


Figure 6. **Schematic diagram of the location of Ty1 element in the genome sequence of S288C strain.** Each of the sixteen chromosome is represented with the long and short arms. The presence of the Ty1 elements are shown by black rectangles. (Image from thesis **Antoine Bridier Nahmias 2015**)

A genome-wide study of retrotransposon in *Saccharomyces cerevisiae* found 331 Ty insertions including full length elements, LTR fragments and solo LTRs (217 Ty1, 34 Ty2, 41 Ty3, 32 Ty4 and 7 Ty5 elements) (**Kim *et al*, 1998).** They mapped the chromosome sequence of the *Saccharomyces cerevisiae* with Ty1-Ty5 LTR query sequences from the yeast genome database (SGD). Around 85% of all insertions were solo-LTRs and LTR fragments. The retrotransposons constituted 3.1% of the yeast genome. The Ty1 elements are most abundant. There are 32 active copies of Ty1 in the genome with 217 solo-LTRs that are generated by homologous recombination between LTRs. This shows that the retrotransposon family is dynamic, and some elements can amplify and survive whereas the others accumulate mutations and are ultimately lost.

On the other hand, nucleotide sequence diversity among the various Ty elements supports the evidence that these elements have entered the genome through a horizontal transfer process. The majority of the Ty1 LTRs except for the full length Ty1 copy, are highly divergent and have been broken by deletions. The range of sequence diversity among Ty1 and Ty5 suggests that they are present in the genome for a long time. The Ty3 and Ty4 LTRs are much more conserved with much lesser divergence suggesting that they are new insertions. The distribution of the insertions of Ty elements are represented in Table No. 2

	Number of insertions[a]					Total Ty insertions on chromosome
Chr. no.	Ty1	Ty2	Ty3	Ty4	Ty5	
I	6 (1)	1	1	0	0	8
II	13 (2)	2 (1)	1	2	0	18
III	11	2 (1)	0	1	2 (1)	16
IV	20 (5)	3 (3)	6	2	0	31
V	18 (2)	5	4	3	2	32
VI	8	1 (1)	0	1	0	10
VII	25 (3)	2 (2)	6 (1)	4	1	38
VIII	14 (1)	1	3	3 (1)	1	22
IX	5	1	2 (1)	1	0	9
X	14 (2)	4	1	3 (1)	0	22
XI	11	1	1	0	1	14
XII	18 (4)	3 (2)	4	1	0	26
XIII	15 (4)	2	0	3	0	20
XIV	5 (2)	1 (1)	4	3	0	13
XV	18 (2)	3 (2)	4	2	0	27
XVI	16 (4)	2	4	3 (1)	0	25
Genome total	217 (32)	34 (13)	41 (2)	32 (3)	7 (1)	331

Table 2: **Distribution of the Ty elements in each chromosome.** The Insertions include full length elements, solo LTRs and LTR fragments. Number of the full-length elements are shown in parentheses (Table adapted from **Kim *et al* 1998**)

The abundance of full length active Ty1 element in the yeast genome makes it a highly useful model to study how retroelements exploit the host proteome to replicate in the cells. The Ty1 RNA is one of the most abundant in the cells accounting to 0.8% of the total RNA content and 10% of polyadenylated RNA (**Curcio et al 1990**). Owing to its homology with retroviral genetic structure and replication cycle, studying how Ty1 integrates into the yeast genome will give us deeper insights as to how retroviruses infect their hosts.

2.1 Homology between the Ty elements in the yeast *Saccharomyces cerevisiae*

2.1.1 Genome structure homology

The genetic structure of the five Ty elements are similar to each other with the only difference being a different orientation of the *POL* gene in Ty3. The only difference between the Ty1 copia and Ty3 gypsy comes from phylogenetic analysis and comparison of the *POL* gene organization. The structures of all the five Ty elements are represented in following figure.

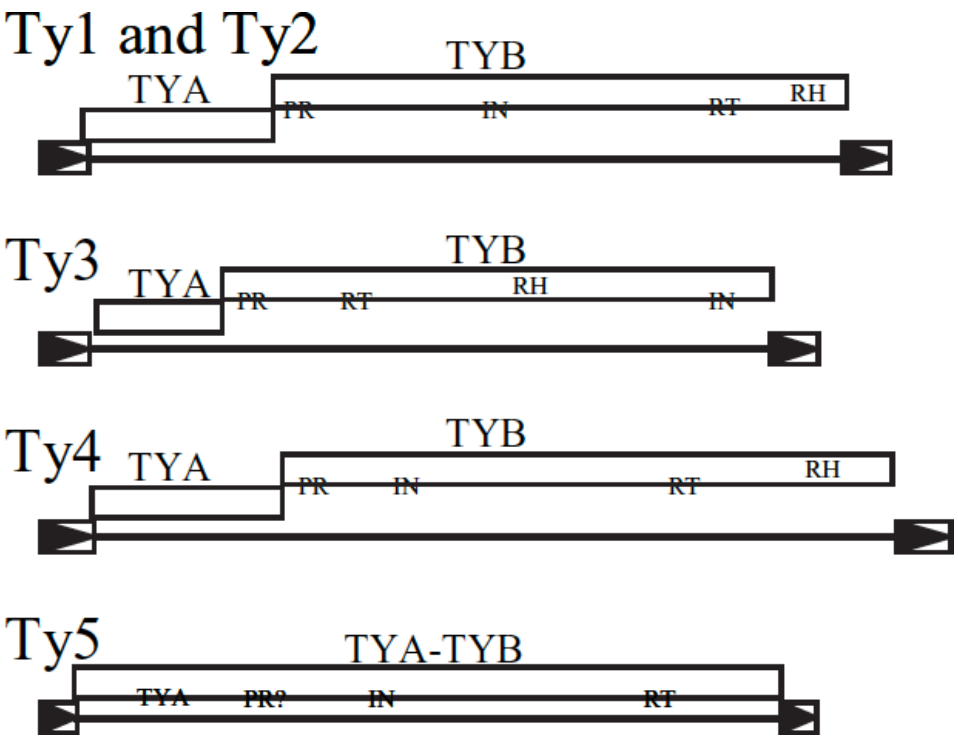


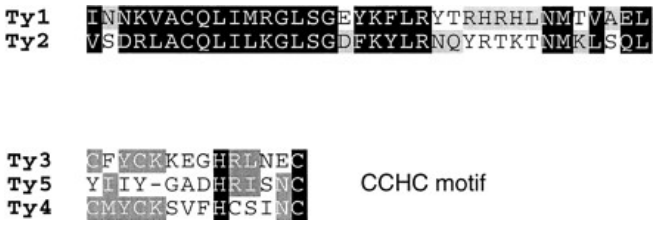
Figure 7. **Schematic diagram of the genetic structure of Ty elements.** The red rectangles mark the LTR regions. The Ty1-Ty4 have a +1 overlapping reading frames in the *TyA* and *TyB* regions. The Ty3 differs in the genomic organization of the *TyB* region. (Adapted from **Lesage and Todeschini 2005**)

2.1.2 Sequence identity among the Ty elements

In the laboratory reference strain, Ty1 and Ty2 are most closely related showing high sequence identity. The Gag region shares 40.9% amino acid sequence identity and the Pol region shares approximately 90% sequence identity. The Ty2 LTR differs by just 1 nucleotide deletion compared to the Ty1 LTR (**Kim *et al* 1998**). Phylogenetic analysis between *S. cerevisiae* strains

and S288C in the Saccharomyces Genome Resequencing Project has revealed that full length Ty1 elements, Ty2 elements and solo Ty2 LTRs in S288C have transposed recently.

In a study conducted by **Jordan and Macdonald, 1999** where they made amino acid sequence alignments between all the Ty elements in yeast demonstrated a high degree of identity and homology among the elements. They observed that the canonical nucleic acid binding motif in retroviruses CCHC motif (**Covey 1986**) was absent in the Gag region of Ty1 and Ty2 elements whereas variants of this canonical motif were present in Ty3, Ty4 and Ty5 elements though Ty3 is distantly related to Ty4 and Ty5 (Figure 7a). The Ty1 and Ty2 Gag has consensus DNA binding domain present in prokaryotes (**Clare and Farabaugh 1985**). The loss of CCHC motif in the lineage probably lead to the Ty1 and Ty2 elements.



A

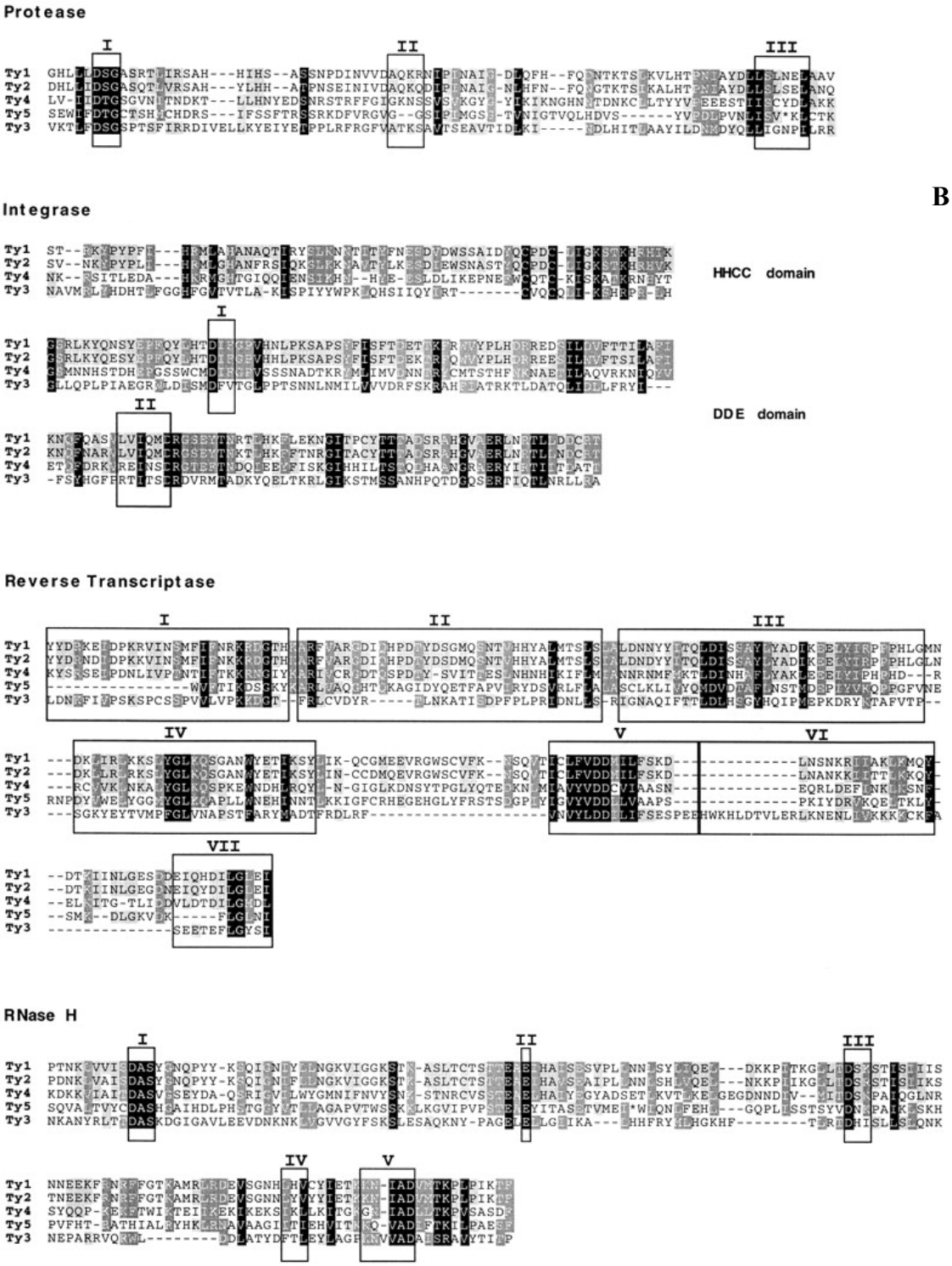


Figure 8. (A) **Sequence alignment of the nucleic acid binding domain of Gag.** The Ty1 and Ty2 have a consensus DNA binding domain found in prokaryotes and Ty3, Ty4 and Ty5 have the canonical HHCC motif (B) **Multiple amino acid sequence alignments of the four coding regions of *POL***. For the protease (PR), integrase (IN), and RNAse H (RH) alignments, boxed regions correspond to conserved motifs likely to be essential to the catalytic activity of the proteins (**McClure 1991**). For the IN alignment, the essential HHCC (Zinc finger domain) and DDE (catalytic region) regions are also indicated to the right of the alignment (Alignment adapted from **Jordan and Macdonald 1999**)

The box regions marked in Figure 7b are the conserved sites identified in previous studies required for the catalytic activity of the proteins in other homologous retroelement proteins and retroviruses (**Xiong 1990, McClure 1991, Capy 1996)**. Though there have been some differences such as the marked II region in the protease sequence which show very less conservation, they have identified regions outside the catalytic regions that are highly conserved among the Ty elements.

Secondly, they even estimated the age of these Ty elements by comparing the nucleotide identity between the 5'-LTR and 3'-LTR. As changes in the LTR are accumulated through evolution, a higher nucleotide sequence identity between either ends of the LTR denotes new insertions (**Swaby and Wickman 1997**). Most of the Ty elements appear to be recent insertions with a high percentage identity between the 5'LTR and 3' LTR. Ty1 99.68%, Ty2 99.42%, Ty3 100% and Ty4 99.55% (**Jordan and Mcdonald 1999**). Ty5 shows 91.6% identity between the two LTRs showing that these insertions are relatively old.

2.2 The Ty1 retroelement in yeast

2.2.1 The genetic structure

Ty1 is 5918 bp in length with LTRs on either sides that are 334 bp long. Ty1 LTRs have dinucleotide inverted repeats 5'-TG…CA-3' at their termini like most LTR retrotransposons and retroviruses. The LTRs are composed of three distinct domains namely, U3, R and U5. They are defined by their positions in the major sense-strand transcript transcribed from the Ty1 RNA. The U5 region which is 38 bp long and U3 region which is 240 bp long are unique to both the ends of the Ty1 RNA. The R region which is 56 nucleotides long is repeated at both ends of the processed transcripts. The functional Ty1 elements contain two overlapping open reading frames; the *GAG (TyA)* and *POL(TyB)* like the Gag-Pol region of retroviruses. The last three nucleotides of the R region of the 5' LTR encodes the first codon of the Gag protein. The *GAG* region is analogous to the retroviral structural capsid protein. There is a +1- ribosomal frame shift *POL* ORF that encodes the three catalytic enzymes required for retrotransposition namely the Protease (PR), Integrase (IN) and reverse transcriptase/RNase H (RT/RH). The last 38 bp of the *GAG* overlaps with the *GAG-POL* region (**Curcio *et al* 2015**). The Ty1 and retrovirus ALV genome structure is represented in figure 8. The *env* encodes the

envelope protein that enables the retrovirus to transport outside the cells and infect neighboring cells.

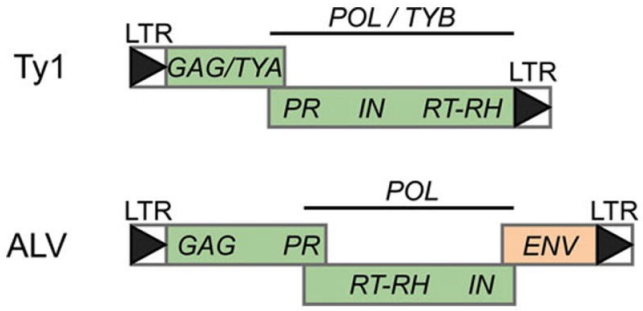


Figure 9: **Structure of the Ty1 element compared to the retrovirus, Avian leukemia virus (ALK)**. The LTRs flank the overlapping open reading frames TYA and TYB analogous to the GAG and POL region of the retroviruses. The functional domains of POL are conserved in retroviruses and LTR retroelements. They are translated as GAG-POL polyprotein which are post translationally processed to form Protease (PR), Integrase (IN) and reverse transcriptase RNAse H (RT-RH). The retroviral encode an extra env protein that makes its replication extracellular (Image adapted from **Curcio *et al,* 2015**).

2.2.2 Characterization of the Gag and Pol proteins

The Ty1 *GAG (TyA)* and *POL (TyB)* regions are translated into the functional proteins required for Ty1 retrotransposition *in vivo*. It was first reported by **Muller et al 1987** that Ty1 undergoes proteolytic cleavage leading to the formation of the virus like particles (VLP). They observed the formation of VLPs after overexpressing the Ty1 genes as the endogenous gene is hard to detect *in vivo* under normal growth conditions. Mutations in the TyB region lead to over production of the immature Gag proteins. Subsequently, **Youngren et al 1988** aimed to characterize the functional role of the Ty1 protease suggesting that it is responsible for catalyzing both the TyA an TyB regions. They observed that mutating the Ty1 protease had severe effects on Ty1 retrotransposition, cDNA synthesis, VLP assembly and formation. **Garfinkel et al 1991** using antibodies against the TyB region identified the mature Gag-Pol proteins namely the p23 protease, p60 reverse transcriptase/RNaseH and p90 Integrase from the 190 kDa Gag-Pol protein. The apparent size of the proteins is derived from their size as seen by SDS-PAGE. **Merkulov et al 1991** showed that the C-terminal domain of the Gag contains a cleavage site which is vital for VLP formation and transposition. The p49 KDa product is processed at the C-terminus to release the mature p45 Gag. We see that the protease is a key enzyme required for VLP formation and Ty1 retrotranspostion. The proteolytic cleavage also occurs in an ordered manner. The GAG/PR site is vital and needs to be processed first for the subsequent PR/IN and IN/RT processing.

It has been shown that mutational block of the GAG/PR site blocks the other processing and severely defects Ty1 retrotransposition (**Merkulov et al 2001**). The amino acid sequence where the processing occurs has also been characterized: GAG/PR (KAH/NIA), PR/IN (TIH/NVH), IN/RT (LIA/AVK). The different protein precursors and the processed products are represented in the following figure 9.

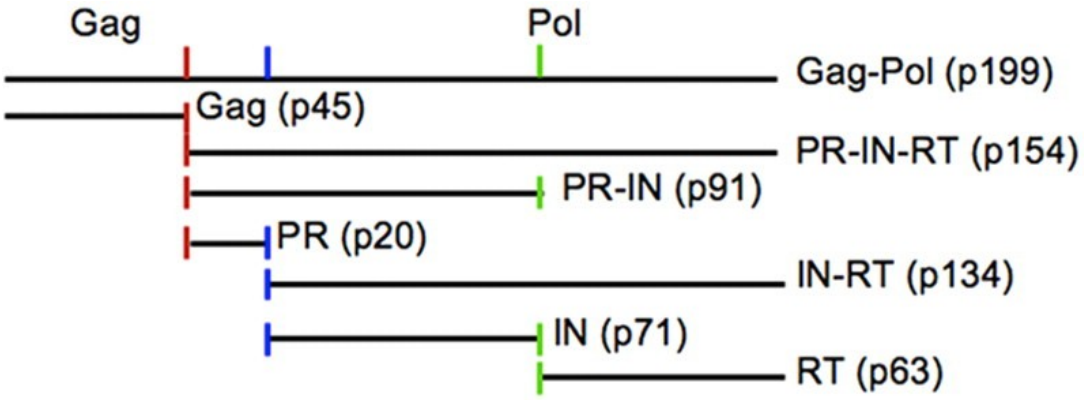


Figure 10: **Gag and Pol processing by the Ty1 Protease.** The Ty1 RNA gives two precursors, the Gag-p49 and Gag-Pol 199. The Gag-p49 is processed to give mature Gag-p45. Then the Gag-Pol is processed sequentially yielding the Protease (p20), IN (p71) and RT/RH (p63). The GAG/PR cleavage is vital for the other two processing to happen. (Image has been adapted from **Matsuda and Garfinkel 2009**)

The **Ty1 Gag** is the major structural component of the VLPs and also regulates Ty1 retrotransposition *in vivo*. It is analogous to the capsid protein of the retroviral VLP. **Cristofari et al 2000** and **Nishida et al 2015** showed that the C-terminal end of the Gag has a nucleic acid chaperon (NAC) activity *in vitro* and *in vivo*. The NAC activity is necessary to anneal the $tRNA_i^{Met}$ to the primer binding site of the Ty1 RNA for its reverse transcription. The Gag p49, Gag-Pol proteins and the Ty1 genomic RNA (gRNA) form cytoplasmic foci known as retrosomes where they assemble into VLPs (**Malagon et al 2008**). After the VLP assembly the Gag p49 are matured by cleavage at the C-terminal end yielding Gag p45. This maturation is vital for Ty1 retrotranspostion. The mature VLPs protect the gRNA from being degraded by the nuclease and the processing of the functional Gag-Pol proteins, RNA packaging and cDNA synthesis happen inside the VLPs (**Weickzorek et al 2016**).

The **Ty1 integrase (IN)** is a 71 KDa protein cleaved from the p199 Gag-Pol region and is responsible to integrate the double stranded Ty1 cDNA into specific sites of the yeast genome.

The IN forms a complex with the cDNA known as intasomes and this pre-integration complex transports into the nucleus interacting with several host factors that mediate the integration site choice. **Moore and Garfinkel 2009** showed that the Ty1 IN dimerizes *in vitro* by drawing analogy from retroviral IN. Ty1 IN might form tetramers *in vivo.* Nevertheless, this requires further investigation. The Ty1 IN forms the focus of my thesis among the other Ty1 proteins and will be explained in detail in the later sections.

The **reverse-transcriptase/RNase H (RT)** is a 63 KDa protein encoded by the C-terminal of the *POL* region and is involved in the reverse transcription of the Ty1 RNA to the double stranded cDNA. The RNase H domain degrades the Ty1 RNA from an RNA: DNA duplex during the reverse transcription process. The Ty1 RT requires a divalent cation, Mg^{2+} ions for its catalytic activity (**Bolton et al 2002**). The RT has three aspartic acid residues in its polymerase catalytic domain, out of which two (YXDD motif) are highly conserved in other eukaryotes (**Xiong et al 1990**). Mutation of the two aspartic acid residues to asparagine stops transposition and polymerase activity except for the D211N mutation ($YXDD_{211}$). Though the D211 mutant has polymerase activity *in vitro* and no transposition *in vivo (***Uzun et al 2001***),* the binding affinity towards metal ions Mg^{2+} is hugely reduced with more preference to Mn^{2+} ions and shows defect in complete polymerization of a DNA substrate *in vitro* (**Pandey et al 2004**). Few years later, **Pandey et al 2008** showed transposition defect in this mutant is due to the inability of the polymerase to bind and release pyrophosphate and translocate on the template DNA. Like retroviral RT activity, the Ty1 RT physically interacts with the Ty1 IN during the reverse transcription process. **Wilhelm et al 2005** characterized the functional role of the IN/RT interaction and showed that deleting the last 115 amino acid residues of the IN-terminal decreases the RT activity, which is consistent with the fact that expression of just the RT ORF in *E.*coli yields inactive RT unless expressed as a IN/RT fusion (**Wilhelm et al 2000**). It was also observed that deletion of IN also stops synthesis of the plus strand cDNA during the RT process. This functional IN/RT interaction also helped us (in this study) to address the insolubility of Ty1 IN when expressed at low temperature. This is discussed in the Results section.

2.3 The Ty1 replication cycle in yeast

As mentioned before, the Ty1 replication cycle is analogous to the retroviral life cycle with the exception that it is exclusively intracellular as it doesn't encode the envelope protein. The pictorial representation of the replication cycle is shown in the following figure 10.

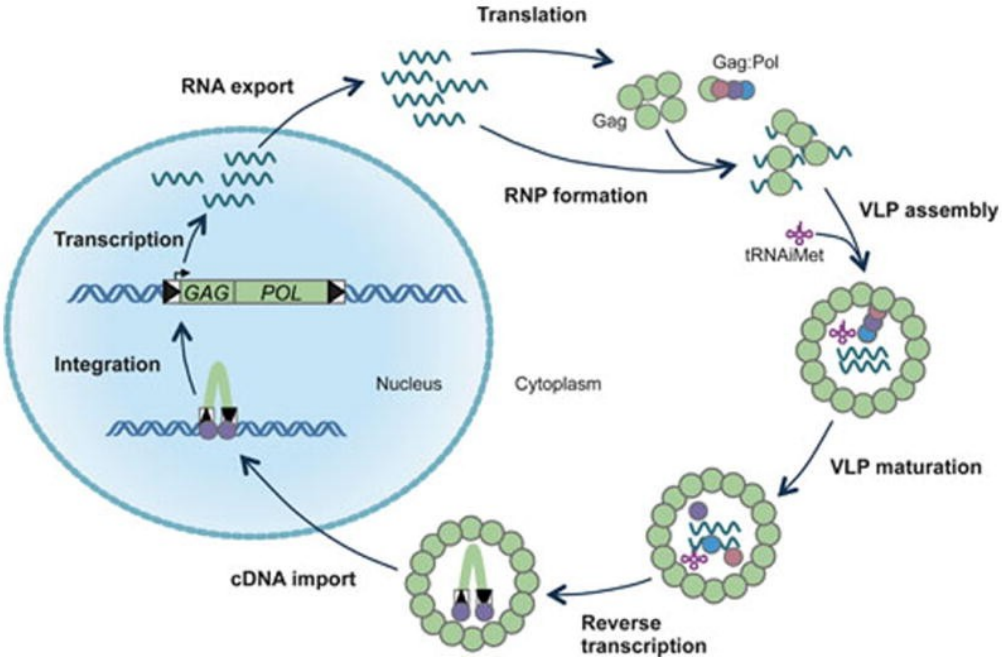


Figure 11. **Ty1 replication cycle.** The Ty1 element in the nucleus is transcribed and Ty1 RNA (wavy lines) is exported to the cytoplasm. The mRNA is translated into Gag and Gag-Pol proteins and the RNA tethers to these proteins known as Ty1 RNPs. The Ty1 RNPs form the virus like particles (VLPs) which incorporates the Ty1 mRNA, $tRNA_i^{Met}$ primer and Pol proteins. Pol proteins are cleaved by PR (maroon ball) to yield mature RT (blue ball) and IN (purple ball). The Ty RNA is reverse transcribed into a double stranded cDNA using the $tRNA_i^{Met}$ primer. The cDNA forms a complex with the IN to form the pre-integration complex, which is imported into the nucleus and IN integrates the cDNA into specific site of the genome (Image adapted from **Curcio *et al* 2015)**

2.3.1 Transcription of the Ty1 element

The nucleotide positions mentioned here are from the Ty1-H3 element, which is the most highly characterized Ty1 element in S288C strain and put into a plasmid vector (**Boeke et al 1985**). The Ty1 element is transcribed by the RNA Polymerase II machinery of the host cell. The Ty1 transcription starts precisely at the 241^{th} position in the R region of the 5' LTR end (**Mules et al 1998**) and ends at the last nucleotide of the R region at the 3' end which is then capped by a Poly-A tail (**Elder et al 1983**). The Ty1 RNA is 5640 bp in length and contains two TATA boxes upstream of the transcription site (**Fulton et al 1988**). Ty1 transcription is completely stopped by mutation of these two TATA boxes. There are two termination sequences at the 3'

U3 and R region namely, TS1 and TS2 (**Yu and Elder et al 1989)**. Presence of just one TS sequence shows partial efficiency of transcriptional termination suggesting that both the TS sequences work in an additive fashion for complete transcription stop. There are several binding sites for transcription factor in a 1 kb window covering the 5'LTR and GAG ORF. Each of these transcription factors regulates Ty1 mobility. Therefore, they will be discussed later under the section "2.3.1 Transcriptional regulation of Ty1 mobility". The Ty1 transcript is then transported into the cytoplasm for translation of the functional enzymes, VLP assembly and maturation. Figure 11 shows the pictorial representation of the transcription start site and termination nucleotide positions.

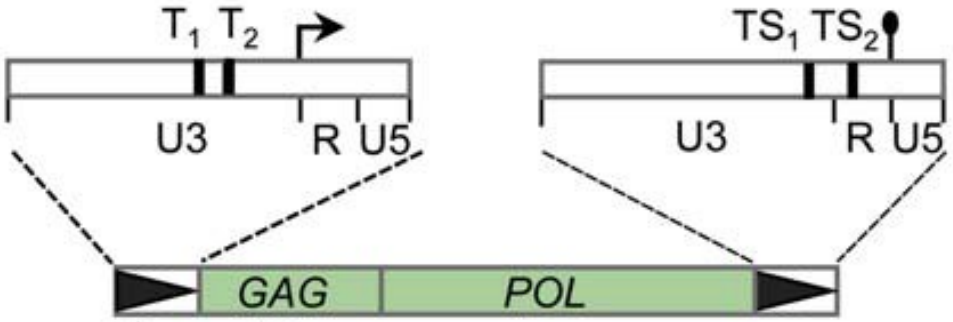


Figure 12. **Ty1 transcription start and termination sites.** The transcription begins at the 5' end of the LTR at the 241th position in the R region marked by an arrow. Termination happens at the last nucleotide of the R region in the 3' LTR at the R/U5 junction. The promoter region has two TATA boxes, T_1(159-165) and T_2 (167-173) and two termination sequences TS_1 (5,776-5,781) and TS_2 (5,837- 5,842). (Image adapted from **Curcio et al 2015**)

2.3.2 VLP assembly and maturation

After the transcription process, the Ty1 mRNA is exported to the cytoplasm for translation and VLP assembly. The VLP assembly and maturation is a critical step in the replication cycle because it protects the Ty1 gRNA from nuclease attack and concentrates all the proteins required for cDNA synthesis. The VLPs are formed by the assembly of the Gag and the Gag-Pol precursors which is then required for RNA packaging. Several Gag segments have been shown to be important in VLP assembly such as amino acid residues 41 to 62, 114 to 147, 223 to 287, 330 to 346 (**Roth 2000**). The N-terminus of the Gag is protruded out of the surface and the C-terminus of both the forms of Gag are embedded in the core. This configuration keeps the RNA binding C-terminus domain inside the capsid cell. The Ty1 VLPs are 14 MDa in size and are composed of a spikey shell of 30-80 nm diameter. The maturation of the Gag p49 to

p45 takes place after the VLP assembly because VLPs formed under the presence or absence of Ty1 PR mediated processing of Gag have the same average size (**Curcio et al 2015**). Following the packaging of the Ty1gRNA, the Gag Pol precursors are processed in an ordered manner by the Ty1 PR. The Gag-Pol precursors and the processing sites have been discussed earlier.

2.3.3 Reverse transcription of Ty1 genomic RNA to cDNA

The reverse transcription of the Ty1 genomic RNA (gRNA) to the cDNA is carried out by the reverse transcriptase (RT) inside the mature VLP. **Boeke et al 1985** first reported that Ty1 transposes through an RNA intermediate. The reverse transcription steps are analogous to the reverse transcription of the retroviral RNA in VLPs. The downstream of 5' LTR contains a primer binding site (PBS) where the initiator primer, $tRNA_i^{Met}$ binds to initiate the reverse transcription process. The DNA extends upto the U5 and R region of LTR. This DNA product linked to the $tRNA_i^{Met}$ is known as the minus strand strong stop cDNA. Once the RT reaches the 5' end of the Ty1 mRNA, the Ty1 RNA portion from the RNA: DNA duplex is degraded by the RNase H domain of the RT. Following this degradation, the minus strand strong stop cDNA moves to the 3' end and hybridizes to the R region. Using this minus strand strong stop cDNA as a primer, a minus strand cDNA until the U5 and R region is synthesized. Following this synthesis, the Ty1 RNA is degraded leaving behind two polypurine tracts (PPT). The PPTs are used as primers to synthesize the plus strand strong stop cDNA until the 5' end of the Ty1 minus strand cDNA. Now, this plus strand strong stop cDNA moves to the 3' end of the minus cDNA strand and hybridizes to the U5 and R regions. This as a primer extends to form the complete plus strand. Lastly, the minus strand cDNA extends to form the U3 region using the plus strand strong stop cDNA as a template. The steps of the reverse transcription process are shown in Figure 12.

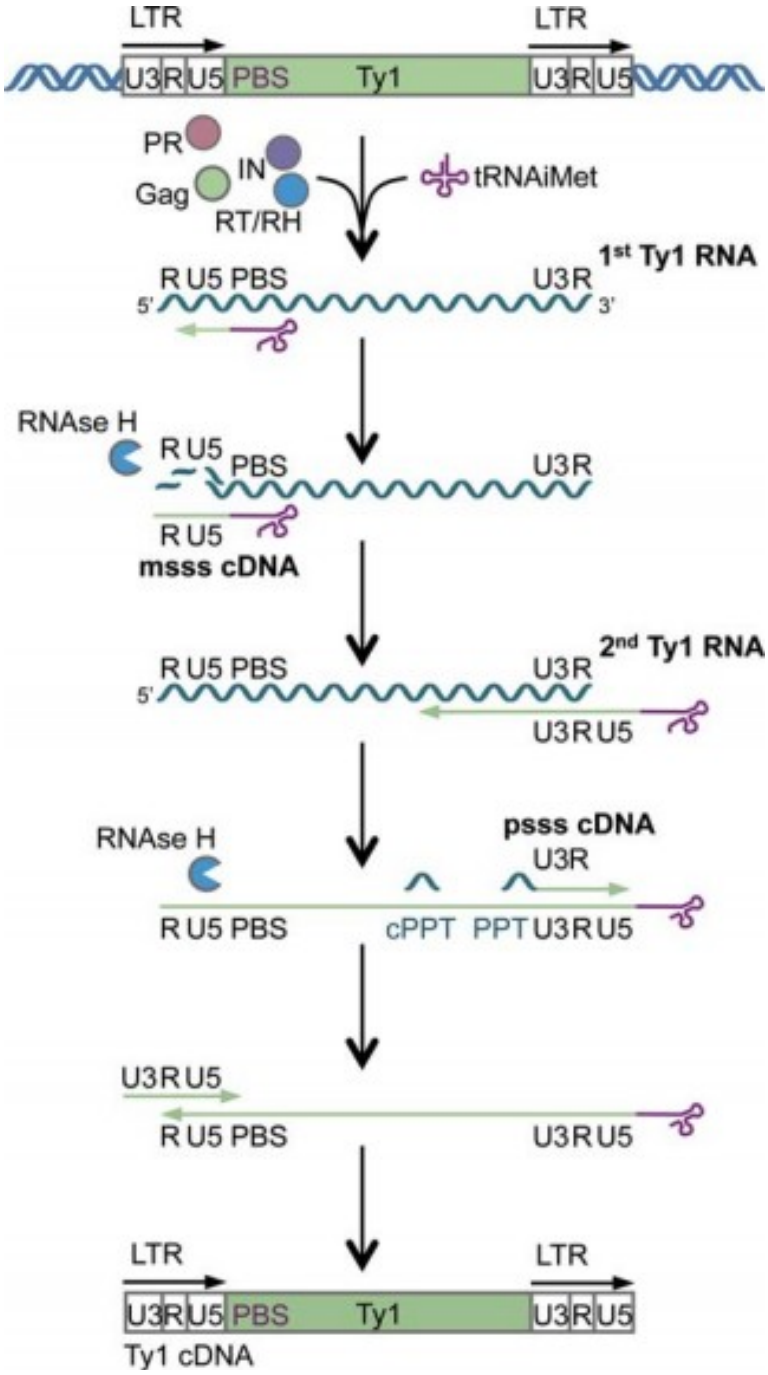


Figure 13. **The reverse transcription steps of the Ty1 mRNA.** Each of the steps are explained in the description above (Adapted from **Curcio et al 2015**).

2.3.4 cDNA integration in the host genome

The final step of the replication step is integration of the double stranded cDNA copy into the host genome by the Ty1 IN. The Ty1 IN forms a complex with the cDNA known as intasomes and travels into the nucleus to gain access into the genome. During this time the pre-integration

complex is formed where multimeric Ty1 IN interacts with several host factors to mediate the integration site specificity. The inverted dinucleotides at the end of the cDNA are absolutely necessary for efficient integration. During the integration process, the Ty1 IN initiates a nucleophilic reaction between the 3'OH of the cDNA and the target DNA site. This is a transesterification reaction known as the strand transfer step. At the end of the integration process, there remains a 5-nucleotide break on each strand of the target DNA with exposed 3'-OH groups. Repair of the DNA breaks leads to a 5 bp target site duplication (TSD). Generally, the Ty1 IN is not involved in the DNA repair indicating that other host factors that might be involved in this repair system.

2.4 Regulation of the Ty1 retrotransposition *in vivo*

Though the Ty1 mRNA accounts to almost 1% of the total mRNA in the cells, the retrotransposition event is an extremely rare event (**Curcio et al 2015**). There have been studies suggesting that a high rate of retrotransposition can cause random insertional mutations which might be detrimental to the cells. Thus, this indicates that the cells have a strong mechanism to regulate the Ty1 element expression in order to avoid the deleterious effect it may cause if it is left rampant. There are strong evidences showing that the active copy number in the cells are a driving factor to the regulatory mechanism. Higher the copy number, stringent is the regulatory process to inhibit retrotransposition. There are several other environmental stress conditions such as nitrogen starvation, ionizing radiation and acute adenine starvation that activate Ty1 transposition by stimulating the transcription of the Ty1 element.

2.4.1 Transcriptional regulation

The Ty1 transcription has been shown to be controlled by at least nine transcription factors namely Gcr1, Ste12, Tec1, Mcm1, Tea1/Ibf1, Rap1, Gcn4, Mot3, and Tye7. The transcription regulatory sequence is laid out in a 1 kb window encompassing the 5'LTR and partially the *GAG ORF*. Unlike budding yeast promoters which are compact and have few regulatory sequences upstream of TATA box, the Ty1 promoter is relatively larger mimicking the promoters in higher eukaryotes.

Morillon et al 2000 showed that that the transcription factors **Ste12** and **Tec1** work in coordination to activate Ty1 transcription through the Kss1 pathway in diploid cells. During

nutrient (nitrogen starvation) limitation, the Kss1-mitogen activated protein kinase (MAPK) pathway is activated leading to invasive filamentous growth enabling the cells to look for nutrients away from the colonizing site. Ty1 promotor region has a filamentation and invasive responsive elements (FRE) site downstream of the TATA box and Tec1/Ste12 complex has been shown to bind to this site *in vitro* (**Baur et al 1997**). Deletion of either Ste12 or Tec1 severely affected Ty1 mRNA levels in haploid cells suggesting that they are required for Ty1 transcription in haploid cells but only required in diploid cells under nutrient limitation. Amongst all the transcription factor, **Mot3** has been identified as a repressor of Ty1 transcription. Deletion of *MOT3* showed a slight increase in Ty1 expression (**Madison et al 1998**). **Servant et al 2012** provides several evidences, showing that **Tye7** activates Ty1 transcription under adenyclic nucleotide stress. They showed that Ty1 contains three Tye7 binding sites in the *GAG ORF* and Tye7 is able to activate Ty1 transcription by dysregulating the Ty1 anti-sense RNA (Ty1 AS) expression, which represses Ty1 retrotranspostion post-transcriptionally. Deletion of Tye7 has been shown to lower Ty1 expression in adenine deprived cells. **Gcr1,** an activator of glycolytic enzymes, recognizes an upstream activator sequence (UAS) in the 5' LTR region of Ty1 and activates Ty1 transcription in the presence of glycerol and lactate (**Turkel et al 1997, Lopez et al 2000). Errede et al 1993** reported that **Mcm1**, which normally binds to mating genes also binds to the regulatory sequences on Ty1 promoter and positively affects Ty1 transcription in a different manner than mating genes. Generally, Mcm1 interacts with Ste12 or alpha1/alpha2 repressor proteins at mating gene promoters but binding of Mcm1 to Ty1 target site is independent of these proteins. **Gray and Fassler 1993** showed that Rap1p binds to the downstream enhancer region of the Ty1 *GAG ORF* which contains a consensus Rap1 binding site, very close to the Mcm1 binding site. Though Rap1 mediated Ty1 activation is not dependent on Mcm1 binding or vice versa, maximum activation of Ty1 transcription requires a complex of three proteins, Rap1, Mcm1 and an unknown factor (known as Ibf1 initially). Subsequently, **Gray and Fassler 1996** isolated and characterized the gene *TEA1*, showing that it binds to short palindromic sequence between the Mcm1 and Rap1 binding site. *TEA1* deletion has been shown to decrease Ty1 mRNA levels by 25-30% and a 2-fold rise when *TEA1* is over expressed. **Morillon et al 2002** showed a variation over a range of 50 folds in the expression of all the 31 individual Ty1 element *in vivo*. They accounted the high expression of several (8 out of 11) Ty1 elements due to the presence of five potential Gcn4 binding sites in the 5' LTR region. Gcn4 is a transcriptional activator of amino acid biosynthesis genes and generally binds to multiple sites on its target genes. Overexpression of Gcn4 consequently increased Ty1 mRNA levels of high expressing Ty1 elements but didn't activate expression of Ty1 elements with a lower number of Gcn4 binding sites.

Gcn4 production is increased under amino acid starvation. Therefore, Ty1 mRNA levels were tested under histidine starvation. Indeed, there was a 3-fold decrease in Ty1 mRNA levels under histidine starvation in mutant *gcn4* cells suggesting the role of Gcn4 mediated Ty1 transcription under amino acid limitations. Figure 13 shows the pictorial representation of the transcription regulatory binding sites for different transcription factors.

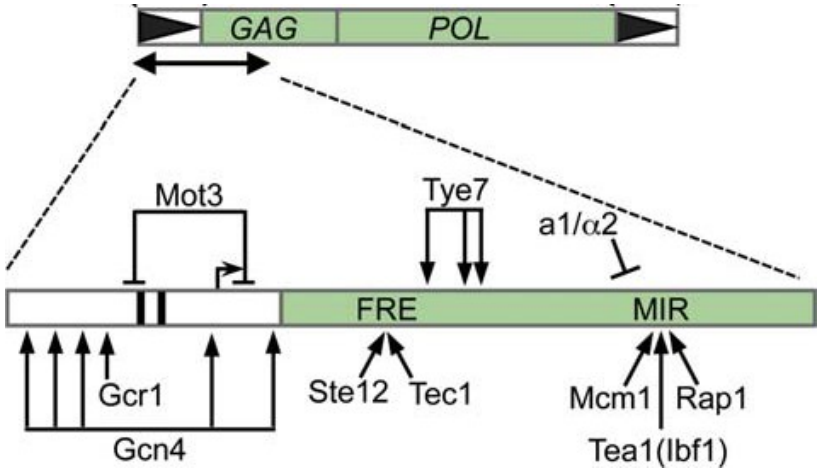


Figure 14. **Ty1 transcription regulatory sequence and binding sites.** The Ty1 promoter extends over 1 kb including the 5′ LTR and part of the *GAG* ORF. The positions of the Ty1 activator binding sites are: Gcn4 (five binding sites: 12 to 17, 79 to 84, 98 to 103, 155 to 160, and 318 to 323), Gcr1 (115 to 119), Ste12 (395 to 401), Tec1 (418 to 422), Tye7 (three binding sites: 463 to 468, 661 to 666, and 727 to 732), Mcm1 (833 to 848), Tea1/Ibf1 (884 to 899), and Rap1 (911 to 923). The filamentous response element (FRE) comprises Ste12 and Tec1 binding sites, MIR comprises Mcm1, Tea1/Ibf1, and Rap1 binding sites. The positions of the Ty1 repressor binding sites are: Mot3 (higher affinity site *in vitro* at positions 147 to 150) and a1/α2 (832 to 863) (Image adapted from **Curcio et al 2015**).

There are at least three other chromatin factors that regulate Ty1 transcription activity, namely SAGA complex, Swi/Snf and ISW1. **Happel et al 1991** showed that the *SNF2, SNF5* and *SNF6* play an important role in activating Ty1 transcription. Initially, these genes were identified as activators of *SUC2* and other glucose repressible genes. They observed that the defect in *SUC2* expression in the three *SNF* mutants were suppressed by mutations in the *SPT6* gene. Since mutations in SPT genes (suppressors of Ty) were identified to suppress Ty transposition in a genetic screen (**Winston et al 1984**), Happel et al decided to examine the role of *SNF* mutants on Ty1 transcription. *SNF* mutants completely abolished Ty1 transcription *in vivo*. The SAGA complex consists of Spt-Ada-Gcn5 proteins which function as a histone acetyltransferase and acetylates histones bound to chromatin.

The Spt3, Spt7, Spt8 and Spt20 are part of the SAGA complex and Ty1 transcription is completely abolished in the absence of these genes (**Grant et al 1997**). The Ada-Gcn5 complex can also activate Ty1 transcription independently similar to that observed in *SNF* mutants (**Pollard et al 1997**). The ISWI is a chromatin remodeling ATPase and is made of two subunits (ISW1 and ISW2). The complex has been shown to repress Ty1 transcription in addition to other elements such as *INO1* and *PHO3* (**Kent et al 2001**). Thus, the involvement of chromatin factors in Ty1 transcription suggests that Ty1 transcription prefers specific chromatin conformation.

2.4.2 Transcriptional co-suppression

The laboratory strain does not contain more than 32 active Ty1 copies. The cells have adapted a regulatory mechanism to control the copy number (CNC) of retroelements and protect themselves from insertional mutagenesis. This regulatory mechanism is proportional to the increase in the copy number of Ty1. One such regulatory mechanism is the transcriptional co-suppression, a homology dependent process which was reported by **Jiang 2002** using a strain containing the Ty1-URA3 reporter (controlled by the native Ty1 promoter) construct and following its expression by selecting and counter-selecting the strain on Ura- and 5-FOA (forms a toxic compound with Uracil) plates respectively. The *URA3* ORF was randomly inserted into a cloned Ty1 element in *E. coli*. Then the Ty1-*URA3* fusion gene was introduced into a Ura- yeast strain by homologous recombination with the endogenous Ty1 element. Cells expressing the Ty1-*URA3* in frame was selected on Ura- plates. If co-suppression exists, then the Ty1-*URA3* expression will be abolished or downregulated in certain cell populations enabling the cells to be 5-FOA resistant. Indeed, they found many colonies which were both Ura+ and 5-FOA resistant suggesting that there exists a co-suppression regulation. Expression of the Ty1 genes existed in either of the two states, complete abolishment of expression or complete expression within a subset of cell population. This was described as a rapid switch between Ura- and Ura+ states. As this regulation requires the presence of the native promoter, it is happening at the transcription level. Another important characteristic of this regulation is the presence of the high copy number of Ty1 elements. This regulation is not observed in a strain containing no active Ty1 copy.

Beretta et al 2008 shows that there is a Ty1 cryptic unstable transcript (CUT), an antisense RNA encompassing the Ty1 5'LTR region that represses Ty1 mobility. The transcriptional co-suppression may be mediated by these CUTs. Xrn1 is a 5'-3' exonuclease that destabilizes the Ty1 CUT and maintains the Ty1 expression. When Xrn1 is deleted there is a 10-fold accumulation of an antisense Ty1RNA (also known as Ty1 ASRNA) mapping to the 5' end of the Ty1 RNA and a 5-fold decrease in the Ty1 full length RNA. They also observed accumulation of a sense oriented strand which is a truncated version of the full length Ty1 RNA in the absence of *XRN1*. Using the Ty1 *His3AI* mobility screen (method explained in the later sections) they even observed a 500-fold reduction in Ty1 retrotranspostion in the absence of *XRN1*. To show how the Ty1 ASRNA directly regulates Ty1 mobility they overexpressed the Ty1 ASRNA using a galactose inducible promoter and observed a decrease in Ty1*URA3* reporter expression *in trans* and mobility compared to the wild type strain. Though there are evidences from much recent publications that Ty1 ASRNA also acts posttranscriptionally to regulate Ty1 mobility.

2.4.3 Posttranscriptional co-suppression

Garfinkel et al 2003 reports a new post transcriptional form of CNC in yeast different from what was proposed by **Jiang 2002**. With the increasing number of the Ty1 elements, there is accumulation of a Gag derived polypeptide encoded by the C-terminal half of the Ty1 GAG which is a potential inhibitor of Ty1 cDNA synthesis and retrotransposition with no effect on the RNA levels. Therefore, this mechanism described is a post-transcriptional process. When they tested for transposition of a genomically tagged Ty1-His3AI in a "Ty-less" *S. paradoxus* strain (containing no Ty1 element) they saw a 22-fold increase in mobility compared to a laboratory strain which normally contains 32 copies of Ty1. Subsequently, with the introduction of more Ty1 elements there was a 4800-fold range decrease of Ty1-His3AI mobility in a copy dependent manner. This suggests that Ty1 regulation is sensitive to copy number and happens irrespective of the fact if the suppressing Ty1 elements is genetically marked or not. They mapped a segment between nucleotides 238-1702 in Ty1 that confers the post transcriptional co-suppression. Deletion of nucleotide 238-281 or 1600-1702 completely stops the repression activity of this segment. There exists a significant difference of this mechanism compared to the transcriptional co-suppression. A native Ty1 promoter is required for the latter regulation whereas transcriptionally silent galactose induced (pGTy1) Ty1 confers post transcriptional CNC. But, when over induced with galactose, the CNC mechanism has been shown to be

overridden by the cells (**Curcio et al 1992**) . Due to a high homology between Ty1 and Ty2, a high copy number of Ty2 and not Ty3 or Ty5 confers CNC on Ty1. **Saha et al 2015** describes a Gag restriction factor encoded by an internal Ty1 transcript that disrupts VLP assembly and inhibits transposition as copy number increases. **Saha et al 2015** observed the presence of a 4.9 kb Ty1 transcript, a shorter transcript than the 5.7 kb Ty1 transcript by northern blotting in different yeast strains (Δ*spt3*, Δ*xrn1,* By4742, Ty-less *S.paradoxus* strain) which was also observed by **Beretta et al 2008** in a Δ*xrn1* background. There are two closely placed AUG1 and AUG2 in the Gag region that marks the transcription start site of Ty1i, and encodes the p22/p18 protein. One or both start codons can be used for the transcription of the restriction factor. Ribosome profiling shows more preference for AUG1 for transcription initiation than AUG2. This p22/p18 form of Gag interfering with the assembly of functional VLP is similar to that of Gag-like restriction factors found in mammals (**Sanz-Ramos et al 2003**). p18 was not visible in the contrary to p22 in the absence of *spt3,* where transcription of the full length Ty1 mRNA is abolished and so is the production of protease. This suggests that the PR processes the p22 to form the p18 product. When the two AUG codons were mutated with the alanine codon, there was an increase in the retrotransposition frequency suggesting the potential role of the p22/p18 Gag restriction factor as a negative regulator of Ty1 retrotransposition. Results also suggest that both the p22 and p18 form can function as negative regulators (**Saha et al 2015**). Processing and stability of Ty1 proteins were observed when Ty1 his3AI and p22 Gag was co-expressed. There was accumulation of more PR-IN polyprotein and less mature IN. VLP morphology also changes upon co-expression with the p22 Gag. When viewed under an electron microscope, VLPs appear open or incomplete suggesting that VLPs were not formed correctly. **Nishida et al 2015** shows that the Gag p45 physical interacts with the p22 Gag restriction factor and blocks its nucleic acid chaperon activity that hampers RNA encapsidation into the VLPs. **Blaszczyk et al 2017** provides evidence that the 5' UTR of the Ty1i mRNA is important in maintaining its translational efficiency to the p22 product. The secondary and tertiary structure of the 5'UTR regulates translation by regulating the recruitment, positioning and movement of the ribosomes. They have identified two high affinity Gag binding sites in the Ty1i RNA and mutating either of the sites decrease the p22 level and destabilizes the Ty1iRNA. The Ty1i RNA secondary structure shows an interaction between the 5'UTR and the coding sequence where the initiation codons are present. Even a slight disruption of this secondary structure can severely defect p22 translation.

In reference to what was explained before about how the Ty1 ASRNA regulates Ty1 retromobility on a transcription level by abolishing Ty1 expression, there have been evidence of the role of Ty1 ASRNA interfering with Ty1 retrotransposition posttranslationally (**Matsuda et al 2009**). Ty1ASRNAs are enriched in the VLPs and negatively regulates Ty1 mobility by reducing integrase level, reverse transcriptase level and cDNA synthesis. The model for CNC proposed by them is that the Ty1ASRNA after packaging in the VLPs base-pairs with the Ty1 mRNA and directly inhibits Pol protein processing. They observed an accumulation of the PR-IN polyprotein in CNC^{+} cells (containing 37 Ty1 elements), suggesting that processing is hampered. It destabilizes mature IN and prevents annealing of the initiator tRNA methione to the primer binding site of the Ty1 mRNA thus blocking the reverse transcription process. **Conte et al 1997** showed that the *FUS3* gene, encoding a mitogen activated protein kinase (MAPK) Fus3 is activated in response to pheromones and negatively regulates Ty1 retromobility post translationally. They indicate that in the absence of *FUS3* the Ty1 proteins in the VLPs are stabilized showing elevated levels of cDNA and Ty1 retromobility.

2.4.4 Environmental stress conditions and genotoxic agents

There are several environmental stress conditions and DNA damaging reagents that elevate Ty1 retrotransposition. It has been already described in the previous section how nitrogen starvation leads to an increase in Ty1 retrotransposition by activating specific transcription activators Ste12/Tec1 via the Kss1 pathway. Acute adenine starvation has been shown to elevate Ty1 mobility by an activation in the Ty1 transcription level. Preferentially, the low expressed endogenous Ty1 levels are highly expressed under adenine starvation (**Todeschini et al 2005**). The Ty1mRNA levels where high in cells devoid of the *Bas1* transcription factor that is required for the activation of *ADE* genes when extracellular adenine is absent. This result suggest that Bas1 may represses Ty1 transcription when Adenine is not available. To show that Ty1 transcription is elevated under acute adenine conditions and is independent of Bas1, they deleted the de novo AMP biosynthesis pathway and grew the cells under very limited conditions of Adenine and observed an even higher Ty1 mRNA levels than in the Δ*bas1* background. This increase in the Ty transcription is correlated with a higher cDNA level and Ty1 mobility. Other DNA damaging conditions such as exposure of cells to UV irradiation caused high Ty1 transcription (**Rolfe et al 1985**). **Bradshaw et al 1989** observed that cells treated with 4NQO (4-Nitroquinoline-1-oxide) or UV cells showed high levels of Ty1 RNA and a high rate of transposition approximately 17-folds more compared to the control untreated cells when they

tested for two target loci in the genome. **Staleva and Venkov 2001** proposed mechanisms involved in the elevation of Ty1 transposition on treating cells with 4NQO or MMS (methyl methane sulphonate). Different signaling pathways might be involved during DNA damage response that activates Ty1 transcription and transposition. Using a mutant to arrest the G1 phase of the cell cycle, they observed a 40-fold increase in Ty1 transposition compared to when G2 arrested cells. This may suggest the dependence of G1 on the induction of Ty1 transposition. Another protein, Rad9 which is checkpoint protein required for a G1 cycle arrest shows a 2-fold increase in Ty1 transposition compared to the 5-folds increase in Ty1 transposition in wild type cells with or without the treatment of 4NQO, suggesting that there may be other proteins along with Rad9, such as Rad24 involved in the signal transduction pathway to increase Ty1 mobility. Gamma rays has been shown to activate Ty1 transposition (**Sacerdot et al 2005**). The rate of transposition rises in a dose dependent manner and higher levels of Ty1 RNA and cDNA were observed. There was no effect on the Gag protein. In cells lacking *Ste12,* where there is no Ty1 expression, gamma irradiation was able to stimulate Ty1 transposition which indicates that there exists a separate regulatory pathway independent of Ste12. The mechanism behind higher levels of Ty1mRNA and Ty1 transcription under gamma radiation is not elucidated yet but **Sacerdot et al 2005** mentions two hypothesis that might explain such scenario. First, the response to gamma irradiation activates the Mec1 dependent environment stress response (ESR). This ESR is controlled by Mec1 through a transcriptional activator Msn4 which binds to stress response elements (SRE) on its target genes to activate them. Ty1 has a consensus SRE site in the promoter region downstream of the TATA box. This might explain the transcription activation of Ty1 by Msn4 during stress by gamma irradiation. Another hypothesis could be that gamma irradiation changes the chromatin structure at the Ty1 promoter region by repressing the expression of both *HTA1/2* and *HTB1/2* (**Gasch et al 2001**). Thus, an aberration in the histones could activate Ty1 transcription. Altogether, this data suggests that ionizing radiation activates Ty1 transposition both on a transcription and post-transcriptional level. The mechanism is complex and yet to be studied. Another important factor that affects Ty1 transposition is temperature. **Pacquin et al 1984** showed that there was a 100-fold increase in Ty1 transposition at temperature lower than 30°C by an assay based on the changes in expression of the Aldehyde dehydrogenase gene (*ADH2*) on insertion of the Ty1 gene upstream of it. They grew cells at four different temperatures (15°C, 20°C, 30°C and 37°C) and observed that the highest rates of transposition were at 15°C and 20°C and completely abolished at 37°C. **Lawler et al 2002** showed that the defect in Ty1 transposition at higher temperatures (>30°C) is due to the inability of the Ty1 PR to process the Gag-Pol precursors. Higher temperatures are

marked with accumulation of the unprocessed Gag-Pol precursors and low cDNA levels indicating a low RT activity in the VLPs. Transposition is maximal at temperatures ranging between 20°C – 25°C but can also be detected at a lower frequency at 30°C. They observed at higher temperature, the Gag-PR fusion which is essential for further Gag-Pol processing is hindered whereas Gag-p49 processing still occurs to some extent. Using monoclonal antibody raised against Ty1 IN, visualization of mature IN was undetectable in cells grown at 37°C compared to 22°C. Ty1 PR is temperature sensitive and loses its protease activity at temperatures higher than 30°C. This suggests the inhibition of the Gag-Pol processing at high temperature as a primary reason behind the inhibition of Ty1 retrotransposition. A more recent study identified two genes *RFX1* and *SML1*, when deleted led to higher Ty1 transposition even at higher temperature. The translated proteins from these two genes are negative regulators of the enzyme, ribonucleotide reductase responsible for regulating the dNTP level in cells. Though no improvement in Gag-Pol processing or higher cDNA levels were observed in these mutants at high temperature, a high homologous recombination efficiency was detected. This suggests that the higher Ty1 transposition was on account of homologous recombination of the already existing Ty1 cDNA with other endogenous Ty1 elements and not because of a Ty1 IN mediated integration (**O'Donnell et al 2010**).

2.5 The Ty1 host partners that affect its mobility

The Ty1 retroelement exploits the host factors for its replication *in vivo*. Each step of the replication cycle requires the presence of several host factors. Several studies have identified genes that influence negatively or positively Ty1 retromobility. The timeline of the genetic studies are as follows; **Scholes et al 2001, Griffith et al 2003, Nyswaner et al 2008** and **Risler et al 2012**. These studies were imperative in indicating several genes that positively or negatively influence Ty1 retromobility but none provide any functional role of the identified partners in the replication cycle.

As the retrotransposition frequency in laboratory strains are extremely low (approximately 10^{-7}/element/generation), assays have been developed were an induction of Ty1 mobility is achieved. **Boeke et al 1985** first isolated a functional Ty1-H3 and cloned it into a plasmid. He expressed the element under the control of a galactose inducible promoter where overexpression of the Ty1 genes led to high retrotransposition frequency. Precise excision of an intron inserted into the Ty1 element formed the rationale behind developing retrotranscript

indicator genes (RIG). RIGs are constructed by putting an artificial intron into the ORF of the selectable marker in a non-spliceable orientation. The RIG is inserted into the Ty1 element at the 3' untranslated region such that the Ty1 and the intron inserted marker gene are in opposite transcriptional orientation. This will enable the intron to be spliced from the Ty1 transcript. After the intron is spliced from the Ty1 transcript, reverse transcription and cDNA integration of the element in the chromosome renders a functional copy of the marker gene. Cells expressing the selectable marker correspond to a functional Ty1 transposition event (**Curcio and Garfinkel 1991**). The pictorial representation of the mobility assay is showed in Figure 15.

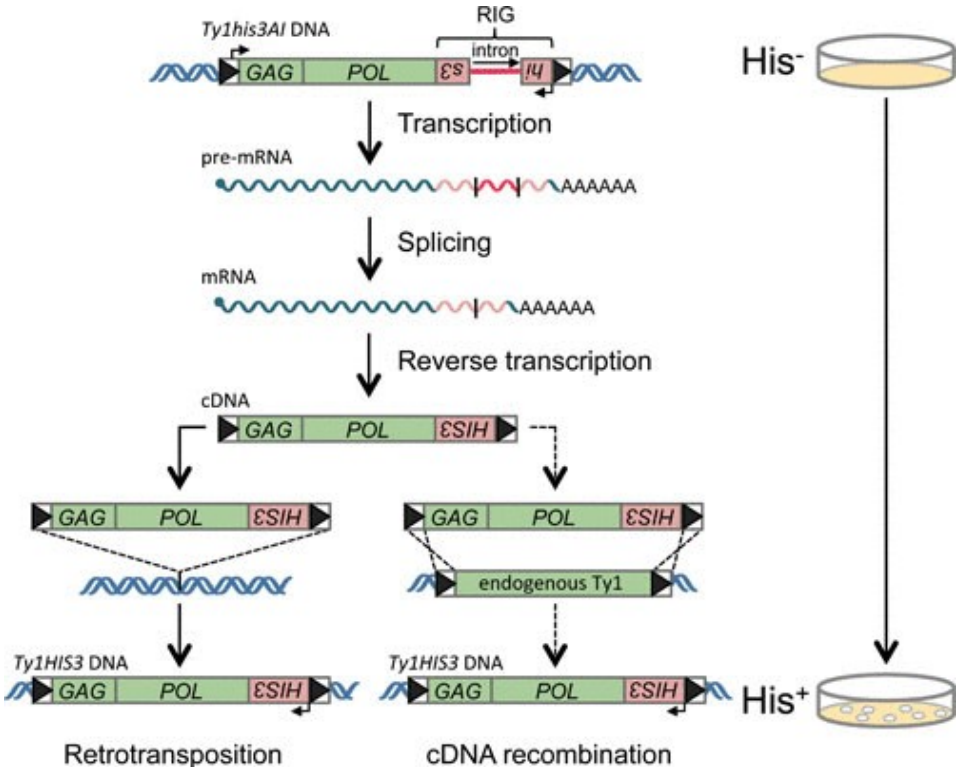


Figure 15. **Retro mobility assay using the RIG**. A chromosomal Ty1 element tagged with the *his3AI* RIG undergoes precise splicing leading to the expression of a functional *HIS3* gene after cDNA integration into the host chromosome. The dashed lines represent the low frequency of Ty1 cDNA recombination in wild-type cells. Cells that sustain a Ty1*HIS3* retromobility event give rise to His^+ colonies (The Image is adapted from **Curcio et al 2015**).

Scholes et al 2001 conducted the first genetic screen using a chromosomally marked Ty1 His3AI expressed from its own native promoter identified 29 negative regulators (*RTT* genes) of the Ty1 mobility. Deletion of the *RTT* genes enhanced the level of Ty1 transposition in a range from 5-211 folds but didn't affect Ty1 mRNA levels. Although they showed high cDNA

level, suggesting that these genes regulate Ty1 replication on a post transcriptional level. Though the retromobility frequency of the chromosomal Ty1His3AI element was detected, it was far too low to be able to identify a large number of mutants for reduced mobility. The list of genes identified by Scholes et al 2001 are provided in the following table.

Gene name	ORF	Gene product function
		Marginal regulators
HSP78	*YDR258C*	ATP-dependent protease, mitochondrial protein chaperone
MCM6	*YGL201C*	Component of hexameric helicase complex involved in DNA replication
MLP2	*YIL149C*	Nuclear envelope protein that interacts with Yku70
RFX1	*YLR176C*	Repressor of DNA damage-inducible genes
RNR1	*YER070W*	Large subunit of ribonucleotide reductase
RTT 102	*YGR275W*	Unknown, not similar to any protein of known function
TIF4632	*YGL049C*	One of two eIF4G homologues in yeast
VAC8	*YEL013W*	Vacuolar protein involved in protein targeting
		Regulators
EST2	*YLR318W*	Catalytic subunit of telomerase
NUT2	*YPR168W*	Component of Pol II transcription mediator complex
RAD50	*YNL250W*	See *MRE11*; previously identified as regulator of Ty1 transposition
RAD5 7	*YDR004W*	RecA homologue stimulating strand-exchange activity of Rad51 during homologous recombination; previousl identified as regulator of Ty1 transposition
RRD2	*YPL152W*	Encodes putative phosphotyrosyl phosphatase activator
RRM3	*YHR031C*	Superfamily I DNA helicase required for replication fork progression in rDNA
RTT101	*YJL047C*	Cullin, putative component of the Skpl-Cullin-F-box complex (SCF) family of E3 ubiquitin ligases
RTT103	*YDR289C*	Unknown, not similar to any protein of known function
RTT105	*YER104W*	Unknown, not similar to any protein of known function
RTT106	*YNL206W*	Unknown, similarity to DNA structure-specific recognition protein (SSRPs)
RTT107	*YHR154W*	Unknown, BRCT-domain protein family
RTT108	*YPR164W*	Unknown, mutant sensitive to diepoxybutane and mitomycin C
RTT109	*YLL002W*	Unknown, mutant sensitive to diepoxybutane and mytomycin C
RTT110	*YOR144C*	Inhibits direct repeat recombination
SAE2	*YGL175C*	Regulator of the activity of the MRX complex
SCH9	*YHR205W*	Kinase in stress response and nutrient-sensing signaling pathway
SGS1	*YMR190C*	RecQ-family helicase that suppresses recombination; previously identified as regulator of Ty1 transposition
TEL1	*YBL088C*	Protein kinase required for telomere length regulation and TM checkpoint pathway

On the other hand, **Griffith et al 2003** exploited the retromobility assay using a plasmid borne Ty1 over-expressed using a Gal promoter. This led to a high induction of retrotransposition frequency (1000-folds more) which made it possible for them to identify a large pool of genes (102 genes) that regulated Ty1 mobility. 99 genes were activators and 2 genes were repressors of Ty1 mobility. Over-expression of Ty1 genes have been shown to over-ride the regulation mechanism of the cells. Of the 102 identified genes, 46 altered the cDNA levels whereas the rest 56 didn't have any effect on cDNA levels. This data suggests that half of the identified genes affected Ty1 transposition on a post transcriptional level until cDNA synthesis whereas the other half affected Ty1 transposition post cDNA production. The identified genes are provided in the table below.

Group/no. of genes	Genes deleted
Chromatin (10)	*ARD1, NAT1, SAP30, SIN1 (SPT2), SIN3, SIN4, SPT4, SPT10, SPT21, STB5*
DNA repair (4)	*APN1, MMS22, RAD52, XRS2*
Miscellaneous (27)	*APG17, APL5, BEM1, BUD6, CHO2, CYK3, DCC1, ERV14, FYV3, HOF1 (CYK2), JNM1, KCS1, KRE24, MAD2, MFT1, PAT1, NUM1, SCP160, SDF1, SEC22, SEC65, SMI1, SWA2, TPM1, TPS2, VPH1, VPS9*
Nuclear transport (2)	*NUP84, NUP133*
Protein folding/modification (8)	*CPR7, DBF2, DOA4, MCK1, NAT3, PFD1, SSE1, TCI1*
Ribosomes/translation (9)	*DBP3, RPL6A, RPL14A, RPL16B, RPL19B, RPL20B, RPL21B, RPP1A, RPS10A*
RNA metabolism (8)	*CBC2, DBR1, LEA1, LSM1, NOP12, RIT1, STO1 (CBC1), YDL033c*
Transcription (10)	*CTK1, DEP1, HAC1, PHO23, POP2, RPA49, RTF1, SRB8, SSN2, SUB1*
Transcription elongation (7)	*ELP2, ELP3, ELP4, ELP6, IKI3 (ELP1), KTI12, THP2*
Unknown (16)	*YBR077c, YDL115c, YDR496c, YFL032w, YGL250w, YGR064w, YKL053c-A, YLR052w, YLR322w, YML010c-B, YNL226w, YNL228w, YNL295w, YOL159c, YOR292c, YPL080c*

Table 5. **Genes identified by Scholes et al 2001 and Griffith et al 2003 as regulators of Ty1 mobility.**

Nyswaner et al 2008 identified 91 genes that repressed Ty1 mobility by estimating the retromobility frequency of the Ty1His3AI element expressed on a plasmid controlled by its native promoter in yeast deletant mutants. Out of the identified genes, 80% are involved in maintaining chromatin structure, DNA repair and transcription. They have characterized the role of 33 genes in Ty1 mobility and demonstrated that only 5 mutants showed higher Ty1 mRNA levels showing that the rest regulated Ty1 transposition on a posttranscriptional level. Out of the rest 58 genes, 20 genes were identified in previous Ty1 genetic screen and the rest

38 genes need to be characterized. The most recent genetic screen was conducted by **Risler et al 2012** using a novel approach called synthetic genetic array (SGA). *rtt1* or *med1* deletant strains harboring a chromosomal Ty1-His3AI element was mated with 4847 gene deleted haploid strains. *rtt1* and *med1* mutations leads to a hypertranposition phenotype. Retrotransposition frequencies were measured in the progeny and 275 genes that lead to repression of the hypertransposition phenotype was identified. Out of the identified 275 retrotransposition host factor (*RHF*) genes, 45 *RHF* genes were already found to influence either Ty1 or Ty3 mobility (**Griffith et al 2003**, **Aye et al 2004, Irwin et al 2005**). The cDNA levels showed a mild change in 181 single mutants compared to the wild type, suggesting that the factors activate Ty1 retromobility post cDNA synthesis. Nearly, 43 genes involved in ribosome biogenesis, RNA degradation, modification and transport when deleted stimulated Ty1 cDNA levels. In fact, previously **Dakshinamurthy et al 2010** identified several genes that affect Gag processing *in vivo*. They showed that *Bud22,* an enzyme involved in ribosome biogenesis is required for efficient ribosome +1 frameshifting. Deletion of *bud22* showed high defects in Gag processing and hindered Ty1 mobility. The lower Ty1 mobility is correlated to the low Ty1 cDNA level in the *bud22* mutant. Due to the massive number of identified genes in **Nyswaner et al 2008** and **Risler et al 2012,** the data will be provided in the supplementary data. Though the genetic screens identify novel host factors affecting Ty1 replication, recent studies are more focused on delineating a functional role of the host factors on Ty1 mobility. **Salinero et al 2018** showed that the Ty1 mobility is either positively or negatively regulated in the absence of non-essential subunits of the mediator complex. Mediator is a transcriptional co-activator made of 25 subunits, containing head, middle and tail regions. The tail domain is involved in association with the transcriptional activator proteins and the other two regions are involved in the association of the RNA Polymerase II and pre-integration complex formation. To determine whether the rise in Ty1 mobility was due to an increase in the Ty1 mRNA, northern blotting results showed that there was no statistical difference in the mRNA levels in the head, middle or tail mutants compared to the WT suggesting that mediator regulates Ty1 by a post transcriptional process. They observed a 600-fold decrease in Ty1 mobility in the tail module triad gene, *med3* deletion mutant. They demonstrated that this decrease in the Ty1 mobility is due to an increase of the Ty1i expression and accumulation of the p22 Gag restriction factor in the *med3* deletion mutant. This suggests that the mediator regulates Ty1 mobility by maintaining the ratio between the full length Ty1 mRNA and the truncated Ty1i RNA.

3. The Ty1 retroelement integration

The Ty1 integration is not random in the genome. Several studies provide evidence that Ty1 is present on specific areas in the genome, specifically at the 1 kb window upstream of the Pol III transcribed genes, principally the 275 tRNA genes (**Ji et al 2003, Devine and Boeke 1993**). These areas are referred to as gene poor regions or safe havens and is thought to be an adaptive mechanism how cells avoid the deleterious effects of the Ty1 retroelement. However, integration is a very random event accounting to 10^{-7} integrations/per element/generation. This makes studying Ty1 integration a difficult process as it is hardly detected under normal growth conditions. To address this issue, a galactose inducible vector was used to clone a functional Ty1 element and study its replication following its over-expression in cells (**Boeke et al 1985**). Recent studies on high throughput screening have provided advanced information regarding the Ty1 preference of nucleosomal surfaces in this 1 kb window region upstream of tRNA genes (**Mularoni et al 2012, Baller et al 2012**).

3.1 The gene structure and characterization of the Ty1 Integrase domains

The Ty1 IN has three domains conserved like other LTR retrotransposons and retroviruses. The N-terminal of the IN contains a highly conserved zinc binding domain (ZBD) HHCC motif. **Moore and Garkinkel 2009** characterized the N-terminal ZBD by analyzing the effects of mutating each of the HHCC and thirteen residues in the sequence between H22 and C55 (X_{32}). They observed a 4000-folds decrease in Ty1 retrotransposition on mutating the zinc coordination amino acid motif and decrease in Ty1 IN and RT stability. The proteolytic processing was also affected. Mutations of the hydrophobic residues (L28, L32, I37 and V45) in the intergenic X_{32} reduced retrotransposition frequency by 85- 688 folds but didn't have any effect in the stability of IN, RT or on proteolytic processing. Instead, they showed reduced IN-IN interaction. This suggests that the HHCC motif is critical for transposition and the hydrophobic residues are required for multimerization of IN.

The catalytic core contains a highly conserved region among eukaryotic and prokaryotic integrases and transposases. It is a DDE motif and binds divalent cations Mg^{2+} ions required for the transesterification process during cDNA integration in the host genome. When a linker amino acid is inserted between the DE motif, it completely abolishes Ty1 retrotransposition

(**Eichinger and Boeke 1988**). Mutating the DDE motif with substitution by alanine has been shown to inhibit transposition. Conservation of the catalytic domain ends approximately 120 amino acids downstream of the DDE motif. The C-terminal region is less conserved and consists of a GKGY motif which is also found in other members of the Pseudoviridae family. (**Peterson and Voytas 2002)**. The extreme end of the C-terminal region contains a nuclear localization signal (NLS) which enables the Ty1 IN to enter the nucleus and complete the integration process. It consists of two basic KKR region separated by 29 amino acids. **Moore et al 1998** and **Kenna et al 1998** identified this NLS and specific mutations (mutations in the basic region) abolishing the inability of the IN to direct into the nucleus abolishes Ty1 transposition whereas no effect in the VLP assembly, RT activity and PR processing is observed. **McLane et al 2008** showed that the Ty1 IN-NLS physically interacts with the α-importin receptor and importin α. Yeast deletant mutants for Ran-GTPase, importin α and importin β that impair the classical nuclear import pathway severely effects Ty1 transposition *in vivo.* This suggests that the Ty1 IN exploits the classical nuclear import machinery to enter the nucleus and is critical for Ty1 retrotransposition. Recent studies have shown that the Ty1 IN interacts with the components of the Nuclear pore complex (NPC) for mobility which will be discussed in the later section "Ty1 host partners". The Ty1 Integrase domains are represented in figure 14.

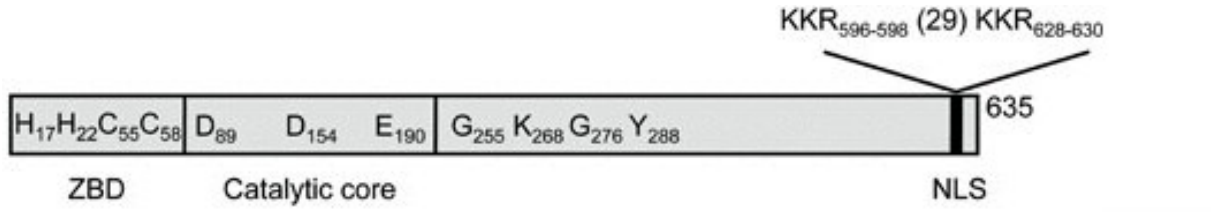


Figure 16. **Domains of the Ty1 Integrase**. Amino acid residues, HHCC motif in the zinc binding motif (ZBD) and the catalytic domain, DDE motif that are conserved in the *Pseudoviridiae* family of long terminal repeat-retrotransposons are shown. Identical clusters of basic residues (KKR) that define the bipartite nuclear localization signal (NLS) separated by 29 amino acids are also showed with black arrows.

3.2 The tethering model of integration

Initially the Ty1 cDNA integration process was explained by the integration model which involves the interaction between the cDNA LTR and the Ty1 IN to mediate integration. With recent studies, the integration process is explained as a tethering model, where specific host

proteins tether with the Ty1 IN to mediate integration at specific sites in the genome. Though the cDNA integration process requires the presence of different host factors *in vivo, in vitro* models of integration using DNA substrates and recombinant IN have been developed (**Braitermann and Boeke 1999**). The "tethering model" was proposed in the 1990s (**Sandmeyer et al 1990**) though specific host factors interacting with the integrase of the LTR retroelements were not identified. Within these several years, tethering factors were identified for many related retroviruses and retroelements.

Studies on Ty3 and Ty5 gave the first implications that Ty integrases tether with host factors near the targeting site for integration. Ty5 integration *in vivo* into the heterochromatin at the silent mating loci requires the interaction between the Ty5 integrase targeting domain and the Sir4p (**Xie et al 2001**). The targeting sequence in the Ty5 IN has been identified as a six amino acid sequence in the C-terminal domain. The Ty5 interaction with the Sir4p requires phosphorylation of a serine residue (S1095) in the targeting sequence. Mutations in the targeting sequence of the Ty5 integrase abolishes Ty5 integration. However, **Brady et al 2007** observed that Esc1, a protein associated with the inner nuclear membrane interacted with the same domain of Sir4 as Ty5 IN and 75% of mutations that disrupted the IN-Sir4 interaction also disrupted Sir4-Esc1 interaction. A small motif is present in Esc1 that interacts with Sir4. To demonstrate a functional role of this motif, they swapped it with the Sir4 -Ty5 IN binding domain and observed that the chimeric Ty5 IN-Esc1 was able to target at the heterochromatin suggesting a molecular mimicry as a mechanism of retroelements to interact with cellular partners.

Likewise, for Ty3 integration site preference to the initiation site of Pol III transcribed genes is due to the interaction between Ty3 integrase and the TFIIIB complex (Brf1 and TBP), demonstrated by *in vitro* approaches (**Kircher et al 1995, Qi X et al 2012**). The Ty3 integration is point specific and integrates 1 to 2 nucleotides at the tRNA transcription site. **Yieh et al 2000** developed an i*n vitro* assay to study Ty3 integration in the context of the role of TFIIIB independent of TFIIIC. They found that the TFIIIB can bound to the TATA box upstream of U6 gene through interaction with TBP independent of TFIIIC. They observed that Brf1-TBP is enough to direct Ty3 integration specifically at the U6 gene promoter. However, physical interaction between the Tfc1 subunit of TFIIIC and Ty3 integrase has been observed by **Aye et al 2001** which influences the Ty3 integration target site at Pol III promoters.

Related retroviruses such as the HIV Integrase has been shown to interact with the cellular lens epithelium-derived growth factor (LEDGF/p75) via the C-terminal region to integrate at active genes transcribed by the Pol II machinery (**Cherepanov et al 2003, Cherepanov et al 2005**). LEDGF is subjected to alternative splicing that produces the p75 and p52 isoforms. Both the forms have chromatin binding ability except for the p75 form which contains the integrase binding domain and regulates HIV integration. There are still some unanswered questions as to how the p75 LEDGF isoform identifies the active transcription units and whether it couples HIV integration with Pol II transcription machinery (**Singh et al 2015**). Similar to Ty5 IN, another host factor JPO2 has been found to interact with LEDGF/p75 and mutations that disrupt LEDGF/p75 -HIV IN interaction also disrupt the LEDGF/p75-JPO2 interaction (**Maertens et al 2006**) suggesting that like retroelements, retroviruses may also tether integration complexes at the target sites by mimicking host factors. However, this needs to be demonstrated further. Another retrovirus, Murine Leukemia virus (MLV) integrates at transcription sites of Pol II genes. **Sharma et al 2013** identified that bromodomain and extraterminal domain (BET) proteins Brd2, Brd3 and Brd4 as cellular binding partners of the MLV integrase. Recombinant Brd4 interacts strongly with MLV integrase *in vitro* and stimulates concerted MLV integration. ChIP seq data also revealed that BET proteins binding sites and MLV integration sites are positively correlated.

However, Ty1 integration at Pol III genes were identified nearly two decades ago but the molecular mechanism behind its integration choice was unknow. Recently, **Bridier-Nahmias et al 2015** demonstrated the Ty1 IN interacts with the AC40 subunit of the Pol III complex to direct Ty1 integration in a 1 kb window upstream of the Pol III transcribed genes. However, developing a loss of this interaction mutant by swapping the AC40 of *S. cerevisiae* with AC40 of *S. pombe* revealed that integration site changed to the subtelomeric regions from the known tRNA genes with no change in the rate of transposition. This data suggests the importance of the Pol III complex as the tethering factor of the Ty1 IN.

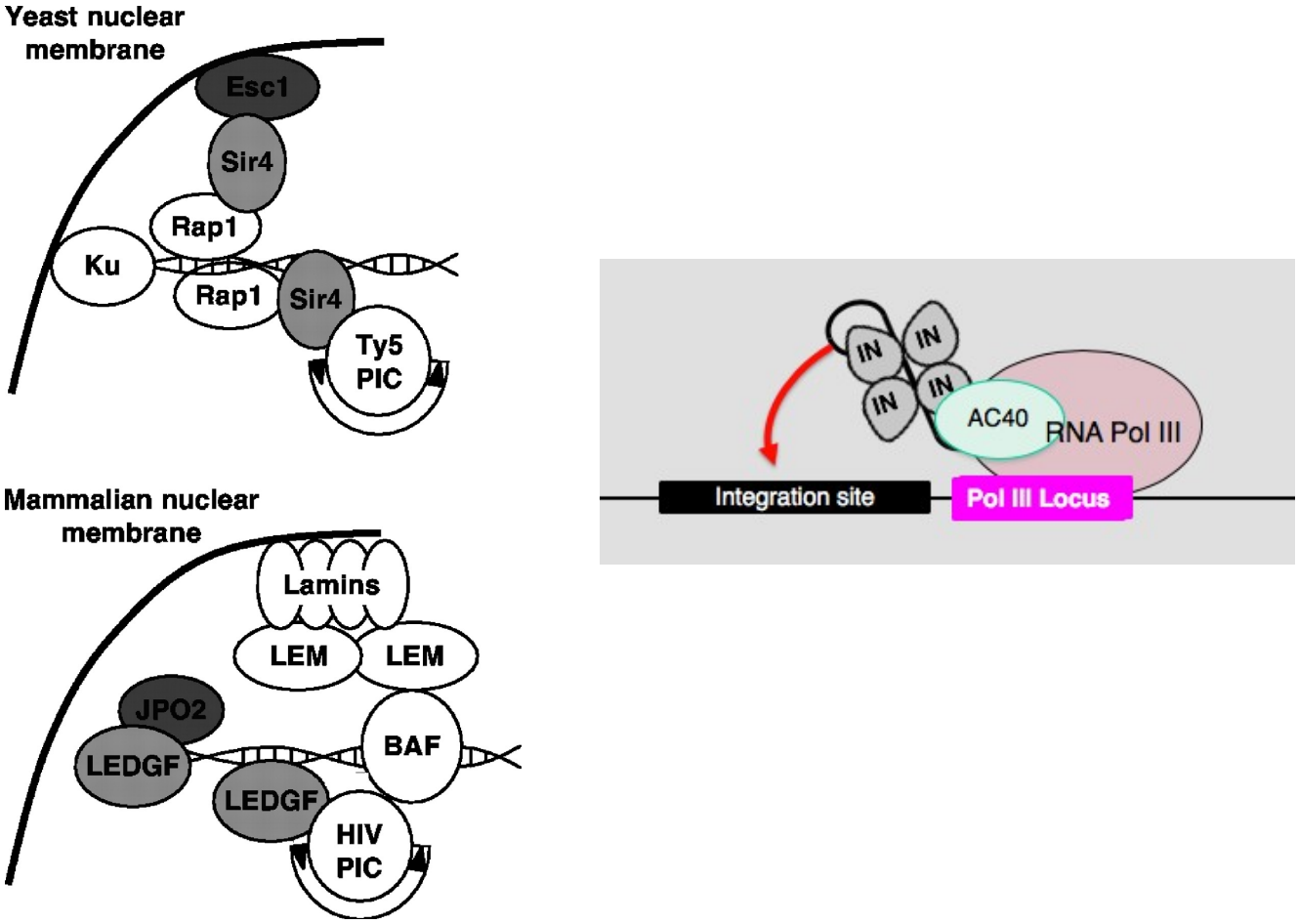


Figure 17. **The tethering model**. The tethering model for Ty5 IN and HIV IN are indicated on the left panel. Ty5 IN interacts with the Sir4p to integrate at the heterochromatin and Esc1 interacts with Sir4 similar to Ty5 IN. The HIV IN interacts with LEDGF/p75 to integrate at the Pol II gene promoters. The LEDGF/p75 also interacts with JPO2 (Adpated from **Brady et al 2007**). The tethering model of Ty1 IN with AC40 subunit of Pol III is indicated in the right panel.

4. The Ty1 Integration specificity

High-throughput screens have shown that integration site choice for all Ty elements are upstream of RNA Pol III transcribed genes except for the Ty5 which integrates into the heterochromatin of the silent mating loci. In this section, I will describe the factors affecting the integration selection site choice for Ty1 retroelements. Retroelements that integrate upstream of RNA Pol III transcribed genes are represented in the following table 3.

Mobile element	Clade	Host[a]	Preferred integration site
LTR Retrotransposons			
Ty1	Ty1-*copia*	*Saccharomyces cerevisiae*	~1 kb window upstream of RNA Pol III-transcribed genes, including tRNA and 5S rRNA genes [13, 14, 21]
Ty2	Ty1-*copia*	*Saccharomyces cerevisiae*	~1 kb window upstream of RNA Pol III-transcribed genes [21]
Ty3	Ty3-*gypsy*	*Saccharomyces cerevisiae*	1–4 bp upstream of tRNA TSS[b] [93]
Ty3p	Ty3-*gypsy*	*Saccharomyces paradoxus*	~6 bp upstream of tRNA TSS [110]
Ty4	Ty1-*copia*	*Saccharomyces cerevisiae*	~1 kb window upstream of RNA Pol III-transcribed genes [21]
Tj1	Ty3-*gypsy*	*Schizosaccharomyces japonicus*	1–10 bp upstream of tRNA TSS [121]
Beta/ Tca8	Ty3-*gypsy*	*Candida albicans*	6-30 bp upstream of tRNA MCS[c] [119]
Tyl3	Ty3-*gypsy*	*Yarrowia lipolytica*	~5 bp upstream of tRNA TSS [116]
Ylt1	Ty3-*gypsy*	*Yarrowia lipolytica*	~5 bp upstream of tRNA TSS [117]
Tyl6	Ty3-*gypsy*	*Yarrowia lipolytica*	~5 bp upstream of tRNA TSS [115]
DGLT-A	Ty3-*gypsy*	*Dictyostelium discoideum*	13–33 bp upstream of tRNA MCS [125]
Skipper-2	Ty3-*gypsy*	*Dictyostelium discoideum*	8–23 bp downstream of tRNA gene [124, 172]
Non-LTR Retrotransposons			
TRE5	L1	*Dictyostelium discoideum*	40–54 bp upstream of tRNA MCS; 37–41 bp upstream of extrachromosomal 5S rRNA genes [125, 134, 135, 174]
TRE3	L1	*Dictyostelium discoideum*	40–150 bp downstream tRNA genes [137, 175]
NLTR-A	L1	*Dictyostelium purpureum*	2–6 bp upstream of tRNA MCS [124]
NLTR-B	L1	*Polysphondylium pallidum*	39–64 bp upstream of tRNA MCS [124]

Table 6. **Retroelements integrating upstream of Pol III transcribed genes.** The Ty 1-Ty4 elements from the budding yeast all integrate upstream of Pol III genes. TSS: tRNA transcription site, MCS: tRNA mature coding sequence (Table adapted from **Cheung et al 2018**)

4.1 Factors affecting Ty1 integration at the RNA Pol III transcribed genes

The *in vitro* Ty1 insertion assay using purified VLPs as source of IN and DNA substrates, show that targeting is random suggesting that *in vivo* several host factors are required for targeting Ty1 mostly at Pol III genes (tRNAs). Genomic copies of Ty1 associated with tRNA genes were observed decades back in 1979 (**Cameron et al 1979).** Approximately, 1 kb region upstream of the 5' promoter region of tRNA genes and the glycine tRNA gene, *SUF16* was demonstrated as a Ty1 integration hotspot (**Ji et al 1993**). For Ty1, Ty2 and Ty4 integration patterns are similar and is region specific upstream of tRNAs. Whereas Ty3 is point specific and integrates at the Pol III transcription start site. Therefore, to understand this interplay between Ty integration and Pol III transcribed genes, it is imperative to delineate the role of Pol III transcription machinery, chromatin structure at Pol III genes or other host factors present on Pol III genes in Ty1 retromobility and integration site selection.

4.1.1 Chromatin remodeling factors

The chromatin structure near the tRNAs have an open structure which are maintained by chromatin remodelers (**Kumar et al 2013**) and might be easily accessible for the Ty1 pre-integration complex to integrate its cDNA. Genome wide mapping studies have provide evidences that Ty1 elements prefer nucleosomal surfaces at the H2A/H2B interface as their sites of integration (**Mularoni et al 2012**, **Baller et al 2012**). These two studies found 90% of the Ty1 insertions were upstream of tRNAs but all the genes were not equally targeted probably because Pol III transcription machinery was not recruited at those genes. The importance of the role of chromatin in Ty1 integration was first indicated by **Bachman et al 2005** when they showed that the N-terminal of Bdp1 subunit of the TFIIIB is involved in recruiting an ATP dependent chromatin remodeler, Isw2 at tRNA genes which is required for the periodic integration of Ty1, with a periodicity of 80 bp upstream of tRNAs. Loss of Isw2 or Bdp1 leads to inhibition of the Ty1 integration periodicity suggesting that Isw2 is required for maintaining a specific nucleosomal structure that favours Ty1 integration periodicity. But, Ty1 is targeted at tRNA genes even in the absence of Bdp1 and Isw2 indicating that TFIIIB is not a host factor of Ty1 IN. By comparing the deep sequencing data from the two publications, Mularoni 2012 and Baller 2012, with genome wide nucleosomal positioning data sets, they have identified two Ty1 integration hotspots on each nucleosome with an interval of 70 bp on the first three nucleosomes upstream of Pol III transcribed genes. Esp1 is a cysteine protease in yeast known

as separase and is involved in the separation of sister chromatids during mitosis. It was demonstrated that Esp1 physically interacted with the Ty1 IN in vivo. During the metaphase of the cell division, this interaction is enriched (**Ho et al 2015**). A yeast mutant with mutation in *Esp1* that reduces its cleavage activity shows reduced Ty1 retromobility and reduced Ty1 insertion upstream of the *SUF16* tRNA gene. On the other hand, mutation in a cohesion protein (Scc1), that is involved in holding sister chromatids together during cell division, showed a higher Ty1 mobility and integration upstream of the *SUF16* tRNA gene. This indicates that a removal of a cohesin probably makes the access of the Ty1 pre-integration complex to the nucleosomes easy. However, the whole model is more complex and the role of Esp1 in Ty1 integration may be just one of the many mechanisms that exist.

4.1.2 Modifications in the histones

Histone modifying enzymes have been shown to impact Ty1 targeting on several instances. Hos2 and Set3, member of the Set3 histone deacetylases complex, are required for Ty1 integration upstream of tRNAs (**Mou et al 2006**). Hos2 has been found to physically interact with tRNA genes by chromatin immunoprecipitation which probably hints to its role in modifying chromatin at Pol III genes. Deletant mutants of Hos2 and Set3 have shown reduced endogenous Ty1 mobility upstream of *SUF16* with no defect in Ty1 mRNA level or Ty1 cDNA level. Deletion of another histone deacetylase, Rpd3 also reduced Ty1 mobility and integration upstream tRNAs (**Nyswaner et al 2008**). Paf1 complex interacts with the chromatin and is a transcriptional elongation factor and has been shown to monoubiquitylate histone H2B by the Bre1-Rad6 ubiquitin ligase. Mutations leading to the disruption of the Paf complex leads to enhanced Ty1 mobility and insertion at tRNA. Genome wide mapping showed that yeast strains harboring mutant Rad6 showed enhanced Ty1 integration in coding regions compared to the wild type strain (**Baller et al 2012**). This indicates that Paf1 mediated histone modification by Bre1-Rad6 represses Ty1 mobility. Genetic screens made by **Scholes et al 2001** identified various negative regulators of Ty1 (*Rtt* genes). Rtt109, a histone acetyltransferase and Rtt106, histone chaperone has been identified as repressors of Ty1 retormobility. Rtt109 acetylates histone H3 and interacts with Rtt106 to promote replication coupled nucleosome assembly. Probably Ty1 mobility is increased due to the halting of DNA replication in the absence of *rtt109* or *rtt106*. This field of research on how histone modifications aid in Ty1 targeting is ongoing. A comprehensive library of H2A and H2B mutants have been developed and it would be a worth testing Ty1 targeting in these mutants (**Jiang et al 2017**).

4.2 Interplay between the Pol III transcription machinery and Ty1 Integration

The Ty elements have evolved to develop their targeting upstream of Pol III transcribed genes because there exists an advantage in doing so. As Pol III transcribed genes are present in multiple copies thus, insertion into one locus will not be lethal to the cells. The promoter regions of the tRNA genes are also enclosed into the coding region and therefore, the 1 kb window upstream of tRNA genes provide enough space for the Ty insertions without affecting the promoter activity. There are 275 copies of the tRNA genes and the 5sRNA are present in a tandem array of 100-200 copies. Thus, there are an abundant amount of target sites for the Ty1-Ty4 elements.

The RNA polymerase III complex plays an important role in the integration site choice upstream of the tRNAs. It has been shown that mutation in the Box B of the *SUF16* promoter region that abolishes transcription also severely defects Ty1 insertion upstream of Pol III transcribed genes suggesting that an active Pol III transcription is necessary for Ty1 targeting (**Devine et al 1996**). In relation with the tethering model described earlier, it was shown that integration upstream of the tRNAs is due a direct interaction between the AC40 subunit of Pol III and the Ty1 IN. The physical association of Ty1 IN and AC40 has been shown by co-immunoprecipitation and yeast 2 hybrid assay (**Bridier Nahmias A et al 2015**). This study also demonstrated that under the loss of this Ty1 IN-AC40 interaction by swapping the AC40 from *S.cerevisiae* with AC40 from *S.Pombe* where the Pol III transcription remained active, the integration site was redistributed to the sub-telomeres with no change in the retrotransposition frequency. This study answered the long-standing question as to why Ty1 targets the Pol III transcribed genes. This also indicates that there are other host factors that interact with the Ty1 IN for Ty1 targeting at the sub-telomeres. Likewise, **Cheung et al 2016** also confirmed that the Ty1 IN co-purifies with the Pol III complex. *In vitro* interaction assays using recombinant proteins purified from *E. coli* they showed there is a direct interaction between the Rpc31, Rpc34 and Rpc53 subunits with Ty1 IN. They have provided some evidence to show that the Rpc53/37 heterodimer can directly influence Ty1 targeting. They showed that 280 amino acid truncations of the N-terminal of Rpc53 significantly reduces Ty1 integration upstream of tRNA genes with no change in the frequency of retromobility. Nevertheless, it is unknown which interactions between Ty1 IN and Pol III subunits are lost in the Rpc53 N-terminal truncated mutant.

5. The Pol III transcription machinery

The RNA Polymerase III enzyme transcribes all the tRNAs, 5SrRNA, U6 small nuclear RNA, the snR52 small nucleolar RNA, RNA subunit of the signal recognizing particle, SCR1 and RNA component of the RNase P, RPR1 (reviewed by **Acker et al 2013**). The RNA Pol III machinery consists of the RNA Pol III complex, two transcription factors TFIIIB and TFIIIC required for the transcription of tRNAs and TFIIIA specifically required for the transcription of 5S ribosomal RNA. The tRNA promoter contains two conserved internal elements, Box A and Box B that are recognized by the transcription factors to recruit the RNA Pol III at the promoters. Unique for 5SrRNA, there is a specific box C recognized by the TFIIIA (**Camier et al 1995**).

5.1 The Pol III transcripts in yeast

5.1.1 Transfer RNA (tRNA)

There are multiple tRNA genes in the eukaryotic genome. It has been reported that there are 275 tRNA genes present in the budding yeast genome (**Goffeau et al 1996**). They encode the tRNAs required for decoding the codon from each RNA to the protein product. tRNAs are 75-95 bp long and are characterized by a secondary structure made of three hairpin loops, terminal helical stem folding into a L-shaped tertiary structure allowing them to fit into the P and A site of the ribosome(**Goodenbour 2006**). The three hairpin loops are known as the D-arm, T-arm and the anticodon arm. The helical terminal stem is the acceptor stem. The acceptor arm is 7-9 bp long ending into a CCA tail on which the amino acid is loaded by the aminoacyl tRNA synthetase through a covalent bond to the 3'- OH group of the tail. This sequence is critical for protein translation (**Sprinzl 1979**). The anticodon arm contains the anticodon in its loop used for reading the codons on the mRNA. There are several tRNA isoacceptor for each of the 20 amino acids and thus the genetic code is referred to as degenerate. Chromatin immunoprecipitation has shown a high occupancy of Pol III complex and TFIIIB on tRNA genes suggesting that these genes are highly transcribed (**Soragni et al 2008**). The tRNA genes consists of two highly conserved intragenic sequences known as the A-box and the B-box which is recognized by the TFIIIC to recruit the TFIIIB assembly (**Galli et al 1981**, **reviewed Acker et al 2013**). Though the primary function of these genes are to provide template for the tRNA molecules, there have been growing evidence about their role in mediating functional genome

organization and genome evolution. It has been demonstrated that tRNA genes can regulate the transcription of neighboring Pol II genes by inhibiting the progress of replication forks (**Hull et al 1994, Deshpande et al 1996**). tRNA genes have also been showed to function as a heterochromatin barrier. For example,the $tRNA^{thr}$ regulates the silencing of genes at the silent mating loci (**Donze et al 2001**). Multiple copies of tRNA genes are foundscattered throughout the whole yeast genome as a single transcriptional unit as shown by genome wide analysis (**Percudani et al 1997**).

5.1.2 5S ribosomal RNA

The 5S rRNA is 120 nucleotides long and is tethered to the large subunit of the ribosome in eukaryotic organisms. It has a molecular mass of 40 KDa. There are 140 5S rRNA genes in the haploid genome of the *S. cerevisiae* (**Elion et al 1984**). The ribosomal DNA (rDNA) of the budding yeast is encoded by the *RDN1* gene and lies in a 1-2 Mb region consisting of 100-200 tandem copies of 9.1 Kb repeats in the right arm of chromosome XII. The ribosome is composed of two subunits, the larger 60S subunits comprise of the 25S, 5.8S and 5S rRNA (precursor of 35S rRNA). The 40S subunit is comprised of the 5S rRNA. All the rRNA species are transcribed by Pol I except for the 5S rRNA which is uniquely transcribed by Pol III. Each of the 5S rRNA genes lie within the ribosomal repeat between the major promoter element and the initiation site of the large rDNA (35S rRNA) and is transcribed in the opposite direction (**White 1998)**. Unlike most rRNAs, it is not transcribed in the nucleolus. Thus, an interaction with the 60S subunit of the ribosomal protein, L5 is needed to assemble at the ribosomal site within the nucleolus. The 5S rRNA gene contains a 14-nucleotide box C in the internal coding region where TFIIIA interacts to recruit TFIIIC, TFIIIB and initiates the Pol III transcription (**Lassar et al 1983**). The tertiary structure of the 5S rRNA has been elucidated as a free molecule or in the ribosomal complex but its function in protein synthesis remains to be elucidated. However, biochemical studies have led to the hypothesis that 5S rRNA may act as a factor to signal communication between the functional centres of the ribosome (**Bogdanov et al 1995**) and thus mediating directly the enzymatic steps of the translation process. The secondary structure of the 5S rRNA consists of five helices (denoted as I-V), four loops (B-E) and one hinge (A) which together forms a Y shaped structure. Loops C and D are the terminal hairpins and the B and E are the internal ones. It has been hypothesized the helix V can form a hairpin structure and form a transcriptional complex with TFIIIA (**Ciganda et al 2011**).

5.1.3 U6 small nuclear RNA (snRNA)

The U6 snRNA is the non-coding small nuclear RNA encoded by the *SNR6* gene in yeast. The U6 small nuclear ribonucleoprotein (snRNP) combines with other snRNPs to assemble into the spliceosome required for precise splicing of the introns in mRNA transcripts. There are five other snRNPs namely U1, U2, U4 and U5 which are all transcribed by the Pol II transcription machinery except for the U6 which is transcribed by the Pol III machinery. The U6 RNA sequence is the most highly conserved sequence across the other species and is an essential gene (**Brow 1998**). **Brow 1990** demonstrated that the yeast U6 promoter is different from the vertebrate U6 gene promoter in its presence of an essential promoter element positioned around 120-130 bp downstream of the 3' end. The sequence at this region, matches closely to the B box of tRNA genes to which TFIIIC binds. The A box is present at the 5' end that maps to +21 nucleotides at this region. Thus, the distance between the transcription initiation site of SNR6 to the downstream B box is the longest than any other Pol III transcribed genes known.

5.1.4 RNA subunit of the signal recognition particle (SCR1)

The signal recognition particle (SRP) is abundantly present and conserved required for targeting the proteins to the endoplasmic reticulum. SRP is made of six subunits and a 7S RNA molecule. The subunits are encoded by *SRP14, SRP21, SRP68, SRP72, SEC65, SRP65* and the RNA is encoded by the *SCR1* gene. Deletion of a subunit inhibits growth but not cell death indicating that SRPs are not essential in yeast. The SCR1 gene is 522 bp long making it the longest Pol III transcribed gene. Although there is a canonical TATA box in the promoter, but its transcription is strictly dependent on the intergenic A and B box and requires the presence of all the Pol III transcription apparatus (**Dieci et al 2002**). The *SCR1* gene is translated highly efficiently from just one copy accounting to 0.2% of the total RNA in yeast (**Felici et al 1989**). Like the tRNA genes which are 90 bp long, Pol III can recycle and re-initiate transcription of the *SCR1,* despite its long length suggesting that recycling is not limited to short length DNA which explains the abundance of this RNA in the cells.

5.1.5 RNA component of the RNase P (RPR1)

The *RPR1* gene is 369 bp long and is an essential gene encoding for the RNase P, an endonuclease that cleaves pre-tRNAs to form mature tRNAs. The RPR1 contains both the A and B box promoters like the tRNA genes but are present considerably further downstream

compared to the tRNAs. Analysis of the RPR1 RNA in yeast cells showed the presence of the 369 bp long transcript and in less abundance another transcript with an extra 84 nucleotides at the 5'end and 16 to 30 nucleotides at the 3' end (**Lee et al 1991**). This longer RNA is believed to be the precursor RNA which undergoes a single 5' cleavage and multiple 3' cleavages to yield the mature RNA. This kind of precursor RNA hasn't been found in humans suggesting that this maturation process may be specific to yeast. Although the RNA is the catalytic subunit of the nuclear RNase P, there are other proteins required to assist the RNA in its function. The proteins and the RNA assemble to form the functional complex. The proteins are Pop4, Pop5, Rpp1, Rpr2 and Snm1 (**Liu and Altman 2009**). The RNase P complex has similar structural features across species suggesting a common function of the conserved structures. The precursor RNA is devoid of the Rpr2 subunit and still shows a steady kinetics in pre-tRNA processing suggesting that Rpr2 is dispensable for maturation despite being essential for growth.

5.2 The Pol III transcription factors

5.2.1 TFIIIA (Transcription factor III A)

TFIIIA is a 40 kDa protein whose only essential function is the transcription of the 5S ribosomal RNA. It displays high affinity towards the internal control region (ICS) of the 5S rRNA. The gene encoding TFIIA is designated as PZF1 and is found adjacent to the *RPB6* gene, encoding the ABC23 subunit shared among all the RNA Polymerases. The PZF1 gene is a single copy gene transcribed into an mRNA, 1.5 kb long and is essential for cell viability. The *PFZ1* and the *RPB6* genes are transcribed divergently and are separated by just 233 bp (**Woychik et al 1992**). The yeast 5SrRNA ICS (C-box) is made up of 14 bp and TFIIIA occupies a 35 bp region encompassing part of the C-Box (**Challice and Segall 1989, Braun et al 1989)**. TFIIIA contains 9 zinc- finger domains out of which the first three are required for DNA binding. Fingers 1 and 7 are important for the assembly of the transcription machinery on the 5SrRNA genes. There are 81 amino acid residues between the zinc finger 8 and 9 which is critical for transcription activity *in vitro* but is not required for DNA binding (**Milne and Segall 1993**). There exists a short leucine rich segment in this region which confers transcriptional activity to TFIIIA by modulating an interaction with TFIIC (**Rowland 1998**).

5.2.1 TFIIIC (Transcription factor III C)

TFIIIC is a large protein complex with a molecular mass of 600 KDa and is made of 6 subunits Tfc1 (t95), Tfc3 (t138), Tfc4 (t131), Tfc6 (t91), Tfc7 (t55) and Tfc8 (t60) (**Ducrot et al 2006**). They are all encoded by separate genes and are essential for cell viability. They are designated as two large globular domains tA (Tfc4, Tfc1 and Tfc7) and tB (Tfc3, Tfc6 and Tfc8) of similar mass each and 10 nm in diameter as viewed by electron microscopy (**Schultz 1989**). Both the globular domains exhibit high flexibility as it can stretch across a long rang separating the A box and B box in the tRNA genes. In humans the sixth subunit of TFIIIC known as TFIII35 was identified by **Dumey-Odelot H et al 2007** related the smallest subunit of the TFIIIC in yeast. Tfc1 which is highly conserved across yeast and human has been shown to interact with the Tfc6 and Tfc4 subunits (**Sabine et al 2003**). It plays a critical role in Pol III initiation complex assembly and recognizes the box A and box B on the promoter elements of the tRNA genes to recruit the TFIIIB complex upstream of the transcription start site which in turn recruits the Pol III and activates transcription (**Geiduschek 2001**). The two subunits Tfc1 and Tfc3 are involved in the recognition of the A and B box and cooperate in DNA binding. A point mutation in the *TFC3* gene affects severely the TFIIIC: DNA complex and this phenotype is suppressed by an amino acid substitution in Tfc1 **(Arrebola et al 1998, Lefebvre et al 1994)**. The Tfc8 which is present in tB could be involved in a functional interaction with the TBP subunit of TFIIIB (**Deprez et al 1999**). Photocrosslinking experiments mapped the Tfc1 on the A box and the Tfc3 on proximity to the B box (**Bartholomew et al 1990**). Tfc1 occupies the same region on the 5SrRNA even in the absence of the A box (**Braun et al 1992**). Several genetic and biochemical approaches have also provided several evidences regarding the functional role of Tfc4 in TFIIIB assembly on the DNA. Tfc4 interacts with the Brf1 and Bdp1 subunits of TFIIIB (**Chaussivert et al 1995, Willis 2002**). The recruitment of TFIIIB to the DNA is correlated with several conformational changes of Tfc4. Firstly, the efficiency of crosslinking to DNA varies with TFIIIB assembly. Secondly, Tfc4 can alter the visibility of TFIIIB binding sites and thirdly mutations in the N-terminal of Tfc4 increased TFIIIB assembly (**Kassavetis et al 1992, Moir et al 2002**).

5.2.2 TFIIIB (Transcription factor III B)

TFIIIB is composed of 3 subunits TBP, Brf1 and Bdp1. TFIIIB has never been shown form a stable complex in the absence of DNA. Purification of TFIIIB yields TBP/Brf1 subcomplexes

and Bdp1 (**Kassavetis et al 1991**). It is the initiation factor that recruits the Pol III complex at the transcription start site of class III genes to activate transcription.

The TATA Binding protein (TBP) is a 27 KDa protein encoded by the *SPT15* gene and is essential. It is a general transcription factor required for driving transcription for all the genes transcribed by the three polymerases in yeast (**Fan X et al 2005**). TBP binds to DNA directly through the TATA region or is recruited as a complex such as SAGA or TFIIIB. It is one of the few transcription factors that are highly conserved among eukaryotes. It is 240 amino acid long with a diverge N-terminal but a highly conserved C-terminal (**Schmidt et al 1989**). The N-terminal region inhibits TBP binding to the TATA region and promotes interaction with its protein partners (**Lee et al 2001**). The C-terminal has two surfaces which are involved in binding with transcriptional activators and formation of inactive homodimers (**Kou et al 2003**). As mentioned earlier, TFIIIB interacts with TFIIIC via an interaction between Tfc8 and TBP (**Deprez et al 1999**).

The Brf1 is a 70 KDa protein encoded by the *BRF1* gene and is essential. The N-terminal region of Brf1 plays an important role in Pol III recruitment to genes and the C-terminal part holds the TFIIIB assembly in a tight configuration (**Kassavetis 1998**). The primary function of Brf1 in TFIIIB assembly is contributed by the C-terminal domain spanning amino acids 435-545 and the transcription directing function by the N-terminal region. Each half of Brf1 interacts with DNA bound TBP; the C-proximal half of Brf1 interacts with the N-proximal pseudo repeat loop of TBP and the N-proximal half of Brf1 interacts with the C-proximal pseudo repeat lobe of TBP (**Kassavetis 1998**). The three components of the TFIIIB are held together by a more stable Brf1-TBP interaction than TBP-DNA interaction. (**Huet et al 1994**). Apart from interacting with TFIIIC, it also functions to dissociate inactive TBP dimers to active monomers (**Alexander et al 2004**).

Bdp1 is a 68 KDa protein encoded by the *BDP1* gene and is essential. It has a 60 residue SANT domain functioning as a histone modulator and interacts with Brf1. TFIIIB cannot bind to promoters which do not contain a TATA box. Therefore, its assembly onto the start site of transcription requires the presence of the TFIIIC complex. The Tcf4 subunit recruits the Brf1 by protein-protein interaction. The TBP bound to Brf1 anchors the complex to the DNA. Then, the Tfc4 subunit again recruits Brf1 to the TFIIIB assembly (**Kassavetis et al 2006**). Bdp1 and Brf1 play an important role in promoter opening required for transcription. Mutations in Brf1

and Bdp1 could still recruit Pol III but interfere with the separation of the DNA strand at the upstream end of the transcription start site.

5.3 The Pol III complex

The Pol III complex is made of 17 protein subunits and has a molecular mass of 0.7 MDa. Among the three RNA Polymerases in the cell, Pol III is the largest and the most complex one. Many subunits present in Pol III are shared subunits among the three polymerases and are highly conserved across human and yeast. An overview of the protein subunits are provided in the table below.

Pol I	Pol II	Pol III
ABC27	ABC27	ABC27
ABC23	ABC23	ABC23
ABC14.5	ABC14.5	ABC14.5
ABC10a	ABC10a	ABC10a
ABC10b	ABC10b	ABC10b
AC40		AC40
AC19		AC19
A190	Rpb1	C160
A135	Rpb2	C128
A12.2	Rpb3	C11
A14	Rpb11	C17
A43	Rpb9	C25
A49	Rpb4	C37
A34.5	Rpb7	C53
		C31
		C34
		C82

Table 7. **The RNA polymerases subunits in yeast**. The shared subunits among all the RNA Polymerases are marked in red. The subunits shared among Pol I and Pol III are in blue and the unique subunits are in black.

All the three polymerases share a horse-shoe shaped core made of 10 subunits. The five subunits are common and the rest five share a sequence similarity in the range 19-36%. The core has two subunits attached to it that forms the stalk (**Cramer et al 2008**). The stalk is important for transcription initiation. There are other 5 specific subunits that can form subcomplexes: C37-C53 can form a heterodimer which is related to the TFIIF of Pol II and the Pol I A49/A34.5. C31-C82-C34 heterotrimeric complex related to the TFIIE of Pol II (**Carter and Drouin, 2010**).

The core contains the two largest Pol III subunits C160 and C128, together forming the active centre and the nucleic acid binding cleft. The other subunits include the Pol I and Pol III shared subunits AC40 and AC19, the five subunits shared among all the RNA polymerases namely ABC 27, ABC 23, ABC 14.5, ABC10a, ABC10b and C11. The C11 subunit shows a strong homology with the TFIIS of Pol II and mutations in this subunit negatively affects transcription termination (**Chedin et al 1998**). C11 is involved in regulating the Pol III enzyme switch between RNA elongation and RNA cleavage mode. This cleavage activity of Pol III is required to remove barriers during the termination process. The overall architecture of the core is conserved among all the polymerases, consisting of a clamp head, jaw and middle cleft. The DNA gains access into the polymerase through the clamp. However, the clamp head of the C160 subunit is larger than its Pol I and Pol III counterparts and its C-terminal protrudes from the core and together with the N-terminus makes contact with the stalk (C17-C25). The C128 also demonstrates conserved folds but contains an extended surface that protrudes out increasing the depth of the DNA binding cleft (**Hoffmann et al 2015**).

The stalk is made up of the two subunits C17 and C25 related to the A14 and A43 subunits of Pol I and Rpb4 and Rpb7 subunits of Pol II. The stalk is attached to the core via the C160 subunit. The structure of the C17-C25 subcomplex show that there exists a C17 N-terminal "tip-associated" domain that packs against the C25 tip domain. The contact between the two proteins are formed by a helix in C17 and two loops in C25 (**Jasiak et al 2006**). Mutations in the loops hamper C17 binding i*n vivo.* The C17 subunit has been shown to be important for cell viability and tRNA synthesis. The complex also interacts with the Brf1 subunit of TFIIB and to the heterotrimer of Pol III suggesting that it might be involved in initiation complex assembly (**Ferri et al 2000, Geiduschek and Kassavetis 2001**). The C17-C25 subcomplex

have high binding affinity to duplex DNA indicating that this complex might be involved in pol III binding to target genes.

The heterotrimer consists of the subcomplex formed by the C82-C34-C31 subunits involved in the initiation of transcription (**Werner et al 1992**). C-terminal deletion of the C31 subunit negatively affects tRNA transcription in vivo but this phenotype is suppressed by overexpression of the C160 subunit suggesting a functional interaction between the two (**Thuillier et al 1995**). Another study by **Brun et al 1997** showed that C34 subunit plays a major role in Pol III recruitment by the pre-initiation complex. Mutations in the gene encoding for C34 showed defects in its interaction with Brf1 subunit of TFIIIB *in vivo* and slows cell growth. Consistent with this data, mutations in C34 are defective in open complex formation and Pol III recruitment at the PIC. In good correlation with Brf1-C34 interaction, **Bartholomew et al 1992**, demonstrated that C34 localized upstream on promoter DNA in the initiation complexes.

The heterodimer is a stable complex formed by the C53 and C37 subunits and functions as a recognizer of the termination sequence to terminate the Pol III transcription process (**Landrieux et al 2006**). They purified a Pol III from cells lacking the 27 C-terminal residues of the C37. *In vitro* transcription was active but transcription termination was hampered. However, **Kassavetis et al 2010** also demonstrated that these subunits are also involved in the transcription elongation and initiation process through interactions with the DNA. C53 is known to interact with DNA at the downstream end of the transcription complex and C37 binds to the upstream non-template strand during elongation (**Tate et al 1998**). **Chedin et al 1998** demonstrated that deleting 27 amino acids in the C-terminal domain of C37 results in the loss of subunits C53 and C11. Likewise, substituting the C11 with the *S.pombe* orthologue causes a loss of C37 and C53 suggesting they might be present in close contact in the Pol III complex.

The Pol III Cryo-EM structure (**Hoffmann et al 2015**) is provided below.

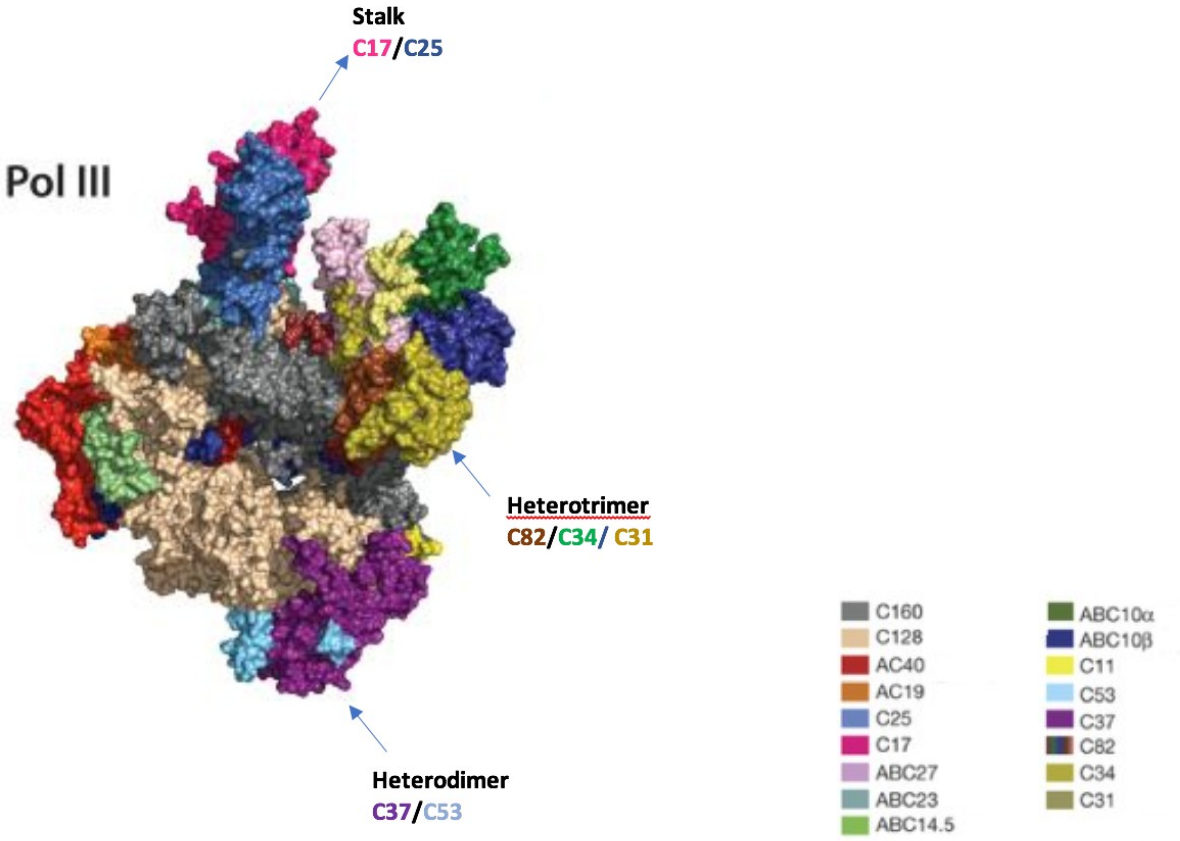


Figure 18. **The Cryo-EM structure of the Pol III complex in yeast.** Each of the subunit is colour coded. The stalk, heterotrimer and the heterodimer are indicated. (Image adapted from **Hoffmann et al 2015**)

5.4 The Pol III transcription mechanism

The transcription of by Pol III can be divided into the following steps: Initiation, elongation and termination. During the initiation stage, the internal promoter elements in the tRNA genes, box A and box B are first recognized by TFIIIC which recruits the TFIIIB complex and helps in its assembly upstream of the transcription start site (TSS). Then, TFIIIB recruits the Pol III complex for transcription initiation through an interaction between Brf1 and the C17 and C34 subunits of Pol III (**Ferri et al 2000, Andrau et al 1999**). During the elongation process TFIIIC remains bound to the DNA as it is not simultaneously displaced from the A and B boxes. Termination of the transcription mainly occurs at a tract of A residues on the DNA template. Minimum length of A_7– A_8 has been shown to terminate transcription in yeast (**Turowski et al 2016**). During termination the destabilization signal is promoted by a weak base pairing

interaction between the oligo(dA) in the template strand and oligo (U) in the nascent transcript. It has been been also shown by **Turowski et al 2016** that RNA Pol III transcription can even continue beyond the canonical termination sequence approximately 50 np from the 3' end of the tRNAs. But these transcripts are degraded by the exosome nuclease complex to prevent their accumulation. TFIIIC also functions in transcription re-initiation. To maintain the high transcriptional activity of the Pol III genes, it is necessary for the Pol III complex to rapidly reload to the transcription start site after transcription termination. **Ferrari et al 2004** suggests that the as TFIIIC is always bound to the DNA during the elongation process and for longer genes, TFIIIC facilitates the interaction between Pol III and TFIIIB, which remains bound to the promoter.

Having described the mechanism, the class III genes are recognized by the Pol III transcription machinery on promoter elements which vary in the location of the A and B box. The Type1 promoter in yeast contains an intergenic sequence (ICS) present in the 5S rRNA gene. It consists of the A box approximately +50-60 bp upstream of the TSS, an intermediate region +67–72bp and a C-box +80-97 bp. TFIIIA binds to the ICS and recruits the TFIIIC which in turn recruits the TFIIIB to initiate transcription process. The Type II promoter is present all the tRNA genes where the distance between the TSS and A box is fixed around +8-19 bp and the B -box approx. +45-62 bp but can vary in distance. The U6 gene in the yeast contains a mixed promoter where there is a TATA box -22 upstream of the TSS, A box +21-31 bp downstream the TSS and a B box +230 -240 bp downstream of the 3'end (**Teichmann et al 2010**). The figure showing the various Pol III genes, their promoter organization and transcription mechanism are depicted in the following figure below.

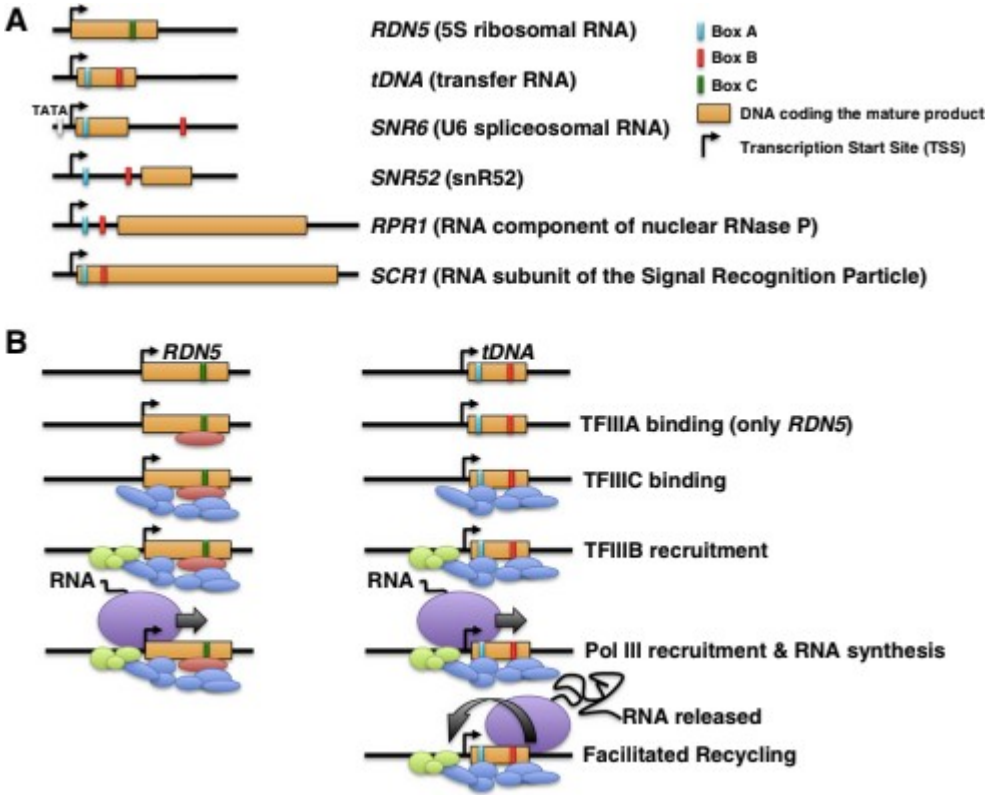


Figure 19. **Structure of the RNA Polymerase transcribed genes and the *in vitro* assembly of the Pol III transcription machinery**. (A) The gene names are indicated on the right. The different boxes are colour coded as indicated and the arrow shows the TSS. (B) The Pol III transcription machinery assemble on *RDN5* (left panel) and tDNA (Right panel). (Image adapted from **Acker et al 2013**)

5.5 Host factors regulating the Pol III transcription

There are several host factors that regulate the Pol III transcription *in vivo*. **Acker et al 2013** reviews all these regulators vividly. Chromatin immunoprecipitation (ChIP) was important in identifying several host factors present on Pol III transcribed genes. As there is a strong relation between Pol III transcription and Ty1 retrotransposition, it might be possible that factors affecting Pol III transcription can directly or indirectly influence Ty1 mobility. In fact, genetic screens have identified Pol III effectors as host factors of Ty1 such as Bud27 and CKb2 both of which regulate the Pol III machinery (**Nyswaner et al 2008**). We have even identified several Ty1 IN host factors that are also present on Pol III genes.

5.5.1 Characterized regulators of Pol III transcription

Maf1 is encoded by a non-essential gene *MAF1* and represses Pol III transcription through various signaling pathways. Pol III inhibition under the influence of drugs, oxidative stress or DNA damage requires Maf1. The phosphorylated form of Maf occurring under favourable growth conditions are located in the cytoplasm and cannot gain access into the nucleus. Under stress conditions, Maf1 is dephoshorylated and is transported to the nucleus to target the class 3 genes and block transcription (**Oficjalska et al 2006, Harismendy et al 2003**). Maf1 has been shown to physically interact with the N-terminal of the C160 subunit of Pol III by pull down experiments. However, the phosphorylation state of Maf1 regulates it binding to C160. The least phosphorylated form of Maf1 binds to Pol III (**Oficjalska et al 2006**). **Desai et al 2005** showed that yeast and human Maf1 have been shown to interact with TFIIIB *in vivo* . *In vitro,* studies have shown that Maf1 binds to Brf1 and inhibits TFIIIB assembly on DNA- TFIIIC complex. Cryo-EM data (**Vannini et al 2010**) has shown that Maf1 causes a conformation change in the heterotrimer C82/C34/31 subcomplex inhibiting the recruitment of the Pol III complex to the TBP-Brf1 promoter complex.

Sub1 has been identified as an activator of Pol III transcription. Sub1 can bind to single stranded DNA and has been implicated in various other cellular processes such as transcription elongation, mRNA 3' end processing, DNA repair and chromatin organization. **Wang and Roeder 1998** first showed that the human orthologue, PC4 co-purified with human TFIIIC and induces transcription *in vitro*. **Tavenet et al 2009** also revealed that Sub1 is present on class III genes and *SUB1* deletion mutants affects Pol III transcription and Pol III/Bdp1 occupancy on Pol III genes *in vivo*. *In vitro* studies have revealed that Sub1 helps the TFIIIB and TFIIIC assemble on the tRNA genes with interacting with both the factors. It regulates distinct steps of the transcription process enhancing both initiation and reinitiation.

5.5.2 Kinases

TFIIIC (Tfc1) and Pol III subunits are phosphorylated (**Conesa et al 1993**, **Chedin et al 1998**). Large scale phospho-proteomics studies also revealed that several components of the Pol III transcription machinery are phosphorylated and have identified the phosphorylation sites (**Mohammed et al 2008, Chi et al 2007, Ptacek et al 2005**). Kinases that regulate Pol III transcription have been identified through genetic and biochemical approaches are TORC1,

PKA, Sch9 and CK2. Ck2 and TORC1 are present on Pol III genes and Ck2 is only detected on tRNA genes (**Graczyk et al 2011**) whereas TORC1 binds specifically to th 5S rRNA gene in yeast.. The TBP subunit of the TFIIIB is phosphorylated by CK2 which is required for efficient Pol III transcription. During DNA damage signals are relayed to the Pol III transcription machinery through an association between TBP and CK2 i*n vivo* (**Ghavidel 2001**). Moreover, when cells are switched from repressive conditions to favorable growth conditions, CK2 is involved in releasing Maf1 from the Pol III-tDNA complex (**Graczyk et al 2011**). It is to be noted that Ckb2 a subunit of CK2 was identified in a genetic screen to negatively affect Ty1 mobility *in vivo* (**Nyswaner et al 2008**). I have particularly studied the role of CK2 phosphorylation on Ty1 IN. TORC1 physically interacts with Maf1 and phosphorylates it *in vitro* (**Wei et al 2009**) suggesting that it might regulate Maf1 phosphorylation directly or by associating with other proteins in the complex. PKA is involved in modulating the localization of Maf1 in the nucleus and Sch9 controls the interaction between Pol III and Maf1 (**Moir et al 2006, Huber et al 2009)** but no studies show the presence of these kinases on Pol III genes.

5.5.3 Chromatin remodellers

The tRNA genes lack nucleosomes whereas the upstream regions have highly ordered nucleosomes. Several chromatin remodelers and chromatin modifying complex have been shown to regulate Pol III transcription. Therefore, specific chromatin structure at the vicinity of Pol III genes is essential for efficient Pol III transcription.

The FACT complex plays a major role in building, maintain and adjusting the chromatin barrier. It is composed of the Spt16, Pob3 stable heterocomplex that is recruited on the nucleosome by the Nhp6 protein (**Formosa 2012**, **Fromosa et al 2001**). **Mahapatra et al 2011** showed that the FACT complex acts as a chaperone for the deposition of histone variant H2A.Z at the *SUP4* tRNA gene. Mutant Spt16 showed a loss of the histone variant at the tRNA gene and acts as a suppressor of tRNA transcription. In human cells, FACT complex is present on class III genes and is involved in the activation of Pol III transcription (**Birch et al 2009**). However, the FACT complex also plays an important role in promoting viral replication in cells. For instance, recently **Matysiak et al 2017** has been identified as a potential target of the HIV integrase. FACT binds to the LEDGF/p75 tethering factor of the HIV IN. *In vivo* and *in vitro* studies have shown that the FACT along with HIV IN and LEDGF/p75 can form a tripartite complex and generates highly favored nucleosomal structures for HIV integration *in*

vitro. FACT mediated chromatin remodeling promotes HIV integration *in vivo*. Moreover, Spt16-Pob3 has been shown to interact with the ALV Integrase and promote its integration *in vitro* (**Winans et al 2017**)

The RSC proteins are chromatin remodeling complexes and approximately 25% of the RSC targets are beside the Pol III transcribed genes (**Ng et al 2002**). The RSC complexes hydrolyses ATP to regulate the chromatin structure. A study demonstrated that RSC models the chromatin structure upstream of *SNR6* gene. *In vivo* transcription of two Pol III transcribed genes *SNR6* and *RPR1* are affected negatively in Rsc4 mutants (**Sontourina et al 2006**). Loss of the catalytic subunit of RSC, Sth1 severely defects global Pol III transcription (**Parnell et al 2008**). It has been hypothesized that the RSC is required to maintain tRNA genes in a low-density nucleosome environment favourable for transcription efficiency (Parnell et al 2008) as tRNA genes have low nucleosome occupancy, high histone turnover and high RSC occupancy. Another chromatin remodeler Isw2 that help deposit and position new nucleosome is recruited to the tRNA genes by tan interaction with TFIIIB (**Bachmann et al 2005**). The Isw2 has been shown to target at the tRNA genes by the N-terminal region of Bdp1p. Mutations in the Bdp1 N-terminus or Isw2 leads to similar disruption of the nucleosome positioning upstream of some tRNA genes.

5.5.3 Chromatin modifying complexes

Chromatin modifications seem to be important also for Ty1 targeting because it prefers nucleosomal surfaces at the H2A-H2B interface. Several chromatin modifying complexes, such as the histone acetylases may also play role in tRNA gene expression on account of their presence on class III genes. Rtt109, a histone acetyltransferase, acetylates the K56 residue of the histone H3 and this modification is enriched during exponential growth on tRNA genes (**Rufiange et al 2007)**. Asf1 and Vps75 are two histone chaperons that activate the Rtt109 catalytic activity. Chromatin immunoprecipitation have shown the occupancy of Asf1 and Vps75 on several class III genes (**Schwabish et al 2006**) suggesting that Rtt109 might influence Pol III transcription via the histone chaperons. Several histone deacetylases like Hda1 and Hos3 are present on tRNA genes (**Mou et al 2006**). Hos3 has been shown to stimulate Ty1 integration *in vivo.* Hos2 deletion mutants have shown reduced Ty1 mobility suggesting that chromatin modifications might influence specific nucleosome structure that is favorable for Ty1 integration.

6. Objective of this study

Bridier Nahmias A et al 2015 showed that AC40, a subunit of Pol III is a major target of Ty1 IN and that this interaction is absolutely necessary for the integration of Ty1 at the tRNA genes. However, without the Ty1 IN/AC40 interaction the integration site is redistributed to the subtelomeres without any change in the rate of retrotransposition. Therefore, after years of active research the tethering model for the Ty1 IN was proposed. However, it was evident that Ty1 IN also interacts with several other host factors to mediate its targeting specificity at the subtelomeres as the Pol III machinery is not present at this region. Moreover, there could be other host partners that might influence Ty1 targeting at the Pol III locus.

As we see from the literature vast number of host factors were identified through genetic screens that regulate Ty1 transposition *in vivo*, but none demonstrates the molecular mechanism of their role. Therefore, to find partners that may directly influence Ty1 mobility it is important to perform a proteomics screen that interacts with the Ty1 IN mediating its integration specificity. However, there have been few proteomic studies in the present to identify Ty IN host factors, but they were unsuccessful in providing an exhaustive and large list. Very few partners were identified out of which most of them were the subunits of the Pol III complex **(Cheung et al 2016)**. In two independent experiments, they found 12 proteins that were enriched in both experiments, namely, Adh2, Bud13, Ckb2, Irc24, Maf1, Reh1, Ura3, Rpc25, Rpc34, Rpc40, Rpc53 and Rpc82.

Therefore, the purpose of this study was to develop a robust proteomics approach to exhaustively identify protein targets of the Ty1 IN *in vivo* and characterize the functional role of the partners in Ty1 retrotransposition.

Chapter 2

RESULTS & DISCUSSIONS

Retrotransposition is a vital mechanism that helps the host to adapt to stress conditions and plays an important role in evolution. Thus, to understand how these eukaryotic hosts regulate retrotransposons to create genome rearrangements for adaptation and survival, it is imperative to identify host factors and delineate their molecular mechanism in retroelement mobility (**Risler *et al*, 2012**). There have been studies in the past to identify host genes by genetic screening that affect LTR retromobility in yeast such as Ty1 (**Scholes et al 2001, Griffith *et al* 2003**, **Nyswaner et al 2008**) and Ty3 (**Irwin et al, 2005**) but the molecular mechanisms as to how these host factors affect retromobility remains largely unexplored. This thesis aims to exhaustively identify the host protein partners that interact with the Ty1 Integrase and characterize their role in Ty1 retrotransposition *in vivo*. To achieve this, Ty1 Integrase (Ty1 IN) was ectopically expressed in yeast cells. After *in vivo* crosslink Ty1 IN was purified with its associated proteins that were identified by mass spectrometry. Upon stringent analysis of the mass spectrometry data, several putative new Ty1 IN partner were identified and among them, we decided to investigate more particularly the casein kinase CK2. We exploited various biochemical and genetic approaches to study its role in Ty1 retrotransposition.

1. Exhaustive identification of Ty1 IN partners

Exhaustive identification of protein partners demands a robust purification process where we are able to purify a significant amount of the bait protein and capture as much associated proteins as possible without losing weak and transient interactors. To achieve this, we deployed the **T**andem **C**hromatin **A**ffinity **P**urification (TChAP) procedure to purify the Ty1 IN as a bait from a large volume of yeast cell culture (8l). This method was successfully developed in our laboratory to identify host factors associated with the RNA Polymerase III transcription machinery in yeast *Saccharomyces Cerevisiae* **(Nguyen *et al*, 2015).** This method was set up like ChIP experiments but were adapted to large volume cultures. The schematic diagram of the workflow is provided in Figure 1. The global approach involves purifying an epitope tagged Ty1 IN as a bait after *in vivo* cross-link and identifying the co-purified protein by mass spectrometry. Along with other host partners, we also expect to find the Pol III complex as a binding partner which will serve as a positive control for the experiment.

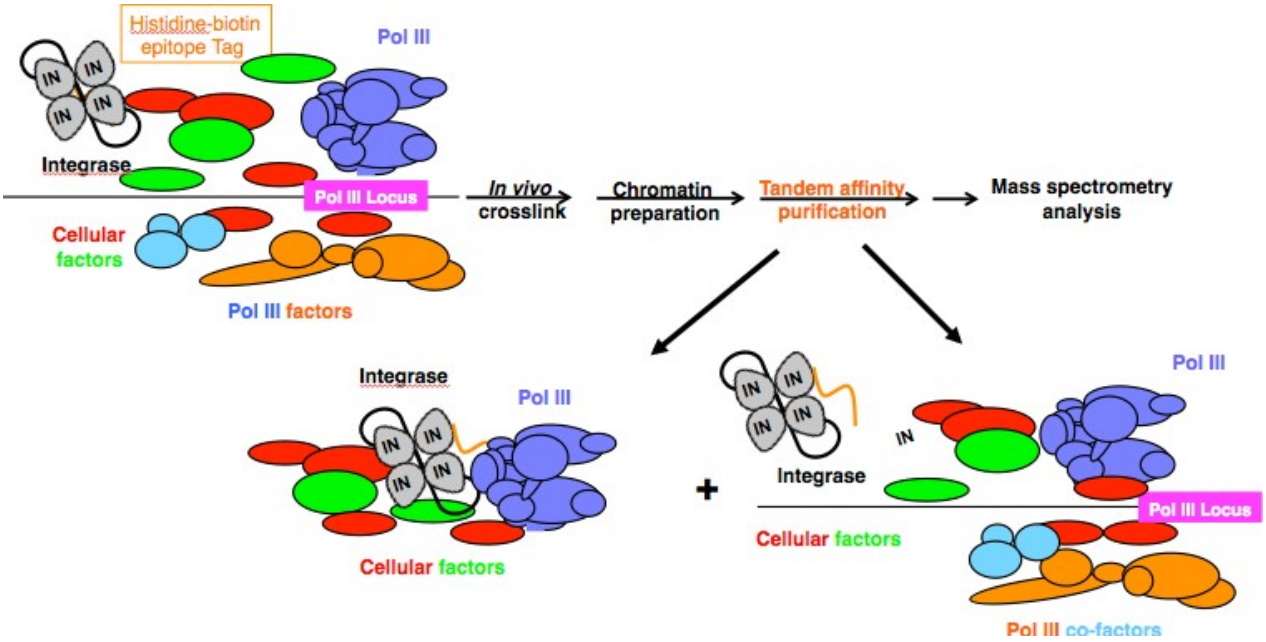


Figure 1: **Schematic diagram of the work flow to identify the targets of Ty1 IN.** The approach involves purification of the Ty1 IN HBH after *in vivo* crosslink by TChAP and identify the targets by mass spectrometry. We expect to find the Pol III transcription machinery and other host factors associated with Ty1 IN.

1.1 Optimization of the TChAP procedure for the purification of Ty1 IN

The TChAP method is a powerful purification method involving three basic steps: *in vivo* crosslink of a large volume of culture, chromatin preparation and tandem purification of the bait protein. It is a week-long process where a large volume of yeast cell culture is grown to exponential phase and crosslinked with 1% formaldehyde for 20 min to freeze all protein-protein interaction *in vivo*. Then cellular extracts are prepared and sonicated to solubilize both DNA and proteins. Eventually, the bait protein in the extract is purified on nickel column and then, streptavidin beads. The co-purified proteins are analyzed by western blot and identified by mass spectrometry.

Similar to the HIV-1 Integrase, Ty1 IN is mainly insoluble. Moreover, endogenous Ty1 IN is almost non-detectable in saccharomyces cerevisiae. Thus, we first had to set up the purification procedure by modifying some steps of the current TChAP protocol to be able to purify Ty1 IN and its associated proteins. The classical procedure and the steps we modified in our experiment is shown in the figure 2 below.

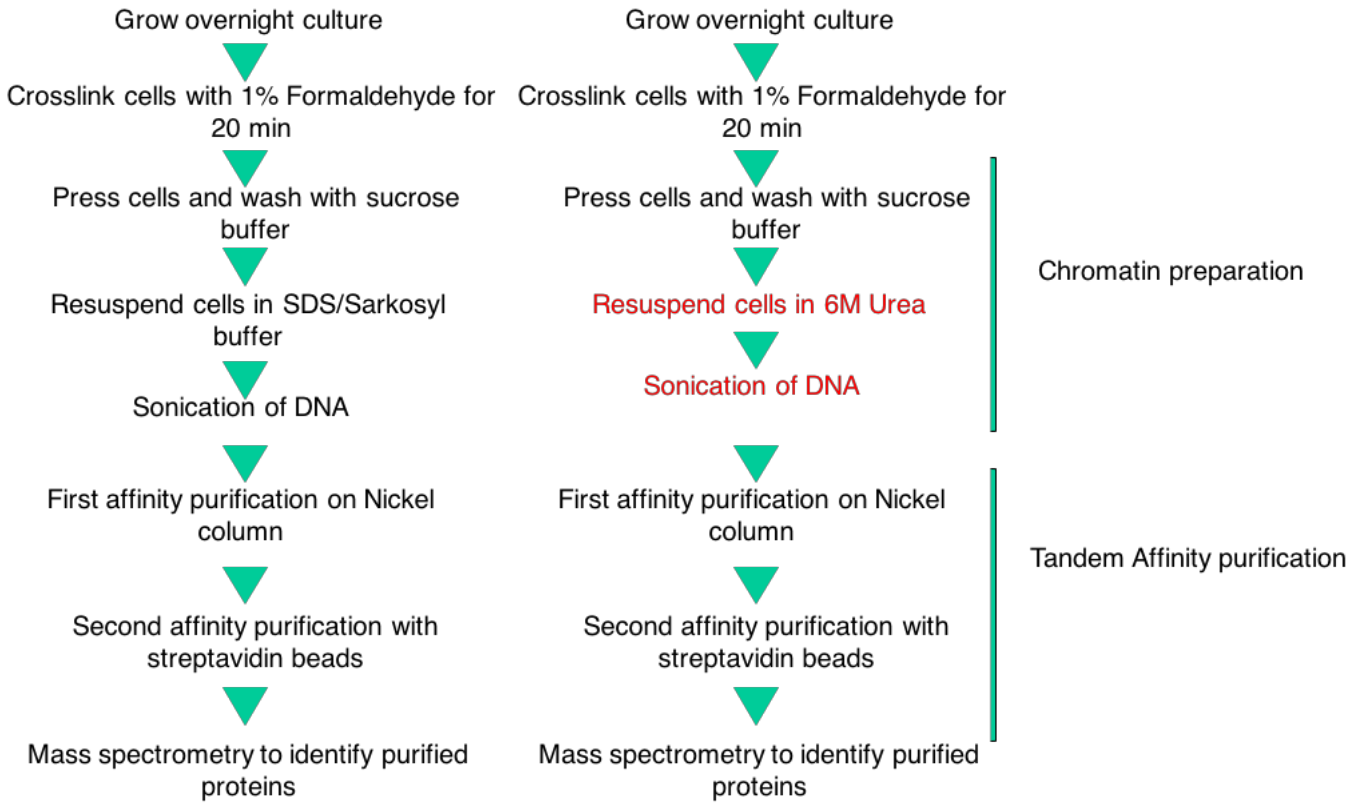


Figure 2: **Flowchart of the TChAP procedure.** The left panel shows the classical procedure of TChAP developed in our lab for the identification of host factors associated with the Pol III machinery (**Nguyen et al 2015)**. The right panel shows the procedure of TChAP used for the purification of Ty1 IN partners. The steps modified are marked in red.

1.1.1 Setting up of Ty1 IN expression *in vivo*

The Ty1 mRNA is abundantly present in the cells accounting to 1% of the total yeast RNA, the level of Ty1 proteins is very low and Ty1 IN is hardly detected in yeast cells under normal growth conditions which makes it difficult to purify the endogenous Ty1 IN by tandem purification. Therefore, we decided to express the Ty IN ectopically on a plasmid. However, there are various factors that govern a good and robust proteomics approach. Firstly, according to our previous experiments, overexpression of the bait protein leads to a high background of non-specific binders and the classical pgal-Ty1 vector used in all the studies is an overexpression vector and thus, is not suitable for our experiment. Secondly, we should be able to purify the bait under stringent conditions and thirdly, we need an inducible system to regulate the Ty1 IN expression *in vivo* as to avoid its deleterious effect on the cells. **Boeke et al, 1985** suggested that high expression of Ty1 element has a deleterious effect on cell growth because

it leads to high Ty1 transposition and the cells accumulate random mutations. In addition, the reverse transcriptase (RT) might function on other mRNAs of the cells leading to fatal consequences. Thus, keeping this is mind we decided to use the ptet-off inducible expression system based on a low copy plasmid (pcm185) developed by **Gari et al. 1997** which would not over-express the Ty1 IN *in vivo*. The advantage of using such a system over other expression systems is the fact that it requires no change of nutrient composition unlike the galactose inducible expression system where addition of galactose induces expression of the gene of interest but changes the metabolism of cells. In order to be able to perform the TChAP procedure, the protein was expressed as a fusion protein with an HBH (histidine-biotin-histidine) tag derived from a bacterial peptide biotinylated *in vivo* by endogenous biotin ligases flanked by two histidine on either side located at its C-terminal end. This tag facilitates tandem purification by nickel affinity chromatography and streptavidin beads batch purification under fully denaturing conditions (**Kaiser *et al* 2008**).

We aimed to identify partners under conditions that regulates retrotransposition. Temperature stress is an important factor that has been shown to stimulate retrotransposition *in vivo*. **Pacquin et al 1985** demonstrated that Ty1 retrotransposition is high at low temperatures such as 15°C and is minimal at 30°C. However, most of the retrotransposition assay are performed at 20°C-22°C. We decided to purify Ty1 IN associated proteins at both temperature conditions that stimulate and repress retrotransposition respectively. This would enable us to make a comparative study of the identified partners that might regulate Ty1 retrotransposition under different conditions of temperature. As a first step, we tested the *in vivo* expression of Ty1 IN-HBH controlled by the tetOff promoter at 20°C and 30° respectively. To perform TChAP, we also had to handle large culture volume and the methodology becomes tedious more particularly below 20°C due to the large replication time of the yeast cells. To be able to manage such experiments we set culture conditions for ON induction. The cells were grown in the classical glucose synthetic medium to stationary phase in the presence of 1µg/ml doxycycline (a tetracycline derivative) at 30°C on the first day to keep the Ty1 IN expression repressed. The following day the stationary cells were diluted accordingly in fresh medium without doxycycline for overnight induction of the Ty1 IN at both temperatures 20°C and 30°C independently.

As shown in figure 3, the expression of Ty1 IN HBH is clearly observed at both 30°C and 20°C and the expression was tightly regulated in the presence of doxycycline at a concentration of

1µg/ml. There was no leakage of the promoter in the presence of the antibiotic. This induction of the Ty1 IN did not affect the growth of the cells either. However, we observed that the expression of the Ty1 IN was higher at 30°C compared to 20°C. Initially, we thought that this is because the optimal expression of yeast proteins is higher at 30°C than at 20°C. This low expression gave low yields of the Ty1 INHBH after purification at 20°C. However, to address the issue we developed a method explained in the section 1.2.

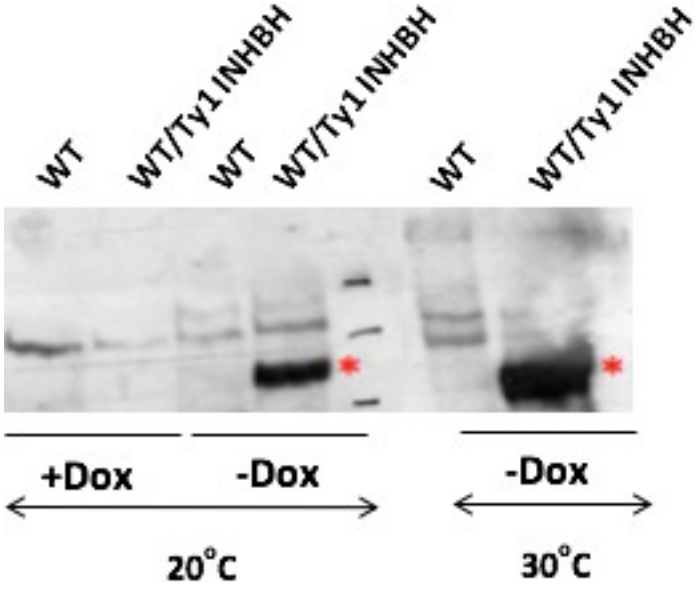


Figure 3: **Setting up of expression of Ty1 IN HBH controlled by a TetOff promoter.** Western blot showing **t**he induction of Ty1 IN HBH as observed in the absence of doxycycline at 30°C and at 20°C. The Ty1 INHBH was visualized using anti-streptavidin antibody.

1.1.2 Optimization of chromatin preparation

Sonication helps in solubilization of the DNA and chromatin bound proteins. This step depends on various parameters such as the amount of the cells, volume of the lysate and the extent of crosslinking (**Nguyen *et al* 2015**). Since Ty1 IN is mainly insoluble, we decided to perform our TChAP procedure in a high denaturant, 6M Urea containing buffer that may help to solubilize the proteins and that was already used with success in tandem affinity purify (**Tagwerker et al 2008**). Thus, we had to optimize the sonication step in Urea buffer because previous TChAP were set up for extracts prepared in non-chaotropic buffer. TChAP studies performed in the lab showed that one of the most important parameters for protein extraction in TChAP method is the time of sonication. It was necessary to optimize the conditions as too long sonication could damage the polypeptides in the bait protein and produce low yields after purification. And, too

less sonication would not be enough for complete solubilization of the proteins in the extract. As implied in **Nguyen et al 2015**, we sonicated the extracts at an amplitude of 70% for 10 secs followed by 50 sec intervals and collected samples at various time points (5 min, 10 min and 15 min) to analyze the amount of DNA and protein in the supernatant (soluble fraction) and pellet (insoluble fraction).

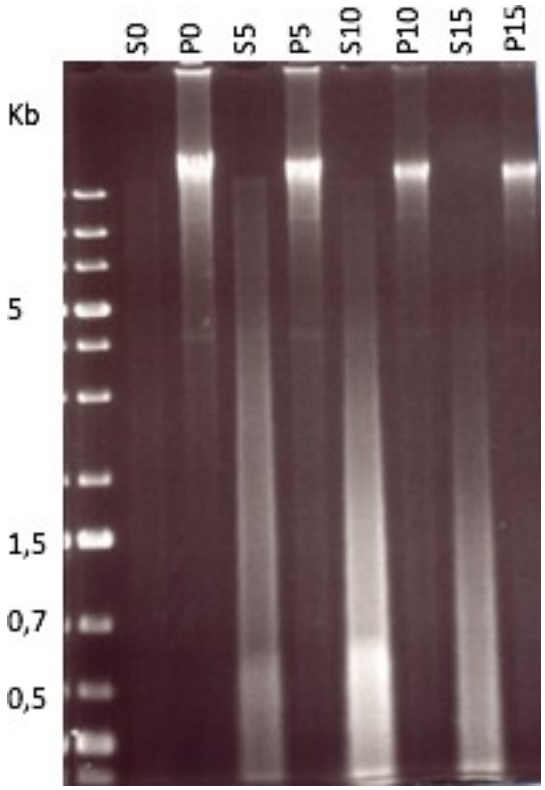


Figure 4: **Solubilization of DNA by sonication.** DNA extracted from the pellet (P) and supernatant (S) after sonication for 0 min, 5 min, 10 min, 15 min for cross-linked cells was resolved on a 1% agarose gel and visualized by ethidium bromide.

Previous experiments have shown that DNA solubilization is corelated with protein solubilization. We observed optimal DNA and protein solubilization after 15 min of sonication which is in good agreement with **Nguyen et al 2015**. Unfortunately, in our first attempt to purify Ty1 IN HBH after 15 min sonication resulted in no yield of the protein suggesting that the sonication time was too long which might have damaged the protein integrity and inhibited its binding to the affinity column (Data not shown). Thus, we chose to sonicate our extracts to the minimal for 5 min as we observed rapid solubilization of both the DNA (Figure 4) and

proteins at this time point (Data not shown). Indeed, we saw that 5 min of sonication did not negatively affect the yield of the protein after purification.

1.2 Large scale purification of Ty1 IN HBH

We purified the Ty1 IN HBH from a large volume of cell culture (8l) grown to exponential phase. The cells were crosslinked by 1% Formaldehyde for 20 min to be able to keep weak and transient interactors **(Nguyen *et al*, 2015)**. As a negative control we purified an untagged Ty1 IN following the same TChAP procedure. A protein pattern was observed for the negative control suggesting the presence of many background proteins (Figure 5C). Previous TChAP experiments in our laboratory suggest the presence of various protein in the control experiment. Firstly, there are yeast biotinylated proteins such as Arc1 which are highly expressed and can bind to the column non-specifically (**Kim et al 2004**). Others include ribosomal proteins which are also highly expressed and some sticky proteins that are very often detected in proteomics experiment.

The purified proteins were resolved on a 4-12% gradient gel and stained with Coomassie blue (Figure 5A, 5B). As observed from the stained gels (Figure 5B), the yield of proteins obtained for the cultures grown at 30°C were detectable. So, the lane in the gel was excided into 3 pieces and sent for mass spectrometry identification of the co-purified proteins. On the other hand, purification of the Ty1 IN from cells grown at 20°C were extremely low and only detected by western blot and thus could not be sent for mass spectrometry analysis.

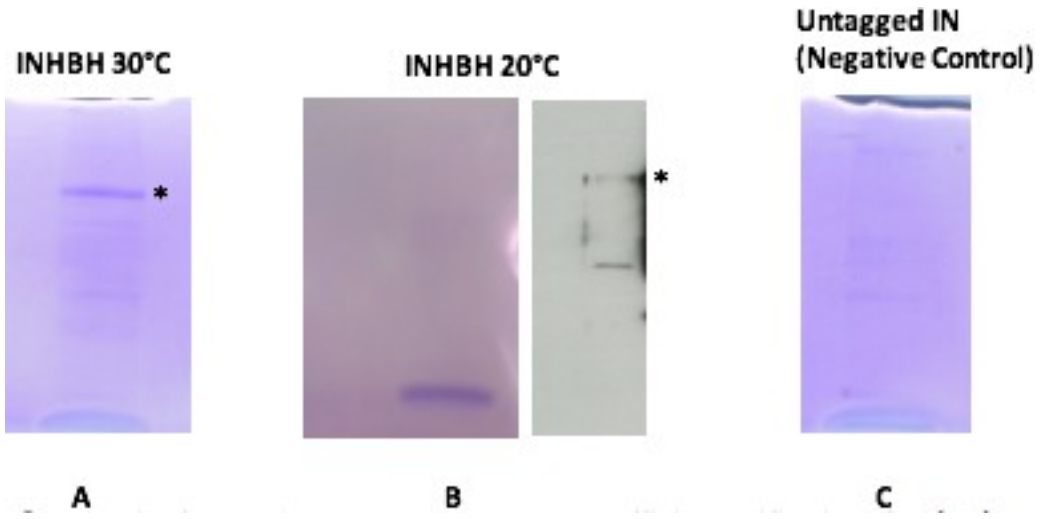


Figure 5: **Analysis of the Ty1 IN samples upon tandem affinity purification after *in vivo* crosslink.** Ectopically expressed Ty1 INHBH or untagged Ty1 IN were induced ON at 30°C and 20°C (A, B). Cell

were crosslinked and the proteins were purified sequentially by Ni-chelate chromatography and streptavidin beads. After purification samples were analyzed by Coomassie blue staining or western blot using anti-streptavidin antibody. Ty1 INHBH is marked with a black asterisk.

The low yield of Ty1 IN HBH was correlated with the lower expression of the Ty1 INHBH protein detected at low temperature (figure 3). Thus, to improve the expression of the protein at low temperature, we co-expressed the reverse transcriptase/RNase H domain (RT-RH) and Ty1 IN HBH *in vivo* based on evidences in the literature**. Wilhelm *et al* 2006** suggests that the IN and RT interact to form a stable complex during the entire reverse transcription step. In fact, the DNA plus strand and minus strand synthesis from the Ty1 mRNA is done by two IN/RT complexes. Also, **Tekeste *et al* 2015** demonstrated that there is a functional role of the IN/RT complex during the retroviral HIV replication. The HIV integrase and the reverse transcriptase physically interact during reverse transcription of the viral DNA. These data indicated that IN and RT/RH form a functional complex that might be required to maintain their stability.

An active Ty1 RT-RH was produced in *E. coli* when the last 115 aa from the C-terminal region of the IN was fused to the N-terminal region of the RT which is important for proper folding of the protein and preserving its activity **(Wilhelm *et al* 2000)**. So, we developed an expression vector expressing a fused IN/RT complex controlled by a TetOff promoter and co-expressed with full length Ty1 IN HBH. Indeed, this approach enabled us to achieve a higher yield of the

Ty1 IN at 20°C (Figure 6) suggesting that co-expressing the two proteins together stabilizes the Ty1 INHBH expression *in vivo*.

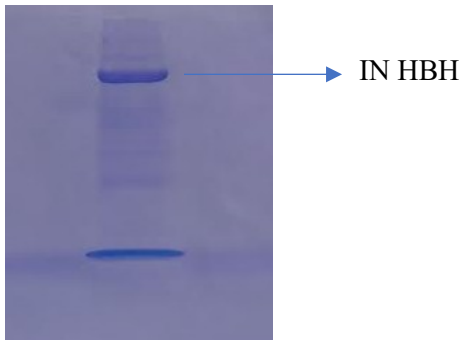


Figure 6: **Purified Ty1 INHBH upon co-expression with RTRH at 20°C.** Ty1 INHBH and RTRH were induced ON at 20°C. Cells were then crosslinked and protein were purified sequentially on Ni column and streptavidin beads. Purified proteins were analyzed by Coomassie blue staining.

1.3 Mass spectrometry data analysis to identify the positive targets of Ty1 IN

Identification of the co-purified proteins were done by mass spectrometry at the Mass spectrometry laboratory, IBB, Warsaw, Poland. The advancement in mass spectrometry has vastly increased the sensitivity of the technique which gives a large dataset of protein hits. The challenging task was to analyze the data and choose the putative targets and discard the false positives or non-specific interactors. To achieve this, we have set criteria to screen the list of MS identified proteins **(Nguyen *et al*, 2015)**. The criteria are as follows:

a) The protein should be identified by at least two independent peptides
b) The protein should not be present in the list of background proteins (present in the protein purify of the untagged protein)
c) The protein should be present in at least 2 replicates to be considered putative

These aforementioned criteria are basic and should be fulfilled by the protein to be considered as significant. But, to even refine the MS data other criteria were followed as well.

a) Protein complexes whose most of the subunits were identified were considered putative partners
b) Proteins that were present in most of the MS data in previous experiments of our laboratory to identify RNA Pol III associated proteins were considered sticky proteins and were considered less significant.
c) Proteins with low molecular weight and high protein coverage indicate that they are significant.

Approximately 500 proteins were identified by mass spectrometry. Thus, after stringent analysis and refinement of the raw data we were able to choose approximately 100 putative partners (**ANNEX**). However, this is a putative list and even deeper analysis and experiments are required to know whether they are true interactors of the Ty1 IN.

The partners mentioned in the following table no.1 seemed most interesting from evidences in the literature. All these proteins were either identified in previous genetic screen studies to identify Ty1 host factors or are present on Pol III genes or may be important for Ty1 targeting at specific sites in the genome. More details are presented later. These partners will be examined

further. However, we did not find any significant partners specific to different temperatures. One observation we made specifically at 20°C was that the Pol III transcription factors TFIIIC and TFIIIB were recovered in our TChAP experiments. This might suggest that at 20°C, Ty1 IN is associated with the Pol III transcription machinery whereas at 30°C Ty1 IN is found associated with free RNA Pol III polymerase. This is just a hypothesis and we need to test this further.

Protein	Coverage 30°C	Coverage 20°C	
RPC40	45,4	27,2	
RPO31	17,5	7,6	
RET1	13,1	13,2	
RPC82	22,5	18	
RPC53	10,4	10,9	
RPB5	30,2	0	RNA Pol III complex
RPC34	10,7	6,9	
RPC25	22,6	3,8	
RPC19	10,6	10,6	
RPA135	38,2	20	
RPA190	34	9,9	
RPA49	23,9	12	RNA Pol I complex
RPA34	19,7	10,7	
RPA43	17,5	8,6	
RPC34	10,7	6,9	
RPC25	22,6	3,8	
RPC19	10,6	10,6	
CKB1	11.5	4.7	
CKB2	15,9	0	Casein kinase II
CKA1	21,2	10,2	
CKA2	31,9	10	
RAP1	14.1	12.6	
REB1	20.1	18.5	Transcription factors
BRE1	11.7	17.3	Ubiquitin ligase
CEG1	10.5	15.9	
CET1	6.9	6	5'-mRNA capping enzyme
TUP1	11.6	19.8	Transcription repressor
CYC8	11.2	11.4	

Table 1: **Curated MS data of putative partners associated with Ty1 IN HBH at both 30°C and 20°C**. List of proteins identified by TChAP of Ty1 INHBH after *in vivo* crosslink. Protein peptides satisfying the criteria were chosen to be examined further. The coverage of the proteins identified by mass spectrometry under both purification conditions 20°C and 30°C are presented.

Nine subunits of RNA Pol III out of 17 subunits have been identified making it a strong interactor of Ty1 IN and serves as a positive control for our experiment. This is in good correlation with the literature where two hybrid assays demonstrate an association between Ty1 IN and Pol III *in vivo* and that this interaction is important for Ty1 integration upstream of RNA Pol III transcribed genes (**Bridier-Nahmias *et al* 2015).** Generally, we expect that large molecular weight subunits of a protein complex should be identified with a high coverage if they are associated with the bait protein and the small molecular weight subunits often don't get detected by mass spectrometry due to their small size. However, in our mass spectrometry data, we observed that the rather small molecular size proteins Rpc40, Rpc19 and Rpb5 were identified with a much higher coverage than the largest Pol III subunit C160 or C128 at both temperatures suggesting that there is a bias in this case and possibly the integrase and these specific Pol III subunits are present in very close proximity to each other in Ty1 IN-Pol III complex. This needs to be tested further with *in vitro* binding assays to know if there exists a direct physical interaction between these subunits and Ty1 IN. In fact, we later demonstrated a direct interaction between Ty1 IN and AC40 by *in vitro* assay which is in good alignment with the TChAP results.

We also observed that we were able to recover 8 out of 10 subunits of Pol I which makes it a putative target. Initially, we thought that Pol I might be a contaminant as it is highly expressed protein *in vivo*. Though Ty1 integration at ribosomal DNAs transcribed by Pol I haven't been studied extensively, one study by **Bryk et al 1997** indicates that Ty1 are targeted to rDNA repeats but a specialized chromatin structure in the rDNA leads to transposition and transcription silencing of Ty1 in this region. Moreover, The AC40 subunit is a shared subunit between Pol III and Pol I, therefore, it was not surprising to recover the Pol I complex in the MS data. It would be interesting to gain deep insights into how Pol I might regulate Ty1 integration *in vivo*.

Rap1 is a DNA binding protein involved in transcription activation, repression, chromatin silencing and telomeric length maintenance **(Shore *et al* 1987).** The Ty1 promoter region has RAP1 binding sites and Ty1 transcription is modulated by Rap1 (**Gray *et al* 1993**) along with eight other transcription factors. But, what is interesting about Rap1 is the fact that it is a highly conserved telomeric protein in yeast, protects chromosomal ends and promotes gene silencing **(Chen *et al* 2011).** Therefore, we hypothesized that may be Rap1 might regulate Ty1 integration in sub-telomeric region under the loss of interaction with Pol III.

Reb1 is yet another transcription factor of Pol I that is also present in sub-telomeric region of the yeast and inhibits telomerase activity and lengthening of telomeres in coordination with Tbf1 (**Berthiau *et al* 2006**). Thus, this serves as a good hypothesis that Reb1 might play a role in Ty1 targeting at rDNAs or might affect Ty1 integration specificity at sub-telomeres.

We also recovered both the subunits of the protein complex Tup1-Cyc8. Interestingly, it also regulates gene in the sub-telomeric region. It is a general transcription repressor and is composed of 1 Cyc8 subunit and 4 subunits of Tup1. The **Tup1-Cyc8** complex works in coordination with the histone deacetylases Hda1 and Rdp3 to repress the Flo1 gene in the sub telomeres (**Fleming *et al* 2014**). Histone deacetylase also have a role in Pol III transcription but haven't been investigated in detail. Thus, this enabled us to choose this for further analysis to know whether it affects Ty1 integration specificity at the tRNAs or sub-telomeres.

Cet1-Ceg1 are 5'-mRNA capping enzymes that have been shown to have higher occupancy on Ty1 retroelements when the RNA Polymerase II C-terminal domain is truncated (**Aristizabal *et al* 2015**). We have recovered both the subunits of the complex with high coverage even though they are small molecular weight proteins RNA Pol II-CTD truncation has been shown to enhance Ty1 expression and retrotransposition. It would be interesting to investigate if these capping enzymes contribute to the stability of the Ty1 mRNA and regulate their expression *in vivo*.

Bre1 is a ubiquitin transferase found in the genetic screen as a negative regulator of both Ty1 and Ty3 mobility (**Nyswaner *et al* 2008, Griffith *et al* 2003, Irwin *et al* 2005**). Bre1 mutants showed enhanced Ty1-His3AI and Ty3-His3AI *in vivo* retromobility. In coordination with Rad6 it is involved in the ubiquitination of H2B (**Nyswaner *et al* 2008)**. As Ty1 elements are targeted to the H2A/H2B interface of the nucleosomes at the tRNA genes, it would be interesting to investigate whether Bre1 ubiquitination alters the structure of the nucleosomes and negatively affects Ty1 retrotransposition *in vivo*.

CK2 is a serine-threonine kinase having myriads of substrate and have been identified as a positive regulator of Pol III transcription in yeast. They are made of four subunits namely Cka1, Cka2, Ckb1 and Ckb2. Out of the many kinases present in yeast, we have recovered all the four subunits of the complex with a high coverage even though each subunit is a low molecular size

protein. This makes the Ck2 complex a significant Ty1 IN binding partner. From evidences in the literature *CKB2* have been found to negatively regulate Ty1 mobility *in vivo* (**Nyswaner et al 2008**).

1.4 Validation of the Ty1 IN targets by *in vivo* co-immunoprecipitation

To further validate whether these proteins are *bona fide* partners of the Ty1 IN we exploited another biochemical assay such as i*n vivo* co-immunoprecipitation (CoIP) to verify whether these proteins are indeed associated with Ty1 IN *in vivo* (Figure 7). For this experiment, strains expressing TAP-tagged proteins were revived from the TAP yeast library. In such strains, the Ty1 INHBH was expressed ectopically through a plasmid controlled by the TetOff promoter. Cells were grown overnight in glucose containing synthetic medium in the presence of doxycycline (1µg/ml) and diluted accordingly in the absence of doxycycline for inducing the expression of the Ty1 INHBH ON so that cells were exponential phase the following day, exactly the same way as was done in the TChAP procedure. CoIP was done on protein extracts prepared from crosslinked cells or from non-crosslinked cells.

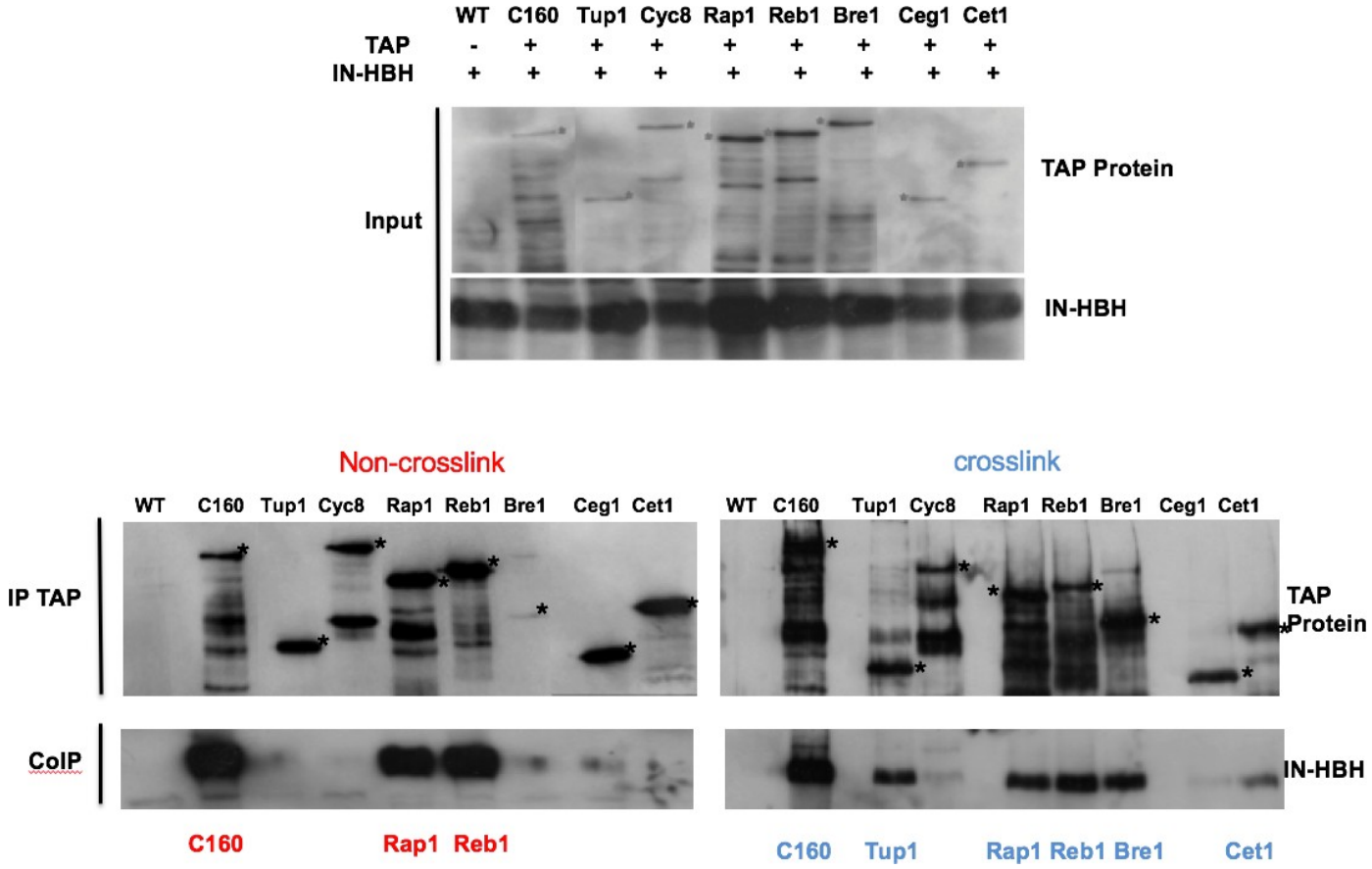


Figure 7: **Validation of Ty1 partners by co-immunoprecipitation.** Ty1 IN HBH was ectopically expressed and induced ON in TAP-tagged strains or WT strains as indicated. Immunoprecipitation of TAP tagged proteins were performed by IgG beads. Immunoprecipitated and co-immunoprecipitated proteins were analyzed by western blotting using anti-Streptavidin and anti-TAP antibodies.

The Pol III complex (C160-TAP) serves as the positive control in this experiment and as expected can be coimmunoprecipitated with Ty1 INHBH in protein extracts prepared from non-crosslinked cells, suggesting that Ty1 IN and Pol III form a complex *in vivo*. A WT strain expressing only the Ty1 INHBH and no TAP tagged protein was used as a negative control showing that the induced Ty1 INHBH in these strains doesn't bind to the IgG beads (used to immunoprecipitated the TAP tagged proteins) non-specifically. However, CoIP performed on protein extracts prepared from crosslinked cells showed much more positive candidates than from non-crosslinked cells. This is in good correlation with TChAP suggesting that this proteomics approach is good and generated positive interactors. Targets that co-purify with the Ty1 INHBH *in vivo* without cross link indicate that they can form a complex with the IN and this association is not disrupted throughout the CoIP procedure. Interestingly, Rap1 and Reb1 appeared to coimmunoprecipitate with Ty1 INHBH *in vivo*. On the other hand, some targets such as Bre1-TAP, Tup1-TAP, Cyc8-TAP and Cet1-TAP that were found positive only when cells were crosslinked suggesting that they might be weak or transient interaction and we lose them during the CoIP procedure. Another possibility, could be that these targets are not present in a complex with Ty1 INHBH, instead could be present on genes in close proximity to Ty1 IN. For example, if we consider Bre1, it interacts with Rad6 to ubiquitinate H2B **(Nyswaner *et al* 2008)**. So, we could wonder that Bre1 is physically associated to histones on nucleosomes and not to Ty1 IN. And as Ty1 prefers nucleosome surfaces for integration, we only observe Bre1 and Ty1 IN association upon cross link because they are present in close proximity.

1.5 Demonstrating direct interaction between Ty1 IN and the targets

In vivo co-immunoprecipitation demonstrates the existence of complexes of these proteins in the cell. To demonstrate a direct physical interaction between two proteins, one method is an *in vitro* binding assay. Therefore, we performed an *in vitro* GST pull down assay for this purpose.

We cloned the protein targets into the pGEX bacterial expression vector expressing a GST fusion protein that could be easily captured on GST sepharose beads. The plasmids were then transformed into BL21(DE3) bacterial cells. The expression of the GST fusion proteins was regulated by a lac operon and a high induction of expression was achieved when BL21(DE3) bacterial cells were treated with IPTG. We found that the induction of the GST fusion proteins was best observed at 30°C when treated with 0.5 mM IPTG for 3 h.

The next step was to express a soluble and active Ty1 IN in *E. coli*. Ty1 IN is mainly insoluble like the HIV1 -IN and thus, we had to find measures to solubilize the protein. We came across two fusion tags in the literature that enhance the solubility of a protein and preserves its activity *in vivo,* namely Fh8 and SSO7d (**Costa et al 2013, Min Li et al 2014**). Fh8 is 8KDa tag that enhances solubility and facilitates binding of the tagged protein on a hydrophobic matrix. And, SSO7D is a 7 kDa non-specific DNA binding protein from *S. solfataricus* which was used by **Min Li *et al* 2014** to construct a soluble and hyperactive HIV Integrase that showed highly efficient concerted integration *in vitro*. The SSO7D HIV IN was present predominantly in monomeric forms compared to WT which mostly aggregated. This tag has also been used on DNA polymerase to increase its processivity and performance *in vitro* (**Wang *et al* 2004**). We expressed the Ty1 IN either as a 6His-Fh8 or 6His-SSO7d N-terminal fusion in *E. coli* from a Ty1 IN codon optimized bacterial plasmid vector. A small C-terminal EPEA tag was also put at the C-terminal end of the tagged Ty1 IN to be able to be detected by anti-EPEA antibody. We achieved a highly soluble 6His-Fh8-Ty1 IN and 6His-SSO7d-Ty1 IN in *E. coli* compared to an untagged Ty1 IN. For the moment, we haven't tested the activity of these forms yet.

The GST fusion protein targets expressed in *E. coli* were captured on GST sepharose beads and thereafter the beads bound to the GST fusion proteins were incubated with a soluble extract containing 6His-SSO7d-Ty1 IN-EPEA or 6His-Fh8-Ty1 IN-EPEA for either 3h or overnight. As a negative control the GST protein was included in the experiment showing that the tagged Ty1 IN doesn't bind to the GST tag non-specifically. Dst1 which was not found to be associated with Ty1 IN *in vivo* was chosen as a negative control also. The eluted fractions from the beads were analyzed by Coomassie blue staining and western blot to see whether the GST tagged targets and SSO7d-Ty1 IN-EPEA or Fh8-Ty1 IN-EPEA can co-elute.

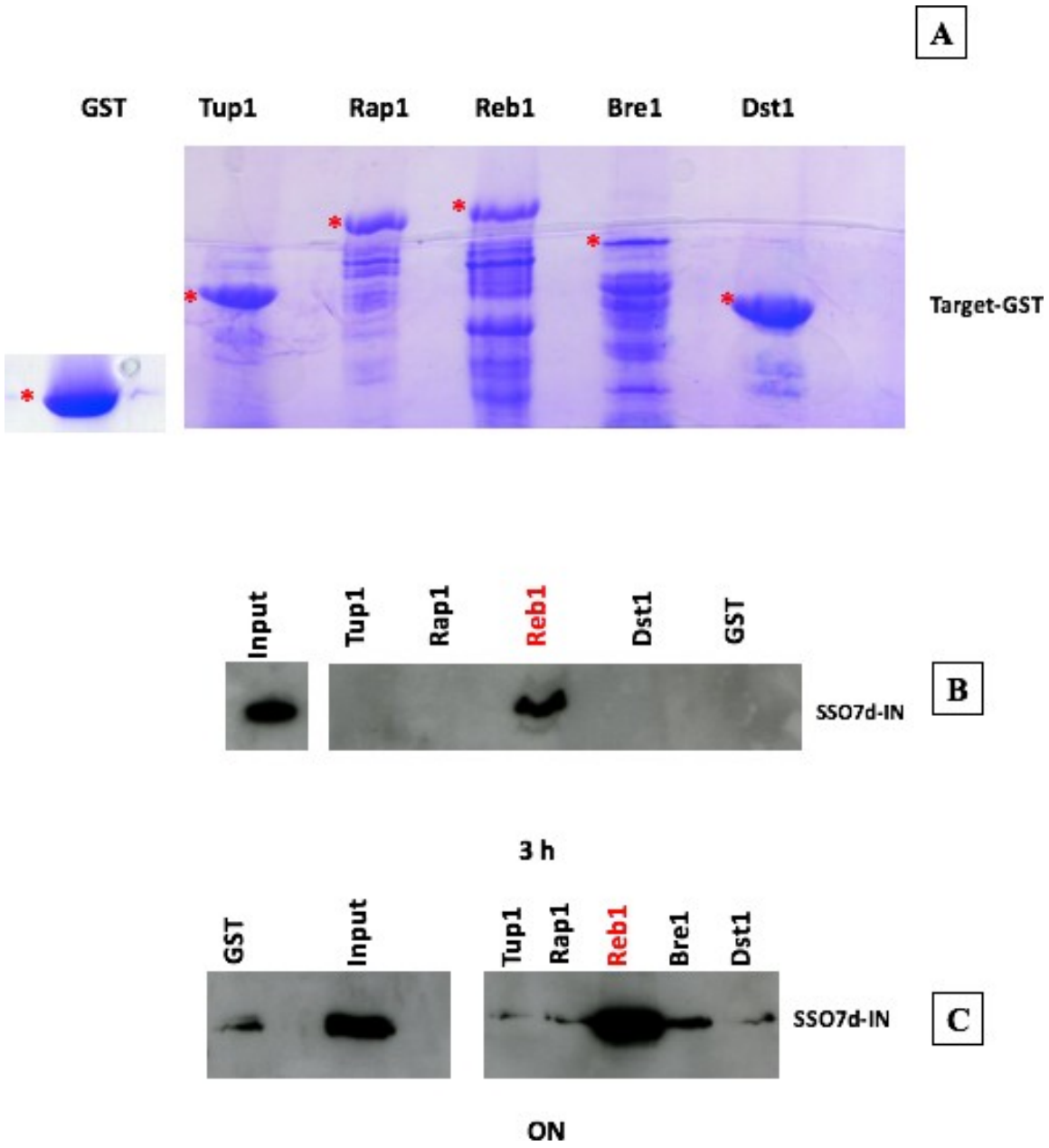


Figure 8: **REB1 directly interacts with Ty1 IN *in vitro*.** Target proteins were expressed as GST fusion proteins in *E.coli* and purified on GST sepharose beads (Upper panel). Purified 6His SSO7d- Ty1 IN-EPEA was incubated with the target bound GST sepharose beads for either 3 h (B) or ON (C). The presence of Ty1 IN -EPEA was detected by western blot using anti-EPEA antibody.

As seen in Figure 8, in our experimental conditions only REB1-GST was found to interact with Ty1 IN *in vitro*. However, this interaction could only be detected using 6His SSO7d Ty1 IN-EPEA and not with the 6His Fh8 Ty1 IN-EPEA (Data not shown). Initially, we thought the discrepancy was due to different oligomeric forms of the two tagged Ty1 IN forms.

During the last few months, extensive experiments were performed to try to obtain a soluble and unaggregated from of Ty1 IN in *E. coli* more particularly by **Celia GOUZERH** and in the lab of **Dr. Juan REGUERA** in Marseille, France who are also involved in the project from a recent ANR grant obtained to decipher the structure of the Ty1 IN. She identified various factors that resulted in the formation of aggregates during the purification procedure such as high protein concentration, presence of DNA and high salt concentration. It is to be noted that Ty1 IN binds strongly to DNA non-specifically which is why the use of benzonase to degrade DNA is necessary while extract preparation. Unfortunately, in the beginning when I performed the GST pull down assay, protein extracts of either 6His-Fh8-Ty1 IN-EPEA or 6His-SSO7d- Ty1 IN-EPEA were not prepared keeping in mind the above-mentioned conditions. Hence, we do not know the structure of the tagged Ty1 IN in the extract (multimers or aggregates). Now, that the conditions are set up, it would be interesting to perform the GST pull down assay with a more defined Ty1 IN complex. However, *in vitro* assays are quite difficult with Ty1 IN**.** In the study by **Cheung et al, 2016,** the direct interactions obtained between C31, C11 and C53 subunits of Pol III and Ty1 IN were not strong and less convincing. In our attempt to demonstrate a direct interaction between the AC40 subunit and Ty1 IN *in vitro* we were unsuccessful using the GST pull down assay. On the other hand, an *in vitro* CoIP on extracts from cells co-expressing the AC40 and Ty1 IN gave a positive interaction. Thus, we see that experimental conditions and approaches affect the outcome of the result. We could even co-express the targets with Ty1 IN and perform a similar *in vitro* CoIP.

CONCLUSION

During the first year of my PhD study, we were able to set up optimal conditions for the TChAP procedure for identifying the associated partners of the Ty1 IN. We have found several protein partners that were also identified in genetic screen studies (**Scholes et al 2001, Nyswaner et al 2008**). However, we also identified some very interesting novel partners never identified before such as Reb1, Sub1, Dst1 and more. Some of these partners are present on tRNA genes and have been shown to regulate Pol III transcription. Therefore, they can be an important link between the Pol III machinery and Ty1 integration *in vivo*. Pol III was found to be a positive interactor as expected and interestingly, we even recovered the Pol I complex as an interactor of Ty1 IN for the first time.

The validation of the targets by biochemical assays prove that the TChAP procedure is a good proteomics approach to identify Ty1 IN host partners. The next step here is to investigate whether these identified partners can play a role in Ty1 retrotransposition *in vivo*. Although the interaction assays seem very interesting with several partners like Rap1 and Reb1, we decided to progress further to study the role of casein kinase II (CK2) in Ty1 retrotransposition because it has already been identified as a Ty1 repressor in a genetic screen study (**Nyswaner et al 2008**) and has been extensively studied as a regulator of RNA Pol III transcription in yeast. As there is a strong interplay between Ty1 integration and Pol III transcription it would be worth deciphering its molecular role in Ty1 mobility. The raw mass spectrometry data is still being reviewed and analyzed meticulously to be able to refine the data so that we don't overlook any significant targets.

2. Characterization of Casein kinase 2 in Ty1 retrotransposition (Collaboration with Dr. Pascale Lesage)

In this part of the project, data has been presented from both the lab of **Dr. Pascale LESAGE** and ours. Within the tenure of my PhD thesis, I focused on delineating the role of Casein kinase 2 (CK2) role on Ty1 retrotransposition. The yeast CK2 is a serine-threonine kinase made of 2 catalytic subunits (Cka1, Cka2) and 2 regulatory subunits (Ckb1, Ckb2) having innumerable substrates in the cells including the RNA Pol III transcription machinery. They are encoded separately by the *CKA1*, *CKA2*, *CKB1* and *CKB2* genes. The CK2 holoenzyme in yeast requires both the regulatory subunits for its catalytic activity (**Kubinski et al 2007**). It has been showed that 4 forms of CK2 exists *in vivo* (**Domanska et al 2007**). The four forms are as follows:

Cka1, Cka2, Ckb1, Ckb2 = $\alpha\alpha'\beta\beta'$

Cka2 dimer = $\alpha'\alpha'$

Cka1, Cka1, CKb1, Ckb2 = $\alpha\alpha\beta\beta'$

Cka2, Cka2, Ckb1, Ckb2 = $\alpha'\alpha'\beta\beta'$

In the budding yeast, deletion of either of the catalytic subunits CKA1 or CKA2 is viable whereas deletion of both the catalytic subunits is synthetic lethal, suggesting that CK2 activity is absolutely necessary for cell survival (**Litchfield 2003**). The regulatory subunits stimulate the catalytic activity, stabilize the CK2 tetramer and acts as a scaffold protein to interact with its partners (**French et al 2007**). The CK2 holoenzyme structure are highly conserved across various species. CK2 have been demonstrated to play a wide role in cellular processes such as gene control, cell cycle progression, cell proliferation, apoptosis, differentiation and transformation (**Olsten and Litchfield 2004**). Though CK2 itself, is regarded as a serine/threonine kinase, it has been shown to phosphorylate tyrosine residues both *in vivo* and *in vitro* (**Wilson et al 1997)** indicating that it might have a dual specificity. CK2 doesn't only phosphorylate its targets. It has been shown that the CK2 regulatory subunit in human are auto phosphorylated at the N-terminal domain and abolishing the mutation have been shown to increase the stability of the regulatory subunit suggesting that auto phosphorylation is linked to the stability of the protein (**Litchfield et al 1991**). In yeast, it has been shown that the Ckb1 regulatory subunit of the CK2 complex is differentially phosphorylated by a kinase Kns1 under stress conditions reducing its occupancy on tRNA genes (**Sanchez Casalongue et al 2015**).

CK2 is likely to be involved in regulating the transcription of all Pol I, Pol II and Pol III genes *in vivo*. CK2 regulates coding genes by acting as a transcriptional activator **(Litchfield 1993)** and Pol I genes by acting on the upstream binding factor (**Voit et al 1995**). CK2 was amongst the very first kinases to positively regulate RNA Pol III transcription in yeast Moreover, CK2 is present on Pol III transcribed genes, specifically the tRNAs in yeast (**Graczyk *et al* 2011**). In human cells, it has been demonstrated that CK2 phosphorylates and physically interacts with the TFIIIB complex *in vivo* to facilitate its recruitment by TFIIIC (**Johnston et al 2002**). In yeast, the TBP subunit of TFIIIB is phosphorylated by CK2 which is necessary for its activity (**Ghavidel et al 1999**). And, during DNA damage, signals are transduced by CK2 to TFIIIB to repress Pol III transcription **(Ghavidel *et al* 2001)**. Maf1, the Pol III repressor was also shown to be directly phosphorylated by CK2 *in vitro*. When yeast strains are subjected to favourable conditions from stress conditions, CK2 is necessary for the release of Maf1 from tRNA genes to initiate Pol III transcription (**Graczyk *et al* 2011).** Thus, we see that the CK2 acts in several ways to positively regulate Pol III transcription *in vivo*.

CKB2 was also identified in a genetic screen that negatively regulated Ty1 retromobility by **Nyswaner *et al* 2008**. The *CKB2* deletant mutant gave a 60-fold increase in Ty1 retromobility compared to the WT. Their data suggests that CK2 might act at post transcriptionally to regulate Ty1 retrotransposition as CKB2 deletion mutant did not have any effect on the Ty1 mRNA level. Then, **Irwin *et al* 2005** also showed that Ty3 retromobility is also enhanced upon deletion of *CKB2*. As both Ty1 and Ty3 are homologous, prefer Pol III transcribed genes as their preferred sites of integration and CK2 regulates both, we wondered that the molecular role of CK2 should be deciphered soon. Recently, in a proteomics screen by **Cheung *et al*, 2015,** Ckb2 was identified as an interactor of the Ty1 IN. Nevertheless, none of the published work shows the molecular mechanism behind the role CK2 plays in regulating Ty1 retrotransposition.

2.1 Protein and genetic interaction between CK2 and Ty1

2.1.1 CK2 is physically associated to Ty1 IN *in vivo*

2.1.1a In vivo coimmunoprecipitation

We identified peptides of all the four subunits of the CK2 in our mass spectrometry data with high coverage and in all of the TChAP experiments. To validate whether Ck2 is a *bona fide* partner of Ty1 IN, we demonstrated that Ty1 IN is associated with the CK2 complex *in vivo* by co-immunoprecipitation (CoIP) (Figure 9) and yeast two hybrid assay (Figure 10). For the CoIP experiment, The Ty1 IN HBH was ectopically expressed through a centromeric plasmid controlled by a tetoff promoter in the strains containing genomically TAP tagged Ck2 subunits. Strains were grown in classical glucose medium at 30°C overnight in the presence of doxycycline (1µg/ml) to repress the IN HBH expression. The, the cells were diluted accordingly to induce the expression of Ty1 INHBH overnight in the absence of doxycycline the same way as that in the TChAP procedure.

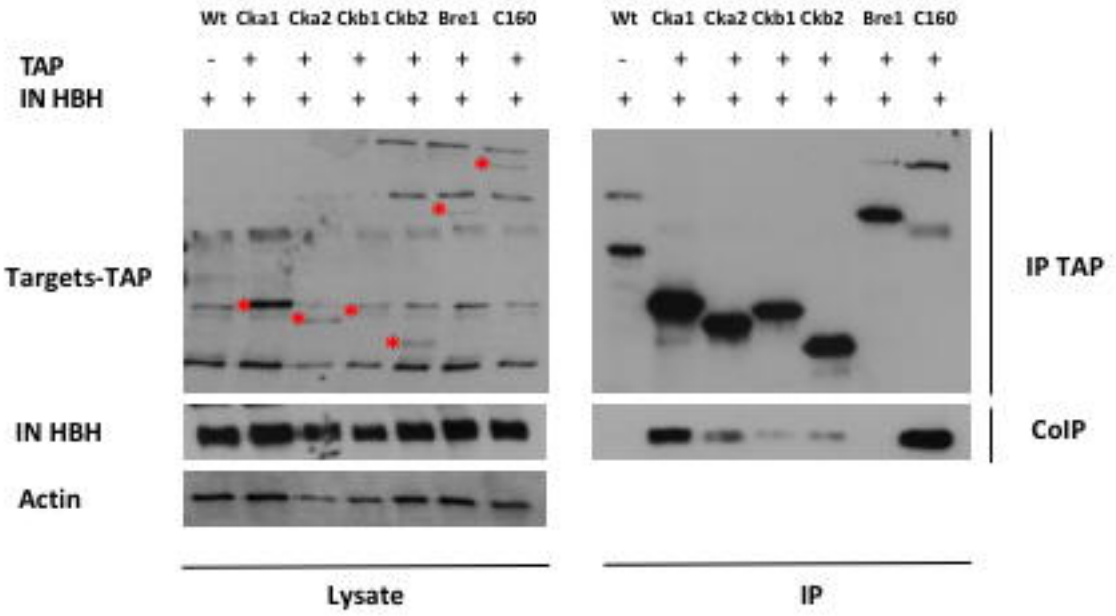


Figure 9. **Ty1 IN is associated with CK2 *in vivo*.** Ty1 IN HBH was ectopically expressed in WT or CK2-TAP tagged strains as indicated. TAP tagged proteins were immunoprecipitated with IgG beads. Immunoprecipitated and co-immunoprecipitated proteins were analyzed by western blotting with anti-TAP and anti-streptavidin antibodies. β-actin was used as a loading control of the input.

Figure 9 shows that under our experimental condition Ty1 IN HBH co-purifies with CK2-TAP *in vivo*. The Pol III complex (C160-TAP) was used as a positive control and Bre1-TAP served as a negative control showing that the Ty1 IN HBH is not a sticky protein and doesn't bind non-specifically to any TAP tagged proteins. The association between Ty1 IN HBH and CK2-TAP is not as strong as Pol III suggesting that it is a stronger binding partner and more molecules of the Ty1 IN HBH and Pol III complexes can be obtained. Since CKA1 does not exist as a free protein *in vivo* (**Domanska et al 2007**) its interaction with Ty1 INHBH suggest that the IN is associated with the tetrameric form of CK2. Thus, the association is also detected when the TAP tag is present on all the other CK2 subunits. However, the discrepancy in the CoIP for all the CK2 subunits suggest that the presence of the TAP tag can influence the formation of complex between CK2 and Ty1 IN.

2.1.1 b Yeast 2 hybrid assay

To determine more precisely which CK2 subunits interact with Ty1 IN we performed a yeast 2 hybrid assay. It showed that there is an interaction between CK2 and Ty1 IN *in vivo* (Figure 10). We have used two reporter genes in the assay to select positive clones, *HIS3* and *LacZ.* The full length Ty1 IN is fused to the Gal4 activation domain and the subunits of CK2 are fused to the Gal4 DNA binding domain. A positive interaction between the two proteins will activate transcription of the *HIS3* gene and the *LacZ* gene. Under the experimental condition, we observed no interaction when the subunits Cka1 and Cka2 were bound to the Gal4 DNA binding domain. Ckb1 gave blue coloration in the negative control, suggesting that the promoter was leaky driving transcription of the reporter gene even in the absence of the bait protein. Only, the Ckb2 subunit gave a positive interaction as the blue coloration was distinctly visible upon interaction with Ty1 IN. This suggests that Ckb2 interaction might be direct with Ty1 IN, though *in vitro* binding data is necessary to confirm it.

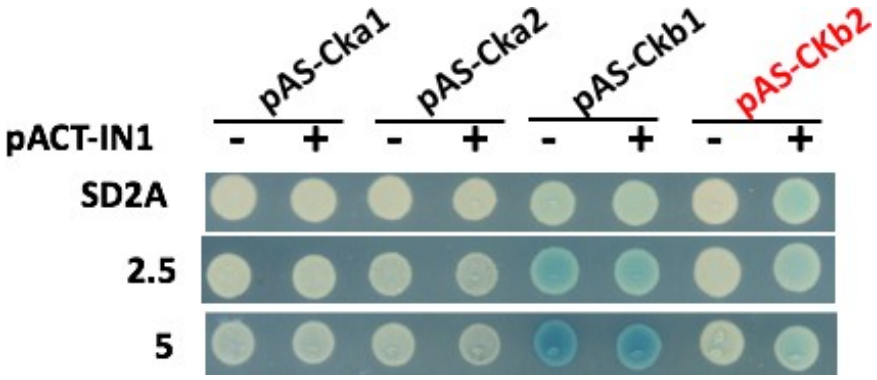


Figure 10. **Ty1 IN interaction with Ckb2 is detected by yeast two hybrid assay.** Yeast strain Y190 carrying empty vector pACT2 (negative control) or pACT2-IN were co-transformed with plasmids encoding fusions of the Gal4 DNA binding domain with the four subunits of CK2. The yeast cultures were spotted on plates having varying concentrations of the *HIS3* gene product competitive inhibitor, 3AT (0 mM, 2.5 mM and 5 mM). Patches of cells were overlaid on X-gal agarose to reveal *β*-galactosidase activity (blue colour). This experiment was performed by **Dr. Christine CONESA** and **Carine CHAGNEAU**)

2.1.2 CK2 has a repressive effect on Ty1 mobility *in vivo*

The next step was to investigate what effect does CK2 have on the rate of Ty1 retrotransposition *in vivo* by the Ty1 retromobility assay (**Curcio 2015**). The Ty1 retromobility assay is explained in a pictorial format (Figure 11).

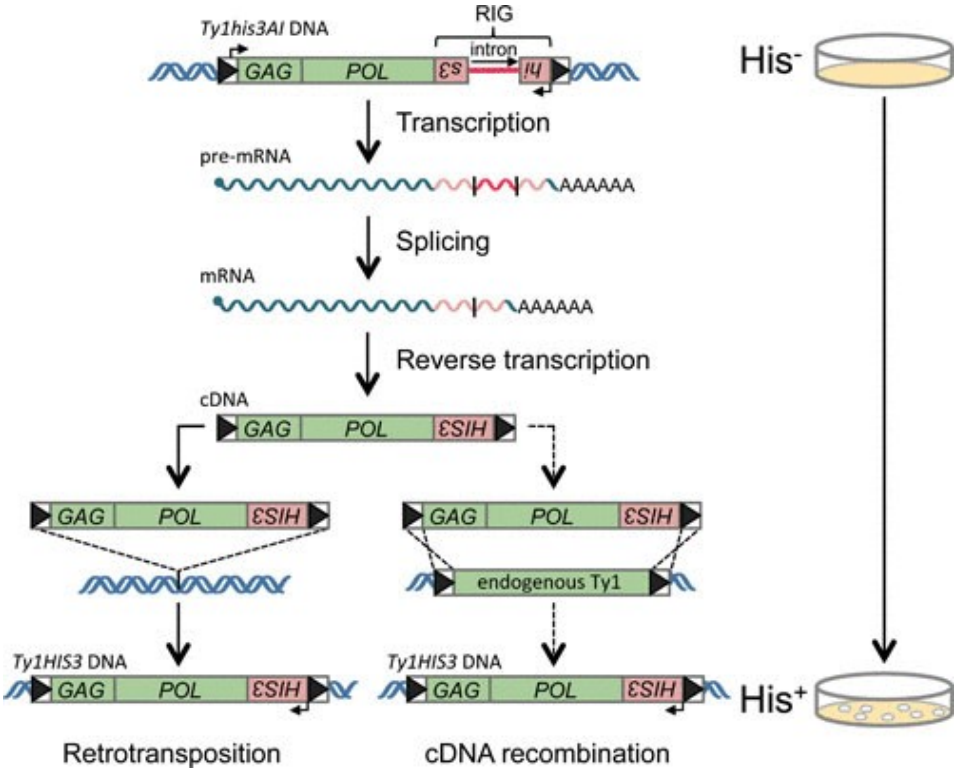


Figure 11. **Retromobility assay using the RIG**. A chromosomal Ty1 element tagged with the *his3AI* RIG undergoes precise splicing leading to the expression of a functional *HIS3* gene after cDNA integration into the host chromosome. The dashed lines represent the low frequency of Ty1 cDNA recombination in wild-type cells. Cells that sustain a Ty1*HIS3* retromobility event give rise to His^+ colonies (The Image is adapted from **Curcio et al 2015**).

His3AI gene is a selectable marker gene known as a reporter indicator gene (RIG) where an artificial intron is introduced. The RIG is inserted into the Ty1 element at the 3' untranslated region such that the Ty1 and the intron inserted marker gene are in opposite transcriptional orientation. Transcription of this element puts the artificial intron in a spliceable orientation. After the intron is spliced from the Ty1 transcript, reverse transcription and cDNA integration of the element in the chromosome renders a functional copy of the *HIS3* gene. Cells harbouring a functional Ty1 *His3* integration are His+ cells and can be selected on His- plates. (**Curcio and Garfinkel 1991**).

Different single and double yeast mutants of the CK2 holoenzyme were developed (by **A. BARKOVA**) containing a single chromosomally integrated functional Ty1-*His3AI* element. It should be kept in mind that other active endogenous Ty1 elements are also present in the genome. Deletion of the two catalytic subunits simultaneously was lethal to the cells whereas all the other mutants were viable. WT strain and mutant CK2 mutant strains containing a single Ty1 *His3AI* element were grown ON at 30°C and the following day diluted and grown at 20°C for 72 h. The cells are then plated on YPD to determine the total number of living cells and His- plates to screen the His+ cells and calculate the rate of Ty1 retromobility *in vivo*.

The assay showed that the frequency of Ty1 mobility altered by several folds in specific CK2 mutants compared to the wild type strain. A difference of 20-fold of Ty1 retrotransposition compared to the WT strain is considered a significant change. The catalytic subunit, Cka1 does not affect the frequency of transposition alone or in combination with either of the regulatory subunits. Cka2 or Ckb1 single deletion mutants doesn't have any effect either. Only, the single mutant Δckb2 gave a 21-fold increase in the Ty1 mobility which is in good correlation with what was shown previously by **Nyswaner et al 2008** where Ckb2 was identified as a repressor of Ty1 mobility. The effect of mobility only in the Ckb2 single mutant probably comes from the fact that Ckb2 is a binding partner of Ty1 IN and might repress Ty1 mobility by forming subcomplexes with the catalytic subunits. However, Cka2 deletion in combination with either of the regulatory subunits enhanced Ty1 mobility by either 115-fold or 170-fold. In the

absence of the regulatory subunit, no stable CK2 holoenzyme was shown to exist *in vivo*. This data together suggests, that Cka2 might repress Ty1 mobility only in the presence of a stable CK2 holoenzyme. As the effect was significantly highest in the Δcka2Δckb2 mutant all further analysis has been done in this particular deletion mutant.

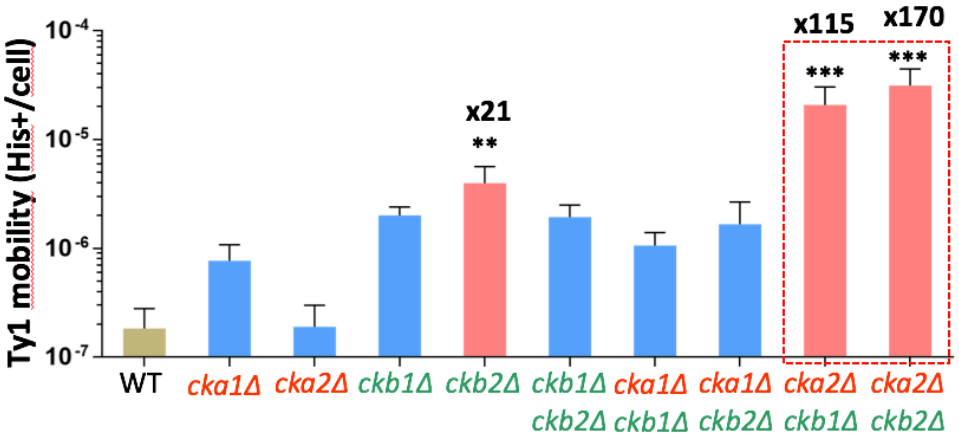


Figure 12. **CK2 represses Ty1 mobility *in vivo***. The transposition rate was determined by monitoring the number of His+ colonies according to the Ty1 *His3AI* mobility assay. Means of median and standard deviations of at least 3 independent experiments are plotted (3 independent cultures per experiment). Mann-Whitney test was used for statistical analysis and significant differences relative to WT cells are indicated (*p≤0.05, **p≤0.01, ***p≤0.001) (Data by Anastasia BARKOVA, PL lab).

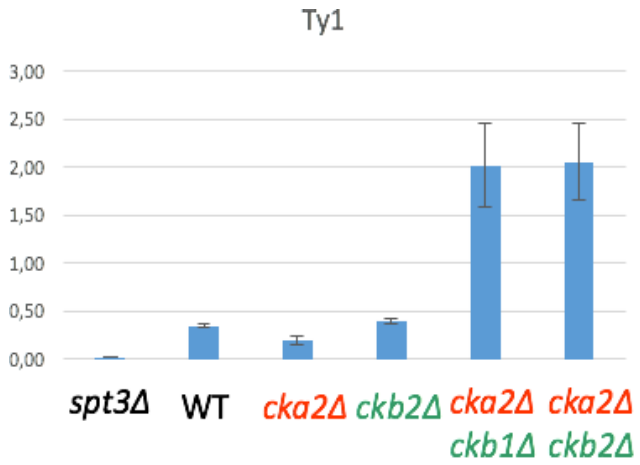
2.2 Ty1 expression in the absence of CK2

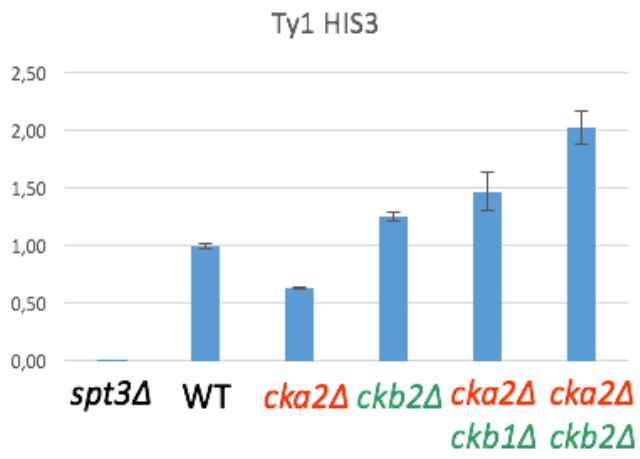
2.2.1 CK2 doesn't affect the Ty1 *His3AI* mRNA level but represses the global Ty1 mRNA level

To investigate further as to know whether the increase in Ty1 mobility is corelated to an increase in Ty1 expression *in vivo*, **A. BARKOVA** performed a qRT-PCR assay to determine the steady-state levels of both the Ty1 His3AI mRNA and the global Ty1 mRNA in the absence of the Ck2 subunits. Three independent cultures of each strain were grown to exponential phase and total RNA was extracted. qRT PCR was performed on the extracted RNA by specific primers namely Ty1-5', Ty1 HIS3 to amplify the total Ty1 mRNA and the Ty1 His3AI mRNA. The location of the primers on the Ty1 RNA are represented in the figure below. It should be noted that the primer used for amplifying the total Ty1 mRNA is located at the 5' end of the

Ty1 element amplifies only the Ty1 mRNA transcript and not the other RNA transcripts from the Ty1 DNA such as the Ty1 antisense RNA or the Ty1i RNA.

Ty1 *His3AI* mRNA level was not significantly stimulated in the CK2 mutants suggesting that CK2 has no effect on transcription of Ty1 *His3AI.* This indicates that CK2 might regulate its expression post transcriptionally. This is in good agreement with the data from **Nyswaner et al 2008** showing that most of the identified Ty1 repressor genes regulate Ty1 mobility post-transcriptionally. They even demonstrated by northern blotting that there was no significant difference in the global Ty1 mRNA or Ty1 *His3AI* mRNA levels in ΔCkb2 compared to the wild type. However, qRT-PCR data of the global Ty1 mRNA level in the CK2 mutants gave a 6-fold increase in Ty1 mRNA specifically in the Δcka2Δckb1 and Δcka2Δckb2 suggesting that CK2 might also regulate Ty1 mobility on a transcription level. The mRNA levels are represented in figure 13. Nevertheless, it is difficult to conclude that just a 6-folds increase in the Ty1 mRNA level accounts for the 170-folds increase in the Ty1 retrotransposition in this mutant suggesting that CK2 is more important in post transcriptional regulation of Ty1 expression *in vivo*.





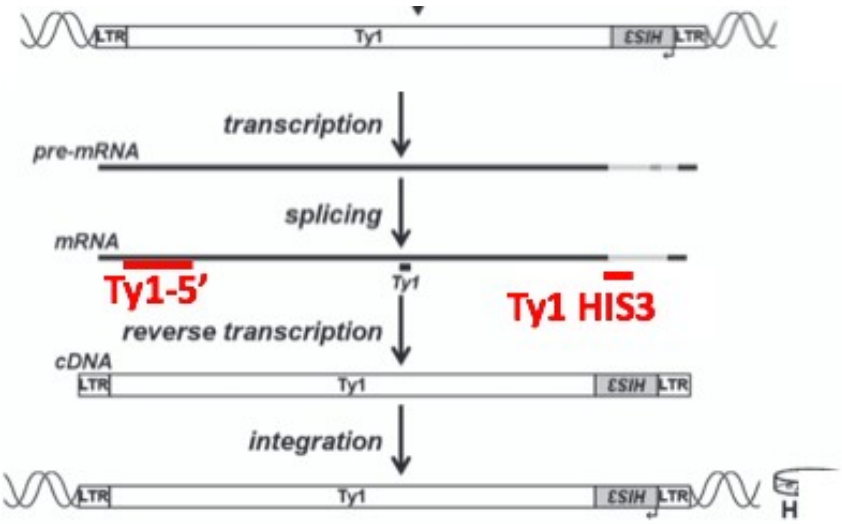


Figure 13.**CK2 affects global Ty1 mRNA level but not Ty1 His3AI mRNA level**. The level of global Ty1 mRNA and Ty1 His3AI mRNA was determined by qPCR on total RNA extracted from three independent cultures grown at 20°C. The Δspt3 serves as the negative control. The lower panel shows the location of the two primers Ty1-5' amplifying the total Ty1 RNA and the Ty1 HIS3 primer amplifying uniquely the Ty1 His3 transcript (Data by **Anastasia BARKOVA, PL lab**).

2.2.2 Endogenous Ty1 IN expression is slightly more abundant in the absence of CK2

After checking on the Ty1 mRNA level *in vivo*, the next step was to see whether the increase in retrotransposition in the ΔCka2ΔCkb2 strain could be related to a change in the Ty1 protein expression. This experiment was performed by **A. BARKOVA** where cells were grown at 20°C in classical glucose synthetic medium to exponential phase. The total protein extracts from cells were prepared by TCA precipitation and Ty1 IN was visualized by western blotting using the anti-IN antibody (8B11). Interestingly, a slight abundance of the endogenous Ty1 IN is observed in the absence of CK2. This suggests that there might exist a problem of stability of the Ty1 IN in the WT strains compared to in the absence of the CK2 holoenzyme. However, this is just a hypothesis and the stability of Ty1 IN needs to be tested in the absence of CK2. The TCA extract showing the level of endogenous Ty1 IN expression in all the CK2 yeast mutants grown at 20°C are shown in figure 13. However, in relation to the stimulation of global Ty1 mRNA it is difficult to conclude that a slight increase in the endogenous Ty1 IN expression in the absence of CK2 leads to such an increase in the Ty1 retrotransposition *in vivo*. Though Ty1 IN is a limiting factor, we couldn't rule out the hypothesis that CK2 might affect the other

polyproteins as well. Currently, the expression of the other Ty protein (Gag, RT) in Ck2 mutants are being investigated. Furthermore, expression of endogenous Ty1 IN was also similarly higher in Ck2 mutant strains grown at 30°C (Data not shown).

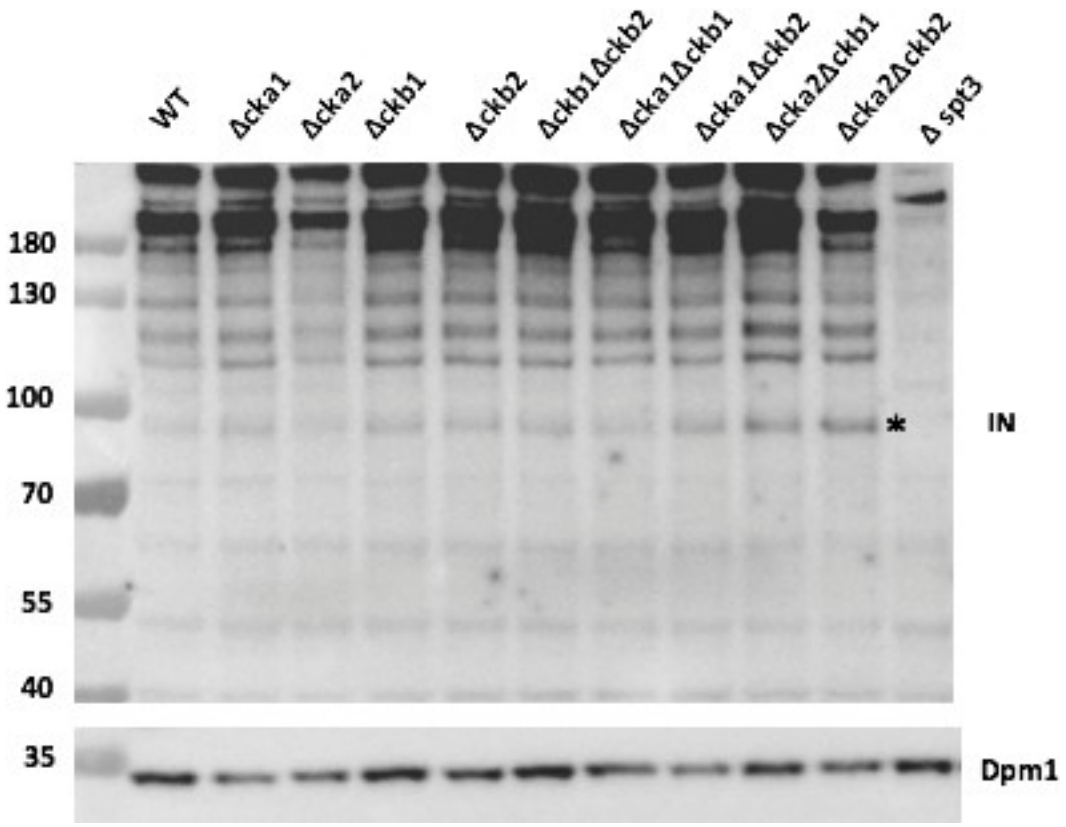


Figure 14. **Endogenous Ty1 IN is slightly more abundant in the absence of CK2**. Total protein extract was prepared by TCA precipitation from cells grown at 20°C to exponential phase. The endogenous Ty1 IN (marked by a black asterisk) was visualized by monoclonal anti-IN antibody (8B11). The Δspt3 strain serves as a negative control. (Data by **Anastasia BARKOVA, PL lab**).

2.2.3 Ty1 IN expressed ectopically is slightly more abundant in the absence of CK2

We also checked the expression of Ty1 IN when ectopically expressed through a centromeric plasmid controlled by a tetOff promoter in the absence of the CK2 holoenzyme. An N-terminally 6His tagged Ty1 IN and an untagged Ty1 IN were cloned into a centromeric vector (pcm185) and transformed in WT and ΔCka2ΔCkb2 strains. Cells were grown at 30°C to stationary phase in the classical glucose synthetic medium in the presence of doxycycline. Then, the cells were diluted accordingly to induce Ty1 IN expression overnight in the absence of doxycycline. Cells were collected an exponential phase the following day for preparing TCA protein extracts. The induction conditions are the same in the TChAP procedure. Protein

extracts were prepared by TCA precipitation and Ty1 IN was visualized using polyclonal anti-IN antibody (prepared in our laboratory) by western blot (Figure 15). Interestingly, we observed that both the 6His-Ty1 IN and Ty1 IN were detected in the ΔCka2ΔCkb2 strains and hardly detected in the WT strain.

The fact that the difference in the level of expression is similar in the absence of CK2 holoenzyme for the endogenous Ty1 IN and the ectopically expressed Ty1 IN which are expressed from two completely different promoters suggests that it is highly likely that CK2 regulates Ty1 on a posttranscriptional level. Therefore, to study Ty1 regulation under more physiological conditions we developed a new tool to express a functional Ty1 element in a centromeric plasmid that could undergo maturation and actively retrotranspose *in vivo*.

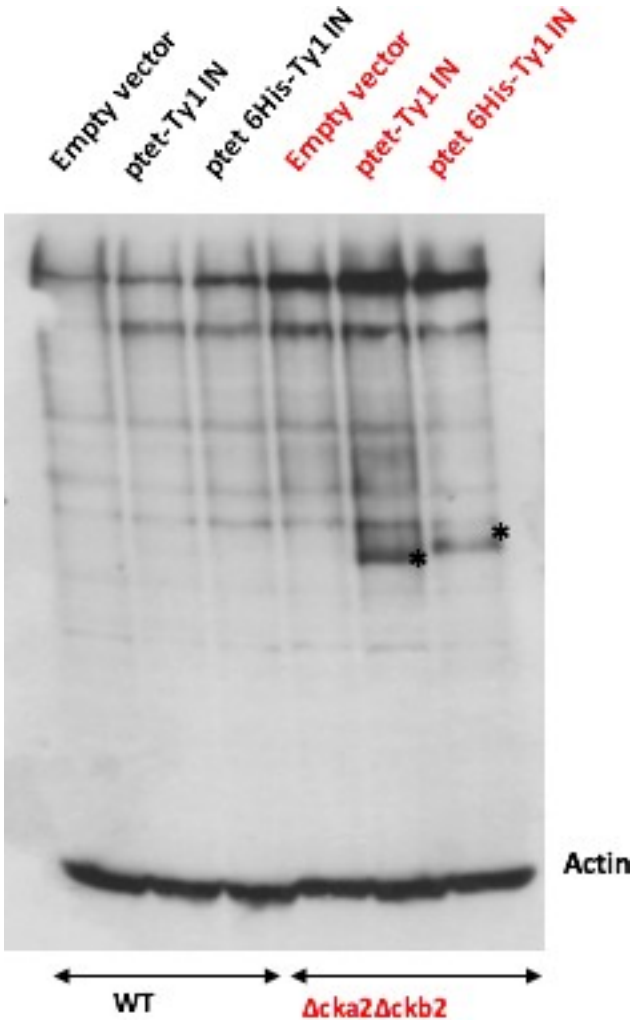


Figure 15. **Ectopically expressed Ty1 IN is slightly more abundant in the absence of CK2**. Total protein extract was prepared by TCA precipitation from cells grown at 30°C to exponential phase and Ty1 IN

was visualized by anti-IN polyclonal antibody (prepared in our laboratory) by western blot. Actin serves as the loading control (Performed by **Carine CHAGNEAU**).

2.3 Designing a novel tool to study the Ty1 retrotransposition *in vivo*

The Ty1 retrotransposition is an extremely rare event under physiological condition approximately 10^{-7}/element/generation (**Curcio et al 1991**). Studies have demonstrated that there exists a strong regulation of Ty1 mobility *in vivo* both at the transcriptional and post-transcriptional level (**Jiang 2002, Garfinkel 2003**) which is evident from the fact that Ty1 proteins are extremely low *in vivo* in comparison to the abundant Ty1 mRNA. This rarity was a problem in studying how Ty1 mobility is regulated *in vivo*. Therefore, to address this issue, an inducible galactose vector expressing the Ty1 functional element was constructed (**Curcio 1991**). Upon induction by galactose, there was a high expression of the Ty1 genes thus, inducing Ty1 retrotransposition. This system was very useful in identifying host factors mediating Ty1 mobility. But, there remained some problems in using this system. Firstly, over expression of Ty1 overrides the posttranscriptional regulation (CNC) *in vivo* (**Tucker et el 2015**). Secondly, use of a galactose in the medium to induce Ty1 expression changes the metabolism of the cells. Therefore, the basic criteria for a suitable Ty1 expression vector are as follows

a. An inducible vector is necessary to regulate the expression of Ty1 otherwise, constitutive expression of the Ty1 might be toxic to the cells.
b. The vector should allow induction of the Ty1 proteins in the classical glucose medium.
c. The vector should not over express the Ty1 proteins.
d. The vector should express a functional Ty1 element that is able to transpose *in vivo*.

Therefore, keeping these criteria in mind, we design a new centromeric vector that would express a functional Ty1 element controlled by a modified TetOff promoter (induction is achieved in the absence of Doxycycline). This system would not over-express the Ty1 proteins and would enable us to study Ty1 retrotransposition in the classical glucose medium. Thirdly, it is more physiological to express the Ty1 IN from a functional Ty1 element rather than just expressing the Ty1 IN alone ectopically. This offers a new system to study *in vivo* Ty1 retrotransposition without disrupting the normal metabolism of the cells.

2.3.1 Designing the promoter of the Ty1 expression vector

One important parameter when cloning the Ty1 element in the centromeric tetOff vector was to see that the vector yields a functional Ty1 element and the transcription start site is exactly at the same site as in the classical Ty1 galactose inducible vector, pgal Ty1.

The pcm185 centromeric vector comprise a $TetO_7$-cyc promoter (**Gari et al 1997**) containing seven TetO repeats and a fragment of the *S. Cerevisiae* cyc1 promoter. **Hahn et al 1985,** showed that there are three TATA regions which are functional and contribute to transcription initiation at the Cyc1 promoter in yeast. They identified two consensus sequences that are present in a window encompassing the TATA regions and close to the transcription initiation site by analysis of 18 yeast promoters transcribed by RNA Pol II. These two sequences account to about 55% of all transcription initiation sites. The first sequence is a **TCGA** sequence where transcription starts at the C or the adjacent nucleotides. The second sequence is a RRYRR, where a pyrimidine is surrounded by two purines. They demonstrated that if the TCGA sequence is introduced more than 50 bp downstream the TATA region, transcription initiates at this point. Interestingly, the Ty1 DNA in the pgal expression vector, contains an Xho1 site CTCGAG which also contains the TCGA sequence at the transcription start site position (Figure 16A). Fortunately, when the vector was first created (**Boeke et al 1985**), the Ty1-H3 was isolated and cloned using the Xho1 site and they found it be functional (☺).

Therefore, the $TetO_7$-cyc promoter was modified accordingly so that the Xho1 restriction site C**TCGA**G (which also contains the TCGA site) is placed 50 bp upstream of the second TATA box in the ptet promoter. Primer extension assay showed that the transcription site for this new vector was same as the pgal Ty1 (data not shown). This vector was designed and constructed by **Dr. Joel ACKER**. The schematic diagram of the modified ptet promoter is shown in figure 16. This vector was reported to be active and functional when tested in the lab of **Dr. Pascale LESAGE**. This vector will be referred to as ptet-TCGA-Ty1. All the further analysis of Ty1 IN expression was done using this vector.

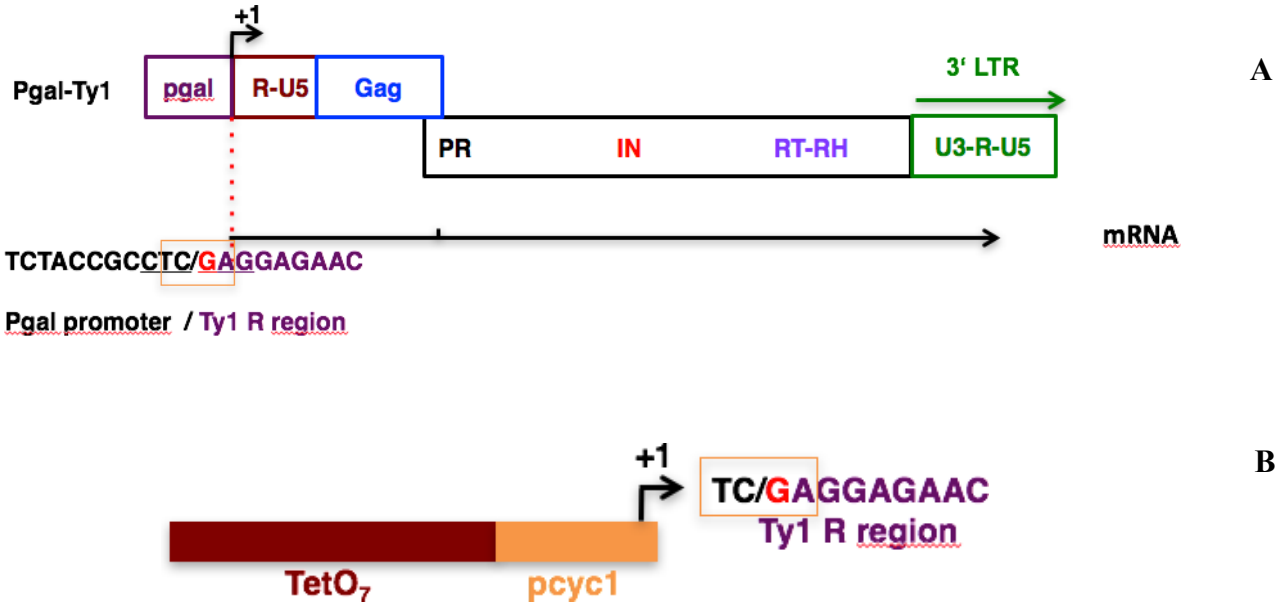


Figure 16. **Designing a physiological Ty1 under the control of a modified TetOff promoter**. (A) The galactose inducible vector (pgal) is presented showing the sequence of the +1-transcription site at the R region of the 5' LTR. (B) The modified TetOff promoter in the pcm185 vector is presented. It contains the $TetO_7$ repeats fused to a pcyc1 promoter. The **TCGA** was inserted 50 bp downstream of the second TATA box in the cyc1 promoter so that the transcription of Ty1 initiates at this site.

2.3.2 Testing the expression of Ty1 IN after maturation from the Ty1 element *in vivo*

We tested the expression of the Ty1 IN expressed from the newly constructed ptet-TCGA-Ty1. We expect to detect Ty1 IN after maturation from the Gag-Pol proteins. The pgal Ty1 was used as a positive control. The vectors were transformed into WT and Δcka2Δckb2 mutant strains and cells were grown to stationary phase in glucose synthetic medium in the presence of doxycycline (1µg/ml). The following day the cells were diluted accordingly to induce the expression of Ty1 overnight in the absence of doxycycline at both 24°C and 30°C. As Ty1 retromobility is relatively higher at low temperature than at 30°C, we decided to check the Ty1 IN expression at both temperatures. Moreover, proteolytic processing is efficient between 22-26°C and is impaired at 34°C (**Lawler et al 2002**). Cells were collected at exponential phase and protein extracts were prepared by TCA protein precipitation. The samples were resolved on an 8% SDS gel and Ty1 IN was visualized by western blot using the anti-IN antibody prepared in our laboratory.

A smear was detected by western blot in the WT strain when Ty1 expressed from the ptet TCGA-Ty1 vector and it was hard to see the Ty1 IN. However, in the Δcka2Δckb2 mutant, there was a high abundance of the Ty1 IN when expressed from the ptet TCGA-Ty1 vector suggesting that effectively Ty1 poly proteins are expressed and matured. Interestingly, we observed that the abundance of the Ty1 IN was much more in the Δcka2Δckb2 mutant when expressed through the ptet-TCGA-Ty1 rather than when Ty1 IN was expressed alone (Figure 14 and Figure 16). This data suggests that CK2 might regulate the Ty1 polyprotein rather than only the Ty1 IN. Currently, this is being tested in the laboratory. Another important observation we made was that over-expressing the Ty1 proteins using the pgal vector showed no difference in the Ty1 IN expression between the WT and Δcka2Δckb2 strains (Figure 16). This might be because overexpression of Ty1 proteins over rides the CNC of the cells which have been demonstrated by past studies (**Jiang 2002, Garfinkel 2003**). This suggests that in this case the pgal Ty1 vector might not be a suitable vector to study the regulation of Ty1 protein expression *in vivo*. However, we have to warn that the difference in the level of Ty1 IN observed in the absence of CK2 is not always as strong as compared to the WT. Currently, we are investigating the discrepancy of the reproducibility of this particular result. This might also be due to the presence of the endogenous Ty1 elements that could act in *trans* (**Curcio et al 2015**). Thus to abolish the expression of the endogenous Ty1, we would test the expression in both strains in the absence of a Ty1 transcription factor, Ste12. We are currently investigating the phenomenon.

Moreover, When Ty1 IN is expressed through the ptet-TCGA-Ty1 vector in the Δcka2Δckb2 strain, an upper band appears near the Ty1 IN band (marked with a red asterisk). Initially, we thought that it might come from the fact that the polyclonal anti-IN antibody (prepared in our laboratory) might have a different sensitivity than the anti-IN monoclonal antibody (8B11) generally used by other laboratories. However, with the pgal Ty1 IN vector we also detect the upper band with our anti-IN antibody. We hypothesis that there might be two possibilities;

(a) The second upper band might be due to an accumulation of a post translationally modified Ty1 IN in the absence of CK2 as it appears very close to the Ty1 IN band (black asterick, figure 17). According to previous studies such as **Duttler et al 2013**, **Swany et al 2013**, it has been shown that Ty1 proteins including Ty1 IN is ubiquitinylated. So, we would investigate using anti-Ub antibodies if the second upper band corresponds to an ubiquitinylated Ty1 IN.

(b) The second upper band might be due to the incomplete processing of the PR-IN polyprotein during maturation in the absence of CK2. In a previous study deciphering the processing of the TyB polyprotein products revealed that the apparent molecular size of the PR-IN pre-processed product is 91 KDa (**Merkulov et al 2001**) which corresponds to the molecular size of this second upper band on a SDS PAGE. To test this hypothesis, we are going to modify the PR-IN cleavage site in the ptet TCGA-Ty1 vector accordingly to abolish cleavage of the PR-IN polyprotein. This mutant Ty1 would produce an unprocessed PR-IN poly protein (91 KDa). The idea is to compare and see whether the apparent molecular size of this mutant corresponds to the molecular size of the second upper band in the WT Ty1 in SDS-PAGE and whether this hypothesis could be excluded or not.

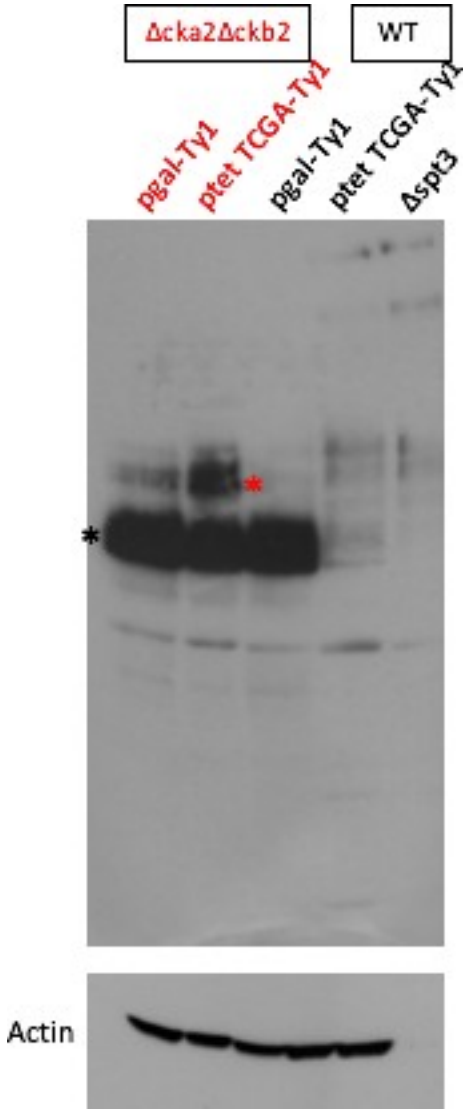


Figure 17. **Expression of Ty1 IN is stimulated in the absence of CK2**. Ectopically expressed Ty1 polyproteins were induced ON at 24°C using a ptetoff vector or pgal vector in the presence or absence

of CK2 as indicated. Proteins were analysed by western blot and Ty1 IN was visualized by polyclonal anti-IN antibody (prepared in our lab). Actin serves as the control for the input. Black asterisk marks the Ty1 IN and the red asterisk marks the putative unprocessed PR-IN polyprotein.

2.4 Ty1 IN is a substrate of CK2

The first question was to know, if CK2 is a binding partner of the Ty1 IN, is it able to phosphorylate Ty1 IN? If yes, what is the functional role of this phosphorylation on Ty1 retrotransposition *in vivo*. In fact, there is evidence in the literature demonstrating that phosphorylation of integrases has a functional role in Ty1 integration and integrase activity. For instance, the Ty5 IN, a homologue of the Ty1 IN, is phosphorylated at the C-terminal targeting sequence (TS) which facilitates its interaction with the Sir4p and subsequent integration at the gene-poor heterochromatin region. However, in the absence of phosphorylation of the TS, integration happens to the random in the genome (**Dai et al 2007**). Recently, **Jaspart et al 2017** showed that the HIV Integrase is a substrate of the Gcn2 serine/threonine kinase *in vitro* and phosphorylation represses HIV replication *in vivo*.

2.4.1 CK2 phosphorylates rTy1 IN *in vitro*

In order to test, whether Ty1 IN is a substrate of CK2, we performed an *in vitro* phosphorylation assay with radiolabeled ATP. As a first step, we affinity purified an epitope-tagged full length 6His-Fh8-Ty1 IN and the 6His-Fh8-Ty1 IN C-terminal (578-635 aa) region from *E. coli.* The Fh8 tag facilitated high solubility and yield of the proteins after purification by affinity chromatography. A TEV protease recognition site was placed just after the tag in the constructs which facilitated removal of the tag by TEV. Following the purification, we performed several small-scale experiments to identify the optimal condition required for efficient TEV processing of the 6His-Fh8 tag. We didn't want the tag to interfere with the phosphorylation assay. Several parameters such as substrate to enzyme ratio, time of the reaction and temperatures were tested for efficient TEV processing. The optimal conditions for TEV processing used were a concentration of substrate to enzyme ratio 1:1 and ON digestion at 4°C.

Then, an active CK2 holoenzyme was purified from a CKA1-TAP yeast strain by magnetic beads coupled to IgG. We also included a purified recombinant Maf1 protein as a positive control as it has been shown previously to be phosphorylated by CK2 (**Graczyk et al 2011**). A highly conserved recombinant human CK2 was also used as a positive control for the Ck2 activity. The Ty1 IN or Ty1 IN C-terminal or Maf1 were incubated with either yeast Ck2 or

human Ck2 in the presence of radiolabeled ATP for 30 min at 30°C. We observed that both the CK2 were able to phosphorylate the full length Ty1 IN and Maf1 suggesting that the purified CK2 from yeast was active and Ty1 IN is a substrate of CK2 *in vitro*. However, we didn't observe phosphorylation of the C-terminal (578-635 aa) region of Ty1 IN *in vitro* under our experimental conditions. Moreover, the fact that CK2 did not phosphorylate this small C-terminal peptide suggests that phosphorylation by CK2 *in vitro* was not non-specific. The SDS gel of the *in vitro* phosphorylation assay is shown in figure 18.

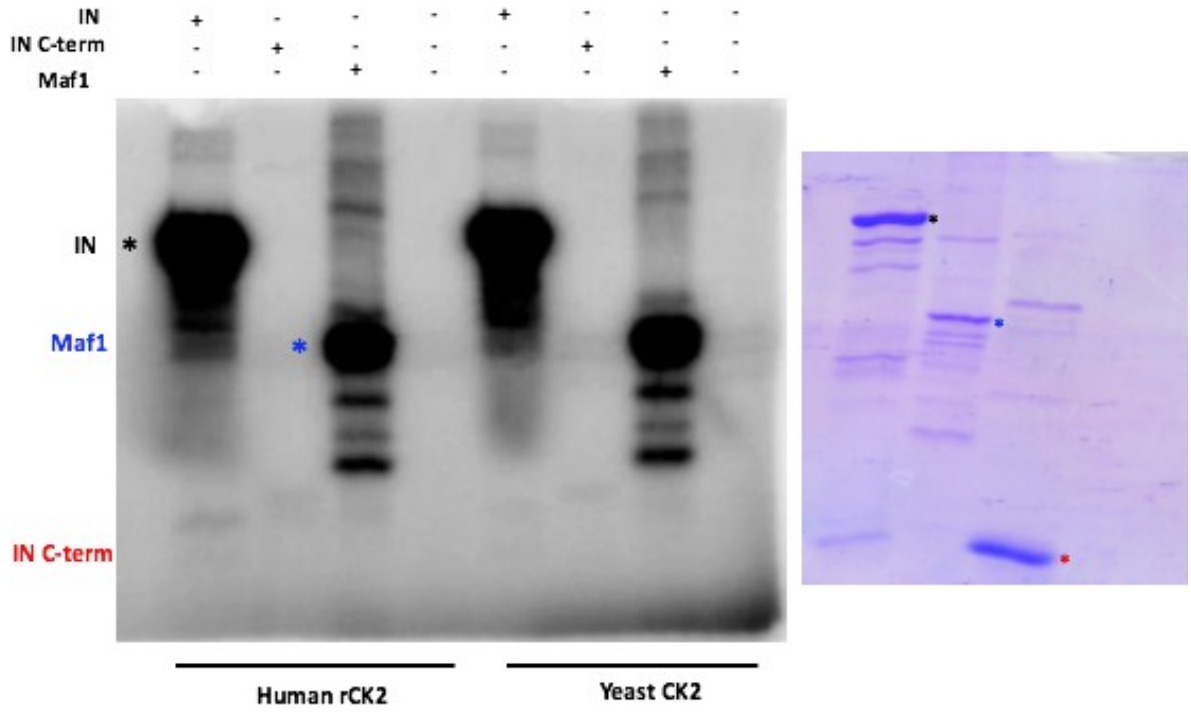


Figure 18. **CK2 phosphorylates Ty1 IN *in vitro***. rTy1 IN, rMaf1 and a C-terminal fragment of the Ty1 IN (rTy1 IN C-term), purified from *E.coli* were used in an *in vitro* phosphorylation assay performed with either human rCK2 or yeast CK2 as indicated, in the presence of [γ-32P]ATP. The proteins were resolved by 10% SDS-PAGE and the incorporated radioactivity was visualized by autoradiography (Left panel). Coomassie blue staining shows the amount of the recombinant proteins used in the assay (Right panel). (Performed by **Dr. Joel ACKER**).

2.4.2 Identification of Ty1 IN phosphorylation sites *in vitro*

The next step was to identify the phosphorylation sites after *in vitro* phosphorylation with CK2. To do this, we performed similarly *in vitro* phosphorylation assay without radioactive ATP.

Proteins were then resolved on a 4-12% gradient gel and stained with Coomassie blue before to slice the Ty1 IN band from the gel and sent it to the mass spectrometry facility in IBB, Warsaw, Poland to identify the phosphorylation sites (6 replicates). The identified phosphorylation sites on the is shown below. Initially, in the first four replicates the coverage of the Ty1 identified by the mass spectrometry was low (65%) and only four amino acids (**S411, T477, T480 and S499**) appeared to be phosphorylated. Thus, we performed two more experiments with ten times more Ty1 IN. This time, the coverage of the Ty1 IN identified was high (89%) and additional residues appeared to be phosphorylated.

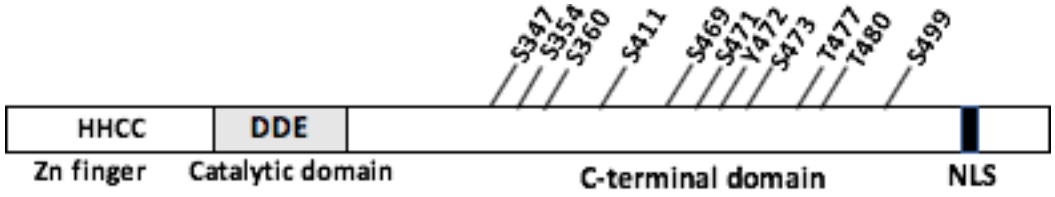


Figure 19. **Phosphorylated sites identified by mass spectrometry after *in vitro* phosphorylation assay.** The table represents the number of phosphorylated peptides identified by mass spectrometry. The percentage coverage of the Ty1 IN identified by mass spectrometry is mentioned. The figure below represents the genome organization of the Ty1 IN. The N-terminal Zn finger, catalytic domain and the nuclear localization signal (NLS) are marked and the identified phosphorylation sites are present in the C-terminal domain.

Firstly, no phosphorylated residues were identified in the Ty1 IN C-terminal region (578-635 aa) which is in good correlation with the *in vitro* phosphorylation assay performed with radiolabeled ATP. One important observation was that specifically the **Y472** out of 21 other tyrosine residues present in the Ty1 IN was found phosphorylated by CK2 both *in vivo* and *in vitro* suggesting that CK2 being a serine/threonine kinase can also phosphorylate tyrosine residues which is consistent with several studies. **Wilson et al 2007** demonstrated that the nucleolar immunophilin Fpr3 in the budding yeast was phosphorylated at Y184 by CK2 *in vivo* and *in vitro*. Moreover, **Y472** is present in a motif where the nearby serine residues (**S469, S471** and **S473**) are also phosphorylated. Probably, this explains why this is a suitable target of phosphorylation than the rest of the tyrosine residue. Interestingly, all the phosphorylated residues were found in the C-terminal domain and none in the N-terminal Zn finger or in the DDE catalytic domain suggesting that the C-terminal might act as a regulating domain. Before

to proceed with any further analysis it was important to show that all the sites phosphorylated *in vitro* are also phosphorylated *in vivo*. This would show that Ty1 IN is indeed a substrate.

2.4.3 Identification of Ty1 IN phosphorylation sites *in vivo*

The Ty1 INHBH was purified by Tandem affinity purification under denaturing conditions in 6M Urea from a yeast culture (2 liters) grown to exponential phase at 30°C. A phosphatase inhibitor cocktail was added to the extract in order to prevent any loss of phosphorylation sites during the purification procedure. The purified protein was resolved on a 4-12% gradient gel and gel slices containing the Ty1 IN HBH was sent for mass spectrometry. We made several replicates and phosphorylated sites identified in most of the experiments were considered significant. We identified several phosphorylation sites in the replicates and interestingly, all the phosphorylation sites identified *in vitro* were present among them which is in good agreement with Ty1 IN being a substrate. The identified phosphorylation sites and the frequency of their presence in 7 replicates are shown below.

Residue	Frequency		Residue	Frequency
S378	7		S360	4
S499	7		S473	4
S382	6		S523	4
S411	6		S558	4
S471	6		T477	4
S513	6		S469	3
T380	6		S511	3
T383	6		T480	3
T424	6		S337	2
S423	5		S347	2
S517	5		S422	2
S354	4		S544	2
			Y472	2

Table 2. **Phosphorylated sites identified in Ty1 IN *in vivo* by mass spectrometry.** The table represents the *in vivo* phosphorylated sites identified by mass spectrometry and they have been ranked according to the frequency of their appearance in the 7 replicates. Residues phosphorylated by CK2 *in vitro* are marked in red that are also present *in vivo*.

We also aimed to identify the *in vivo* phosphorylation sites of the Ty1 IN in the Δcka2Δckb2 yeast mutant, in order to see whether we lose some phosphorylation sites under this condition as compared to the wild type. We purified Ty1 INHBH (2 replicates) in the Δcka2Δckb2 yeast mutant and sent it for mass spectrometry identification of phosphorylation sites. However, under our experimental conditions it would be difficult to identify specific CK2 sites *in vivo* because other kinases present in the cell can phosphorylate the same residues in the absence of CK2. Thus, this requires quantitative mass spectrometry which we haven't performed yet. In particular, the S499 residue was never detected to be phosphorylated in the absence of CK2 whereas it was always found to be phosphorylated under wild type condition suggesting that it might be a specific CK2 target site.

2.4.4 Absence of phosphorylation by CK2 interferes with the expression of Ty1 IN *in vivo*

The next goal was to construct different mutants mimicking a non-phosphorylated amino acid and to test their retromobility in order to determine whether there is any direct correlation between the phosphorylated state of some specific amino acid and the level of Ty1 IN expression *in vivo*. Before to start writing of my manuscript, not all the MS data of the phosphorylation sites ientification were fully analysed. However, at that moment few sites (**S411, S473, T480, S499**) were already identified. As the three residues S473, S499 and T480 lie in close proximity to each other, we constructed a triple mutant in the ptet TCGA-Ty1 vector by substituting the serine and threonine residues with alanine to check for its expression *in vivo.* The plasmid was transformed in the WT strain and grown in classical glucose synthetic medium to exponential phase at 24°C. The Ty1 IN phosphomutant was induced overnight the same way as in all previous experiments in the absence of doxycycline.

Preliminary data showed that the expression of this particular Ty1 IN phosphomutant (ptet TCGA-Ty1 **S473AT480AS499A**) was higher than the wild type Ty1 IN. We need to perform additional experiments with introducing more multiple mutations in order to confirm the result and determine more precisely which amino acids are involved in this regulation. Moreover, we could not conclude that CK2 might not only regulate the expression Ty1 IN alone but also affect the Ty1 polyprotein *in vivo* which accounts for the slightly more abundance of Ty1 IN in the Δcka2Δckb2 mutant. Further analysis of the expression of Gag and RT in this mutant should be investigated.

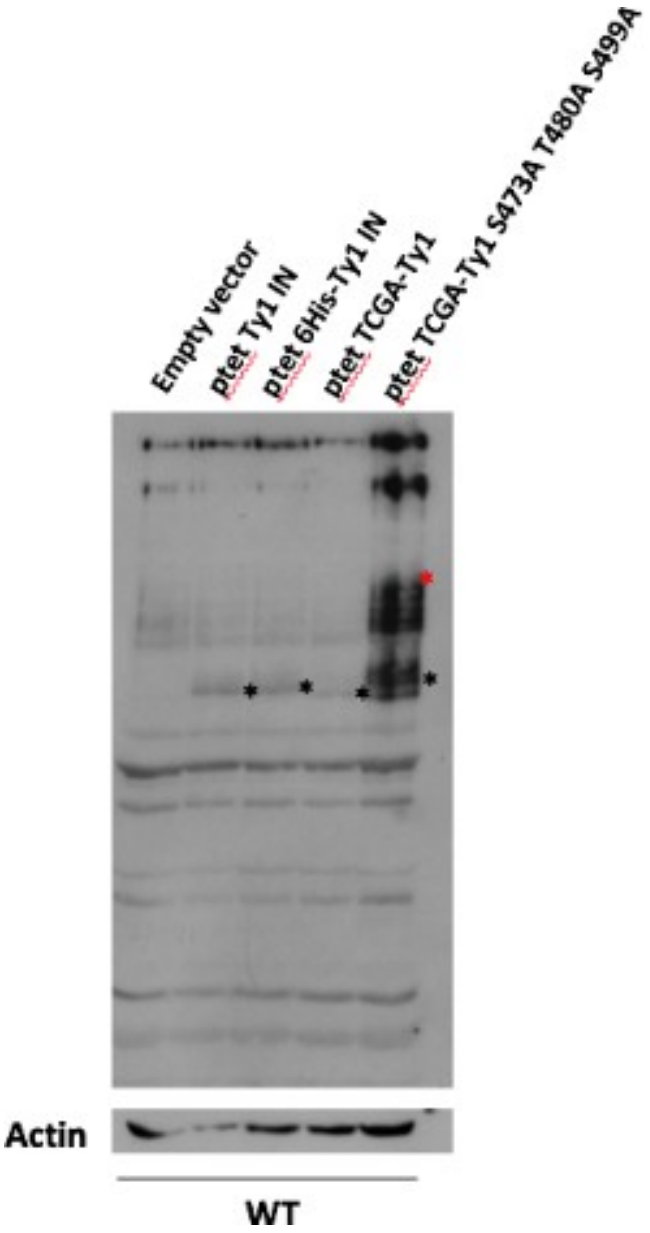


Figure 20. **The absence of CK2 phosphorylation interferes with the expression of Ty1 IN *in vivo***. Ectopically expressed Ty1 IN, Ty1 and Ty1 mutant were induced ON at 24°C using a ptetoff vector. Proteins were analysed by western blot and Ty1 IN was visualized by polyclonal anti-IN antibody (prepared in our lab). Actin serves as the control for the input. Black asterisk marks the Ty1 IN and the red asterisk marks the putative unprocessed PR-IN polyprotein.

2.5 Ty1 IN is degraded by the proteasome

Our data suggests that the Ty1 IN might be a short-lived protein and unstable under normal growth conditions. However, to demonstrate it we need to perform a stability assay to determine the half-life of Ty1 IN in the presence of a translational inhibitor, cycloheximide. On the other hand, it is known that in mammalian cells short-lived and long-lived proteins are both degraded by the proteasome. Whereas, in the budding yeast short-lived proteins are mainly targeted by the proteasome and the long-lived proteins by the vacuolar proteases (**Goldberg et al 1997**). Previous studies such as **Kaake et al 2010** showed that Ty1 proteins including the Ty1 IN are binding partners of Rpn11, a subunit of the 26S proteasome. **Mayor et al 2007,** showed that Ty1 proteins are ubiquitinylated and enriched in the absence of Rpn10, another subunit of the proteasome. We wondered whether the high expression of Ty1 IN in the Δcka2Δckb2 mutant is due to an inhibition of the Ty1 IN degradation pathway by the proteasome machinery of the cells. As ubiquitination dependent proteasomal degradation of proteins is a well-known phenomenon, we wanted to investigate whether the Ty1 IN is a substrate of the proteasome and if yes, whether in the absence of CK2 holoenzyme, Ty1 IN was not able to be degraded by the proteasome machinery any longer.

To test this hypothesis, the expression of Ty1 IN was followed over time in both the wild type strain and Δcka2Δckb2 mutant strain by inhibiting the proteasome with MG132, a potent inhibitor of the 26s proteasome limiting the degradation of ubiquitin-conjugated proteins. As Ty1 IN is hardly detected after induction in the WT strain, an accumulation of the protein over time upon the inhibition of the proteasome would suggest that it is a substrate of the proteasome. However, it has been shown that chemical inhibitors such as MG132 are impermeable to the yeast cells wall and various mutants such as *ΔERG6* and *ΔPDR5* are required to enhance permeability (**Lee et al 1996**). Such mutations can compromise other cellular processes and interfere with protein stability. Later, **Liu et al 2007** established a method to use MG132 in WT cells without the need of any mutations. The protocol involves growing cells in a medium containing L-proline as the source of nitrogen instead of ammonium sulphate and 0.003% SDS that leads to opening of the yeast cell wall facilitating uptake of the MG132 by the cells. Following this protocol, expression of Ty1 IN through the ptet TCGA-Ty1 vector in both WT and Δcka2Δckb2 were induced overnight at 24°C to exponential phase in the absence of doxycycline and grown in synthetic medium containing L-proline. Cells were then treated with MG132 in the presence of 0.003% SDS.

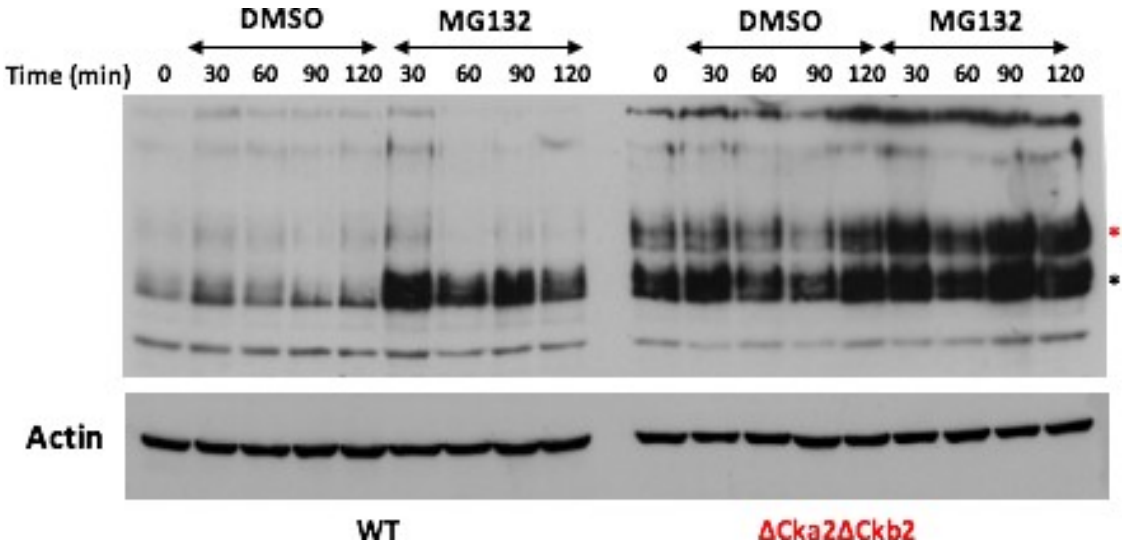


Figure 21. **The Ty1 IN is a substrate of the proteome**. (Performed by **Carine CHAGNEAU**). Ty1 IN expression was induced ON in cells grown to exponential phase at 24 °C in WT and CK2 mutant yeast strains. Following which, proteasome inhibitor, MG132 or DMSO (control) was added and samples were collected at several time points to detect the presence of Ty1 IN by western blot using anti-IN antibodies (prepared in our laboratory). Ty1 I and the second upper band are marked with a black asterisk and the red asterisk respectively. Actin was used as a loading control. Black asterisk marks the Ty1 IN and the red asterisk marks the putative unprocessed PR-IN polyprotein. (Performed by **Carine CHAGNEAU**)

Indeed, we saw an increase of the Ty1 IN in the WT strain after 30 min following the treatment of MG132 suggesting that the Ty1 IN was dependent on the proteasomal degradation pathway. In the control cells treated with DMSO, there was no change in the expression of Ty1 IN over time indicating that the increase in Ty1 IN under MG132 treatment is solely due to the inhibition of proteasome. After 120 min, the signal of Ty1 IN appears to decrease probably because MG132 degrades in the cells after certain time and more addition is required for longer experiments. Another important observation was the appearance of only the lower band corresponding to Ty1 IN and not the second upper band in the WT cells under MG132 treatment unlike Δcka2Δckb2 mutant. This suggests that may be the only proteasome target is the Ty1 IN and not the form in the upper band. There was also a slight difference in the accumulation of Ty1 IN in the Δcka2Δckb2 strain suggesting that CK2 might not be the only protein regulating the Ty1 IN expression under this condition. The MG132 proteasome inhibition assay was performed by **Carine CHAGNEAU**.

This data is in good agreement with the previous studies mentioned earlier demonstrating that Ty1 IN is a binding partner of the proteasome. In fact, from our TChAP experiments in the past, we recovered 9 proteins which are subunits of either the 26S or 20S proteasome as Ty1 IN binding partner by mass spectrometry. But initially we thought it was because the epitope tagged Ty1 IN HBH we expressed *in vivo* for purification was subjected to degradation due to improper folding of this tagged IN. At present, we believe it was possibly because Ty1 IN is a substrate of the proteasome.

PROJECT 2

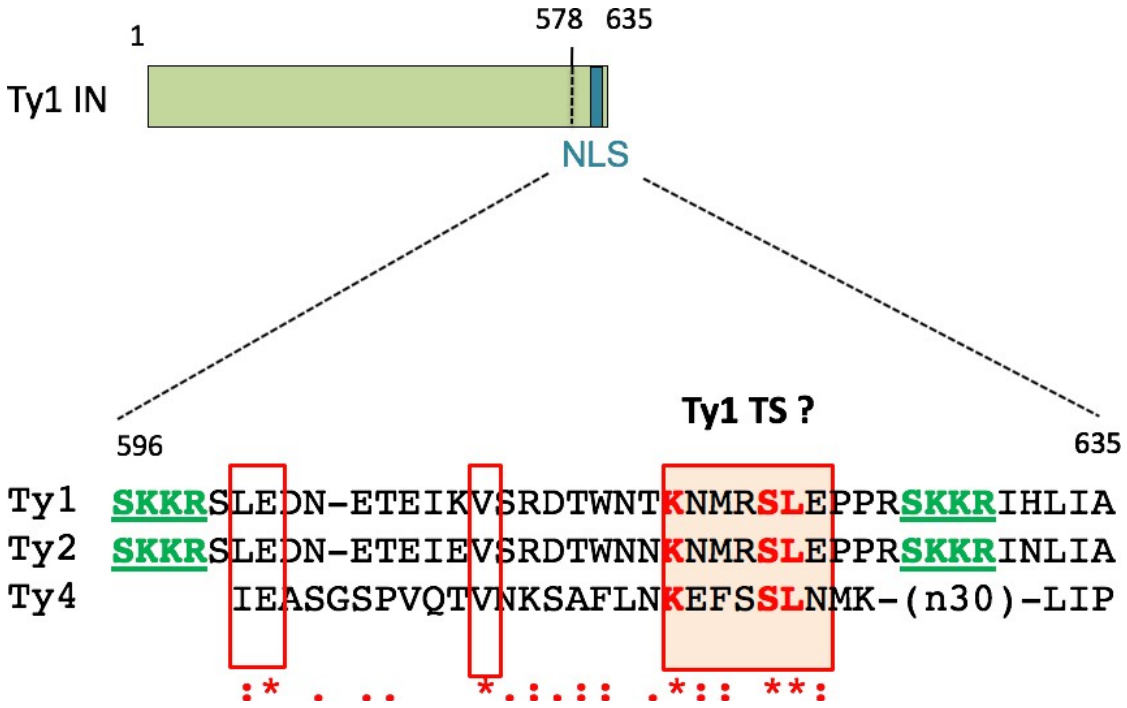
Investigate the interaction between the Pol III/Pol I machinery with the Ty1 IN

Collaboration with Dr. Pascale Lesage

3. Investigating the interaction between Pol and Ty1 IN *in vivo*

Earlier, in 2015 the group **of Dr. Pascale LESAGE** demonstrated that the C-terminal of the Ty1 Integrase (578-635 aa) is absolutely necessary for interacting with the RNA Polymerase III complex. This interaction is important for Ty1 elements to integrate at the tRNA genes (**Bridier Nahmias et al 2015**). They showed by a yeast two hybrid assay using various Ty1 IN truncation constructs as the bait and AC40 as the prey that the Ty1 IN C-terminal (578-635 aa) region is indispensable for associating with AC40 which explains that the targeting sequence (TS) is present within this region. Next, to determine the functional role of this IN-AC40 interaction, they developed a loss of this interaction mutant to study Ty1 integration specificity in the genome. As AC40 is an essential gene and deletion was lethal to the cells, they swapped the AC40 gene of *S. cerevisiae* with the AC40 gene from *S. pombe*. The cells were viable under this condition and displayed active Pol III transcription but no interaction between Ty1 IN and $AC40_{sp}$ was observed in yeast two hybrid assay. Ty1 integration assay showed that under this loss of Ty1 IN/AC40 association the Ty1 integration site was redistributed to the subtelomeres and integration upstream of tRNAs was highly disrupted. However, this mutant showed no significant change in the rate of Ty1 retrotransposition *in vivo*. Thus, this was the first time in many years that Pol III was shown to be a tethering factor of the Ty1 IN in determining the integration site choice upstream of Pol III transcribed genes.

Sequence alignments between the integrase C-terminal regions of the homologous Ty1, Ty2 and Ty4 retroelements revealed the presence of several conserved sites in this region (Figure 23). There lies a motif of 6 amino acids, **KNMRSLE** where the K, S and L are conserved in all the three integrases. This prompted them to mutate each of the six amino acids to investigate their effect on interaction with AC40 and Ty1 integration specificity *in vivo*. Under their experimental conditions, yeast two hybrid assay showed that for the Ty1 IN K617A, R620A, S621A and L622A mutants, the association with AC40 was disrupted in a similar manner as observed for the $AC40_{sp}$ mutant suggesting that these amino acids are necessary for association with AC40. At this point, our group was then involved in this project and I was particularly involved in characterization of the interaction between the Pol machinery and Ty1 IN by *in vivo* and *in vitro* biochemical assays.



TS Mutant	IN -AC40 interaction (Y2H)
K617A	-
N618A	+
M619A	+
R620A	-
S621A	-
L622A	-
E623A	+

Figure 23. **The Ty1 IN mutants lose association with AC40 *in vivo***. The upper panel shows the alignment of the Ty1, Ty2 and Ty4 IN sequence. The KNMRSLE putative TS site is marked. The lower panel shows the results of the yeast 2 hybrid screen performed in the lab of **Dr. Pascale LESAGE.** The K, R, S and L mutants lose association with AC40 *in vivo* (Data from **Dr. Pascale LESAGE**).

3.1 Interaction between Pol III, Pol I and Ty1 IN *in vivo*

3.1.1 *In vivo* coimmunoprecipitation

From our TChAP experiments we have recovered the Pol III complex and have also validated its interaction by *in vivo* CoIP. Interestingly, we have also recovered the Pol I subunits indicating that it is a putative Ty1 IN partner. Moreover, there are several subunits of Pol III that are common to the Pol I machinery and specifically AC40, a major target of the Ty1 IN, is

shared between the Pol I and Pol III complex. Therefore, it is not surprising for Pol I to be able to associate with the Ty1 IN. However, we did not recover any Pol II subunits in our TChAP experiment. We performed a CoIP following the same protocol for Ty1 INHBH induction as in the previous CoIP experiments in strains expressing a TAP tagged A190 (Pol I largest subunit). We observed that Ty1 IN was able to coimmunoprecipitate with the Pol I complex but not with the Pol II complex *in vivo* under the experimental conditions used (Figure 24).

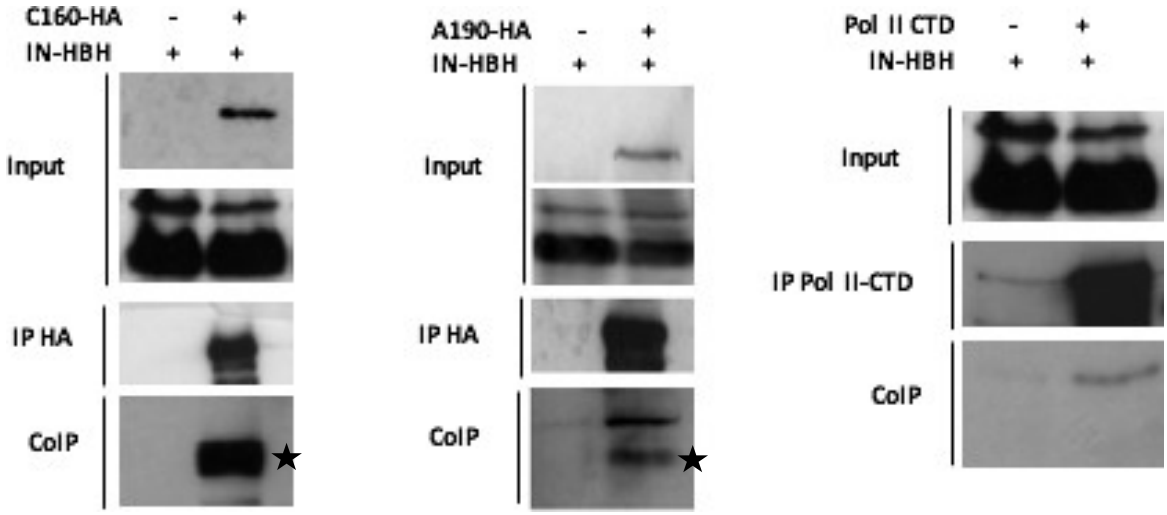


Figure 24. Ty1 IN interacts with Pol III and Pol I but not with Pol II *in vivo*. Ty1 IN-HBH were induced in strains either expressing a HA tagged C160 or HA tagged A190 in the genomic loci or not as indicated grown at 30°C. Immunoprecipitation were done on IgG bead coated with anti-HA (12CA5) antibody (Left and middle panel). Pol II was immunoprecipitated using anti-Pol II C-terminal domain antibody in the same strains expressing a tagged C160. The Ty1 INHBH was visualized by anti-strep antibody. The black asterisk marks the Ty1 INHBH. **In the Pol II CoIP (right panel), the band visible is a non-specific band and should not be confused as a Ty1 INHBH band.**

3.1.2 Yeast 2 hybrid assay

Following the validation by CoIP, we performed yeast two hybrid assay to determine more precisely which RNA Pol subunits might interact with Ty1 IN. Different plasmids were constructed expressing each of the subunits of Pol I and Pol III and Ty1 IN-EPEA were fused to both the Gal4 DNA binding domain (pAS-Pol III/Pol I) and Gal4 activating domain (pACT-Ty1 IN-EPEA) to be able to perform the assay in both directions. Experimental conditions were similar as described in section 2.1.1b. Two reporter genes *HIS3* and *LacZ* were used in the assay. The cells were grown on His- plates with varied concentrations of 3AT toscreen for

strong interactions and thereafter treated with X-gal to screen for blue/white colonies indicating a positive interaction. As a control the plasmid expressing the Ty1 IN was co-transformed with an empty pAS or pACT2 vector.

We observed that under our experimental conditions, the C31 and C11 subunits of Pol III, A14 subunit of Pol I showed positive interactions between pACT- Ty1 IN-EPEA and pAS-Pol III/ Pol I subunits (Figure 25A). Whereas, C25 subunits of Pol III and A190 subunit of Pol I showed positive interactions between pAS- Ty1 IN-EPEA and pACT- Pol III/Pol I subunits (Figure 25B). Most of the time it has been seen for yeast two hybrid assays that the outcome varies depending on whether the Gal4 DNA binding domain or activating binding domain is present on the bait or the prey and vice-versa. Thus, it is important to perform the assay on both directions. However, in our lab the interaction between Ty1 IN and AC40 was not detected by the X-gal blue/white screening suggesting that experimental conditions can affect the final outcome. The interaction between AC40 and Ty1 IN was showed by using an untagged Ty1 IN as a prey in yeast two hybrid assay against AC40 as a bait in the team of **Dr. Pascale LESAGE**. On the contrary, we used a Ty1 IN with a C-terminal EPEA tag as a prey/bait and the tag might interfere with the interaction. However, we observed a direct interaction between the Ty1 IN and AC40 by *in vitro* CoIP (Figure 30). Therefore, concluding a positive interaction is difficult with just one kind of experiment. Several types of interaction assays are required to come to a robust conclusion.

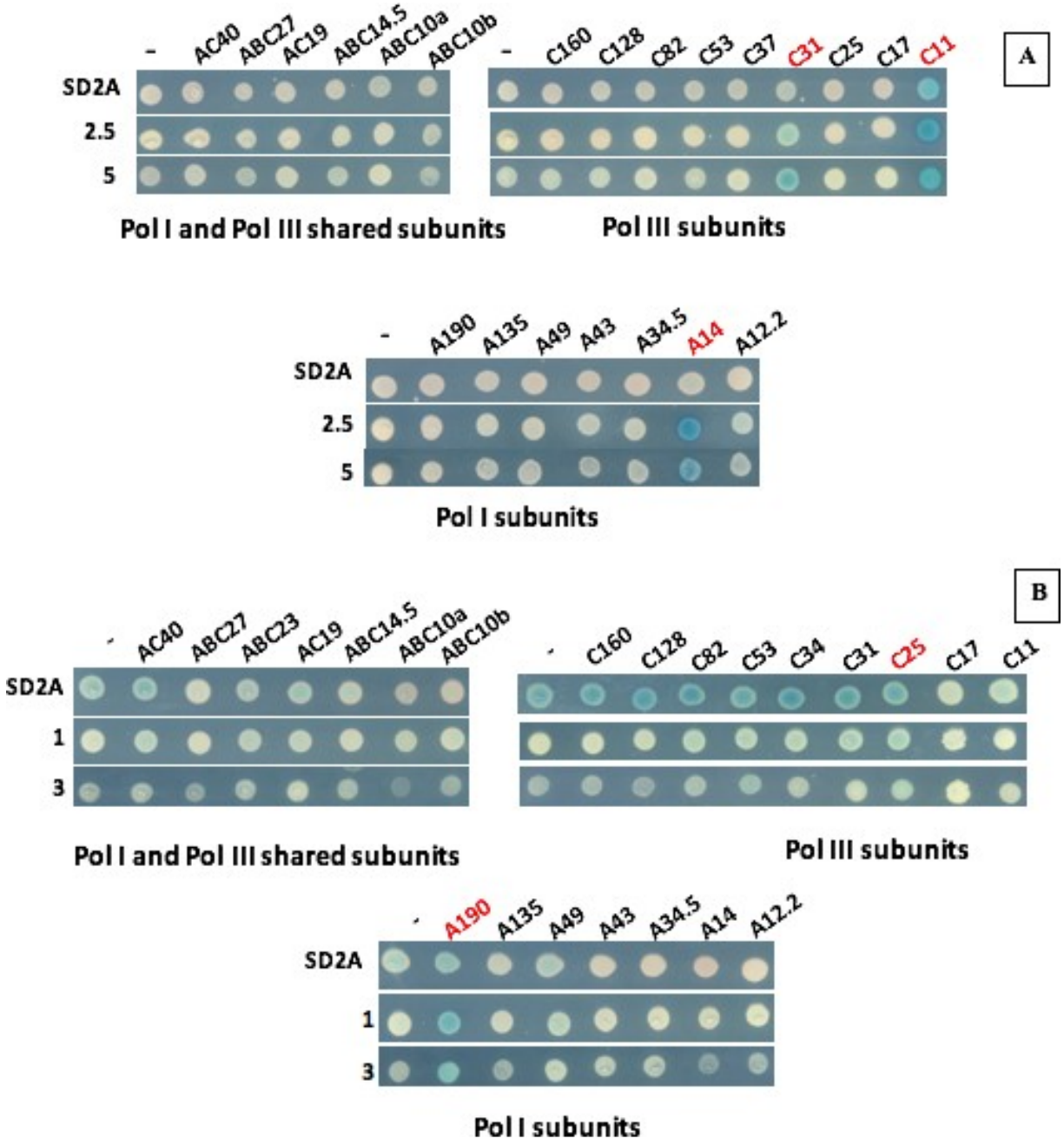


Figure 25. **Pol I and Pol III interaction with Ty1 IN in vivo.** (A) Yeast strain Y190 carrying empty vector pACT2 (negative control) or pACT2- Ty1 IN-EPEA were co-transformed with plasmids encoding fusions of the Gal4 DNA binding domain with the subunits of Pol I and Pol III. (B) Yeast strain Y190 carrying empty vector pAS (negative control) or pAS- Ty1 IN-EPEA were co-transformed with plasmids encoding fusions of the Gal4 DNA activating domain with the subunits of Pol I and Pol III. Patches of cells were overlaid on X-gal agarose to reveal β-galactosidase activity (blue colour). This experiment was performed by **Dr. Christine CONESA** and **Carine CHAGNEAU**)

Recapitulating on our TChAP results, we have observed that the AC40, AC19 and ABC27 being small molecular size proteins have been identified with a high coverage in our mass spectrometry data suggesting that these subunits might be present in close proximity to the Ty1 IN in the complex.

However, to confirm a direct interaction *in vitro* binding assay is absolutely necessary. **Cheung et al 2016,** showed that there is a direct interaction between the C31, C34 and C53 subunits of the Pol III with the Ty1 IN by *in vitro* GST pull down assay. However, the results are not very convincing since the binding of the Ty1 IN with these proteins are hardly detected compared to the input used and a high background signal is present. Nevertheless, identifying direct interactions with Ty1 IN *in vitro* seems to be difficult because when expressed in *E. coli*, different oligomeric forms of Ty1 IN might be present and can also form aggregates easily. Thus, to isolate Ty1 IN multimers and inhibit aggregation few precautions are required to be undertaken as mentioned earlier in section 1.5. Even, under our experimental conditions we were not able to detect a direct interaction between Ty1 IN and AC40 by *in vitro* pull down assays. Therefore, another possibility was to co-express both the proteins in *E. coli* and then perform a CoIP (Figure 30) .

One might wonder why does the Ty1 IN interact with so many subunits of Pol III? One possible reason could be that the oligomeric Ty1 IN being a large protein complex might interact with several subunits. Whereas, interactions between some subunits might be dynamic explaining why they are difficult to be detected. The Pol III cryo-EM structure (**Hoffmann et al 2015**) shows that there are the ABC14.5, AC19 and C128 subunits in close proximity to the AC40 subunit and we wonder whether these subunits could stabilize the Ty1 IN-AC40 interaction or might directly interact with Ty1 IN (Figure 26). We did demonstrate that there is a direct interaction between the AC40 subunit and Ty1 IN. However, the observed interaction was weak. Probably, a stronger interaction could be observed if the AC40-AC19 heterodimer is co-expressed with Ty1 IN *in vitro*. This is being tested in our lab at the moment.

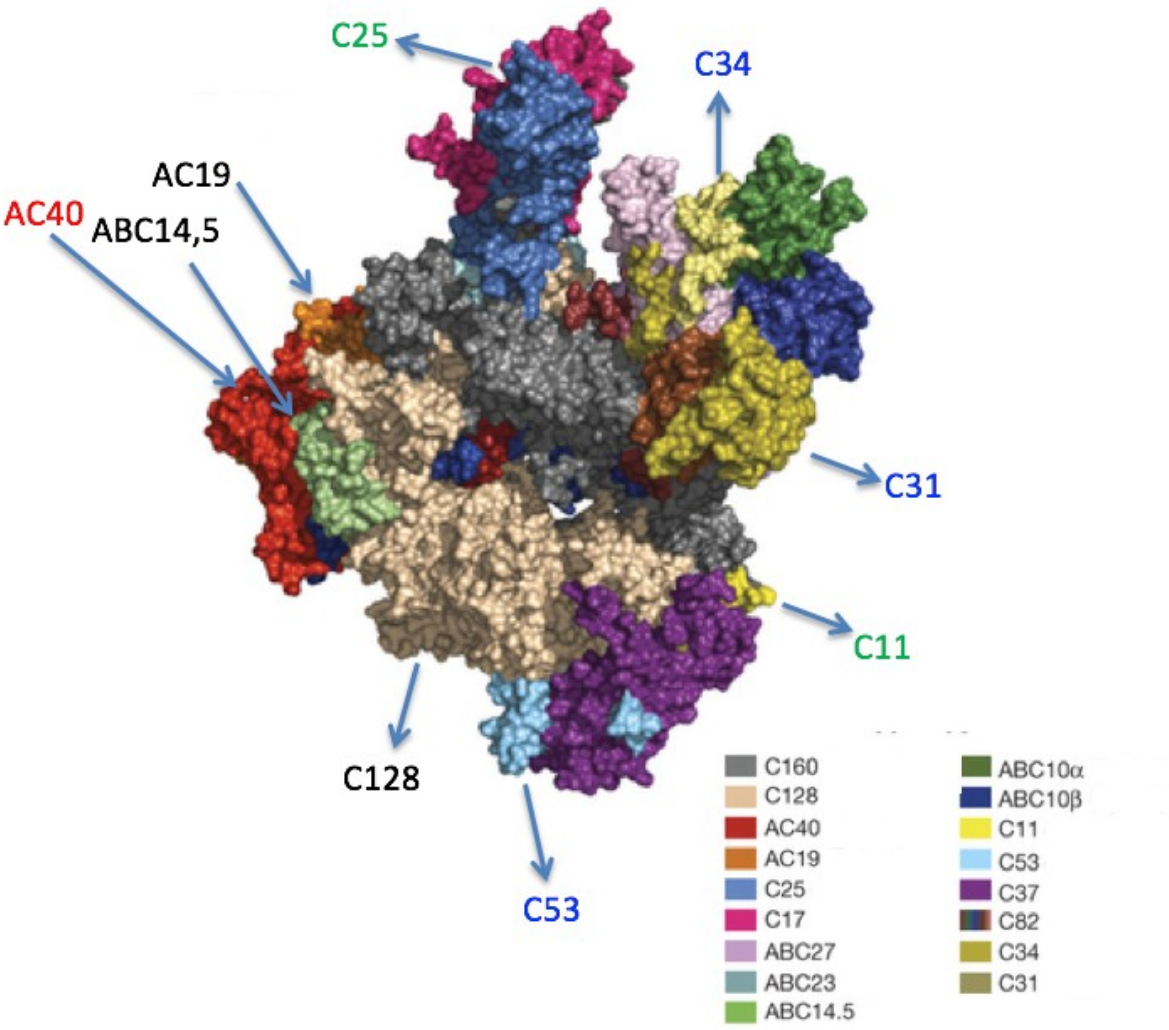


Figure 26. **The Cryo-EM structure of the RNA Pol III complex**. The AC40 subunit is an important target of the Ty1 IN. Nevertheless, ABC14.5, AC19 and C128 lie in close proximity and might stabilize the Ty1 IN-AC40 interaction. The C34, C31 and C53 subunits were identified to bedirectly interacting with the Ty1 IN in vitro (**Cheung et al 2016**) and C11, C25 and C31 were also found interacting with Ty1 IN in our yeast two hybrid assay (Image adapted from **Hoffmann et al 2015**).

3.3 Ty1 IN was found to be specifically associated to specific DNA regions *in vivo*

Studies in the past have shown that the Ty elements have specific integration sites in the genome. The Ty1-Ty4 elements all target the class III genes, majorly the tRNA genes. According to the integration model, the Ty1 IN binds to the target DNA to initiate the integration process. However, such interaction between Ty1 IN to its target DNA *in vivo* has never been shown before. Thus, we wondered if a binding between the Pol III genes and Ty1 IN could be detected by ChIP experiments. We expressed Ty1 IN as an N-terminal HA epitope tagged protein through a centromeric plasmid controlled by the tetOff promoter. The protein was

induced ON at 30°C in the absence of doxycycline exactly the same way as in the CoIP experiments. Cells were collected at exponential phase and extracts were prepared after *in vivo* crosslinking of the cells with 1% formaldehyde. Upon extensive experiments to optimize the ChIP procedure we were able to detect the interaction of Ty1 IN with Pol III genes.

Three Pol III transcribed genes $tRNA^{ileu}$, $tRNA^{leu}$, *SCR1* were analyzed as to see whether Ty1 IN binds to these target genes. The ChIP data showed occupancy of Ty1 IN on all the target genes tested for. The *GAL1* serves as the negative control. Strikingly, we observed that the Ty1 IN binding to the target genes was very strong. The % occupancy appears to be approx. 2.3 for SCR1 binding to Ty1 IN which is much higher than % occupancy of Sub1, a Pol III transcriptional activator binding to SCR1 (**Tavenet et al 2009**). All the ChIP experiments were designed, performed and analysed by **Dr. Christine CONESA**.

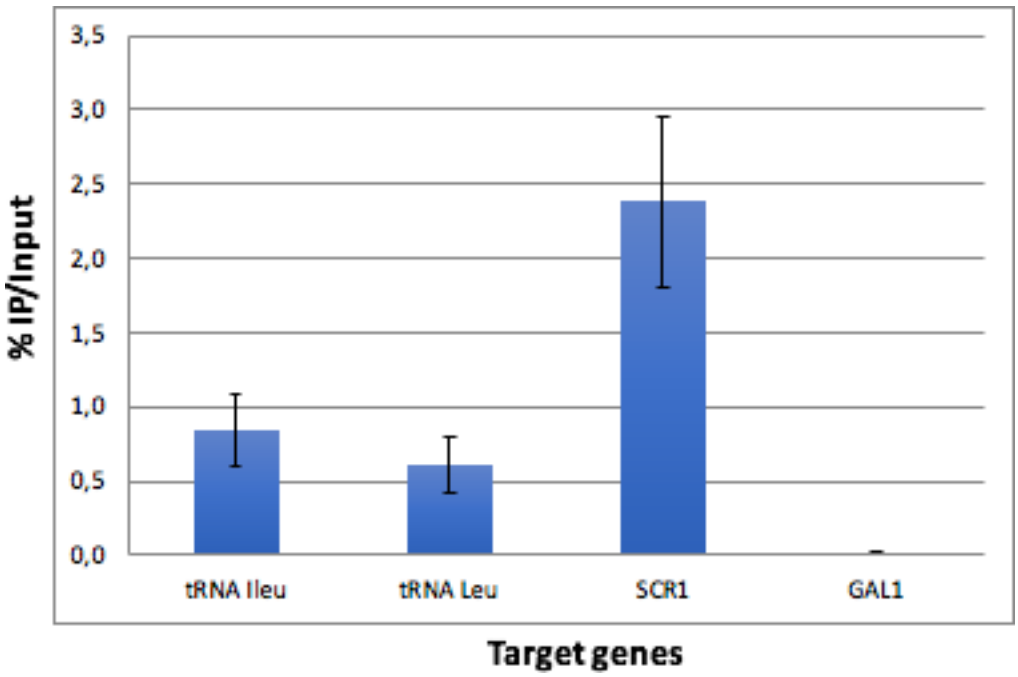


Figure 27. Ty1 IN is found to associate with its target genes *in vivo*. A HA-epitope Ty1 IN was induced in cells grown at 30°C and immunoprecipitated on IgG beads coated with anti-HA (12CA5) antibody. The DNA was analyzed by qPCR with specific oligos in the target genes. *GAL1* was used as a negative control. The Y-axis represents the percentage of the amount of Ty1 IN bound to the target gene compared to the total amount present in the input. The X-axis represents the name of the target genes (Performed by **Dr. Christine CONESA**).

3.4 Characterization of the TS in the Ty1 IN

3.4.1 Single amino acid mutation in the TS disrupt the interaction between Ty1 IN and its partners *in vivo*

Now that we have established the conditions for CoIP for the WT Ty1 IN to co-immunoprecipitate with its targets *in vivo* when expressed through the tetoff inducible system, we introduced the TS mutants, K617A, S621A and L622A in our centromeric tetOff vector to investigate its effect on the binding to Pol III and Pol I *in vivo*. The Ty1 mutant vectors were transformed into strains expressing either C160 or A190 as a HA tag or TAP tag fusion respectively. All the Ty1 IN mutants were induced in the absence of doxycycline. Either of the single TS mutants did not coimmunoprecipitate with RNA Pol III and RNA Pol I complex *in vivo.* (Figure 28 and Figure 29). The fact that a single amino acid mutation is enough to disrupt the association between Ty1 IN and Pol III/Pol I in *vivo* strongly indicate the importance of these amino acids as a TS. Kindly note that the *in vivo* CoIP between the INK617A and PolIII was performed in the lab of Pascale LESAGE showing that this mutant doesn't interact with Pol III also (Data not shown).

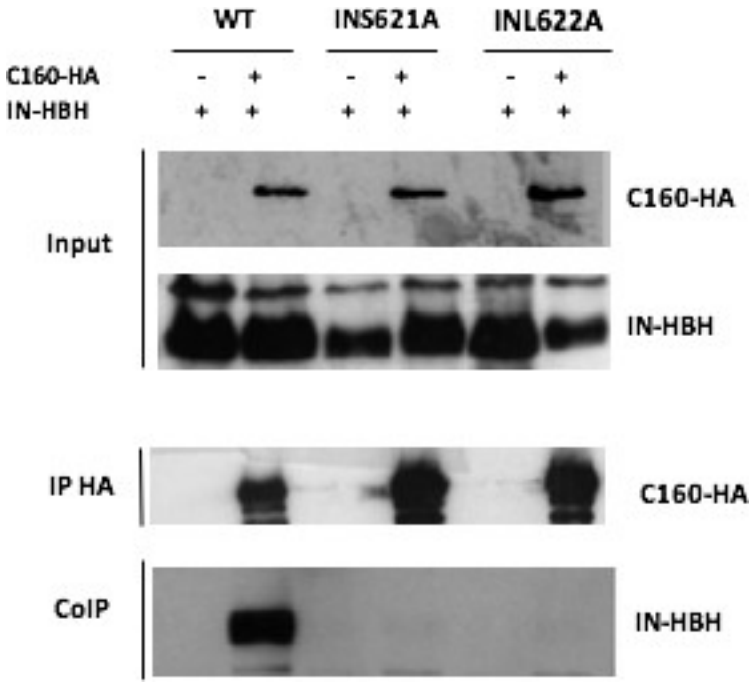


Figure 28. Single amino acid mutation disrupts interaction between Ty1 IN and RNA Pol III *in vivo*. Ty1 IN-HBH (WT), INS621A-HBH, INL622A –HBH were induced in strains expressing a HA tagged C160 in the genomic loci grown at 30°C. Immunoprecipitation were done on IgG bead coated with 12CA5 (Anti-HA) antibody. The Ty1 IN HBH and C160-HA were visualized by anti-Streptavidin and anti-HA (12CA5) antibodies.

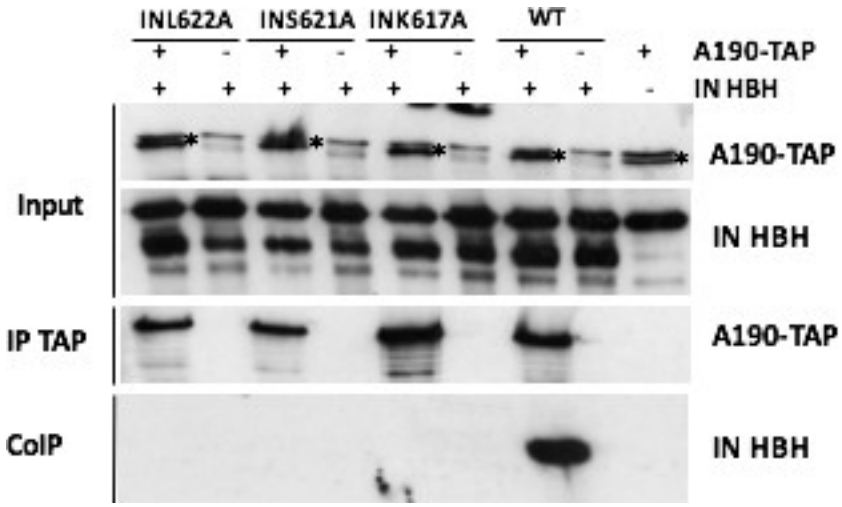


Figure 29. Single amino acid mutation disrupts interaction between Ty1 IN and RNA Pol I *in vivo*. Ty1 IN-HBH (WT), INS621A-HBH, INL622A –HBH were ectopically expressed and the A190 subunit of RNA Pol I was TAP tagged on the genomic loci. Immunoprecipitation of Pol I complex from the extracts were done by Pan mouse magnetic beads coupled to IgG coated. The Ty1 IN HBH and A190-TAP were visualized by anti-Streptavidin and anti-TAP antibodies.

3.4.2 Single amino acid mutation in the TS disrupt the interaction between Ty1 IN and AC40 *in vitro*

Though several evidences have shown there exists an interaction between the RNA Pol III complex and Ty1 IN by *in vivo* CoIP and yeast two hybrid assay showed that AC40 interacts with Ty1 IN, additional experiments are required to confirm a direction interaction. To do this, we performed a pull down assay by producing an epitope tagged Ty1 IN-EPEA and AC40-Streptagtwin separately in *E. coli.* The streptagtwin tag is an affinity tag used for capturing proteins by magnetic strep tactin beads. However, we were not able to detect any interaction under such condition due to the presence of a high background signal (Data not shown).

Therefore, we decided to investigate the interaction with another *in vitro* approach. We performed an *in vitro* CoIP by co-expressing AC40-streptagtwin and Ty1 IN-EPEA in *E. coli*. Upon this experimental condition, we observed a direct interaction between the two proteins and this interaction was also disrupted by the different Ty1 mutants (Figure 30).

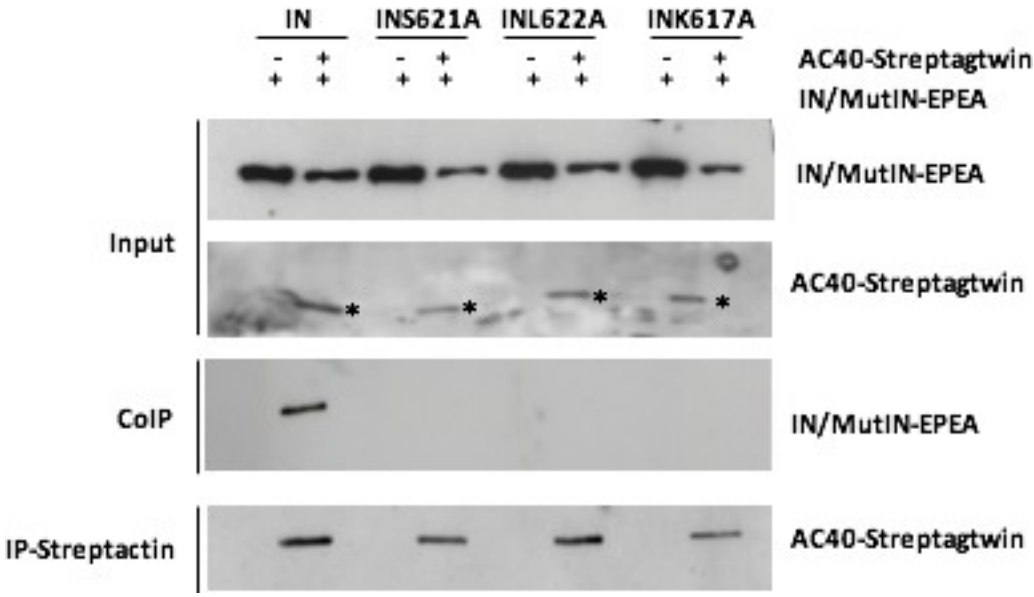


Figure 30. Single amino acid mutation disrupts interaction between Ty1 IN and AC40 *in-vitro*. Co-immunoprecipitation of IN-EPEA, INS621A-EPEA, INL622A-EPEA and AC40-Streptagtwin expressed in *E. coli*. Immunoprecipitation of WT IN and mutant IN in the protein extracts were performed by strep-tactin magnetic beads. Proteins were revealed with anti-EPEA and anti-streptactin antibodies. **For the input, the migration of the SDS gel was not proper at the lower end due to which the bands corresponding to the AC40-Streptagtwin appears distorted.**

Although an interaction between Ty1 IN and AC40 is observed, it is weak and may be the different oligomeric forms of Ty1 IN present in the extract might explain the discrepancy. On the other hand, to obtain a homogenous form of Ty1 IN in the extract a high salt concentration is required during protein extract preparation otherwise it aggregates. On the other hand, in vitro interaction assays are performed at low salt concentrations. So, maintaining this balance of salt concentration is difficult. However, AC19 and AC40 forms a heterodimer in vivo (**Vannini et al 2012**) and probably co-expressing the two proteins with Ty1 IN might show a stronger interaction.

3.4.3 Single amino acid mutations in the TS disrupts the association of Ty1 IN with the target genes *in vivo*

Although no interaction between Pol III, Pol I and TS mutant Ty1 IN has been observed by *in vivo* CoIP, we wondered whether these mutations were sufficient to prevent the interaction between IN and the target DNA. These Ty1 IN mutants were expressed as a HA fusion and following the same experimental condition as for the WT Ty1 IN, ChIP was performed on

extracts prepared from crosslinked cells. Data showed that all the Ty1 IN mutants (INK617A, INS621A, INL622A) were severely defected in their occupancy on the target genes as compared to the WT Ty1 IN (Figure 31). This suggests that this residues in the TS are indispensable for interacting with the target genes. In order to have a genome wide view of the occupancy of the Ty1 IN and the TS mutants on the target genes, we sent the immunoprecipitated DNA for sequencing at the I2BC platform in Gif-sur-yvette, France. Bioinformatics analysis by **Dr. Amna Asif-Laidin** in the lab of **Dr. Pascale LESAGE** revealed that in good agreement to the ChIP experiments, the WT Ty1 IN was found highly occupied at a 1 kb window upstream of the Pol III transcribed genes. Whereas, for the Ty1 TS mutants the occupancy was drastically reduced at the Pol III transcribed genes and were mostly found at the chromosomal ends (Data not shown). In particular the Ty1 INK617A mutant was most severely affected in its occupancy at the Pol III genes.

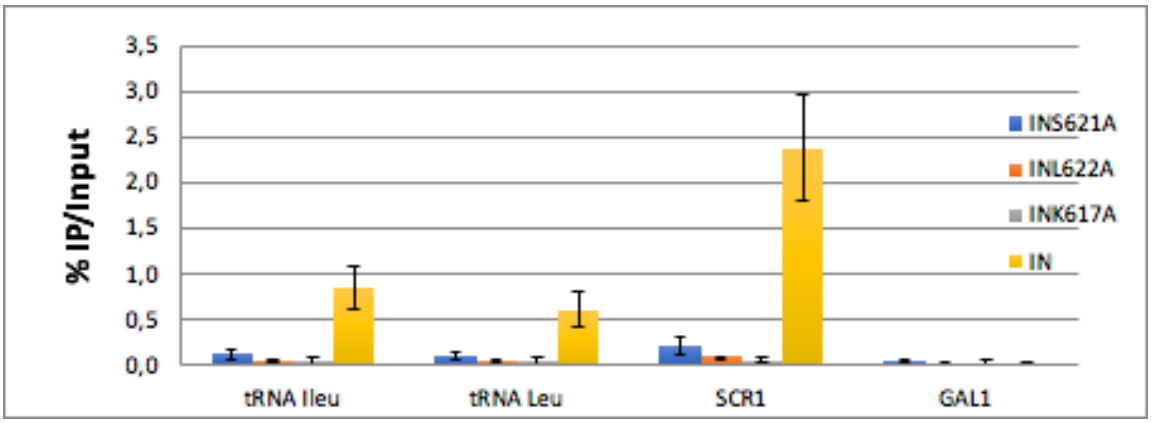


Figure 31. Single amino acid mutations disrupt association of Ty1 IN with the target genes *in vivo*. A HA- epitope tagged mutant Ty1 IN was induced in cells grown at 30°C. Proteins were immunoprecipitated by IgG beads coated with anti-HA (12CA5) antibody. The DNA was analyzed by qPCR with specific oligos in the target genes. *GAL1* was used as a negative control. The Y-axis represents the percentage of the amount of Ty1 IN/mutant Ty1 IN bound to the target gene compared to the total amount present in the input. The X-axis represents the name of the target genes (Performed by **Dr. Christine CONESA**)

Chapter 3

CONCLUSION AND PERSPECTIVES

Our main goal was to identify novel Ty1 IN partners and for the first time we have been able to present a large list of Ty1 IN interacting proteins *in vivo* using a large-scale proteomics approach. There have been many attempts in the past to identify Ty1 host factors and most of them were based on genetic screening. Although, many interesting proteins were identified through genetic studies, a proteomics approach was required to be able to find direct partners or targets closely associated with Ty1 IN. Recently, a proteomics screening was performed by **Cheung et al 2016** to identify Ty1 IN protein partners *in vivo*. In contrast to our approach, they performed purification after inducing Ty1 IN expression through a galactose inducible vector. This over expression of IN led to high background proteins and all the experiments were done in galactose thus changing the cellular metabolism. Very few partners of Ty1 IN were identified in this study. Among the 12 putative partners, 5 of them were RNA Pol III subunits. Moreover, none of the studies investigated the molecular role of the identified partners except for RNA Pol III in Ty1 retrotransposition *in vivo*. On the other hand, we have identified novel partners and validated that they are *bonafide* Ty1 IN interacting partners indicating the strength of our proteomics approach. In particular, we have studied the role of CK2 in Ty1 retrotransposition *in vivo* by various genetic and biochemical approaches.

Secondly, we have been closely involved in characterizing the targeting sequence (TS) of the Ty1 IN with the team of **Dr. Pascale LESAGE**. We have provided convincing evidences about the importance of the TS through biochemical assays. We have demonstrated by *in vivo* CoIP that mutations in the TS of Ty1 IN exhibits no interaction with either Pol III or Pol I in vivo and with AC40 subunit by *in vitro* CoIP. ChIP experiments have shown that mutant Ty1 IN lose their binding to target genes as compared to the WT Ty1 IN. In addition, ChIP seq data showed that the occupancy of the TS Ty1 IN mutants was highly disrupted at the target genes.

Currently, some additional experiments are performed in the lab to complete the analysis but the writing of two research articles involving the role of CK2 in Ty1 retrotransposition *in vivo* and characterization of the targeting sequence of Ty1 IN is ongoing. The former study will be my first co-authorship publication with **Anastasia BARKOVA** (Lab of Pascale LESAGE) to be submitted in 2019 and I will also be a co-author on the latter study.

1.1 Identifying Ty1 IN partners *in vivo* under specific conditions

Although our proteomics approach allowed the identification of true Ty1 partners, there are still some improvements we could bring in our methodology. For the moment, the targets of the Ty1 IN have been identified for an episomally expressed Ty1 IN HBH *in vivo*. It would be interesting to perform TChAP on the physiological Ty1 IN as a bait. We have developed a new vector that was able to express a functional Ty1 element where Ty1 IN was detected upon proteolytic processing of the TyB polyproteins (Results and discussions, Figure 16). Now, to be able to perform TChAP on physiological Ty1 INHBH we have modified the pet TCGA-Ty1 vector and introduced an HBH tag at the C-terminus of IN. In this vector, the orientation of the IN and RT in the TyB region has been reversed so the rather long C-terminalHBH tag doesn't interfere with the IN-RT processing site. We have been able to detect mature Ty1 IN HBH just like the WT Ty1 IN after proteolytic processing of the Ty1 poly proteins at 24°C *in vivo*. It might be possible to encounter new partners for the physiological Ty1 ININHBH.

Secondly, it would be interesting to identify partners specific to some particular conditions. Although, retrotransposition is almost minimal at 30°C compared to 20°C, TChAP conducted at 20°C did not lead to the identification of any specific partner at this temperature compared to 30°C. However, using a physiological Ty1 INHBH as a bait might give some specific partners at 20°C. Retrotransposition is higher when cells are subjected to other environmental stress such as low nitrogen starvation and acute adenine stress. It could be a good idea to perform TChAP under such stress conditions.

Thirdly, it will be interesting to perform TChAP of physiological Ty1 IN in *S. paradoxus* strain that contains no endogenous Ty1 elements. Lab strains that we used for our experiments contain 32 active copies of Ty1 and have a mechanism to inhibit extra copies of Ty1 through a mechanism called copy number control (CNC) (**Garfinkel 2003**). However, in *S. paradoxus* this CNC doesn't exist due to absence of any active Ty1 copies. We have observed that the physiological Ty1 IN is well detected by western blot under normal growth conditions in *S.paradoxus* compared to *S. cerevesiae.* Therefore, TChAP of Ty1 IN in this strain might enable us to identify novel partners in the absence of CNC.

Since, Ty1 IN modifications such as ubiquitinylation and phosphorylation appear to be important according to our data and in the literature (**Swany et al 2013**, **Duttler et al 2013**, **Mayor et al 2007**), it would be interesting to perform TChAP of physiological Ty1 INHBH upon conditions that preserve such modifications or inhibit them. For instance, interacting partners could be screened for Ty1 IN phosphomutant (absence of CK2 phosphorylation) or in cells treated with MG132 to inhibit proteasome and preserve ubiquitinylation.

It has been well established that Ty1 elements prefer the tRNA genes and there are multiple copies of tRNA genes *in vivo.* Ty1 integration at this locus doesn't pose danger to the cell's survival. Thus, these regions have been referred to as safe havens and it has been proposed that this a mechanism by which cells handle active copies of retroelements by avoiding genome damage (**Cheung et al 2018**). During our study, a tethering model was proposed by **Bridier Nahmias et al 2015** that integration occurs at tRNA genes due to an interaction between Ty1 IN and Pol III. However, they also demonstrated that under the loss of Ty1 IN-Pol III interaction, Ty1 integrates at subtelomeres without any change in the overall rate of retromobility *in vivo*. There are many uncharacterized genes in the subtelomeric region, but some genes are expressed in response to stress (**Ai et al 2002**). As the subtelomeric regions are devoid of the Pol III transcription machinery, one possibility could be the presence of another tethering factor at the subtelomeres. The fact that we have recovered some subtelomeric proteins in our TChAP results are very promising. Especially, Rap1 have been shown to associate with Ty1 IN *in vivo* and Reb1 both *in vivo* and *in vitro.* Though Rap1 is one of the Ty1 trancription factor and Reb1 is a Pol I transcription factor, we speculate that they might be involved in the Ty1 targeting at telomeres. Both Rap1 and Reb1 have been shown to bind to telomeres and regulate genes at the silent mating loci (**Giraldo et al 1994, Wahline et al 2000, Berthiau et al 2006**). For instance, several silencing factors at the subtelomeres Sir2, Sir3 and Sir4 bind to telomeric repeats via interaction with the telomeric binding protein Rap1 (**Rusche et al 2003**). Another study by **Smith et al 2011** demonstrated by ChIP experiments that occupancy of Rap1 and Reb1 proteins increases at subtelomeric regions in response to stress. Thus, these two targets should be tested for their role in Ty1 retrotransposition in our laboratory soon. To identify targets that might be involved in Ty1 targeting at the subtelomeres, we may perform a TChAP on a Ty1 IN TS mutant (Ty1 INK617A) which is most severely affected in both interaction with the Pol III complex and binding to tRNA genes. This mutant could be introduced in the physiological ptet TCGA-Ty1-HBH vector. It is possible that under the loss of Pol III association, new subtelomeric proteins might be identified.

1.2 Investigating the role of CK2 on Ty1 IN expression *in vivo*

We have investigated the role of CK2 on both Ty1 retrotransposition and Ty1 IN expression *in vivo.* In good agreement with **Nyswaner et al 2008,** we have shown that CK2 has a repressive effect on Ty1 mobility and might affect Ty1 expression post transcriptionally. We also demonstrated by *in vitro* phosphorylation assay that CK2 phosphorylates Ty1 IN and identified phosphorylated sites by mass spectrometry. Interestingly, in the lab of **Pascale LESAGE** they observed that the expression of endogenous Ty1 IN was slightly more abundant in the absence of the CK2 holoenzyme at both 30°C and 20°C whereas it is hardly detected in the WT cells. We detected the same phenotype when Ty1 IN was expressed from the ptet TCGA-Ty1 vector in the absence of CK2. Protein abundance is a balance between protein synthesis and protein degradation. Ubiquitin mediated proteasome degradation is predominant in eukaryotic cells and we have demonstrated that Ty1 IN is a substrate of proteasome and its inhibition by MG132 leads to a rapid accumulation of substantial amount of this protein which is in good agreement with previous studies that have also shown that Ty1 proteins are targeted by the proteasome. **Kaake et al 2010** identified targets of the Rpn11 (a subunit of the 26 proteasome) in yeast cells at various stages of the cell cycle (G1, S and M phase) and found several TyA and TyB proteins to be present in their mass spectrometry data suggesting they might be potential binding partners of the proteasome. In addition, several subunits of the proteasome were identified as putative Ty1 IN partners in our TChAP experiments.

All these data suggested that Ty1 IN might be unstable and short-lived under normal growth conditions but further experiments are required to clearly demonstrate it. Protein stability in eukaryotic cells could be investigated by a cycloheximide chase assay followed by western blot analysis that will allow to determine the half-life of the protein. Cells are treated with this translational inhibitor that will inhibit new protein synthesis and expression of the protein is followed over a time course. In eukaryotic cells, most short-lived proteins are degraded by the proteasome and before degradation are linked to ubiquitin chains that provide signal for degradation (**Chau et al 1989, Fuertes et al 2003**).

The ubiquitin mediated protein degradation happens in a series of enzymatic reactions (**Burger et al 2004**). The ubiquitin (Ub) is attached to the substrate protein by three enzymes called E1 (UB-activating enzyme, E2 (UB-conjugating enzyme) and E3 (UB ligase) which work sequentially. The first step involves the activation of the ubiquitin by the E1 enzymes. In an

ATP-dependent reaction, E1 adenylates the C-terminus of Ubiquitin and forms a thioester bond between each other. Full activity is achieved when the E1 binds non-covalently to and adenylates a second UB molecule. Then, the E1enzyme transfers the thioester linked ubiquitin to the next enzyme in the signaling cascade, E2 (Ub-conjugating enzyme). The active ubiquitin then links to the E2 by a thioester bond. Finally, E2 is able to interact with a third enzyme, E3 ligase with the help of which the ubiquitin is transferred from the E2 enzyme to the lysine residue on the substrate by forming an isopeptide bond between the lysine of the substrate the C-terminus of Ub. The E3 ligases are the only enzymes in the pathway that are subjected to regulation and bind to specific substrate sequences known as “degrons”. Different types of other E3 ligases, sometimes referred to as E4 (Ub- elongating enzyme) can add more Ub molecules to the first Ub to form a poly-Ub chain. There are seven lysine residues in ubiquitin that can be used to link Ub molecules to the target. Polyubiqutin chains at different positions affect the fate of the substrate, for example, if K11, K29 and K48 polyubiquitin chains are present, the protein is degraded by the proteasome and if K63 or K6 polyubiquitin chains mediates protein localization and trafficking. Thus, the E3 ligases play a critical role in regulating the fate of the proteins. It decides which proteins should be targeted with Ub, the number of Ub molecules needed to be linked and the positions where the polyubiquitin chains will be linked. Finally, the ubiquitinylated substrate is targeted to the 26S proteasome which is composed of two 19S regulatory caps on either ends and a 20S cylinder shaped proteolytic core. The 19S cap recognizes the ubiquitinylated substrate and recycles the ubiquitin molecules after which it is directed into the 20S core driving its degradation.

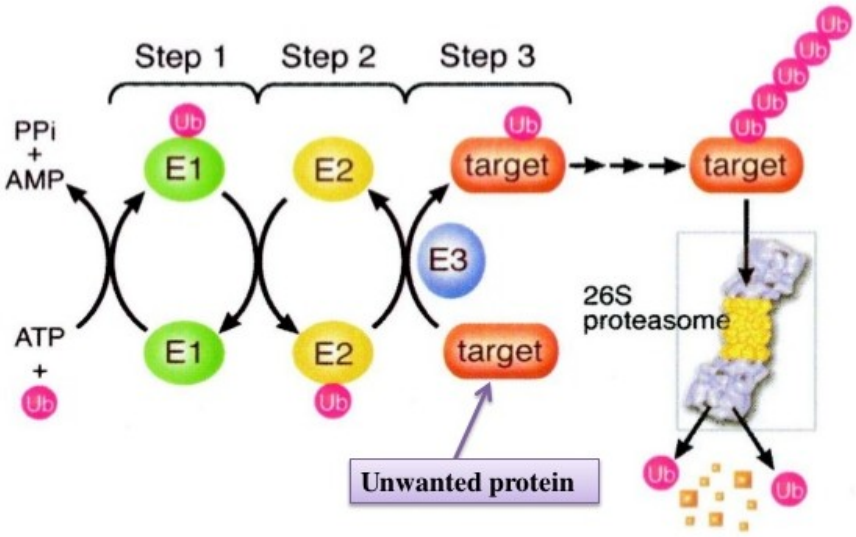


Figure 1. **Ubiquitination mediated proteasome degradation cycle**. The different steps of the pathway are provided (Image adapted from **Arpit Joshi, S.P university**).

There have been quite many studies in the past showing that Ty1 proteins are ubiquitinylated. **Duttler et al 2013** identified several Ty1 polypeptides that are ubiquitinylated cotranslationally from ribosome nucleoprotein complex by quantitative mass spectrometry. In another study by **Mayor et al 2007**, showed that the Ty1 IN proteins are enriched in a *Δrpn10* background by quantitative mass spectrometry. Rpn10 is a component of the 26S proteasome and binds to polyubiquitinylated proteins. Therefore, upon deletion of Rpn10 *in vivo*, 123 candidate proteins were found to be enriched among which 13 were Ty1/Ty2 proteins. All these data suggested that Ty1 proteins including the Ty1 IN are subjected to proteasome in a ubiquitin dependent manner. However, we need to demonstrate that upon our experimental conditions, Ty1 IN undergoes the same modification. To investigate this, we will induce its expression in WT strains, followed by pull down of ubiquitinylated proteins on Tandem ubiquitin binding entity (TUBE) agarose beads and visualize the presence or not of Ty1 IN; GAG and RT by western blot using specific antibodies. However, ubiquitinylation is reversible process and therefore, pull down assay should be done under denaturing conditions such as Urea containing buffer in order to inhibit loss of ubiquitin chains by deubiquitinating enzymes (DUBs).

The next step would be to identify the lysine residues that are ubiquitinylated by mass spectrometry following the purification of a physiological Ty1 INHBH under denaturing conditions. The identified lysine sites would be mutated to arginine in order to abolish ubiquitinylation and check for expression to determine whether the low abundance of Ty1 IN under normal growth condition is a result of ubiquitinylation.

On the other hand, we have also shown that Ty1 IN is phosphorylated and is a substrate of CK2 by *in vitro* assay. The phosphorylated sites of Ty1 IN by CK2 have been identified by mass spectrometry and all these sites also appear to be phosphorylated *in vivo*. In the absence of CK2, we observe an accumulation of Ty1 IN. All these data suggested that CK2 phosphorylation might be related to regulating the ubiquitinylation of Ty1 IN and degradation *in vivo*. Firstly, to confirm such hypothesis we will be investigating the expression of the Ty1 IN phosphomutants *in vivo*. We need to demonstrate that some specific modifications mimicking non-phosphorylated amino acid may be responsible for the accumulation. In addition, we could develop mutants that mimic constitutive phosphorylation and check whether they are less abundant even in the absence of CK2. This would confirm a direct effect of CK2 on Ty1 IN

abundance regulation. It would be also interesting to determine the influence of this CK2 phosphorylation on Ty1 IN stability.

Moreover, there are studies describing a strong interplay between phosphorylation and ubiquitinylation. Phosphorylation regulates ubiquitinylation in three different ways;

1. Phosphorylation might regulate both positively and negatively the activity of the E3 ligase responsible for the transfer of Ub.
2. Phosphorylation of a substrate can create a short sequence known as phosphodegron, thus promoting recognition of the substrate by E3 ligase
3. Phosphorylation can influence interaction between the ligase and the substrate (**Hunter 2007**).

We wonder which possibility suits better to our data. We do not believe that phosphorylation might regulate the activity of the E3 ligase since our preliminary data suggests a diret effect of phosphorylation on the Ty1 IN expression *in vivo*. On the other hand, if a phosphodegron in Ty1 IN is generated by CK2 then deletion of this region might lead to a higher accumulation of Ty1 IN *in vivo* and putative lysine ubiquitin sites should be present in close proximity to the phosphorylation sites. **Swaney et al 2013** conducted a global proteomics assay to identify ubiquitinylated proteins and ubiquitinylated lysine sites on such substrates *in vivo* in yeast and observed three lysine residues (**K10, K62, K531**) in the Ty1 IN to be ubiquitinylated which nowhere falls in close proximity to the identified CK2 phosphorylation sites of Ty1 IN *in vitro* (Figure 2). However, we need to confirm the ubiquitinylation sites of our physiological Ty1 IN by mass spectrometry.

NVHTSESTRKYPYPFIHRMLAHANAQTIRYSLKNNTITYFNESDVDWSSAIDYQ
CPDCLIGKSTKHRHIKGSRLKYQNSYEPFQYLHTDIFGPVHNLPNSAPSYFISFT
DETTKFRWVYPLHDRREDSILDVFTTILAFIKNQFQASVLVIQMDRGSEYTNRTL
HKFLEKNGITPCYTTTADSRAHGVAERLNRTLLDDCRTQLQCSGLPNHLWFSAIE
FSTIVRNSLASPKSKKSARQHAGLAGLDISTLLPFGQPVIVNDHNPNSKIHPR
GIPGYALHPSRNSYGYIIYLPSLKKTVDTTNYVILQGKESRLDQFNYDALTFDEDL
NRLTASYHSFIASNEIQESNDLNIESDHDFQSDIELHPEQPRNVLSKAVSPTDST
PPSTHTEDSKRVSKTNIRAPREVDPNISESNILPSKKRSSTPQISNIESTGSGGM
HKLNVPLLAPMSQSNTHESSHASKSKDFRHSDSYSENETNHTNVPISSTGGTNN
KTVPQISDQETEKRIIHRSPSIDASPPENNSSHNIVPIKTPTTVSEQNTEESIIAD
LPLPDLPPESPTEFPDPFKELPPINSRQTNSSLGGIGDSNAYTTINSKKRSLEDNE
TEIKVSRD TWNTKNMRSLEPPRSKKRIHLIA

Figure 2. **Ubiquitination and phosphorylation sites in Ty1 IN**. The different lysine residues observed by **Swaney et al 2013** to be ubiquitinylated are marked in red and the phosphorylation sites identified by in vitro phosphorylation by CK2 are marked in blue.

This leaves us with the third possibility that seem to be the most convincing one as previous study have shown that Ck2 phosphorylation can stimulate the interaction between E3 ligases and the substrate. For instance, **Klein et al 2017** showed that CK2 phosphorylation of the binding region of PDX-1, a regulator of insulin production in pancreatic cells, promotes binding of the E3 ligase PCIF1 and in turn proteasomal degradation. **Donghong Ju et al 2007** demonstrated that CK2 phosphorylation of the N-terminal domain of Rpn4, a transcriptional activator of the proteasome gene in yeast, enhances the binding between an E3 ligase Ubr2 and Rpn4 that relays a ubiquitin signal for further degradation of Rpn4 by the proteasome. Thus, based on these data, we have proposed a model where CK2 phosphorylation might have a direct effect on Ty1 IN stability *in vivo* presented below. Our model describes that under normal growth conditions, CK2 phosphorylation of the Ty1 IN leads to the recruitment of an E3 ligase that ubiquitinylates Ty1 IN leading to further degradation by the 26S proteasome. The link between CK2 phosphorylation and ubiquitinylation still needs to be investigated. For the moment, we do not know which E3 ligase might be involved in the ubiquitin transfer. There are many E3 ligases in the cell and it is difficult to examine all of them. Thus, an approach would be to identify ubiquitin ligases as binding partners of Ty1 IN under conditions that would preserve ubiquitination. TChAP experiments after treating cells with the proteasome inhibitor, MG132? If such approach would facilitate identification more precisely some ubiquitin ligase, it would be interesting to investigate whether their deletion could lead to higher Ty1 IN expression *in vivo*.

However, from our TChAP results and in vivo coimmunoprecipitation experiments, we have found an association between Bre1 and Ty1 IN *in vivo* in crosslinked cells. Bre1 is an E3 ubiquitin ligase and we speculate that it might have a role in ubiquitinylating Ty1 IN which needs to be investigated. Generally, Bre1 in cooperation with Rad6 is known to monoubiquitinate H2B and have also been identified as a repressor of Ty1 mobility by **Nyswaner et al 2008**. In another study, **Liu et al 2009**, showed that the human Bre1 is able to polyubiquitinylate Ebp1, a tumor suppressor gene, in cancer cells and leads to its degradation by the proteasome. So, it would be a good idea to test whether mutants mimicking constitutive phosphorylation, or no phosphorylation is able to associate with Bre1 *in vivo* by in vivo CoIP and yeast 2 hybrid assay.

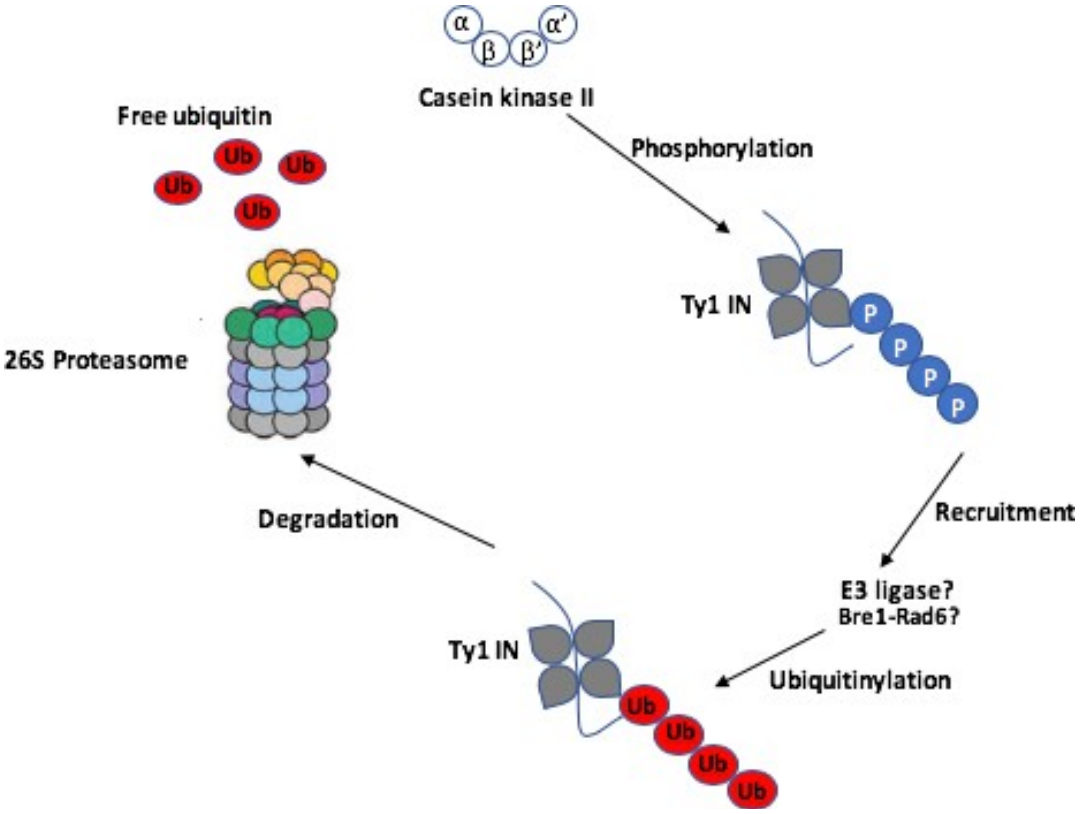


Figure 3. Model showing CK2 phosphorylation leads to ubiquitinylation of Ty1 IN and degradation by 26S proteasome. CK2 phosphorylates Ty1 IN at specific sites that might recruit an E3 ligase to ubiquitinylate Ty1 IN which is further directed to the proteasome for degradation.

1.3 Characterization of the Ty1 IN TS

During my PhD tenure, I also participated in the project involving the characterization of the targeting sequence (TS) of Ty1 IN. Initially, this TS was identified in the lab of Dr. Pascale LESAGE and we confirmed that this TS comprising of six amino acids in the C-terminal domain of Ty1 IN "KNMRSLE" is critical to be able to interact with its partner and also binding on target genes. Under our experimental conditions, mutations in K617, S621 and L622 highly disrupt interaction with Pol III and Pol I *in vivo* and AC40 *in vitro*. These mutations even severely hamper the binding of Ty1 IN on Pol III genes as seen in ChIP experiments. In good agreement, with these data, the group of Dr. Pascale LESAGE also showed that Ty1 elements harbouring these Ty1 IN TS mutants integrated mostly at the subtelomeres compared to the WT Ty1 with no change in retrotransposition frequency, a similar phenotype observed in the $AC40_{sp}$ loss of Ty1 IN-Pol III mutant (**Bridier Nahmias et al 2015**). The effect was more important in the Ty1 IN K617A mutant. All these data suggested that Ty1 IN is recruited specifically to the

Pol III genes through the interaction between the TS and its partner at this region confirming the tethering model.

An open question still arises as to why the Ty1 retroelements integrate in a 750 bp window upstream of Pol III genes when the Ty1 IN is directly recruited on the Pol III genes. Unlikely, the homologous retroelement, Ty3 has been seen integrating specifically at the transcription initiation sites of Pol III transcribed genes (**Chalker et al 1992**). This integration site choice is mediated by an interaction between the Brf1 subunit of TFIIIB and Ty3 IN shown by *in vitro* integration assay (**Yieh et al 2000**). Thus, the mechanism how Ty1 integrates at a long stretch of 750 bp window upstream of the Pol III genes needs to be elucidated. Two studies **Baller et al 2012** and **Mularoni et al** 2012 provide evidence that Ty1 prefers nucleosomal surface at the H2A/H2B interface through high-throughput screening studies. However, the transcription initiation site of Pol III genes is devoid of nucleosomes (**Helbo et al 2017**) suggesting the Ty3 prefers a nucleosome free region for integration. One possibility could be that unlike the interaction between Ty3 IN and TFIIIB which might be static as TFIIIB is tightly bound to the Pol III gene promoter, the interaction between Ty1 IN and Pol III complex might be fairly dynamic which is why the Ty1 element can access a long stretch of chromatin upstream of the tRNA. Another possibility could be that Ty1 IN which is a multimer is a large molecule and might interact with specific histones at the nucleosomes important for integration at these sites. However, these are all hypotheses and need to be investigated deeply.

On the other hand, such kind of study opens another very interesting field of research. This study may allow the possibility to modify the integration preference of other retroelements or other retroviruses. To demonstrate the feasibility of the project, the first step would be to modify the integration site of a related retroelement and investigate whether the modification affects its integration specificity. For this purpose, the Ty5 serves as very good candidate because compared to the other Ty elements in yeast which integrate mostly at Pol III transcribed genes, Ty5 integrates into the heterochromatin and extensive research on Ty5 led to the identification of a short 6 amino acid sequence, LXSSXP in the C-terminal domain of Ty5 IN that interacts with its tethering partner Sir4p to integrate at the heterochromatin (**Gai et al 1998, Xie et al 2001**). Moreover, in a study by **Zhu et al 2003** has demonstrated an approach to control integration site specificity of Ty5. They have engineered Ty5 elements by modifying the 6 amino acid TS of Ty5 IN with peptide sequences of either Rad9 or NpwBP that interacts with their partners Rad53 and Npw38 respectively. They have observed that the two mutant Ty5 IN

interact with their respective partners by yeast two hybrid assay with a similar strength like WT Ty5 IN and Sir4p. Integration assays showed that these newly engineered Ty5 elements integrated in close proximity to DNA sites where the protein partners were present without change in the Ty5 retromobility level *in vivo*. Currently, this project is under development in the lab of **Dr. Pascale LESAGE** and they have very promising preliminary results.

Chapter 4

MATERIALS AND METHODS

In this section, I will be describing the materials I have used particularly and the experiments I have performed. For instance, the ChIP experiments were designed and performed by **Dr. Christine CONESA** and the yeast two hybrid assay by both **Carine CHAGNEAU** and **Dr. Christine CONESA**. The MG132 assay was performed by **Carine CHAGNEAU**. **Dr. Joel ACKER** performed the *in vitro* phosphorylation assay and developed all the plasmids used in this project. The retromobility assay for determining the rate of retrotransposition *in vivo* and the qRT-PCR assay for measuring the steady state level Ty1 mRNA in WT and CK2 mutant yeast strains were performed by **Anastasia BARKOVA** in the lab of **Dr. Pascale LESAGE.**

1. Materials

1.1 Yeast strains

Strain	Genotype
BY4742/Trp-	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 Trp-
YPH500 C160-HA	MATα ura3-52, lys2-801, ade2-101, trp1-Δ63, his3-Δ200, leu2-Δ1,RPC160-HA
BY4742/Trp- Δcka2Δckb2	MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 Trp-, CKA2:kanMX, CKB2:natMX

All the TAP-tagged strains used for the co-immunoprecipitation experiments were obtained from the **CELLZOME** TAP-tagged yeast library.

1.2 Plasmids

Plasmid name	Description
pcm185-Ty1 IN	Centromeric yeast expression vector expressing Ty1 IN
pcm185-Ty1 IN HBH	Centromeric yeast expression vector expressing HBH tagged WT Ty1 IN
pcm185-Ty1 IN S621A HBH	Centromeric yeast expression vector expressing HBH tagged Ty1 INS621A
pcm185-Ty1 IN L622A HBH	Centromeric yeast expression vector expressing HBH tagged Ty1 INL622A

pcm185-Ty1 IN K617A HBH	Centromeric yeast expression vector expressing HBH tagged Ty1 INK617A
pcm185-6His-Ty1 IN	Centromeric yeast expression vector expressing His tagged Ty1 IN
pcm185-TCGA-Ty1	Centromeric yeast expression vector expressing Ty1 element
pcm185-TCGA-Ty1S473AT480AS499A	Centromeric yeast expression vector expressing mutant Ty1 element
pgal Ty1	Galactose inducible yeast expression vector expressing Ty1 element
pet28-IN- EPEA	Bacterial expression vector expressing EPEA tagged WT IN
pet28-INS621A-EPEA	Bacterial expression vector expressing EPEA tagged mutant INS621A
pet28-INL622A- EPEA	Bacterial expression vector expressing EPEA tagged mutant INL622A
pet28-INK617A- EPEA	Bacterial expression vector expressing EPEA tagged mutant INK617A
pacyc184-AC40-streptagtwin	Bacterial expression vector expressing Streptagtwin tagged AC40 subunit
pgex 3x-RAP1-GST	Bacterial expression vector expressing GST tagged RAP1
pgex 3x-REB1-GST	Bacterial expression vector expressing GST tagged REB1
pgex 3x-BRE1-GST	Bacterial expression vector expressing GST tagged BRE1
pgex 3x-DST1-GST	Bacterial expression vector expressing GST tagged DST1
pgex 3x-TUP1-GST	Bacterial expression vector expressing GST tagged TUP1
pet28-SSO7d-IN-EPEA	Bacterial expression vector expressing N-term SSO7d and C-term EPEA tagged Ty1 IN
pet28-Fh8-IN-EPEA	Bacterial expression vector expressing N-term Fh8 and C-term EPEA tagged Ty1 IN

2. Methods

2.1 Transformation in yeast

1. Pre-culture of yeast strain was set up overnight at 30°C in 15 ml of YPD (Yeast peptone dextrose) media.
2. The following day cells were diluted at OD_{600} =0.2 and grown until OD_{600} 0.8 -1 (exponential phase).
3. 10 ml of cells at exponential phase were harvested in a 15 ml falcon by centrifuging at 3500 rpm for 3 min at room temperature and the cell pellet was resuspended in 50µl of LiAc/TE buffer (100 mM LiAc, 10 mM Tris-HCl pH=8, EDTA 1 mM). The mix was transferred to a new Eppendorf tube.
4. 5µl of 10 mg/ml of herring sperm DNA was denatured and put on ice for 5 min before use. The sperm DNA was added to the mix following which 250 ng of plasmid DNA and 350µl of LiAc/TE/PEG buffer (LiAc/TE, Polyethylene glycol 40%) were added.
5. The tube was incubated at 30°C for 30 min, and then heat shocked at 42°C for 15 min.
6. The buffer was then removed by centrifuging rapidly and cells were resuspended in 100µl sterile water before plating on selective media agar plates.

2.2 Tandem Chromatin Affinity purification

The TChAP method described here has been optimized for the purification of Ty1 INHBH and its associated partners. This method was developed in our laboratory for the identification of partners associated with the Pol III transcription machinery. The classical procedure is described in **Nguyen et al 2015**.

Day-1 (Setting up pre -culture)

1. The yeast strain transformed with the pcm185-Ty1 INHBH plasmid was inoculated in 10 ml of minimal growth medium (CASA medium, 4X excess amino acids and doxycycline 10µg/ml) and grown overnight at 30°C.

Day-2 (Inducing the expression of Ty1 INHBH)

1. The cells were diluted accordingly in a large volume of fresh medium (8 litres in our case) in the absence of doxycycline so as to induce the Ty1 INHBH expression

overnight. Note that the following day cells need to be at exponential phase (OD_{600} = 1–1.2) before harvesting

Chromatin preparation

Day-3 (*In vivo* crosslinking of cells, extract preparation)

This step here is discussed for a single culture of 2 litres volume. For large volume multiple 2 liters chromatin extracts were pooled together.

1. The cells at exponential phase were crosslinked with 1% Formaldehyde (prepared in 1X PBS) for 20 min.
2. The cells were then harvested by centrifuging at 3500 rpm at 4°C for 5 min. The rest of the formaldehyde were removed by washing the cell pellets 3 times with ice cold PBS (1X) and once in sucrose buffer (300mM sucrose, 1%Triton X-100, PBS 1X) with centrifugations between each wash.
3. Finally the cells were washed in Urea buffer 1 (20 mM NaH_2PO_4/Na_2HPO_4 pH=7.5, 6M Urea, 1M NaCl, 0.1% SDS, 0.1% Sarkosyl, 10mM Imidazole and 5mM β-mercaptoethanol, pH 8) and then resuspended in 15ml Urea buffer 1 containing EDTA-free protease inhibitor mixture and 100mM PMSF.
4. The cells were then frozen at -80°C in an Eaton press for 1hr.
5. The frozen cells were passed through a small orifice in an Eaton press subjected to 6 tones of pressure. The cell lysates were collected in a 50 ml falcon tube and thawed at room temperature.
6. The chromatin was solubilized and sheared using a Q700 sonicator with a microtip probe (Qsonica). The sonicator was set to 5 cycles of 10 sec ON followed by 50 sec OFF with 70% amplitude. The probe and the extract were kept cold on ice and temperature was not allowed to reach beyond 20°C. Samples, 500µl were taken before and after sonication to assay the amount of DNA and protein in the pellet and supernatant.
7. The resulting lysates were clarified by centrifuging for 30 min at 10000 rpm at 10°C. The supernatant was collected in a separate 50ml falcon to be subjected to the purifier.

Day-4 (Tandem Affinity Purification)

1. Chromatin extract prepared from 8 litres culture approximately 30 ml of sample was subjected to Nickel affinity purification using an AKTA purifier system (GE Healthcare). Before injecting the sample the system was equilibrated with Urea buffer 1.
2. To avoid non-specific binders the sample was first loaded on to a 5ml sepharose fast flow column screwed to another 5 ml nickel sepharose column allowing pre-clearing of the chromatin extract before metal affinity capture. The sample was injected for a total of 3 times.
3. After the complete injection, the system was washed extensively with Urea buffer 2 (20 mM NaH_2PO_4/Na_2HPO_4 pH=7.5, 6M Urea, 300mM NaCl, 0.1% SDS, 0.1% Sarkosyl, 10mM Imidazole and 5mM β-mercaptoethanol, pH 6.3)
4. The proteins from the nickel column were finally eluted with Urea buffer 3 (20 mM NaH_2PO_4/Na_2HPO_4 pH=7.5, 6M Urea, 300mM NaCl, 0.1% SDS, 0.1% Sarkosyl, 10mM EDTA and 5mM β-mercaptoethanol, pH 4.3). The elution fractions were collected in approximately 12 ml volume neutralized with Tris-HCl pH=8.8
5. In parallel, streptavidin magnetic sepharose slurry beads, 300µl were equilibrated with Urea buffer 1 by washing the beads for 5 min on a rotor at room temperature. The eluted fraction from the nickel column (approximately 12 ml) was added to the streptavidin beads and incubated on a rotor at 10°C overnight.

Day-5 (Elution of the proteins from streptavidin beads and analysis of the proteins)

1. The following day the supernatant was cleared from the beads using a magnetic separator. The beads were then washed once with Urea buffer 1 and 3 times with PBS(1X) for 5 min (each wash) on a rotor at room temperature.
2. The beads were then resuspended in 500µl of reversal buffer (250mM Tris HCl pH=8.8, 2% SDS and 0.5M β-mercaptoethanol) and boiled at 95°C for 15 minutes (3 * 5 min at 95°C with 30 sec on ice).
3. The final elution of 500µl was transferred to a low-protein binding Eppendorf tube and 20% Trichloro-acetic acid was added to it for precipitation of the total protein. The mixture was vortexed and incubated on ice for 10 min.

4. Following which the protein pellet was collected by centrifuging the mixture at 4°C at 15000rpm for 15 min.
5. The protein pellet was washed twice with 1ml of ice cold acetone to remove the excess TCA and dried under the hood for 5 min. Finally, the pellet was resuspended in 20µl of reversal buffer and boiled for 5 min at 95°C.
6. The protein sample was then diluted with a 5X sample buffer (Thermofisher scientific) according to the supplier's protocol and resolved by SDS PAGE on a 4-12% gradient NuPAGE gel.
7. The bands were visualized by staining the gel with Imperial blue and de-staining with water overnight.
8. The following day the bands were sliced from the gel and sent to the mass spectrometry facility, IBB, Warsaw, Poland to identify the proteins.

2.3 *In vivo* co-immunoprecipitation (CoIP)

Harvesting of cells

1. Overnight grown yeast cultures (in 1µg/ml doxycycline) were diluted in fresh 100 ml minimal growth medium (CASA medium + correct selectable marker) in the absence of doxycycline.
2. The following day the cells were collected at exponential phase and harvested by centrifugation at 3500 rpm for 3 min at room temperature.
3. The pellet was washed with water and transferred to an Eppendorf tube.

Preparation of extracts

1. The cell pellets were re-suspended in 500µl of IP buffer (50 mM HEPES-KOH pH 7.5, 300mM NaCl, 1mM EDTA, 0.05% NP40, 0.5mM DTT, 5% glycerol, 1mM PMSF and protease inhibitor cocktail tablet).
2. 250 µl of glass beads were then added to the pellets and lysed at 4°C for 45min using a Vibrax at maximum speed.
3. Cell lysate was collected and the extracts were centrifuged at 15000 rpm, 4°C for 15 min.
4. The protein extract, 500µl was transferred to a separate Eppendorf tube.

Preparation of IgG magnetic beads

1. 50µl of magnetic Dynabeads PanMouse (Invitrogen) were washed 2-3 times with 1ml of PBS-BSA buffer (PBS+BSA 0.5 %) in a 2-ml eppendorf tube.
2. 50µL of PBS-BSA buffer was added to the beads with 1-2 µg of the antibody (Anti-HA in our case). Incubate on a rotor at 4°C for 1hr.
3. Following which the beads were washed once with 1 ml PBS-BSA buffer and 1 ml of IP buffer respectively to equilibrate them.

Incubation of protein extracts with Ig-magnetic beads

1. The protein extracts (500µl) were incubated with 50µl of 12CA5 (anti-HA) bound IgG magnetic PanMouse Dynabeads at 4°C on a rotor for 1hr.
2. After incubation, wash the beads 3 times with 1ml of IP Buffer and transfer the beads to a new Eppendorf tube before elution.

Elution of the bound proteins

1. The proteins were eluted from the beads by adding 20µl of sample buffer(50mM Tris-HCl, 2% SDS, 10% glycerol, 2% β-mercaptoethanol) and boiling it at 95°C for 5-10 min.

2.4 TCA extraction

1. 200 µl of 20%TCA and 100µl acid washed glass beads were added to the cell pellet, vortexed for 1 min and kept on ice.
2. The cell lysate was transferred to another tube and 400µl of 5% TCA was added to the beads. The beads were then vortexed again for 30sec.
3. The final cell lysate was centrifuged at high speed for 10 min at 4°C.
4. The supernatant was removed completely and 10µl of Tris 1M pH=8.8 + 40µl l of TCA sample buffer (2X) was added. If the solution was yellow 5µl of 1M Tris-HCl (pH=8) was added to neutralize the acid.
5. The sample was then boiled at 95°C for 5 min before loading on a SDS gel.

2.5 *In vitro* coimmunoprecipitation

This procedure was used to demonstrate a direct interaction between AC40 and the Ty1 IN.

Induction of proteins and harvesting of cells

1. BL21(DE3) bacterial cells were transformed with both the plasmids, pet28-IN-EPEA and pacyc184-AC40-Streptagtwin and grown at 30°C on LB agar plates supplemented with the appropriate antibiotics (Ampicillin and Chlorophenicol) and 1% glucose.
2. The following day, bacterial clones were inoculated in 15 ml LB medium+1% glucose and grown at 30°C overnight.
3. Overnight grown cells (OD_{600} = 3) were diluted to OD_{600} 0.05 in fresh 100 ml LB media (without glucose) and grown to OD_{600} 0.5 at 30°C.
4. The cells were then transferred to 24°C for 30 min before adding IPTG to a final concentration of 0.5 mM to induce expression of the proteins for 3hr.
5. Cells were then harvested by centrifugation at 3500 rpm for 15 min at room temperature, washed with water and transferred to a 2ml Eppendorf tube.

Protein extract preparation

1. Cell pellets were resuspended in 1 ml of the extraction buffer (20mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol, 0.1% NP40) before adding 1µg/ml lysozyme. The mix was kept on ice for 10 min.
2. Cells were then sonicated using a Q700 sonicator with a microtip probe (Qsonica). The sonicator was set to 5 cycles of 10% amplitude for 2 sec ON followed by 40 sec OFF. The tubes were always kept on ice while sonication.
3. The cell lysate was centrifuged at 15000 rpm at 4°C for 30 min and the supernatant (approximately 1ml) was collected in a separate Eppendorf tube.
4. The total protein concentration was estimated by Bradford assay.

Co-immunoprecipitation

1. Magnetic Strep tactin beads (for IP of the AC40-Streptagtwin), 40µl was used for each experiment. The beads were washed twice with PBS (1X) and then incubated with PBS-

BSA buffer (PBS+0.5%BSA) for 30 min on a rotor at 4°C for blocking. Following which, the beads were washed once with the extraction buffer to equilibrate.

2. The protein extract, 800µl was added to the beads and incubated at 4°C on a rotor for 3hr.
3. After the incubation, the magnetic beads were separated from the extract using a magnetic separator and washed once with 1 ml of extraction buffer. The beads were then transferred to a fresh 2 ml Eppendorf tube before to elute the proteins with 50µl sample buffer (2X) by boiling at 95°C for 10 min.
4. The eluted fraction, 25µl was taken for western blot analysis and visualization of the proteins.

2.6 GST pull down assay

This procedure was used to demonstrate a direct interaction between the protein partners and the Ty1 IN.

Preparation of extracts of GST-tagged protein targets

1. BL21(DE3) bacterial cells were transformed with the plasmid, pgex 3x-X-GST (X is the protein partner) grown at 30°C on LB agar plates supplemented with the appropriate antibiotics (Ampicillin).
2. The following day, bacterial clones were inoculated in 15 ml LB medium and grown at 30°C overnight.
3. Overnight grown cells (OD_{600} = 3) were diluted to OD_{600} 0.05 in fresh 100 ml LB media (without glucose) and grown to OD_{600} 0.5 at 30°C.
4. The proteins were induced by IPTG to a final concentration of 0.5 mM for 3hr.
5. Cells were then harvested by centrifugation at 3500 rpm for 15 min at room temperature, washed with water and transferred to a 2ml Eppendorf tube.
6. 1 ml extraction buffer (20mM Tris-HCl pH 7.5, 300 mM NaCl, 10% glycerol, 0.1% NP40) and 1µg/ml lysozyme were added to the cell pellet. The mix was kept on ice for 10 min.
7. Cells were then sonicated using a Q700 sonicator with a microtip probe (Qsonica). The sonicator was set to 5 cycles of 10% amplitude for 2 sec ON followed by 40 sec OFF. The tubes were always kept on ice while sonication.

8. The cell lysate was centrifuged at 15000 rpm at 4°C for 30 min and the supernatant (approximately 1ml) was collected in a separate Eppendorf tube.
9. The total protein concentration was estimated by Bradford assay.

Preparation of extracts of either SSO7d-IN-EPEA or Fh8-IN-EPEA

1. The same procedure for extraction as described above has been followed with the exception being that after the sonication, the cell lysate was incubated for 1hr on a rotor at 4°C for 1hr before centrifugation (this helps in better solubilization of the IN in the extract).

Pull down assay

1. The protein extracts, 1ml was incubated with sepharose slurry for 1 hr at 4°C on a rotor. This removes the non-specific binders and allows pre-clearing. The mix was centrifuged hard for 5 min to separate the slurry and recuperate the extract.
2. In parallel, GST sepharose slurry (Thermofisher beads), 60µl was taken and incubated with PBS-BSA buffer (PBS+0.5% BSA) for 1hr, following which they were washed once with 1 ml extraction buffer for equilibration.
3. The pre-cleared extract was incubated with the GST beads for 1hr on a rotor at 4°C.
4. After incubation, the beads were washed 3 times with 1 ml extraction buffer for 5min each on a rotor at 4°C.
5. Now, this protein-GST bound beads were incubated with the pre-cleared extract containing either SSO7d-IN-EPEA or Fh8-IN-EPEA for 3hr or overnight at 4°C on a rotor.
6. After incubation, the beads were washed 3 times with 1 ml extraction buffer for 5min each on a rotor at 4°C and then proteins were eluted by 50µl of sample buffer (2X) and boiling at 95°C for 10min.
7. The elution, 25µl (50%) was taken for SDS-PAGE and visualization of the proteins by western blot.

2.7 Western blotting

Primary/Secondary Antibody	Dilution	Blocking buffer
Anti-Streptavidin-HRP	**1:15000**	**PBST+2.5%BSA**
Anti-TAP /Anti-Rabbit	**1:5000/1:10000**	**TBST+5% Milk**
Anti-EPEA/Anti-streptavidin	**1:5000/1:10000**	**PBST+2.5%BSA**
Anti-HA/Anti-Mouse	**1:100000/1:10000**	**TBST+5% Milk**
Anti-IN (serum)/anti-Rabbit	**1:10000/1:10000**	**TBST+5% Milk**
Anti-IN (purified)/anti-Rabbit	**1:25000/1:10000**	**TBST+5% Milk**
Anti-Streptactin-HRP	**1:5000**	**PBST+2.5%BSA**
Anti-Actin/Anti-Mouse	**1:10000/1:10000**	**TBST+5% Milk**

1. Protein samples were denatured using SDS sample buffer and boiling at 95°C for 5 min before loading on an 8% SDS gel and resolved by PAGE in a migration buffer (192mM Glycine, 25 mM Tris and 0.1%SDS).
2. Proteins from the gel were transferred to a nitrocellulose membrane in a transfer buffer (36mM glycine, 48mM Tris, 20% ethanol and 0.04% SDS) in an electrotransfer tank at 10V and 1mA/cm^2.
3. The membrane was blocked in the blocking buffer depending on the primary antibody used (see table above) for 1hr at room temperature on a shaker.
4. The membrane was washed once with TBST (TBS+0.1%Tween-20) and incubated with the primary antibody for 1hr.
5. The membrane was then washed twice with the TBST buffer for 5 min each before adding the secondary antibody dilution and incubated for 30 min.
6. The membrane was washed at least 3 times with TBST before revealing in an autoradiography machine using ECL substrate (ECL Kit Thermofisher scientific).

ANNEXE

Table 1. The proteins identified as putative partners after curation of mass spectrometry data

Protein			Protein				Protein	
RPC40			TRM82	Methyltransferase			PRE6	
RPO31			TRM82				PUP2	
RET1							PRE9	20S proteasome
RPC82							PRE10	
RPC53			TFC4				SCL1	
RPB5	RNA Pol III complex		TFC3				PRE4	
RPC34			TFC1	Pol III transcrition factors identifed at 20°C				
RPC25			TFC7					
RPC19			BRF1				RPN3	
			BDP1				RPN7	
RPA135			SPT15				RPT3	
RPA190							RPN8	
RPA49	RNA Pol I complex		SPT16	FACT			RPN6	26S proteasome
RPA34			POB3				RPN5	
RPA43							RPT4	
RPC34			SMT3	Sumo			RPN11	
RPC25							RPN1	
RPC19			UBP6	Deubiquitinase			RPN2	
			UBP14				RPT1	
CKB1			UBP12				RPT5	
CKB2	Casein kinase II						RPT6	
CKA1			UBA2	Sumo activating enzyme				
CKA2			DBP3	RNA helicase				
RAP1			RVB2	Chromatin remodeller				
REB1	Transcription factors		RVB1					
DST1								
			DUS3	Dihydrouridine synthase				
BRE1	Ubiquitin ligase							
CEG1								
CET1	5'-mRNA capping enzyme							
TUP1	Transcription repressor							
CYC8								

BIBLIOGRAPHY

Acker, J., Conesa, C., and Lefebvre, O. (2013). Yeast RNA polymerase III transcription factors and effectors. Biochim Biophys Acta *1829*, 283–295.

Ai, W., Bertram, P.G., Tsang, C.K., Chan, T.F., and Zheng, X.F.S. (2002). Regulation of subtelomeric silencing during stress response. Mol. Cell *10*, 1295–1305.

Alexander, D.E., Kaczorowski, D.J., Jackson-Fisher, A.J., Lowery, D.M., Zanton, S.J., and Pugh, B.F. (2004). Inhibition of TATA binding protein dimerization by RNA polymerase III transcription initiation factor Brf1. J. Biol. Chem. *279*, 32401–32406.

Aristizabal, M.J., Negri, G.L., and Kobor, M.S. (2015). The RNAPII-CTD Maintains Genome Integrity through Inhibition of Retrotransposon Gene Expression and Transposition. PLoS Genet. *11*, e1005608.

Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G., and Llyin, Y.V. (1986). The steps of reverse transcription of Drosophila mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. Cell *44*, 555–563.

Arrebola, R., Manaud, N., Rozenfeld, S., Marsolier, M.C., Lefebvre, O., Carles, C., Thuriaux, P., Conesa, C., and Sentenac, A. (1998). Tau91, an essential subunit of yeast transcription factor IIIC, cooperates with tau138 in DNA binding. Mol. Cell. Biol. *18*, 1–9.

Ayarpadikannan, S., and Kim, H.-S. (2014). The Impact of Transposable Elements in Genome Evolution and Genetic Instability and Their Implications in Various Diseases. Genomics Inform *12*, 98–104.

Aye, M., Dildine, S.L., Claypool, J.A., Jourdain, S., and Sandmeyer, S.B. (2001). A Truncation Mutant of the 95-Kilodalton Subunit of Transcription Factor IIIC Reveals Asymmetry in Ty3 Integration. Molecular and Cellular Biology *21*, 7839–7851.

Bachman, N., Gelbart, M.E., Tsukiyama, T., and Boeke, J.D. (2005). TFIIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of Isw2p to S. cerevisiae tDNAs. Genes Dev. *19*, 955–964.

Baller, J.A., Gao, J., Stamenova, R., Curcio, M.J., and Voytas, D.F. (2012). A nucleosomal surface defines an integration hotspot for the Saccharomyces cerevisiae Ty1 retrotransposon. Genome Res. *22*, 704–713.

Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: New perspectives on an old relation. PNAS *101*, 14572–14579.

Baur, M., Esch, R.K., and Errede, B. (1997). Cooperative binding interactions required for function of the Ty1 sterile responsive element. Mol. Cell. Biol. *17*, 4330–4337.

Bennetzen, J.L. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. Curr. Opin. Genet. Dev. *15*, 621–627.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. *22*, 615–626.

Berthiau, A.-S., Yankulov, K., Bah, A., Revardel, E., Luciano, P., Wellinger, R.J., Géli, V., and Gilson, E. (2006). Subtelomeric proteins negatively regulate telomere elongation in budding yeast. EMBO J *25*, 846–856.

Biémont, C., and Vieira, C. (2006). Genetics: Junk DNA as an evolutionary force.

Birch, J.L., Tan, B.C.-M., Panov, K.I., Panova, T.B., Andersen, J.S., Owen-Hughes, T.A., Russell, J., Lee, S.-C., and Zomerdijk, J.C.B.M. (2009). FACT facilitates chromatin transcription by RNA polymerases I and III. EMBO J *28*, 854–865.

Błaszczyk, L., Biesiada, M., Saha, A., Garfinkel, D.J., and Purzycka, K.J. (2017). Structure of Ty1 Internally Initiated RNA Influences Restriction Factor Expression. Viruses *9*.

Bogdanov, A.A., Dontsova, O.A., Dokudovskaya, S.S., and Lavrik, I.N. (1995). Structure and function of 5S rRNA in the ribosome. Biochem. Cell Biol. *73*, 869–876.

Bolton, E.C., Mildvan, A.S., and Boeke, J.D. (2002). Inhibition of Reverse Transcription In Vivo by Elevated Manganese Ion Concentration. Molecular Cell *9*, 879–889.

Bradshaw, V.A., and McEntee, K. (1989). DNA damage activates transcription and transposition of yeast Ty retrotransposons. Mol. Gen. Genet. *218*, 465–474.

Brady, T.L., Fuerst, P.G., Dick, R.A., Schmidt, C., and Voytas, D.F. (2008). Retrotransposon Target Site Selection by Imitation of a Cellular Protein. Molecular and Cellular Biology *28*, 1230–1239.

Braiterman, L.T., and Boeke, J.D. (1994). In vitro integration of retrotransposon Ty1: a direct physical assay. Mol Cell Biol *14*, 5719–5730.

Bridier-Nahmias, A., Tchalikian-Cosson, A., Baller, J.A., Menouni, R., Fayol, H., Flores, A., Saïb, A., Werner, M., Voytas, D.F., and Lesage, P. (2015). Retrotransposons. An RNA polymerase III subunit determines sites of retrotransposon integration. Science *348*, 585–588.

Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature *334*, 213–218.

Brow, D.A., and Guthrie, C. (1990). Transcription of a yeast U6 snRNA gene requires a polymerase III promoter element in a novel position. Genes Dev. *4*, 1345–1356.

Brun, I., Sentenac, A., and Werner, M. (1997). Dual role of the C34 subunit of RNA polymerase III in transcription initiation. The EMBO Journal *16*, 5730–5741.

Bryk, M., Banerjee, M., Murphy, M., Knudsen, K.E., Garfinkel, D.J., and Curcio, M.J.(1997). Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. Genes Dev. *11*, 255–269.

Burger, A.M., and Seth, A.K. (2004). The ubiquitin-mediated protein degradation pathway in cancer: therapeutic implications. Eur. J. Cancer *40*, 2217–2229.

Bushman, F. (2002). Targeting Retroviral Integration? Molecular Therapy *6*, 570–571.

Cameron, J.R., Loh, E.Y., and Davis, R.W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. Cell *16*, 739–751.

Capy, P., Vitalis, R., Langin, T., Higuet, D., and Bazin, C. (1996). Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? J. Mol. Evol. *42*, 359–368.

Carr, M., Bensasson, D., and Bergman, C.M. (2012). Evolutionary genomics of transposable elements in Saccharomyces cerevisiae. PLoS ONE *7*, e50978.

Carter, R., and Drouin, G. (2010). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. Mol. Biol. Evol. *27*, 1035–1043.

Chalker, D.L., and Sandmeyer, S.B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. *6*, 117–128.

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., and Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science *243*, 1576–1583.

Chaussivert, N., Conesa, C., Shaaban, S., and Sentenac, A. (1995). Complex Interactions between Yeast TFIIIB and TFIIIC. J. Biol. Chem. *270*, 15353–15358.

Chédin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. Genes Dev. *12*, 3857–3871.

Chen, Y., Rai, R., Zhou, Z.-R., Kanoh, J., Ribeyre, C., Yang, Y., Zheng, H., Damay, P., Wang, F., Tsujii, H., et al. (2011). A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. Nat. Struct. Mol. Biol. *18*, 213–221.

Chénais, B., Caruso, A., Hiard, S., and Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. Gene *509*, 7–15.

Cherepanov, P., Ambrosio, A.L.B., Rahman, S., Ellenberger, T., and Engelman, A. (2005). Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc. Natl. Acad. Sci. U.S.A. *102*, 17308–17313.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. *278*, 372–381.

Cheung, S., Ma, L., Chan, P.H.W., Hu, H.-L., Mayor, T., Chen, H.-T., and Measday, V. (2016). Ty1 Integrase Interacts with RNA Polymerase III-specific Subcomplexes to Promote Insertion of Ty1 Elements Upstream of Polymerase (Pol) III-transcribed Genes. J. Biol. Chem. *291*, 6396–6411.

Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E.P., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., and Hunt, D.F. (2007). Analysis of phosphorylation sites on

proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. *104*, 2193–2198.

Ciganda, M., and Williams, N. (2011). Eukaryotic 5S rRNA biogenesis. Wiley Interdiscip Rev RNA *2*, 523–533.

Clare, J., and Farabaugh, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. U.S.A. *82*, 2829–2833.

Clare, J.J., Belcourt, M., and Farabaugh, P.J. (1988). Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. Proc. Natl. Acad. Sci. U.S.A. *85*, 6816–6820.

Conesa, C., Swanson, R.N., Schultz, P., Oudet, P., and Sentenac, A. (1993). On the subunit composition, stoichiometry, and phosphorylation of the yeast transcription factor TFIIIC/tau. J. Biol. Chem. *268*, 18047–18052.

Conte, D., Barber, E., Banerjee, M., Garfinkel, D.J., and Curcio, M.J. (1998). Posttranslational Regulation of Ty1 Retrotransposition by Mitogen-Activated Protein Kinase Fus3. Molecular and Cellular Biology *18*, 2502–2513.

Costa, S.J., Coelho, E., Franco, L., Almeida, A., Castro, A., and Domingues, L. (2013). The Fh8 tag: a fusion partner for simple and cost-effective protein purification in Escherichia coli. Protein Expr. Purif. *92*, 163–170.

Covey, S.N. (1986). Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. *14*, 623–633.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., et al. (2008). Structure of eukaryotic RNA polymerases. Annu Rev Biophys *37*, 337–352.

Cristofari, G., Ficheux, D., and Darlix, J.L. (2000). The GAG-like protein of the yeast Ty1 retrotransposon contains a nucleic acid chaperone domain analogous to retroviral nucleocapsid proteins. J. Biol. Chem. *275*, 19210–19217.

Curcio, M.J., and Garfinkel, D.J. (1991). Single-step selection for Ty1 element retrotransposition. Proc. Natl. Acad. Sci. U.S.A. *88*, 936–940.

Curcio, M.J., Hedge, A.-M., Boeke, J.D., and Garfinkel, D.J. (1990). Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in Saccharomyces cerevisiae. Molec. Gen. Genet. *220*, 213–221.

Dakshinamurthy, A., Nyswaner, K.M., Farabaugh, P.J., and Garfinkel, D.J. (2010). BUD22 Affects Ty1 Retrotransposition and Ribosome Biogenesis in Saccharomyces cerevisiae. Genetics *185*, 1193–1205.

Deprez, E., Arrebola, R., Conesa, C., and Sentenac, A. (1999). A Subunit of Yeast TFIIIC Participates in the Recruitment of TATA-Binding Protein. Molecular and Cellular Biology *19*, 8042–8051.

Devine, S.E., and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev. *10*, 620–633.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. *35*, 41–48.

Dieci, G., Giuliodori, S., Catellani, M., Percudani, R., and Ottonello, S. (2002). Intragenic Promoter Adaptation and Facilitated RNA Polymerase III Recycling in the Transcription of *SCR1*, the 7SL RNA Gene of *Saccharomyces cerevisiae*. Journal of Biological Chemistry *277*, 6903–6914.

Dingermann, T., Frank-Stoll, U., Werner, H., Wissmann, A., Hillen, W., Jacquet, M., and Marschalek, R. (1992). RNA polymerase III catalysed transcription can be regulated in Saccharomyces cerevisiae by the bacterial tetracycline repressor-operator system. EMBO J. *11*, 1487–1492.

Domańska, K., Zieliński, R., Kubiński, K., Sajnaga, E., Masłyk, M., Bretner, M., and Szyszka, R. (2005). Different properties of four molecular forms of protein kinase CK2 from Saccharomyces cerevisiae. Acta Biochim. Pol. *52*, 947–951.

Duttler, S., Pechmann, S., and Frydman, J. (2013). Principles of Cotranslational Ubiquitination and Quality Control at the Ribosome. Molecular Cell *50*, 379–393.

Eichinger, D.J., and Boeke, J.D. (1988). The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. Cell *54*, 955–966.

Elder, R.T., Loh, E.Y., and Davis, R.W. (1983a). RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc Natl Acad Sci U S A *80*, 2432–2436.

Elder, R.T., Loh, E.Y., and Davis, R.W. (1983b). RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc Natl Acad Sci U S A *80*, 2432–2436.

Elion, E.A., and Warner, J.R. (1984). The major promoter element of rRNA transcription in yeast lies 2 kb upstream. Cell *39*, 663–673.

Emiliani, G., Paffetti, D., and Giannini, R. (2009). Identification and molecular characterization of LTR and LINE retrotransposable elements in Fagus sylvatica L. IForest - Biogeosciences and Forestry *2*, 119.

Errede, B. (1993). MCM1 binds to a transcriptional control element in Ty1. Mol. Cell. Biol. *13*, 57–62.

Fan, X., Shi, H., and Lis, J.T. (2005). Distinct transcriptional responses of RNA polymerases I, II and III to aptamers that bind TBP. Nucleic Acids Res. *33*, 838–845.

Felici, F., Cesareni, G., and Hughes, J.M. (1989). The most abundant small cytoplasmic RNA of Saccharomyces cerevisiae has an important function required for normal cell growth. Mol. Cell. Biol. *9*, 3260–3268.

Ferri, M.L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70. Mol. Cell. Biol. *20*, 488–495.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends in Genetics *5*, 103–107.

Finnegan, D.J. (2012). Retrotransposons. Current Biology *22*, R432–R437.

Finnegan, D.J., and Fawcett, D.H. (1986). Transposable elements in Drosophila melanogaster. Oxford Surveys on Eukaryotic Genes *3*, 1–62.

Fleming, A.B., Beggs, S., Church, M., Tsukihashi, Y., and Pennings, S. (2014). The yeast Cyc8-Tup1 complex cooperates with Hda1p and Rpd3p histone deacetylases to robustly repress transcription of the subtelomeric FLO1 gene. Biochim. Biophys. Acta *1839*, 1242–1255.

Formosa, T. (2012). The role of FACT in making and breaking nucleosomes. Biochim. Biophys. Acta *1819*, 247–255.

Formosa, T., Eriksson, P., Wittmeyer, J., Ginn, J., Yu, Y., and Stillman, D.J. (2001). Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. EMBO J. *20*, 3506–3517.

French, A.C., Luscher, B., and Litchfield, D.W. (2007). Development of a stabilized form of the regulatory CK2beta subunit that inhibits cell proliferation. J. Biol. Chem. *282*, 29667–29677.

French, S.L., Osheim, Y.N., Schneider, D.A., Sikes, M.L., Fernandez, C.F., Copela, L.A., Misra, V.A., Nomura, M., Wolin, S.L., and Beyer, A.L. (2008). Visual Analysis of the Yeast 5S rRNA Gene Transcriptome: Regulation and Role of La Protein. Molecular and Cellular Biology *28*, 4576–4587.

Fuertes, G., Villarroya, A., and Knecht, E. (2003). Role of proteasomes in the degradation of short-lived proteins in human fibroblasts under various growth conditions. Int. J. Biochem. Cell Biol. *35*, 651–664.

Fujino, K., Hashida, S.-N., Ogawa, T., Natsume, T., Uchiyama, T., Mikami, T., and Kishima, Y. (2011). Temperature controls nuclear import of Tam3 transposase in Antirrhinum. Plant J. *65*, 146–155.

Fulton, A.M., Rathjen, P.D., Kingsman, S.M., and Kingsman, A.J. (1988). Upstream and downstream transcriptional control signals in the yeast retrotransposon, TY. Nucleic Acids Res *16*, 5439–5458.

Gai, X., and Voytas, D.F. (1998). A Single Amino Acid Change in the Yeast Retrotransposon Ty5 Abolishes Targeting to Silent Chromatin. Molecular Cell *1*, 1051–1055.

Garfinkel, D.J., Boeke, J.D., and Fink, G.R. (1985). Ty element transposition: reverse transcriptase and virus-like particles. Cell *42*, 507–517.

Garfinkel, D.J., Nyswaner, K., Wang, J., and Cho, J.-Y. (2003). Post-transcriptional cosuppression of Ty1 retrotransposition. Genetics *165*, 83–99.

Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A Set of Vectors with a Tetracycline-Regulatable Promoter System for Modulated Gene Expression inSaccharomyces cerevisiae. Yeast *13*, 837–848.

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J., and Brown, P.O. (2001). Genomic Expression Responses to DNA-damaging Agents and the Regulatory Role of the Yeast ATR Homolog Mec1p. Mol Biol Cell *12*, 2987–3003.

Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus11Edited by P. E. Wright. Journal of Molecular Biology *310*, 1–26.

Ghavidel, A., and Schultz, M.C. (2001). TATA binding protein-associated CK2 transduces DNA damage signals to the RNA polymerase III transcriptional machinery. Cell *106*, 575–584.

Giraldo, R., and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J *13*, 2411–2420.

Goldberg, A.L., Akopian, T.N., Kisselev, A.F., Lee, D.H., and Rohrwild, M. (1997). New insights into the mechanisms and importance of the proteasome in intracellular protein degradation. Biol Chem *378*, 131–140.

Goodenbour, J.M., and Pan, T. (2006). Diversity of tRNA genes in eukaryotes. Nucleic Acids Res *34*, 6137–6146.

Goodier, J.L. (2016). Restricting retrotransposons: a review. Mobile DNA *7*, 16.

Goodwin, T.J.D., Butler, M.I., and Poulter, R.T.M. (2003). Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. Microbiology (Reading, Engl.) *149*, 3099–3109.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. U.S.A. *89*, 5547–5551.

Graczyk, D., Debski, J., Muszynska, G., Bretner, M., Lefebvre, O., and Boguta, M. (2011). Casein kinase II-mediated phosphorylation of general repressor Maf1 triggers RNA polymerase III activation. Proceedings of the National Academy of Sciences *108*, 4926–4931.

Grandbastien, M.-A., Audeon, C., Bonnivard, E., Casacuberta, J.M., Chalhoub, B., Costa, A.-P.P., Le, Q.H., Melayah, D., Petit, M., Poncet, C., et al. (2005). Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet. Genome Res. *110*, 229–241.

Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev. *11*, 1640–1650.

Gray, W.M., and Fassler, J.S. (1993). Role of Saccharomyces cerevisiae Rap1 protein in Ty1 and Ty1-mediated transcription. Gene Expr. *3*, 237–251.

Gray, W.M., and Fassler, J.S. (1996). Isolation and analysis of the yeast TEA1 gene, which encodes a zinc cluster Ty enhancer-binding protein. Molecular and Cellular Biology *16*, 347–358.

Gray, W.M., and Fassler, J.S. (2018). Role of Saccharomyces cerevisiae Rap1 protein in Ty1 and Ty1-mediated transcription. Gene Expr *3*, 237–251.

Griffith, J.L., Coleman, L.E., Raymond, A.S., Goodson, S.G., Pittard, W.S., Tsui, C., and Devine, S.E. (2003). Functional Genomics Reveals Relationships Between the Retrovirus-Like Ty1 Element and Its Host Saccharomyces cerevisiae. Genetics *164*, 867–879.

Guio, L., Barrón, M.G., and González, J. (2014). The transposable element Bari-Jheh mediates oxidative stress response in Drosophila. Molecular Ecology *23*, 2020–2030.

Guo, Y., and Levin, H.L. (2010). High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in Schizosaccharomyces pombe. Genome Res. *20*, 239–248.

Hammer, S.E., Strehl, S., and Hagemann, S. (2005). Homologs of Drosophila P transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. Mol. Biol. Evol. *22*, 833–844.

Han, J.S. (2010). Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA *1*, 15.

Happel, A.M., Swanson, M.S., and Winston, F. (1991). The SNF2, SNF5 and SNF6 genes are required for Ty transcription in Saccharomyces cerevisiae. Genetics *128*, 69–77.

Haren, L., Ton-Hoang, B., and Chandler, M. (1999). Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. *53*, 245–281.

Havecker, E.R., Gao, X., and Voytas, D.F. (2004a). The diversity of LTR retrotransposons. Genome Biol *5*, 225.

Havecker, E.R., Gao, X., and Voytas, D.F. (2004b). The diversity of LTR retrotransposons. Genome Biology *5*, 225.

Helbo, A.S., Lay, F.D., Jones, P.A., Liang, G., and Grønbæk, K. (2017). Nucleosome Positioning and NDR Structure at RNA Polymerase III Promoters. Sci Rep *7*.

Hickey, A., Esnault, C., Majumdar, A., Chatterjee, A.G., Iben, J.R., McQueen, P.G., Yang, A.X., Mizuguchi, T., Grewal, S.I.S., and Levin, H.L. (2015). Single-Nucleotide-Specific Targeting of the Tf1 Retrotransposon Promoted by the DNA-Binding Protein Sap1 of Schizosaccharomyces pombe. Genetics *201*, 905–924.

Ho, K.L., Ma, L., Cheung, S., Manhas, S., Fang, N., Wang, K., Young, B., Loewen, C., Mayor, T., and Measday, V. (2015). A role for the budding yeast separase, Esp1, in Ty1 element retrotransposition. PLoS Genet. *11*, e1005109.

Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W.J.H., Sachse, C., and Müller, C.W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. Nature *528*, 231–236.

Huet, J., Conesa, C., Manaud, N., Chaussivert, N., and Sentenac, A. (1994). Interactions between yeast TFIIIB components. Nucleic Acids Res *22*, 3433–3439.

Irwin, B., Aye, M., Baldi, P., Beliakova-Bethell, N., Cheng, H., Dou, Y., Liou, W., and Sandmeyer, S. (2005). Retroviruses and yeast retrotransposons use overlapping sets of host genes. Genome Res *15*, 641–654.

Jasiak, A.J., Armache, K.-J., Martens, B., Jansen, R.-P., and Cramer, P. (2006). Structural Biology of RNA Polymerase III: Subcomplex C17/25 X-Ray Structure and 11 Subunit Enzyme Model. Molecular Cell *23*, 71–81.

Ji, H., Moore, D.P., Blomberg, M.A., Braiterman, L.T., Voytas, D.F., Natsoulis, G., and Boeke, J.D. (1993). Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell *73*, 1007–1018.

Jiang, S., Liu, Y., Xu, C., Wang, Y., Gong, J., Shen, Y., Wu, Q., Boeke, J.D., and Dai, J. (2017). Dissecting Nucleosome Function with a Comprehensive Histone H2A and H2B Mutant Library. G3 (Bethesda) *7*, 3857–3866.

Jiang, Y.W. (2002). Transcriptional cosuppression of yeast Ty1 retrotransposons. Genes Dev. *16*, 467–478.

Johnston, I.M., Allison, S.J., Morton, J.P., Schramm, L., Scott, P.H., and White, R.J. (2002). CK2 forms a stable complex with TFIIIB and activates RNA polymerase III transcription in human cells. Mol. Cell. Biol. *22*, 3757–3768.

Jordan, I.K., and McDonald, J.F. (1998). Evidence for the role of recombination in the regulatory evolution of Saccharomyces cerevisiae Ty elements. J. Mol. Evol. *47*, 14–20.

Jordan, I.K., and McDonald, J.F. (1999). Tempo and mode of Ty element evolution in Saccharomyces cerevisiae. Genetics *151*, 1341–1351.

Jourdain, S., Acker, J., Ducrot, C., Sentenac, A., and Lefebvre, O. (2003). The τ95 Subunit of Yeast TFIIIC Influences Upstream and Downstream Functions of TFIIIC·DNA Complexes. J. Biol. Chem. *278*, 10450–10457.

Jurka, J., and Kapitonov, V.V. (2001). PIFs meet Tourists and Harbingers: A superfamily reunion. Proc Natl Acad Sci U S A *98*, 12315–12316.

Kaiser, P., Meierhofer, D., Wang, X., and Huang, L. (2008). Tandem Affinity Purification Combined with Mass Spectrometry to Identify Components of Protein Complexes. Methods Mol Biol *439*, 309–326.

Kapitonov, V.V., and Jurka, J. (2003). Molecular paleontology of transposable elements in the Drosophila melanogaster genome. PNAS *100*, 6569–6574.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. *3*, e181.

Kapitonov, V.V., and Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. Nature Reviews Genetics *9*, 411–412.

Kassavetis, G.A., Blanco, J.A., Johnson, T.E., and Geiduschek, E.P. (1992). Formation of open and elongating transcription complexes by RNA polymerase III. Journal of Molecular Biology *226*, 47–58.

Kassavetis, G.A., Kumar, A., Ramirez, E., and Geiduschek, E.P. (1998). Functional and Structural Organization of Brf, the TFIIB-Related Component of the RNA Polymerase III Transcription Initiation Complex. Mol Cell Biol *18*, 5587–5599.

Kemble, D.J., McCullough, L.L., Whitby, F.G., Formosa, T., and Hill, C.P. (2015). FACT disrupts nucleosome structure by binding H2A-H2B with conserved peptide motifs. Mol Cell *60*, 294–306.

Kempken, F., and Windhofer, F. (2001). The hAT family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma *110*, 1–9.

Kenna, M.A., Brachmann, C.B., Devine, S.E., and Boeke, J.D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. Mol. Cell. Biol. *18*, 1115–1124.

Kent, N.A., Karabetsou, N., Politis, P.K., and Mellor, J. (2001). In vivo chromatin remodeling by yeast ISWI homologs Isw1p and Isw2p. Genes Dev. *15*, 619–626.

Kidwell, M.G. (2002). Transposable elements and the evolution of genome size in eukaryotes. Genetica *115*, 49–63.

Kim, H.S., Hoja, U., Stolz, J., Sauer, G., and Schweizer, E. (2004). Identification of the tRNA-binding protein Arc1p as a novel target of in vivo biotinylation in Saccharomyces cerevisiae. J. Biol. Chem. *279*, 42445–42452.

Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., and Voytas, D.F. (1998). Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. Genome Res. *8*, 464–478.

Kirchner, J., Connolly, C.M., and Sandmeyer, S.B. (1995). Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science *267*, 1488–1491.

Klein, S., Meng, R., Montenarh, M., and Götz, C. (2016). The Phosphorylation of PDX-1 by Protein Kinase CK2 Is Crucial for Its Stability. Pharmaceuticals (Basel) *10*.

Kou, H., Irvin, J.D., Huisinga, K.L., Mitra, M., and Pugh, B.F. (2003). Structural and functional analysis of mutations along the crystallographic dimer interface of the yeast TATA binding protein. Mol. Cell. Biol. *23*, 3186–3201.

Kubiński, K., Domańska, K., Sajnaga, E., Mazur, E., Zieliński, R., and Szyszka, R. (2007). Yeast holoenzyme of protein kinase CK2 requires both beta and beta' regulatory subunits for its activity. Mol. Cell. Biochem. *295*, 229–236.

Laloux, I., Dubois, E., Dewerchin, M., and Jacobs, E. (1990). TEC1, a gene involved in the activation of Ty1 and Ty1-mediated gene expression in Saccharomyces cerevisiae: cloning and molecular analysis. Mol. Cell. Biol. *10*, 3541–3550.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. The EMBO Journal *25*, 118–128.

Lassar, A.B., Martin, P.L., and Roeder, R.G. (1983). Transcription of class III genes: formation of preinitiation complexes. Science *222*, 740–748.

Lawler, J.F., Haeusser, D.P., Dull, A., Boeke, J.D., and Keeney, J.B. (2002a). Ty1 Defect in Proteolysis at High Temperature. Journal of Virology *76*, 4233–4240.

Lawler, J.F., Haeusser, D.P., Dull, A., Boeke, J.D., and Keeney, J.B. (2002b). Ty1 Defect in Proteolysis at High Temperature. Journal of Virology *76*, 4233–4240.

Lee, D.H., and Goldberg, A.L. (1996). Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in Saccharomyces cerevisiae. J. Biol. Chem. *271*, 27280–27284.

Lee, J.Y., Rohlman, C.E., Molony, L.A., and Engelke, D.R. (1991). Characterization of RPR1, an essential gene encoding the RNA component of Saccharomyces cerevisiae nuclear RNase P. Mol Cell Biol *11*, 721–730.

Lee, M., and Struhl, K. (2001). Multiple functions of the nonconserved N-terminal domain of yeast TATA-binding protein. Genetics *158*, 87–93.

Lefebvre, O., Rüth, J., and Sentenac, A. (1994). A mutation in the largest subunit of yeast TFIIIC affects tRNA and 5 S RNA synthesis. Identification of two classes of suppressors. J. Biol. Chem. *269*, 23374–23381.

Li, M., Jurado, K.A., Lin, S., Engelman, A., and Craigie, R. (2014). Engineered Hyperactive Integrase for Concerted HIV-1 DNA Integration. PLOS ONE *9*, e105078.

Litchfield, D.W. (2003). Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. Biochem. J. *369*, 1–15.

Litchfield, D.W., Lozeman, F.J., Cicirelli, M.F., Harrylock, M., Ericsson, L.H., Piening, C.J., and Krebs, E.G. (1991). Phosphorylation of the beta subunit of casein kinase II in human A431 cells. Identification of the autophosphorylation site and a site phosphorylated by p34cdc2. J. Biol. Chem. *266*, 20380–20389.

Liu, F., and Altman, S. (2009). Ribonuclease P (Springer Science & Business Media).

Liu, Z., Oh, S.-M., Okada, M., Liu, X., Cheng, D., Peng, J., Brat, D.J., Sun, S., Zhou, W., Gu, W., et al. (2009). Human BRE1 is an E3 ubiquitin ligase for Ebp1 tumor suppressor. Mol. Biol. Cell *20*, 757–768.

López, M.C., and Baker, H.V. (2000). Understanding the Growth Phenotype of the Yeastgcr1 Mutant in Terms of Global Genomic Expression Patterns. Journal of Bacteriology *182*, 4970–4978.

Maertens, G.N., Cherepanov, P., and Engelman, A. (2006). Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. Journal of Cell Science *119*, 2563–2571.

Malagon, F., and Jensen, T.H. (2008). The T body, a new cytoplasmic RNA granule in Saccharomyces cerevisiae. Mol. Cell. Biol. *28*, 6022–6032.

Malik, H.S., Burke, W.D., and Eickbush, T.H. (1999). The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. *16*, 793–805.

Mateo, L., Ullastres, A., and González, J. (2014). A Transposable Element Insertion Confers Xenobiotic Resistance in Drosophila. PLOS Genetics *10*, e1004560.

Matsuda, E., and Garfinkel, D.J. (2009). Posttranslational interference of Ty1 retrotransposition by antisense RNAs. PNAS *106*, 15657–15662.

Matysiak, J., Lesbats, P., Mauro, E., Lapaillerie, D., Dupuy, J.-W., Lopez, A.P., Benleulmi, M.S., Calmels, C., Andreola, M.-L., Ruff, M., et al. (2017). Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. Retrovirology *14*, 39.

Mayor, T., Graumann, J., Bryan, J., MacCoss, M.J., and Deshaies, R.J. (2007). Quantitative Profiling of Ubiquitylated Proteins Reveals Proteasome Substrates and the Substrate Repertoire Influenced by the Rpn10 Receptor Pathway. Molecular & Cellular Proteomics *6*, 1885–1895.

McClure, M.A. (1991). Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol. Biol. Evol. *8*, 835–856.

McLane, L.M., Pulliam, K.F., Devine, S.E., and Corbett, A.H. (2008). The Ty1 integrase protein can exploit the classical nuclear protein import machinery for entry into the nucleus. Nucleic Acids Res *36*, 4317–4326.

Meggio, F., Lorenzo, and Pinna, A. (2003). One-thousand-and-one substrates of protein kinase CK2. Faseb J 349–368.

Merkulov, G.V., Lawler, J.F., Eby, Y., and Boeke, J.D. (2001). Ty1 Proteolytic Cleavage Sites Are Required for Transposition: All Sites Are Not Created Equal. J Virol *75*, 638–644.

Mohammed, S., Lorenzen, K., Kerkhoven, R., Breukelen, B. van, Vannini, A., Cramer, P., and Heck, A.J.R. (2008). Multiplexed Proteomics Mapping of Yeast RNA Polymerase II and III Allows Near-Complete Sequence Coverage and Reveals Several Novel Phosphorylation Sites. Anal. Chem. *80*, 3584–3592.

Moir, R.D., Puglia, K.V., and Willis, I.M. (2002). A Gain-of-Function Mutation in the Second Tetratricopeptide Repeat of TFIIIC131 Relieves Autoinhibition of Brf1 Binding. Molecular and Cellular Biology *22*, 6131–6141.

Moore, S.P., and Garfinkel, D.J. (2009). Functional Analysis of N-Terminal Residues of Ty1 Integrase. Journal of Virology *83*, 9502–9511.

Morillon, A., Bénard, L., Springer, M., and Lesage, P. (2002). Differential Effects of Chromatin and Gcn4 on the 50-Fold Range of Expression among Individual Yeast Ty1 Retrotransposons. Molecular and Cellular Biology *22*, 2078–2088.

Morillon, A., Springer, M., and Lesage, P. (2000). Activation of the Kss1 Invasive-Filamentous Growth Pathway Induces Ty1 Transcription and Retrotransposition inSaccharomyces cerevisiae. Molecular and Cellular Biology *20*, 5766–5776.

Mou, Z., Kenny, A.E., and Curcio, M.J. (2006). Hos2 and Set3 promote integration of Ty1 retrotransposons at tRNA genes in Saccharomyces cerevisiae. Genetics *172*, 2157–2167.

Mularoni, L., Zhou, Y., Bowen, T., Gangadharan, S., Wheelan, S.J., and Boeke, J.D. (2012). Retrotransposon Ty1 integration targets specifically positioned asymmetric nucleosomal DNA segments in tRNA hotspots. Genome Res. *22*, 693–703.

Mules, E.H., Uzun, O., and Gabriel, A. (1998). In Vivo Ty1 Reverse Transcription Can Generate Replication Intermediates with Untidy Ends. J Virol *72*, 6490–6503.

Müller, F., Brühl, K.-H., Freidel, K., Kowallik, K.V., and Ciriacy, M. (1987). Processing of TY1 proteins and formation of Ty1 virus-like particles in Saccharomyces cerevisiae. Mol Gen Genet *207*, 421–429.

Nguyen, N.-T.-T., Saguez, C., Conesa, C., Lefebvre, O., and Acker, J. (2015). Identification of proteins associated with RNA polymerase III using a modified tandem chromatin affinity purification. Gene *556*, 51–60.

Nishida, Y., Pachulska-Wieczorek, K., Błaszczyk, L., Saha, A., Gumna, J., Garfinkel, D.J., and Purzycka, K.J. (2015). Ty1 retrovirus-like element Gag contains overlapping restriction factor and nucleic acid chaperone functions. Nucleic Acids Res *43*, 7414–7431.

Nyswaner, K.M., Checkley, M.A., Yi, M., Stephens, R.M., and Garfinkel, D.J. (2008). Chromatin-Associated Genes Protect the Yeast Genome From Ty1 Insertional Mutagenesis. Genetics *178*, 197–214.

O'Donnell, J.P., Gehman, M., and Keeney, J.B. (2010). Regulators of ribonucleotide reductase inhibit Ty1 mobility in saccharomyces cerevisiae. Mob DNA *1*, 23.

Oficjalska-Pham, D., Harismendy, O., Smagowicz, W.J., Gonzalez de Peredo, A., Boguta, M., Sentenac, A., and Lefebvre, O. (2006). General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. Mol. Cell *22*, 623–632.

Olsten, M.E.K., and Litchfield, D.W. (2004). Order or chaos? An evaluation of the regulation of protein kinase CK2. Biochemistry and Cell Biology *82*, 681–693.

Pachulska-Wieczorek, K., Błaszczyk, L., Gumna, J., Nishida, Y., Saha, A., Biesiada, M., Garfinkel, D.J., and Purzycka, K.J. (2016). Characterizing the functions of Ty1 Gag and the Gag-derived restriction factor p22/p18. Mob Genet Elements *6*.

Pandey, M., Patel, S., and Gabriel, A. (2004). Insights into the Role of an Active Site Aspartate in Ty1 Reverse Transcriptase Polymerization. J. Biol. Chem. *279*, 47840–47848.

Pandey, M., Patel, S.S., and Gabriel, A. (2008). Kinetic Pathway of Pyrophosphorolysis by a Retrotransposon Reverse Transcriptase. PLOS ONE *3*, e1389.

Paquin, C.E., and Williamson, V.M. (1984). Temperature effects on the rate of ty transposition. Science *226*, 53–55.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J *27*, 100–110.

Peterson-Burch, B.D., and Voytas, D.F. (2002). Genes of the Pseudoviridae (Ty1/copia retrotransposons). Mol. Biol. Evol. *19*, 1832–1845.

Piégu, B., Bire, S., Arensburger, P., and Bigot, Y. (2015). A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. Molecular Phylogenetics and Evolution *86*, 90–109.

Pollard, K.J., and Peterson, C.L. (1997). Role for ADA/GCN5 products in antagonizing chromatin-mediated transcriptional repression. Mol. Cell. Biol. *17*, 6212–6222.

Poulter, R.T.M., and Goodwin, T.J.D. (2005). DIRS-1 and the other tyrosine recombinase retrotransposons. Cytogenet. Genome Res. *110*, 575–588.

Pritham, E.J., Feschotte, C., and Wessler, S.R. (2005). Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. Mol. Biol. Evol. *22*, 1751–1763.

Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., et al. (2005). Global analysis of protein phosphorylation in yeast. Nature *438*, 679–684.

Qi, X., and Sandmeyer, S. (2012). In vitro targeting of strand transfer by the Ty3 retroelement integrase. J. Biol. Chem. *287*, 18589–18595.

Radford, S.J., Boyle, M.L., Sheely, C.J., Graham, J., Haeusser, D.P., Zimmerman, L., and Keeney, J.B. (2004). Increase in Ty1 cDNA recombination in yeast sir4 mutant strains at high temperature. Genetics *168*, 89–101.

Rai, S.K., Sangesland, M., Lee, M., Esnault, C., Cui, Y., Chatterjee, A.G., and Levin, H.L. (2017). Host factors that promote retrotransposon integration are similar in distantly related eukaryotes. PLoS Genet *13*.

Risler, J.K., Kenny, A.E., Palumbo, R.J., Gamache, E.R., and Curcio, M.J. (2012). Host co-factors of the retrovirus-like transposon Ty1. Mob DNA *3*, 12.

Rolfe, M., Spanos, A., and Banks, G. (1986). Induction of yeast Ty element transcription by ultraviolet light. Nature *319*, 339–340.

Rolling-circle transposons in eukaryotes | PNAS.

Rusche, L.N., Kirchmaier, A.L., and Rine, J. (2003). The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae. Annu. Rev. Biochem. *72*, 481–516.

Sacerdot, C., Mercier, G., Todeschini, A.-L., Dutreix, M., Springer, M., and Lesage, P. (2005). Impact of ionizing radiation on the life cycle of Saccharomyces cerevisiae Ty1 retrotransposon. Yeast *22*, 441–455.

Saha, A., Mitchell, J.A., Nishida, Y., Hildreth, J.E., Ariberre, J.A., Gilbert, W.V., and Garfinkel, D.J. (2015). A trans-Dominant Form of Gag Restricts Ty1 Retrotransposition and Mediates Copy Number Control. Journal of Virology *89*, 3922–3938.

Salinero, A.C., Knoll, E.R., Zhu, Z.I., Landsman, D., Curcio, M.J., and Morse, R.H. (2018). The Mediator co-activator complex regulates Ty1 retromobility by controlling the balance between Ty1i and Ty1 promoters. PLOS Genetics *14*, e1007232.

Sanchez-Casalongue, M.E., Lee, J., Diamond, A., Shuldiner, S., Moir, R.D., and Willis, I.M. (2015). Differential phosphorylation of a regulatory subunit of protein kinase CK2 bytarget of rapamycin complex 1 signaling and the Cdc-like kinase Kns1. J. Biol. Chem. *290*, 7221–7233.

Sandmeyer, S., Patterson, K., and Bilanchone, V. (2015). Ty3, a Position-specific Retrotransposon in Budding Yeast. Microbiol Spectr *3*, MDNA3-0057–2014.

Sandmeyer, S.B., Hansen, L.J., and Chalker, D.L. (1990). Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. *24*, 491–518.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., and Bennetzen, J.L. (1998). The paleontology of intergene retrotransposons of maize. Nat. Genet. *20*, 43–45.

Sanz-Ramos, M., and Stoye, J.P. (2013). Capsid-binding retrovirus restriction factors: discovery, restriction specificity and implications for the development of novel therapeutics. J. Gen. Virol. *94*, 2587–2598.

Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M., and Collins, F.H. (2003). Molecular evolutionary analysis of the widespread piggyBac transposon family and related “domesticated” sequences. Mol. Genet. Genomics *270*, 173–180.

Sawby, R., and Wichman, H.A. (1997). Analysis of orthologous retrovirus-like elements in the white-footed mouse, Peromyscus leucopus. J. Mol. Evol. *44*, 74–80.

Schmidt, M.C., Kao, C.C., Pei, R., and Berk, A.J. (1989). Yeast TATA-box transcription factor gene. Proc. Natl. Acad. Sci. U.S.A. *86*, 7785–7789.

Scholes, D.T., Banerjee, M., Bowen, B., and Curcio, M.J. (2001). Multiple regulators of Ty1 transposition in Saccharomyces cerevisiae have conserved roles in genome maintenance. Genetics *159*, 1449–1465.

Schultz, P., Marzouki, N., Marck, C., Ruet, A., Oudet, P., and Sentenac, A. (1989). The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. EMBO J. *8*, 3815–3824.

Servant, G., Pinson, B., Tchalikian-Cosson, A., Coulpier, F., Lemoine, S., Pennetier, C., Bridier-Nahmias, A., Todeschini, A.L., Fayol, H., Daignan-Fornier, B., et al. (2012). Tye7 regulates yeast Ty1 retrotransposon sense and antisense transcription in response to adenylic nucleotides stress. Nucleic Acids Res *40*, 5271–5282.

Shao, H., and Tu, Z. (2001). Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics *159*, 1103–1115.

Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessl, J.J., Shkriabai, N., Coward, E., Aiyer, S.S., et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. Proc. Natl. Acad. Sci. U.S.A. *110*, 12036–12041.

Sharon, G., Burkett, T.J., and Garfinkel, D.J. (1994). Efficient homologous recombination of Ty1 element cDNA when integration is blocked. Mol. Cell. Biol. *14*, 6540–6551.

Shore, D., and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell *51*, 721–732.

Singer, M.F. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell *28*, 433–434.

Singh, P.K., Plumb, M.R., Ferris, A.L., Iben, J.R., Wu, X., Fadel, H.J., Luke, B.T., Esnault, C., Poeschla, E.M., Hughes, S.H., et al. (2015). LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. Genes Dev. *29*, 2287–2297.

Smith, J.J., Miller, L.R., Kreisberg, R., Vazquez, L., Wan, Y., and Aitchison, J.D. (2011). Environment-responsive transcription factors bind subtelomeric elements and regulate gene silencing. Molecular Systems Biology *7*, 455.

Staleva Staleva, L., and Venkov, P. (2001). Activation of Ty transposition by mutagens. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *474*, 93–103.

Sun, C., Shepard, D.B., Chong, R.A., López Arriaza, J., Hall, K., Castoe, T.A., Feschotte, C., Pollock, D.D., and Mueller, R.L. (2012). LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol Evol *4*, 168–183.

Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., and Villén, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nature Methods *10*, 676–682.

Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L., and Kaiser, P. (2006). A Tandem Affinity Tag for Two-step Purification under Fully Denaturing Conditions: Application in Ubiquitin Profiling and Protein Complex Identification Combined with in vivoCross-Linking. Molecular & Cellular Proteomics *5*, 737–748.

Tate, J.J., Persinger, J., and Bartholomew, B. (1998). Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. Nucleic Acids Res. *26*, 1421–1426.

Tavenet, A., Suleau, A., Dubreuil, G., Ferrari, R., Ducrot, C., Michaut, M., Aude, J.-C., Dieci, G., Lefebvre, O., Conesa, C., et al. (2009). Genome-wide location analysis reveals a role for Sub1 in RNA polymerase III transcription. PNAS *106*, 14265–14270.

Teichmann, M., Dieci, G., Pascali, C., and Boldina, G. (2010). General transcription factors and subunits of RNA polymerase III. Transcription *1*, 130–135.

Tekeste, S.S., Wilkinson, T.A., Weiner, E.M., Xu, X., Miller, J.T., Le Grice, S.F.J., Clubb, R.T., and Chow, S.A. (2015). Interaction between Reverse Transcriptase and Integrase Is Required for Reverse Transcription during HIV-1 Replication. J. Virol. *89*, 12058–12069.

Teng, L., Firpi, H.A., and Tan, K. (2011). Enhancers in embryonic stem cells are enriched for transposable elements and genetic variations associated with cancers. Nucleic Acids Res. *39*, 7371–7379.

Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. (1995). A mutation in the C31 subunit of Saccharomyces cerevisiae RNA polymerase III affects transcription initiation. The EMBO Journal *14*, 351–359.

Todeschini, A.-L., Morillon, A., Springer, M., and Lesage, P. (2005). Severe Adenine Starvation Activates Ty1 Transcription and Retrotransposition in Saccharomyces cerevisiae. Molecular and Cellular Biology *25*, 7459–7472.

Türkel, S., Liao, X.B., and Farabaugh, P.J. (1997). GCR1-dependent transcriptional activation of yeast retrotransposon Ty2-917. Yeast *13*, 917–930.

Uzun, O., and Gabriel, A. (2001). A Ty1 Reverse Transcriptase Active-Site Aspartate Mutation Blocks Transposition but Not Polymerization. J Virol *75*, 6337–6347.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell *45*, 439–446.

Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassavetis, G.A., and Cramer, P. (2010). Molecular basis of RNA polymerase III transcription repression by Maf1. Cell *143*, 59–70.

Voit, R., Kuhn, A., Sander, E.E., and Grummt, I. (1995). Activation of mammalian ribosomal gene transcription requires phosphorylation of the nucleolar transcription factor UBF. Nucleic Acids Res. *23*, 2593–2599.

Voytas, D.F., and Boeke, J.D. (1992). Yeast retrotransposon revealed. Nature *358*, 717.

Wahlin, J., and Cohn, M. (2000). Saccharomyces cerevisiae RAP1 binds to telomeric sequences with spatial flexibility. Nucleic Acids Res *28*, 2292–2301.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. *32*, 1197–1207.

Werner, M., Denmat, S.H.-L., Treich, I., Sentenac, A., and Thuriaux, P. (1992). Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. Molecular and Cellular Biology *12*, 1087–1095.

White, R.J. (2013). RNA Polymerase III Transcription (Springer Science & Business Media).

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics *8*, 973–982.

Wilhelm, M., and Wilhelm, F.-X. (2006). Cooperation between Reverse Transcriptase and Integrase during Reverse Transcription and Formation of the Preintegrative Complex of Ty1. Eukaryot Cell *5*, 1760–1769.

Wilhelm, M., Boutabout, M., and Wilhelm, F.X. (2000). Expression of an active form of recombinant Ty1 reverse transcriptase in Escherichia coli: a fusion protein containing the C-terminal region of the Ty1 integrase linked to the reverse transcriptase-RNase H domain exhibits polymerase and RNase H activities. Biochem. J. *348 Pt 2*, 337–342.

Williamson, V.M. (1983). Transposable Elements in Yeast. In International Review of Cytology, G.H. Bourne, J.F. Danielli, and K.W. Jeon, eds. (Academic Press), pp. 1–25.

Willis, I.M. (2002). A universal nomenclature for subunits of the RNA polymerase III transcription initiation factor TFIIIB. Genes Dev. *16*, 1337–1338.

Wilson, L.K., Dhillon, N., Thorner, J., and Martin, G.S. (1997). Casein Kinase II Catalyzes Tyrosine Phosphorylation of the Yeast Nucleolar Immunophilin Fpr3. J. Biol. Chem. *272*, 12961–12967.

Winans, S., Larue, R.C., Abraham, C.M., Shkriabai, N., Skopp, A., Winkler, D., Kvaratskhelia, M., and Beemon, K.L. (2017). The FACT Complex Promotes Avian Leukosis Virus DNA Integration. J. Virol. *91*.

Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics *107*, 179–197.

Woychik, N.A., and Young, R.A. (1992). Genes encoding transcription factor IIIA and the RNA polymerase common subunit RPB6 are divergently transcribed in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *89*, 3999–4003.

Xie, W., Gai, X., Zhu, Y., Zappulla, D.C., Sternglanz, R., and Voytas, D.F. (2001). Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Mol. Cell. Biol. *21*, 6606–6614.

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J *9*, 3353–3362.

Yieh, L., Kassavetis, G., Geiduschek, E.P., and Sandmeyer, S.B. (2000). The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. J. Biol. Chem. *275*, 29800–29807.

Youngren, S.D., Boeke, J.D., Sanders, N.J., and Garfinkel, D.J. (1988). Functional organization of the retrotransposon Ty from Saccharomyces cerevisiae: Ty protease is required for transposition. Mol Cell Biol *8*, 1421–1431.

Yu, K., and Elder, R.T. (1989). Some of the signals for 3'-end formation in transcription of the Saccharomyces cerevisiae Ty-D15 element are immediately downstream of the initiation site., Some of the signals for 3'-end formation in transcription of the Saccharomyces cerevisiae Ty-D15 element are immediately downstream of the initiation site. Mol Cell Biol *9*, *9*, 2431, 2431–2444.

Zhang, X.H.-F., and Chasin, L.A. (2006). Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc. Natl. Acad. Sci. U.S.A. *103*, 13427–13432.

Zhu, Y., Dai, J., Fuerst, P.G., and Voytas, D.F. (2003). Controlling integration specificity of a yeast retrotransposon. PNAS *100*, 5891–5895.

Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., Collura, K., Kudrna, D., and Wing, R.A. (2007). Transposable element distribution, abundance and role in genome size variation in the genus Oryza. BMC Evol. Biol. *7*, 152.

Titre : Identification exhaustive de partenaires de l'intégrase du rétroélément Ty1 chez la levure *Saccharomyces cerevisiae*. Caractérisation du rôle de la caséine kinase II dans la rétrotransposition de Ty1 *in vivo*

Mots clés : Ty1, Intégrase, TChAP, CK2, Rétroélément

Les rétrotransposons LTR sont des éléments transposables très répandus chez les eucaryotes. Comme les rétrovirus, ils se répliquent par transcription inverse de leur ARN en ADNc, qui est intégré dans le génome hôte par leur propre intégrase (IN). Des études de séquençage à haut débit ont clairement établi que l'intégration ne se fait pas de façon aléatoire dans l'ensemble du génome de la cellule hôte. Des connaissances approfondies sur la biologie rétrovirale ont été acquises grâce à leur étude sur la levure utilisant le Ty1 LTR-retrotransposon comme modèle de travail. Le rétrotransposon Ty1 de la levure Saccharomyces cerevisiae intègre en amont des gènes de classe III, les gènes transcrits par l'ARN polymérase III (Pol III). Des données récentes ont révélé l'importance de l'AC40, une sous-unité de Pol III dans ce ciblage. Une interaction entre le Ty1 IN et l'AC40 est nécessaire pour le choix du site d'intégration des gènes Pol III. Néanmoins, le mécanisme moléculaire reste largement inconnu. Afin d'obtenir une vision globale de l'ensemble du phénomène qui se produit sur le site d'intégration, nous aimerions déterminer de manière exhaustive les protéines qui interagissent avec Ty1 IN et analyser leur rôle dans l'intégration de Ty1 et la transcription de l'ARN Pol III. Pour atteindre cet objectif, nous avons développé des approches protéomiques pour identifier de nouveaux partenaires cellulaires Ty1 intégraux. Nous avons identifié plusieurs nouveaux partenaires Ty1 IN qui semblent intéressants et leur rôle moléculaire dans la rétrotransposition de Ty1 sera étudié. Cependant, dans le cadre de mon doctorat, j'ai particulièrement travaillé à déchiffrer le rôle moléculaire de la protéine caséine kinase II dans la rétrotransposition de Ty1.

Title: Exhaustive Identification of the retroelement Ty1 Integrase partners in yeast *Saccharomyces cerevisiae*. Characterization of the role of Casein kinase II in Ty1 retrotransposition *in vivo*

Keywords: Ty1, Integrase, TChAP, CK2, Retroelement

LTR-retrotransposons are widespread transposable elements in eukaryotes. Like retroviruses, they replicate by reverse transcription of their RNA into cDNA, which is integrated into the host genome by their own integrase (IN). High-throughput sequencing studies clearly established that integration does not occur randomly throughout the host-cell genome. Deep insights on retroviral biology have been gained by their study in yeast using the Ty1 LTR-retrotransposon as a working model. The Ty1 retrotransposon of the yeast Saccharomyces cerevisiae integrates upstream of class III genes, the genes transcribed by RNA polymerase III (Pol III). Recent data revealed the importance of AC40, a Pol III subunit in this targeting. An interaction between the Ty1 IN and AC40 is necessary for integration site choice at the Pol III genes. Nevertheless, the molecular mechanism remains largely unknown. To obtain a global view of the entire phenomenon that occurs on the integration site we would like to exhaustively determine the proteins that interact with Ty1 IN and analyze their role in both Ty1 integration and RNA Pol III transcription. To achieve this goal, we have developed proteomic approaches to identify new Ty1 integrase cellular partners. We have identified several novel Ty1 IN partners that seem interesting and their molecular role in Ty1 retrotransposition will be studied. However, in the tenure of my PhD, I have particularly worked to decipher the molecular role of the casein kinase II protein in Ty1 retrotransposition.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France