Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Emilie Benrabah

le vendredi 19 septembre 2014

Titre :

Mécanismes moléculaires de la régulation du facteur de transcription Shavenbaby par les peptides Pri, chez la drosophile.

École doctorale et discipline ou spécialité :

ED BSB : Gènes, cellules et développement

Unité de recherche :

Centre de Biologie du Développement (UMR5547/CNRS)

Directeur(s) de Thèse :

François Payre
Serge Plaza

Jury :

Pr.Laurent Paquereau, Président
Dr. Olivier Coux, Rapporteur
Dr. François Schweisguth, Rapporteur
Dr. Pascal Thérond, Rapporteur
Dr. Marie-Odile Fauvarque, Rapporteur
Dr. François Payre, Directeur de thèse
Dr. Serge Plaza, Directeur de thèse

RESULTATS

Partie 1 : Les peptides Pri : un interrupteur de l'activité transcriptionnelle de Svb

Partie 2 : Les peptides Pri sont requis pour l'interaction entre Svb et Ubr3, dirigeant ainsi son ubiquitination et sa dégradation partielle par le protéasome

Partie 3 : Discussion et résultats additionnels

CONCLUSION

Partie 1 : Les sPEPs, des molécules essentielles à la vie

Partie 2 : L'ère-peptidique, le nouveau chapitre dans la quête de la compréhension de la vie

REFERENCES

Préambule

L'analyse de l'expression des génomes depuis la fin des années 1990 nous a fait réaliser l'immense part d'ignorance qu'il reste encore à dissiper. En effet, nous ne comprenons le rôle que d'une mineure partie de ceux-ci.

Lors de ma thèse, j'ai étudié le rôle d'une molécule un peu particulière. Il s'agit d'un peptide, d'une très petite protéine (11 résidus) exprimée en tant que telle à partir d'un cadre ouvert de lecture de 33 nucléotides. Ce peptide est essentiel au développement de la drosophile, un organisme modèle très utilisé en génétique. Sans lui, la drosophile meurt en fin d'embryogenèse. Il s'avère que ce peptide pourrait être l'un des représentants d'une nouvelle classe de molécules, qui, si on l'identifie, contribuera à réduire cette part d'ignorance.

J'ai donc choisi de présenter en première partie de mon introduction l'histoire de cette famille de petites protéines, sous forme d'une rétrospective de ces 20 dernières années. J'y exposerai les points clefs (dogmes, dates et techniques) permettant de comprendre comment et pourquoi elle a d'abord été ignorée, puis suscité l'intérêt et donc le déploiement de nombreux efforts dans le monde de la recherche.

Au travers de mes travaux de thèse, que j'introduirai en seconde partie, j'espère avoir contribué (si besoin est) à démontrer la nécessité de poursuivre sur cet élan.

INTRODUCTION

Partie 1 : A la conquête des peptides

La fin des années 1990 a marqué l'entrée dans l'ère post-génomique. Le grand projet de séquençage des génomes a rendu disponible les séquences quasi complètes des génomes de nombreux organismes, référencés sur la Genomes OnLine Database GOLD (http://genomesonline.org), permettant ainsi leur analyse informatique : la génomique. Le nombre de génomes séquencés est en expansion permanente, mais cette avancée n'apparaît que comme une étape dans la compréhension des phénomènes biologiques.

Nous allons voir dans cette première partie d'introduction comment et pourquoi les premiers travaux d'annotation des génomes ont conduit à occulter une catégorie complète du protéome : les petits peptides encodés par les petits cadres ouverts de lecture [Short Peptides (sPEPs) encoded by Small Open Reading Frames (smORFs)]. Je me propose ici de scinder l'ère post-génomique en deux. Je qualifierai cette première phase où les sPEPs ont été ignorés, jusqu'en 2005, « **d'ère pré-peptidique** ». Nous verrons ensuite comment cette période a été incubatrice de techniques et de stratégies progressivement mises en place afin d'identifier de manière systématique smORFs et sPEPs et d'entrer dans « **l'ère peptidique** ».

Cette rétrospective de l'ère post-génomique considérée du point de vue de l'émergence de cette nouvelle classe de molécules a pour objectif de montrer comment mon projet de thèse s'est bien ancré dans – et a participé à – cette dynamique.

I. Définitions générales : smORF et sPEP

Un cadre ouvert de lecture (ORF) est la séquence d'un ARN messager (ARNm) (correspondant à un gène) à partir de laquelle les ribosomes vont traduire une protéine. Il débute par un codon initiateur de la transcription, presque exclusivement un AUG (codant une Méthionine), et se termine par un codon-stop (TAA, TAG ou TGA). Entre eux, le nombre de codons correspondra au nombre d'acides-aminés qu'aura la protéine (Fig. 1). Un smORF ne se différencie d'un ORF que par sa taille : il est par définition, et nous allons voir pourquoi, d'une taille inférieure à 100 codons. Un sPEP est une protéine exprimée à partir d'un smORF et, par conséquence, se différencie aussi des protéines par sa taille, inférieure à 100 résidus.

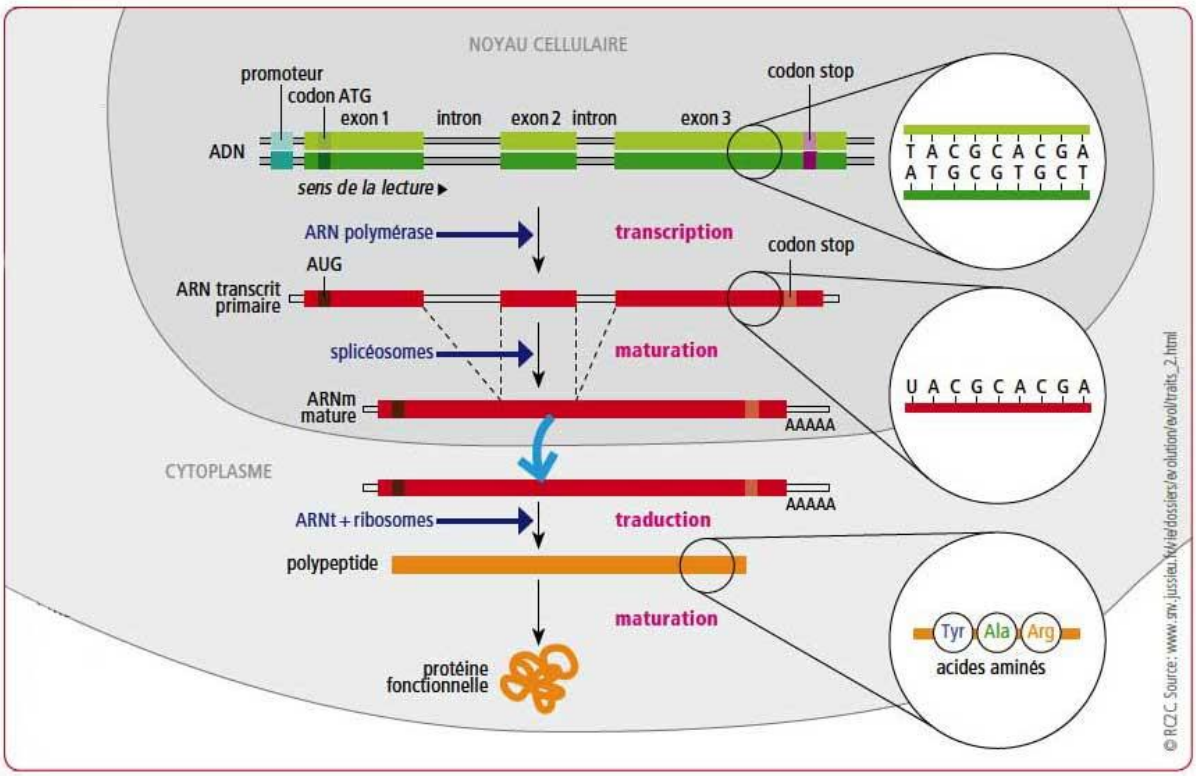


Figure 1 : Modèle simplifié de la structure d'un gène eucaryote et de son expression en protéine. Un gène est une séquence d'ADN qui va être transcrite par la machinerie de transcription en ARN pré-messager. Cet ARN est généralement composé d'exons et d'introns. Sa maturation en ARNm comprend l'épissage (élimination d'un ou plusieurs introns), coiffe en 3'et polyadénylation en 5'. L'ARNm mature est transporté dans le cytoplasme où il va pouvoir être traduit en protéine. La traduction se fait par la lecture de sa séquence par les ribosomes et les ARN de transferts, du codon initiateur (AUG, dans le premier exon) au codon stop dans le dernier exon. Les codons entre eux déterminent le cadre ouvert de lecture (Open Reading Frame, ORF), et leur nombre détermine la taille de la protéine traduite.

Les protéines de petite taille sont généralement appelées peptides, mais pour marquer l'opposition avec les peptides issus de la dégradation d'un précurseur protéique de grande taille, plusieurs nomenclatures ont été introduites, sans qu'une seule ne se généralise. J'ai décidé d'employer « sPEP » dans ce manuscrit. C'est la petite taille des smORFs ou de leur sPEP correspondant, qui, rendant leur analyse difficile, explique pourquoi leur existence n'a d'abord pas pu être étudiée pendant presque une dizaine d'années après les premiers séquençages complets de génomes.

II. Identification

A. L'ère pré-peptidique : L'exclusion par les méthodes classiques d'identification des gènes

Voyons d'abord les raisons théoriques et techniques expliquant comment et pourquoi l'existence des smORFs (et sPEPs) n'a pendant si longtemps pas été mise à jour.

1. L'annotation des génomes et la limite de 100 codons

Lorsque que la séquence des génomes de multiples organismes a été publiée a eu lieu une véritable course à l'annotation. En effet, ces séquences enfin disponibles, on pensait pouvoir « cracker » le code de la vie d'un point de vue génétique, via le décryptage minutieux de ces livres de plusieurs millions ou milliards de caractères, écrits dans un alphabet à 4 lettres que sont les génomes. Selon le dogme en place au moment, l'expression des génomes se fait par l'expression des protéines à partir des gènes qu'il contient. Il était donc désormais primordial de les analyser afin d'y déceler TOUS les gènes, et, le croyait-on, TOUT comprendre : le développement des organismes, la vie des cellules, et donc bien évidemment les maladies. On saurait ainsi voir et corriger les fautes de frappe du livre.

L'approche privilégiée a donc été de chercher, de prédire, les ORFs dans les génomes : ainsi on saurait où sont les gènes et quelles protéines ils expriment en théorie. Mais il s'est avéré nécessaire de définir un seuil de spécificité. En effet, des modèles théoriques démontrent que sur une séquence d'ADN randomisée (contenant 25% de chacune des 4 bases), un ORF d'au moins 250nt sera présent par kb, et un ORF d'au moins 300nt par 36kb (Claverie, 1997). Le seuil de 300nt, soit 100 codons, a donc communément été adopté par les bio-informaticiens comme critère pour définir un ORF fonctionnel. Les ORFs de taille inférieure ne seront pas considérés lors des prédictions de gènes.

Ainsi, dans le génome de *Saccharomyces cerevisiae* (12Mpb), premier génome eucaryote dont le séquençage a été achevé et publié en 1997, environ 6000 ORFs ont été prédits (dont « seulement » 500 peuvent être dus au hasard) (Oliver et al., 1992), (Dujon et al., 1994), (Goffeau et al., 1996). Si dans ce même génome étaient annotés tous les ORFs potentiels dont la taille est comprise entre 2 et 1000 codons, 260000 d'entre eux auraient une taille comprise entre 2 et 99 codons (Fig. 2) (Basrai et al., 1997). Cette zone de tailles n'est donc pas exploitable puisque la probabilité que ces ORFs n'aient pas de signification biologique est extrêmement haute.

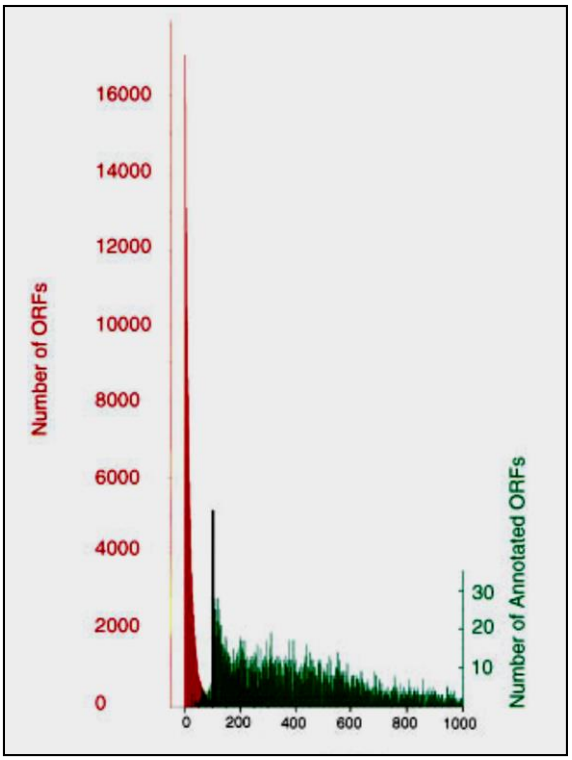


Figure 2 : ORFs du génome de *Saccharomyces cerevisiae*. Le nombre total d'ORFs de la taille indiquée dans le génome de *Saccharomyces cerevisiae* est représenté en rouge. Les ORFs annotés sont en vert. L'échelle du nombre total d'ORFs (en rouge) est compressée 100 fois par rapport à l'échelle des ORFs annotés (vert). La ligne verticale noire indique le seuil de 100 codons choisi pour l'annotation des gènes (Basrai et al., 1997).

Pourtant, des protéines inférieures à 100 résidus sont déjà connues (Basrai et al., 1997), suggérant que cette « zone d'ombre » recèle encore beaucoup de gènes à identifier. Il faudra cependant attendre le développement de techniques permettant de les identifier de manière systématique, et nous verrons comment le grand projet de séquençage des génomes y aura contribué.

2. Les cribles de mutagenèse aléatoire, statistiquement défavorisants

Une autre méthode historique d'identification de gènes est le crible génétique, avec notamment les cribles de mutagenèse aléatoire. On utilise un agent mutagène à des doses qui permettront d'engendrer statistiquement une seule mutation par chromosome, dans le but d'obtenir au final des organismes dont le génome ne présentera qu'une seule mutation. Ainsi, si un phénotype d'intérêt est observé chez l'individu mutant, il n'y a pas d'ambigüité quant à son origine. En 1995, Christiane Nüsslein-Volhard et Eric Wieschaüs ont reçu le Prix Nobel de physiologie et de médecine pour leur crible historique sur l'embryon de drosophile, où l'agent mutagène méthanosulfonate d'éthyle (EMS) a permis d'identifier les acteurs essentiels de l'embryogenèse (Nüsslein-Volhard and Wieschaus, 1980) . Mais il est aisé de comprendre que plus une protéine est grande (la taille moyenne des protéines eucaryotes a été estimée en 2005 à 361aa (Brocchieri and Karlin, 2005), plus la probabilité qu'une mutation unique ait

lieu dans l'ORF y correspondant est grande, et inversement. Ainsi, ces cribles n'ont pour ainsi dire que très peu de chance d'identifier des smORFs.

3. La protéomique, non adaptée aux petites protéines

a. Définition

La protéomique est une approche différente qui a pour but l'identification directe des protéines présentes dans un échantillon biologique. C'est un terme récent, de l'ère post-génomique, puisqu'il a été inventé en 1997 (James, 1997). Mais les techniques de protéomique classique (Gulcicek et al., 2005) sont souvent limitées à l'analyse des protéines >10kDa (90 résidus), ignorant ainsi la grande majorité des sPEPs (Fälth et al., 2006).

b. Inadéquations à l'identification des sPEPs

Voyons pourquoi la protéomique ne permettait pas de voir les sPEPs dans un extrait protéique classique.

i. Dégradation protéolytique

Premièrement, lors des étapes d'extraction et de purification, les protéines et les peptides sont sujets à une dégradation protéolytique, qui sera plus rapide pour les peptides du fait de leur petite taille (Svensson et al., 2003) et de leur sous-représentation dans les cellules. En effet, une analyse à partir de cerveau de porc (optimisée pour minimiser la dégradation protéolytique) a pu montrer que les peptides de taille inférieure à 6kDa (55aa) ne représentent que 0,1% des protéines totales (elles-mêmes ne représentant que 10% du poids du tissu) (Minamino et al., 2003). Ainsi, les extraits protéiques sont déjà appauvris en sPEPs.

ii. Génération de fragments protéiques

Ensuite, la plupart des études protéomiques utilisent des enzymes de digestion (le plus souvent la trypsine) pour générer des fragments protéiques qui peuvent être séquencés par spectrométrie de masse en tandem (MS/MS). Les sPEPs encore présents seront dilués par les peptides issus des grosses protéines. De plus, contrairement aux protéines où il n'est pas nécessaire d'analyser tous les fragments pour en valider l'identification, les sPEPs ne seront souvent représentés que par un seul fragment.

iii. Bases de données pour l'identification des protéines

Enfin, les séquences obtenues sont confrontées aux bases de données préexistantes afin d'identifier les protéines présentes. Mais celles-ci ayant été générées par les précédentes études protéomiques et les prédictions de gènes lors des annotations des génomes, elles ne référencent pas ou peu de sPEPs et ne permettent pas de relier de nouveaux sPEPs à des gènes.

iv. Et la peptidomique?

En 2001 on voit pourtant apparaître la peptidomique, où les méthodes de protéomique sont dérivées et optimisées pour l'étude des protéines dont le poids moléculaire est compris entre 0,5 et 15kDa (Schulz-Knappe et al., 2001). Mais ce domaine est surtout développé dans l'objectif d'analyser des hormones, neuropeptides ou cytokines, qui ont en commun avec les sPEPs leur taille, mais sont quant à eux issus de la dégradation de protéines précurseurs. De ce fait, ces techniques ont surtout été développées et utilisées pour l'analyse de fluides tels que le plasma sanguin ou le liquide cérébro-spinal (Schrader and Schulz-Knappe, 2001), mais peu pour des analyses de protéome à base de tissus.

4. L'ère pré-peptidique : un cercle vicieux

L'ensemble de ces données nous démontre que dans les premières années de l'ère post-génomique, devant la quantité astronomique d'informations que représente la séquence brute des génomes, il n'était pas possible d'être exhaustif. Dans un premier temps, les chercheurs n'ont donc pas eu d'autre choix que d'ignorer l'existence les sPEPs. Ceci a permis d'être plus spécifique sur l'étude des gènes et protéines « classiques », en attendant que les techniques s'améliorent. De plus, des études pionnières sur le rôle de sPEPs particuliers (Andrews and Rothnagel, 2014), dont les travaux de notre équipe auxquels j'ai contribué en master et en thèse, sont venues rappeler l'importance d'une classe de molécules encore dans l'ombre.

Ces premières années ont donc été en défaveur des sPEPs : n'étant présents dans aucune base de données, ils ne permettent pas d'en identifier de nouveaux. Nous allons voir maintenant comment la tendance a commencé à s'inverser depuis le milieu des années 2000.

B. L'entrée dans l'ère peptidique

Le grand projet des séquençages des génomes aura eu une conséquence qui a révolutionné la biologie : le développement de techniques de séquençage de l'ADN à haut débit, appelées NGS (Next Generation Sequencing), qui ont vu le jour dès 2005 et sont de plus en plus performantes, en ce sens qu'elles permettent de séquencer toujours plus, toujours plus vite, et toujours moins cher (Pickrell et al., 2012). Les principales plateformes de séquençage sont référencées dans le tableau 1. Dans les différents points traités je vais expliquer comment cette envolée a impacté sur le monde des sPEPs.

Company	Sequencing platforms	Year	Read length	Data/run/time	References
Roche	454	2005	100 bp	40 Mb/run/10 h	www.my454.com
	GS-FLX Titanium	2008	650 bp	400-650 Mb/run/10-20 h	
Illumina/Solexa	GA(II)	2008	35-75 bp	35 Gbp/run/14 days	www.illumina.com
	HiSeq 2500	2012	2x100 bp	600 Gb/run/11 days	
Applied Biosystems	ABI SOLiD	2008	20-30 bp	3 Gb/run/7 days	www.appliedbiosystems.com
	SOLiD 4	2009	2x50 bp	30-80 Gb/run/6-16 days	
Complete Genomics	NA	2009	70 bp	18 x 20-60 Gb/run/12 days	www.completegenomics.com
Helicos	HeliScope	2012	> 25 bp	28 Gb/run/8 days	www.helicosbio.com
Pacific Bioscience	PacBio RS	2012	1000 bp	100 Gb/1 h	www.pacificbiosciences.com
Oxford Nanopore	GridIon	2012	10000 bp	100 Gb/run/5 h	www.nanoporetech.com

Tableau 1 : résumé des plateformes de séquençage d'ADN de nouvelle génération.

1. Meilleure annotation des génomes

Les NGS ont permis le séquençage d'un nombre de génomes en complète expansion depuis 1997 (Fig. 3). Ceci va permettre un gain extraordinaire de précision en génomique comparative. En effet, la robustesse des paramètres en rapport avec la conservation évolutive augmente avec le nombre de génomes disponibles, et cela a engendré de grands progrès en génomique.

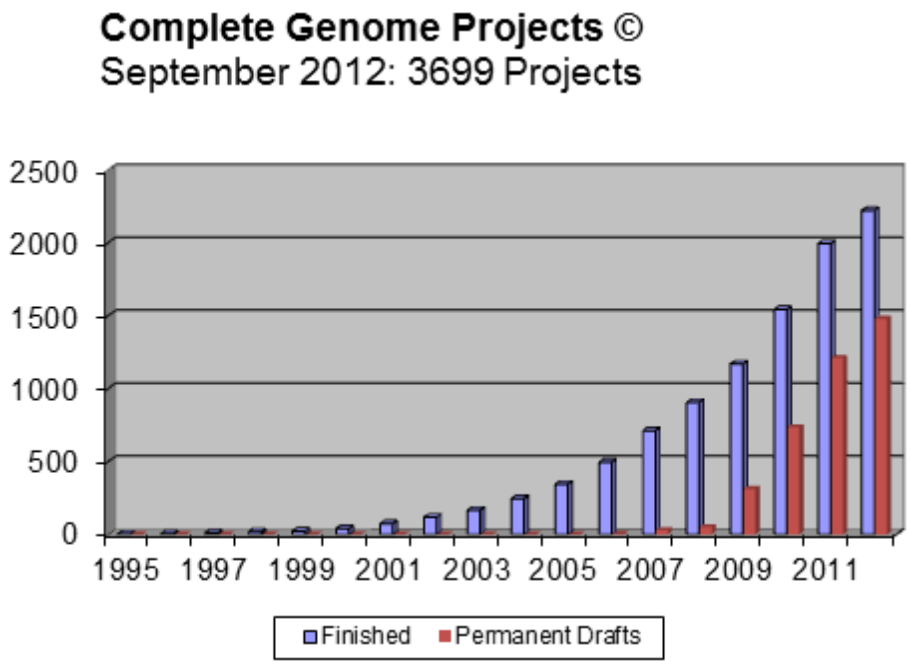


Figure 3 : Nombre de génomes séquencés entre 1995 et 2012, selon les données référencées sur la base de données Genomes OnLine Database GOLD (http://genomesonline.org)

a. Deux stratégies de prédiction de gènes

Les programmes de prédiction de gènes utilisent deux stratégies en combinaison : la recherche **empirique** et la recherche ***ab initio*** (ou *de novo*) (Sleator, 2010). (Fig. 4)

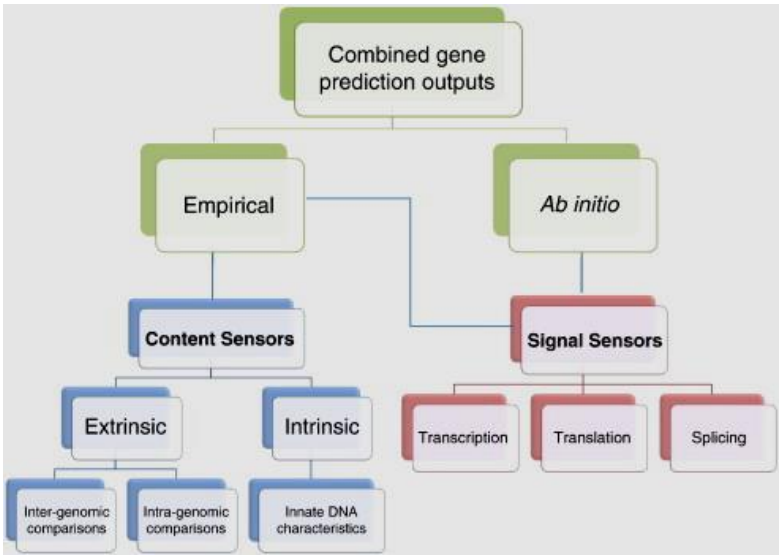


Figure 4 : Vue schématique des méthodes de prédiction de gènes eucaryotes et les senseurs utilisés en routine pour localiser les gènes dans les séquences génomiques (Sleator, 2010).

i. La prédiction empirique

Les prédicteurs empiriques permettent d'identifier les gènes sur la base d'homologies de séquences avec des bases de données génomiques, transcriptomiques [i.e. banques d'ADNc (ADN complémentaires) et d'EST (Expressed Sequence Tag)], ou protéomiques. Ils permettent de différencier les séquences codantes et non-codantes en cherchant des informations sur la composition de la séquence considérée : les « **content sensors** ». Ceux-ci peuvent être extrinsèques lorsqu'ils sont basés sur la comparaison inter- ou intra-génomique (sur le principe qu'une séquence codante présente une meilleure conservation évolutive qu'une non-codante, (Mathé et al., 2002)), ou intrinsèques lorsqu'ils sont basés sur l'analyse de la séquence elle-même.

Plusieurs critères permettent de juger si une séquence est codante. Par exemple, la composition en nucléotides (le pourcentage de GC est généralement plus haut dans les séquences codantes en comparaison avec le génome dans sa totalité) ou l'usage des codons sont des content sensors intrinsèques très utilisés (Archetti, 2004).

Afin d'automatiser cette recherche de content sensors intrinsèques, des modèles mathématiques sont élaborés à partir de sets de séquences d'apprentissage (connues comme codantes ou non-codantes). Ils vont ainsi permettre de dire si, sur des bases probabilistes, une séquence est codante ou pas. Le plus utilisé est le modèle de Markov (Krogh et al., 1994). Il en existe plusieurs catégories que je ne détaillerai pas ici, mais dont les principes sont brièvement expliqués dans la Box1.

A Markov model (MM) is a stochastic model which assumes that the probability of a particular nucleotide occurring at a given position depends only on the k previous nucleotides. In this case k is the order of the MM, the larger k the finer the MM can characterize dependencies between adjacent nucleotides. Such a model is defined by the conditional probabilities P(X|k previous nucleotides), where X = A, T, G or C. In order to build a Markov model, a learning set of sequences, on which these probabilities will be estimated, is required.
The most frequently used categories of MMs in eukaryote gene prediction methods are outlined below:

Positional weight matrices (PWM)	The simplest MMs are homogeneous zero order MMs which assume that each base occurs independently with a given frequency. Such simple models are often used for non-coding regions.
Weight array model (WAM)	An inhomogeneous higher order MM capable of capturing potential dependencies between adjacent positions of a signal.
Three-periodic Markov model	Characterize coding sequences. Coding regions are defined by three MMs, one for each position inside a codon.
Interpolated Markov models (IMM)	IMMs combine statistics from several MMs, from order zero to a given order k (typically k = 8), according to the information available.
Hidden Markov models (HMM)	HMMs allow for insertions and deletions and so variation in signal length.
Generalized Hidden Markov models (GHMM)	GHMMs allow a string, rather than a single symbol, as the output of a state.
Semi-Markov conditional random field (SMCRF)	A more flexible variation of GHMM which allows a wider range of biological features to be incorporated with fewer technical concerns (Bernal et al., 2007)
Evolutionary Hidden Markov Models (EHMM)	EHMMs model molecular evolution as a Markov process in two dimensions: a substitution process over time at each site in the aligned genomes, which is guided by a phylogenetic tree; and a process by which the rate of evolution changes from one site to the next (Brent and Guigo, 2004)

Box 1 : Vue d'ensemble des modèles de Markov utilisés en analyse de séquences et prédiction de gènes (Sleator, 2010).

ii. La prédiction ab initio

Les prédicteurs *ab initio* vont quant à eux chercher des séquences spécifiques, des signatures, d'un gène : les « **signal sensors** » (Fig. 5). Ces signaux peuvent être des signes de **transcription** (site de la coiffe de l'ARN, site de polyadénylation, boîte TATA dans le promoteur), de **traduction** [séquence Kozak en amont d'un codon initiateur de la traduction (Kozak, 1996)] et d'épissage [sites donneurs et accepteurs et point de branchement(Brent and Guigó, 2004)].

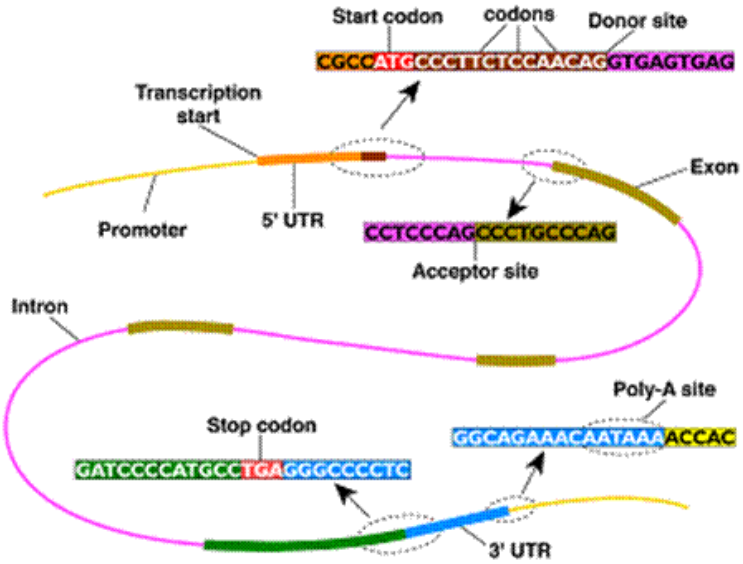


Figure 5 : Schéma d'un gène eucaryote avec les principaux « signal sensors ».

b. La naissance des prédicteurs « multi-génomes »

De manière générale, la précision et la spécificité de l'annotation augmentent avec le nombre de génomes séquencés et disponibles. D'une part cela permet d'affiner mathématiquement les algorithmes, d'autre part cela permet d'intégrer des critères d'évolution et de conservation qui vont donner une nouvelle robustesse aux prédictions.

i. Nouveaux modèles mathématiques

On voit apparaître de nouveaux modèles mathématiques auxquels est ajoutée la dimension de phylogénie : l'Evolutionary Hidden Markov Model (EHMM) (Holmes, 2003; Pedersen and Hein, 2003) et le phylo-HMM (Siepel, 2003; Siepel and Haussler, 2004), sur lesquels sont basés les prédicteurs de gènes *ab initio* « multi-génomes ». On notera que ces sophistications ont été permises par les nombreux progrès dans le domaine en pleine expansion de l'informatique. C'est parce que les calculateurs sont de plus en plus performants que l'on peut traiter de plus en plus de données en parallèle. A titre d'exemple, les travaux de

Boffelli *et al.* montrent que la divergence collective de génomes d'espèces proches (les primates dans son étude) permet d'identifier des gènes qu'on ne trouve pas en comparant deux espèces éloignées (humain et poisson, ou humain et souris dans ses études), avec des « dual-genome predictors » (Boffelli et al., 2003; Boffelli et al., 2004a; Boffelli et al., 2004b).

ii. Meilleure exploitation des mutations

Ces nouveaux algorithmes vont donc intégrer la conservation évolutive (nombre et nature de mutations entre plusieurs génomes) comme nouveau critère de discrimination entre régions codantes et non-codantes.

- Profil Indel

Le profil défini par les mutations correspondant à des insertions ou à des délétions (profil indel) est différent selon si la région est codante ou non. Ces mutations sont en effet généralement plus nombreuses dans les régions non-codantes. Celles qui sont présentes dans les régions codantes le sont généralement par multiple de trois (pour préserver la phase de lecture) (Siepel et al., 2005).

- Test du Ka/Ks

Les mutations correspondant à des substitutions sont elles aussi exploitables grâce à un test développé par Nekrutenko : le ratio du nombre de substitutions non –synonymes (Ka) sur celui de substitutions synonymes (Ks) (Nekrutenko et al., 2002). En effet, dans une région codante il y aura plus de Ks puisque la substitution n'induit pas de changement de l'acide aminé correspondant au codon considéré. Ce test du Ka/Ks est donc un bon critère de distinction entre les régions codantes ou non : un score faible montre que la pression de conservation s'est faite sur la séquence protéique (et non nucléotidique) et qu'il s'agit sans doute d'une région codante.

c. Le seuil de 100 codons n'est plus nécessaire

Des outils bioinformatiques prenant en compte ces différents paramètres, accessibles grâce au nombre grandissant de séquences génomiques disponibles, permettront petit à petit de s'affranchir du seuil des 100 codons en deçà duquel un smORF avait peu de chance d'être biologiquement significatif. On voit apparaitre des programmes dédiés à la recherche et à l'analyse des smORFs (Cheng et al., 2011), tels que sORF Finder (Hanada et al., 2010), BIAUCAS (Takahashi et al., 2012) ou uPEPeroni (Skarshewski et al., 2014) qui facilitent leur identification.

2. L'impact des NGS dans l'analyse des gènes codants

Depuis 2005, on voit se multiplier les travaux de recherches systématiques de smORFs dans les organismes modèles de génétiques (Tableau 4). La plupart de ces études sont basées sur l'analyse informatique de la conservation et de la nature des smORFs, désormais possible. Mais grâce à deux nouvelles techniques engendrées par le grand projet des séquençages des génomes, il est maintenant possible de tester la valeur de ces prédictions et de valider expérimentalement la transcription et la traduction des smORFs identifiés : le **RNA-sequencing** et le **ribosome profiling**, que je vais décrire dans les sections suivantes.

a. La révolution du RNA sequencing en transcriptomique

Dans la section traitant des stratégies d'annotation des génomes, je disais que les prédicteurs empiriques de gènes en permettent l'identification sur la base d'homologies de séquences avec des bases de données génomiques, transcriptomiques ou protéomiques. Nous allons voir ici comment les progrès en transcriptomique ont permis de meilleures prédictions de gènes.

Le transcriptome correspond à la totalité des transcrits dans un échantillon biologique à un instant donné. Avoir une connaissance exhaustive du transcriptome (nature et quantité de chaque transcrit) permet donc d'identifier tous les éléments fonctionnels du génome nécessaires dans une condition donnée.

i. Les anciennes méthodes de transcriptomique

En 1997, simultanément à la publication du génome de *Saccharomyces cerevisiae*, des micro-puces à ADN contenant la totalité de ce génome ont été développées, permettant ainsi l'analyse de sa transcription au cours du temps ou en fonction des conditions environnementales (DeRisi et al., 1997). Cette méthode, créée 2 ans auparavant, (Schena et al., 1995) est basée sur l'hybridation d'ADNc marqués (synthétisés par transcription reverse des ARNs de l'organisme) avec un oligonucléotide spécifique d'un gène fixé sur la puce. Ainsi, on sait en mesurant les signaux correspondant à chaque oligonucléotide quels gènes sont exprimés. De plus, il est possible de créer des puces qui permettront de voir les variants d'épissage (Clark et al., 2002). Cette méthode, bien qu'à haut débit et peu chère, comporte des inconvénients, dont nous citerons surtout un haut bruit de fond, des hybridations non-spécifiques et donc une quantification limitée et souvent compliquée (Okoniewski and Miller, 2006; Royce et al., 2007).

En parallèle étaient développées d'autres méthodes ne présentant pas les problèmes de bruit de fond et d'hybridations non-spécifiques (conséquences d'une méthode basée sur l'hybridation), où les ADNc sont directement séquencés avec la méthode de Sanger (Boguski et al., 1994; Gerhard et al., 2004). Ces méthodes ont en revanche pour défaut leur bas débit, leur prix et généralement la non-possibilité de quantification.

ii. Les approches "Tag-based"

Cependant, de nombreuses méthodes « Tag-based » (Harbers and Carninci, 2005), qui visent à générer des étiquettes de chaque ARN, sont développées pour générer des banques d'ADNc non-biaisées et exhaustives. Par exemple, la méthode CAGE (Cap Analysis Gene Expression, (Kodzius et al., 2006), basée sur la méthode de « Cap-trapping » (Carninci et al., 1996) qui permet de générer l'ADNc à partir de la coiffe d'un ARNm, et après marquage en 5' avec un linker qui introduit un site de restriction Mme1 (enzyme qui coupe à 20nt de son site), permet de générer des étiquettes spécifiques de l'ADNc. (Fig. 6). Ces méthodes permettent une bonne quantification des transcrits, mais resteront coûteuses et fastidieuses tant que le séquençage des étiquettes se fera par la méthode Sanger.

iii. Le RNA sequencing

Et en 2008, on voit apparaître les premières analyses de transcriptomes où les ADNc (ou les banques d'étiquettes générées via des approches « Tag-based ») vont bénéficier des NGS (Cloonan et al., 2008; Morin et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Vera et al., 2008; Wilhelm et al., 2008), permettant ainsi, de la même manière que pour les génomes, une plus grande vitesse d'analyse, un coût toujours plus bas, et une bonne capacité de quantification (permettant de voir jusqu'à un changement de 9000 fois dans l'expression d'un ARN entre deux conditions (Nagalakshmi et al., 2008). C'est le RNA-sequencing (RNA-seq). L'augmentation de la capacité de séquençage permet une plus grande profondeur d'analyse. Ainsi, la présence de transcrits peu abondants peut être révélée. Dans le tableau 2, Wang *et al* ont résumé les avantages du RNA-seq par rapport aux autres méthodes de transcriptomique (Wang et al., 2009).

Table 1 | **Advantages of RNA-Seq compared with other transcriptomics methods**

Technology	Tiling microarray	cDNA or EST sequencing	RNA-Seq
Technology specifications			
Principle	Hybridization	Sanger sequencing	High-throughput sequencing
Resolution	From several to 100 bp	Single base	Single base
Throughput	High	Low	High
Reliance on genomic sequence	Yes	No	In some cases
Background noise	High	Low	Low
Application			
Simultaneously map transcribed regions and gene expression	Yes	Limited for gene expression	Yes
Dynamic range to quantify gene expression level	Up to a few-hundredfold	Not practical	>8,000-fold
Ability to distinguish different isoforms	Limited	Yes	Yes
Ability to distinguish allelic expression	Limited	Yes	Yes
Practical issues			
Required amount of RNA	High	High	Low
Cost for mapping transcriptomes of large genomes	High	High	Relatively low

Tableau 2: Avantages du RNA-Seq comparé aux autres méthodes de transcriptomique (Wang et al., 2009).

En 2009, une nouvelle méthode permet cette fois de directement séquencer l'ARN sans passer par des étapes de transcription reverse pouvant induire de nombreux biais : le **Direct RNA Sequencing** (DRS) (Ozsolak et al., 2009).

iv. Le consortium FANTOM

Ces progrès en transcriptomique vont permettre une meilleure annotation et compréhension des génomes, mais vont surtout révéler des surprises et soulever de nouvelles questions. Depuis 2000, le consortium FANTOM (http://fantom.gsc.riken.jp/), à Riken, bénéficie des progrès continuels en transcriptomique et séquençage, dans son projet d'annotation fonctionnelle de tous les ADNc de la souris (Kawai et al., 2001) (banque de données qui a été utilisée pour la prédiction de gènes dans le génome humain publié en 2001, (Lander et al., 2001; Venter et al., 2001). En 2002, FANTOM2 travaillait sur une banque d'ADNc pleine taille générés « classiquement » : ils étaient au nombre de 60770 (Okazaki et al., 2002).

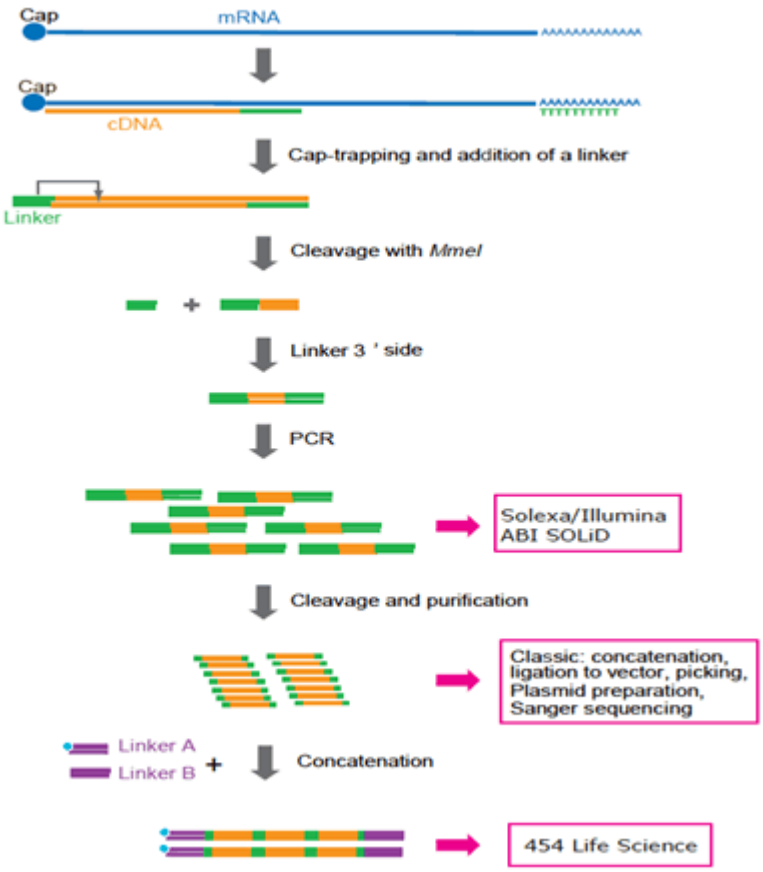


Figure 6 : Représentation du protocole de préparation des ADNc avec la méthode CAGE adaptée à plusieurs plateformes de séquençage. Cette méthode a été développée à Riken dans le cadre du consortium FANTOM (http://fantom.gsc.riken.jp/), qui vise à annoter fonctionnellement tous les ADNc de souris.

En 2005, FANTOM3 a bénéficié de la méthode CAGE (développée à Riken) : les transcrits identifiés étaient alors 102801 (Carninci et al., 2005) (Fig. 6). Cette meilleure sensibilité a permis de mettre en évidence que la majorité du génome est transcrit, à partir des deux brins (alors que les ARNm ont été estimés à 2% du génome, (Frith et al., 2005), Tableau 3). De plus, 33% des ADNc analysés (34030 / 102801) ne contiennent pas de cadre ouvert de lecture >100 codons et vont être annotés comme non-codants. En 2006, Frith *et al*. partent du constat qu'il existe une « cassure » à 100 codons sur la répartition des tailles de toutes les protéines annotées sur les 102801 ADNc (Fig. 7), avec seulement 3,3% de celles-ci <100 résidus. Ils utilisent donc un programme, CRITICA (Coding Region Identification Tool Invoking Comparative Analysis, (Badger and Olsen, 1999)) pour ré-analyser les 102801 ADNc de souris générés par FANTOM. Ce programme, pour prédire les ORFs, n'utilise plus de seuil de taille mais des données telles que le test du Ka/Ks et la composition en nucléotides (qui je le rappelle sont différents entre régions codantes ou non et peuvent être reconnus par

des algorithmes entrainés à partir de sets de séquences avérées codantes ou non). Grâce à cet outil, la répartition de tailles des ORFs prédits ne présente plus cette cassure, avec 10% (1240) des protéines <100 résidus.

Organism	No. of protein-coding genes	Genome size (Mb)	Coding sequences		UTR sequences		Total transcribed noncoding sequences		Ratio of noncoding to coding sequences
			Mb	%	Mb	%	Mb	%	
Whole genome									
Human	~20–25 000	2851	34	1.2	32	1.1	1619	57	47:1
Mouse	~20–25 000	2490	31	1.3	26	1.1	1339	54	43:1
Fruit fly	~13 500	120	22	18	6.4	5.3	53	44	2.4:1
Nematode	~19 000	100	26	26	0.4	0.4	33	33	1.3:1
Nonrepetitive portion of genome only									
Human		1455	33	2.3	26	1.8	867	60	27:1
Mouse		1422	29	2.0	22	1.6	811	57	28:1
Fruit fly		109	21	20	6.2	5.7	48	44	2.2:1
Nematode		86	25	29	0.3	0.4	26	31	1.1:1

Tableau 3 : Pourcentages des séquences codantes et des transcrits non-codants de génomes d'organismes modèles (Frith et al., 2005)**.**

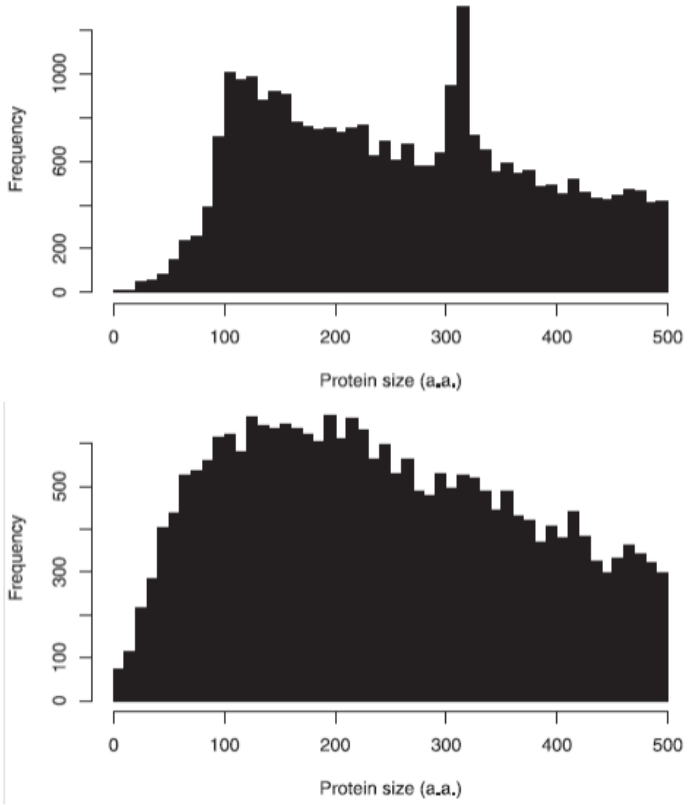


Figure 7 : Distributions de la taille des protéines. **En haut** : distribution de la taille de 40865 protéines de souris référencées sur la base de données International Protein Index (IPI). **En bas** : distribution de la taille de 31035 protéines de souris prédites à partir de la banque d'ADNc FANTOM en utilisant le programme CRITICA (Frith et al., 2006).

En accord avec ce résultat, en 2008, Dinger *et al.* modélisent la taille d'un ORF obtenu sur un transcrit de séquence randomisée et de taille croissante et démontrent que pour un transcrit d'1kb, la taille 300nt pour un ORF est supérieure de deux déviations standard à la taille moyenne obtenue au hasard (Fig. 8) (Dinger et al., 2008). C'est donc un seuil de

spécificité très mal choisi pour des ARNm de cette taille. Par ailleurs, cela démontre aussi que pour les longs transcrits de plus de 6kb, le seuil choisi de 300nt est inférieur à la taille moyenne d'un ORF escompté par hasard, il est donc inadapté pour des longs ARNm.

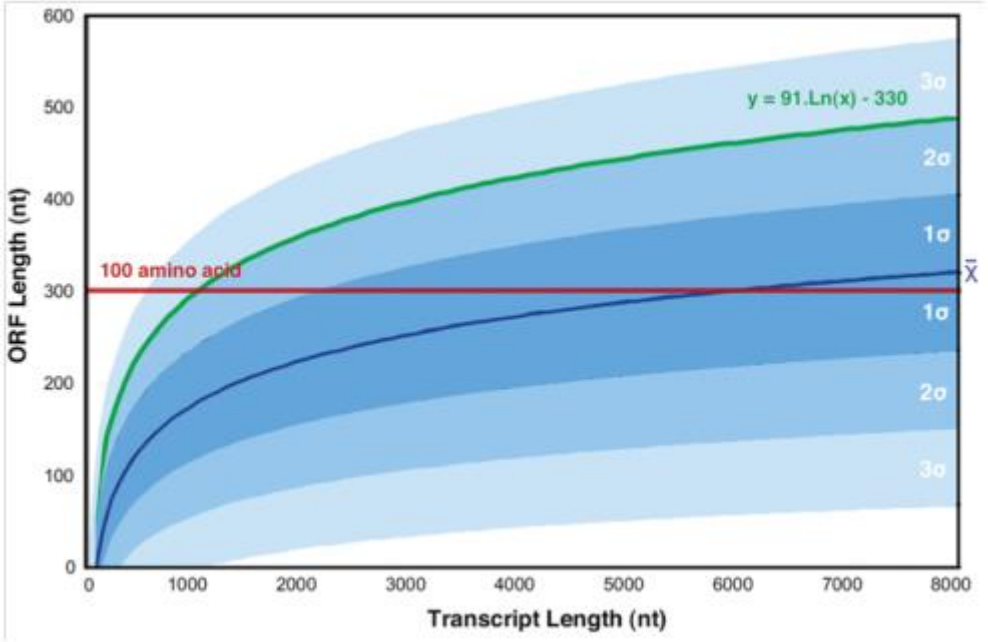


Figure 8 : Fréquence d'ORFs dans des transcrits générés au hasard de taille croissante. 20000 transcrits de taille et composition nucléotidique variables ont été virtuellement générés et scannés. L'ORF maximal et la taille des transcrits sont reportés et suivent une courbe logarithmique. Les régions bleutées représentent les fréquences des ORFs à 1, 2 ou 3 déviations standards de la moyenne (ligne bleue). La ligne rouge indique le seuil de 300nt utilisé en annotation des génomes. La ligne verte représente le seuil qu'il aurait fallu utiliser (moyenne + 2 déviations standards), qui est dépendant de la taille du transcrit considéré (Dinger et al., 2008).

Les travaux de Frith et Dinger démontrent donc que les protéines < 100 résidus ont été largement sous-estimées. Il devient alors essentiel de différencier les ARN codants et non-codants, non plus sur la seule présence d'un ORF, mais sur la traduction de celui-ci. Et ceci devient possible grâce à une technique développée en 2009 dans le laboratoire de Jonathan S. Weissman : le **ribosome profiling** (Ingolia et al., 2009).

b. Le ribosome profiling

La technique du ribosome profiling, aussi appelée ribosome footprinting, est une discipline nouvelle, conceptuellement à cheval entre la transcriptomique et la protéomique puisqu'elle va permettre de prédire le protéome à partir de la partie du transcriptome en cours de traduction. Elle tire avantage du RNA-seq qui permet de séquencer les transcrits à haut débit, et part du principe que les ribosomes présents sur un ARNm en cours de traduction le protègent d'une dégradation à la nucléase sur une région d'environ 30nt (Steitz, 1969). Ainsi, les régions protégées, une fois séquencées, révèleront au codon près le profil exact des régions en cours de traduction (Fig. 9). En 2011, ils vont notamment coupler cette technique à

l'utilisation d'une drogue, l'harringtonine, un inhibiteur de la synthèse protéique (Huang, 1975; Tscherne and Pestka, 1975; Fresno et al., 1977) qui cause l'accumulation des ribosomes au codon d'initiation des ORFs et confère ainsi à cette technique une meilleure sensibilité (Ingolia et al., 2011).

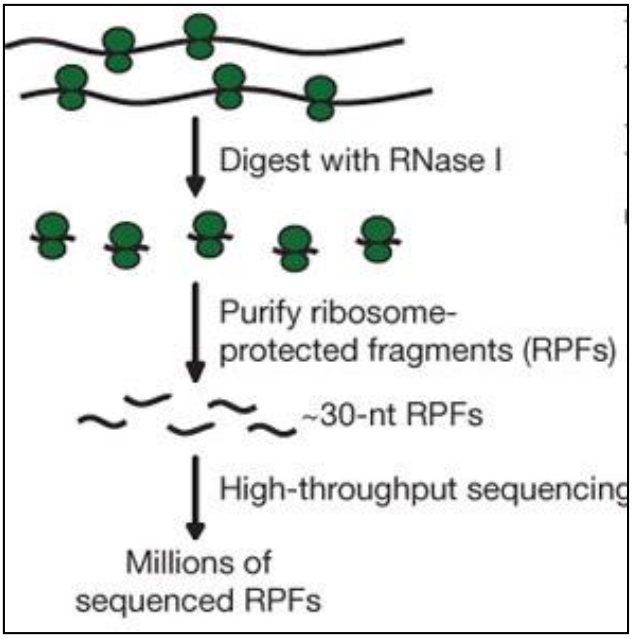


Figure 9: Schéma du principe du ribosome profiling.

Cette technique est en train de révolutionner la génomique : **un gène codant est un gène traduit**. On peut enfin s'affranchir de la nature du gène, de sa conservation évolutive, de la nature des codons d'initiation, et surtout de la taille de son (ses) ORF(s). C'est cependant une technique récente encore controversée, la preuve absolue de la traduction d'un sPEP étant sa détection directe. C'est désormais possible grâce à la peptidomique, initialement développée pour l'analyse de peptides issus de la dégradation d'un précurseur protéique, qui a petit à petit trouvé sa place dans la recherche des sPEPs et l'identification des smORFs (Slavoff et al., 2013).

3. L'ère peptidique : le début d'un cercle vertueux

L'ensemble de ces nouvelles techniques et de ces nouvelles données, couplé à la conviction qu'il existe un grand nombre – peut-être encore autant que les grandes protéines déjà identifiées – va donner lieu à de plus en plus d'études qui visent à caractériser de manière systématique les sPEPs dans de nombreux organismes modèles, en combinant plusieurs des approches décrites (Tableau 4). Par exemple, le groupe de Petra Van Damme a développé une approche de protéogénomique où sont couplées des analyses en ribosome profiling et des annotations génomiques avec les résultats de spectrométrie de masse obtenus avec des

nouveaux équipements ultra-sensibles (UltiMate® 3000 RSLC Nano System) (Menschaert et al., 2013).

L'ère pré-peptidique est révolue, il y a une véritable émulation autour des sPEPs : 1/ les progrès informatiques et le nombre de génomes séquencés permettent des prédictions de gènes beaucoup plus sensibles et robustes, 2/ les séquençages nouvelle génération ont été détournés au profit de la transcriptomique et du ribosome profiling, qui comme nous l'avons vu ont engendré un véritable bond en avant dans l'identification des gènes et des ORFs, 3/ les techniques d'analyses protéomiques sont de plus en plus sensibles et commence à permettre de valider les prédictions de sPEPs.

Location of smORFs or data set of sequences analysed	Number of transcripts or sORFs analysed	Approaches						Expression		Number of smORFs with coding potential identified	References
		Sequence similarity with other species	K_a/K_s analysis	Length or position similarity with other	Initiation context	Nucleotide composition	Protein domains, motifs and clusters	Transcription analysis	Mass spectrometry		
Humans											
5′ UTR	27660	✓	–	–	–	–	–	–	–	43	Iacono *et al*, 2005
5′ UTR	21768	✓	✓	–	✓	–	–	–	–	204	Crowe *et al*, 2006
Overlapping with major ORF	14159	✓	✓	–	–	✓	–	–	–	40	Chung *et al*, 2007
Overlapping with major ORF	9163	✓	–	✓	✓	✓	✓	–	–	217	Ribrioux *et al*, 2008
Overlapping with major ORF	26009	✓	–	–	✓	–	–	–	–	168	Xu *et al*, 2010
Overlapping with major ORF	76000	–	–	–	✓	–	–	–	–	24547	Vanderperre *et al*, 2012
Whole genome	NA	–	–	–	–	–	–	✓	✓	4	Oyama *et al*, 2004
Whole genome	NA	–	–	–	–	–	–	✓	✓	8	Oyama *et al*, 2007
Whole genome	NA	–	–	–	–	–	–	✓	✓	237	Ma *et al*, 2014
Whole genome	NA	–	–	–	–	–	–	✓	✓	90	Slavoff *et al*, 2013
Whole mRNA	83886	–	–	–	–	–	–	–	✓	1259	Vanderperre *et al*, 2013
Mice											
Whole genome	NA	✓	–	–	✓	–	–	✓	–	35	Crappé *et al*, 2013
Intergenic	102801	–	✓	–	–	✓	–	–	–	1240	Frith *et al*, 2006
Drosophila melanogaster											
5′ UTR	19389	✓	✓	–	–	–	–	–	–	44	Hayden and Bosco, 2008
Intergenic	593586	✓	✓	✓	–	–	–	✓	–	401	Ladoukakis *et al*, 2011
Arabidopsis thaliana											
5′ UTR	34000	✓	✓	–	–	–	–	–	–	19	Hayden and Jorgensen, 2007
5′ UTR	23036	✓	✓	✓	–	–	–	–	–	18	Takahashi *et al*, 2012
5′ UTR	10122	✓	✓	–	–	–	–	–	–	18	Vaughn *et al*, 2012
Intergenic	570948	✓	✓	–	–	✓	–	✓	–	3241	Hanada *et al*, 2007
Intergenic	96358	✓	✓	–	–	✓	–	✓	–	2302	Hanada *et al*, 2013
Intergenic	606285	✓	–	–	–	–	✓	✓	–	1044	Lease and Walker, 2006
Zebrafish											
Intergenic	2450	✓	–	–	–	–	–	✓	–	190	Bazzini *et al*, 2014
Rice											
5′ UTR	32127	✓	–	–	–	–	–	–	–	29	Tran *et al*, 2008
Cottonwood											
Intergenic	12852	✓	–	–	–	–	✓	✓	✓	611	Yang *et al*, 2011
Common bean											
Intergenic	31576	✓	–	–	–	–	✓	✓	–	776	Guillén *et al*, 2013
Yeast											
Intergenic	NA	✓	–	–	–	–	–	✓	✓	299	Kastenmayer *et al*, 2006
5′ UTR	5542	✓	–	✓	–	–	–	✓	–	15	Zhang and Dietrich, 2005
5′ UTR	5602	✓	–	✓	–	–	–	–	–	252	Cvijovic *et al*, 2007
5′ UTR	2167	✓	✓	✓	–	–	–	–	–	12	Neafsey *et al*, 2007

Tableau 4 : Etudes qui ont identifié des smORFs codants ou potentiellement codants (Andrews and Rothnagel, 2014).

III. Deux grandes classes de smORFs

Les études visant à identifier les smORFs et à démontrer leur expression en sPEPs ont permis de montrer qu'il en existe deux grands groupes : ceux présents sur un ARNm contenant un ORF déjà connu (de taille > à 300nt) et ceux présents sur des ARN a priori non-codants puisque qu'aucun ORF n'y avait été annoté précédemment.

A. Dans les ARN messagers

Les smORFs présents sur les ARNm sont classés en deux catégories que je vais brièvement présenter ici : les smORFs en amont de l'ORF majeur (upstream smORFs, uORFs) présents dans la séquence non-codante en 5' (5' UnTranslated Region, 5' UTR), et les smORFs en aval (dowstream smORFs), dans le 3' UTR ou chevauchant l'ORF majeur (overlapping smORFs).

1. En amont de l'ORF majeur

L'existence des uORFs a été relevée pour la première fois en 1987 par Marilyn Kozak qui a analysé le 5' UTR de 700 ARNm (Kozak, 1987). Elle montre que 10% d'entre eux possèdent un ATG « non-fonctionnel » en amont de l'ORF majeur, conduisant à des smORFs de 20 à 100nt. Leur rôle est maintenant partiellement élucidé : ils permettent de réguler la traduction de l'ORF majeur située en aval (Fig. 10) (Kozak, 1986; Kozak, 1987; Morris and Geballe, 2000). Cette régulation peut se faire soit 1/ par ré-initiation : une première initiation de la traduction est faite par le ribosome sur le uORF qui est traduit en sPEP. Au codon stop, la sous-unité 60S se détache tandis que la 40S scanne jusqu'au codon initiateur de l'ORF majeur où il y a une ré-initiation et traduction de la protéine, ou 2/ par leaky scanning : la sous-unité 40S va scanner l'ARN et s'arrêter soit au codon initiateur de l'uORF, soit scanner jusqu'à celui de l'ORF majeur. Ainsi il y aura soit production de sPEP, soit de la protéine. Ces deux mécanismes permettent donc de réduire la quantité de protéines produite.

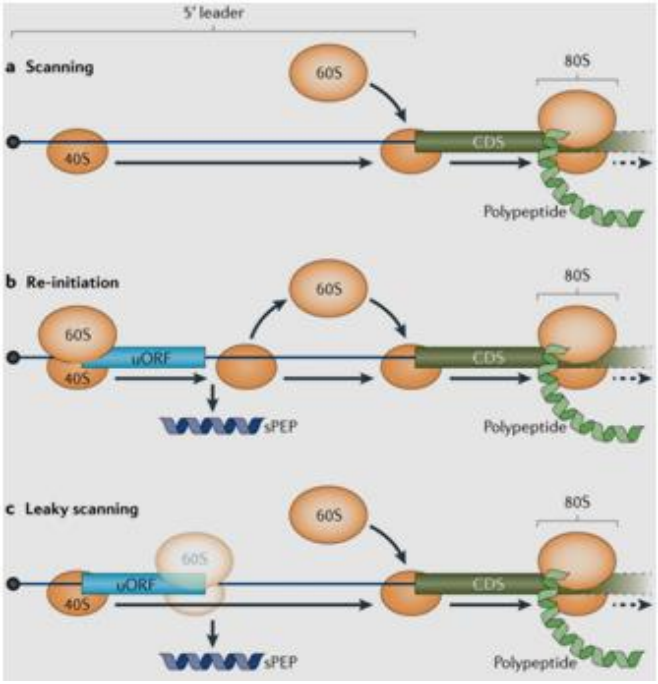


Figure 10 : Schémas des modèles de traduction d'ARNm contenant ou non un uORF. **a**. Modèle standard de traduction d'un ARNm eucaryote. La sous-unité 40S du ribosome, se lie à la coiffe (rond gris) et scanne l'ARN jusqu'au premier codon d'initiation. La sous-unité 60S se combine avec la 40S pour former le ribosome 80S, qui traduit l'ORF. **b**. On parle de ré-initiation quand 40S initie la traduction au codon initiateur d'un ORF après avoir traduit un uORF en amont, sans se dissocier de l'ARNm. **c**. Dans le leaky scanning, 40S peut soit initier la traduction à l'uORF et produire un sPEP, soit scanner au travers et initier à l'ORF majeur et produire une protéine (Andrews and Rothnagel, 2014).

Leur traduction a été largement démontrée notamment grâce au ribosome profiling, que ce soit chez *Saccharomyces cerevisiae* (Ingolia et al., 2009) ou chez la souris (Ingolia et al., 2011). Un sPEP est donc produit, même si le but n'est pas cette expression mais de réguler négativement l'expression de la protéine principale. Cependant, la conservation évolutive de la séquence en acides-aminés de certains de ces uORFs (Crowe et al., 2006) montre qu'il ne faut pas exclure que les sPEPs produits puissent être bioactifs, et non seulement la conséquence de ce mécanisme de régulation.

2. Chevauchant ou en aval de l'ORF majeur

Concernant les smORFs qui vont chevaucher ou être en aval de l'ORF majeur, aucun rôle régulateur n'a pu être montré. Pour l'instant, l'hypothèse qui prime est qu'ils permettent tout simplement de produire un autre produit protéique à partir de l'ARNm (qui du coup est polycistronique). La traduction des smORFs chevauchants a pu être montrée par ribosome profiling (Ingolia et al., 2011; Michel et al., 2012). En revanche, d'autres études montrent que les 3' UTR des ARNm sont presque dépourvus de ribosomes (Ingolia et al., 2011; Chew et al., 2013; Guttman et al., 2013). Mais des analyses en spectrométrie de masse à partir de cellules humaines ont tout de même pu identifier des sPEPs exprimés à partir de downstream smORFs (Oyama et al., 2007; Slavoff et al., 2013).

B. Dans les ARNs non-codants

1. Les lincRNAs, aussi nombreux que les ARNm

La deuxième classe de smORFs correspond à ceux trouvés dans les ARN a priori non-codants. Les ARN non-codants sont classés en 2 groupes (Dogini et al., 2014) : les petits et les longs (Tableau 5). Les petits ARN non-codants étant trop petits pour permettre la traduction de smORFs (et leur fonction étant pour la plupart déjà caractérisée), il n'y a pour l'instant pas de débat sur le fait qu'ils soient non-codants. En revanche, il en va autrement pour les longs ARN non-codants intergéniques (lncRNA ou lincRNA). On sait qu'ils sont aussi nombreux que les ARN codants (ARNm) déjà répertoriés (Okazaki et al., 2002; Carninci et al., 2005), et par définition peuvent contenir des smORFs < 300nt. Dans leur cas, le débat sur l'expression ou non de sPEPs est ouvert (Kageyama et al., 2011). En effet, les ARNm codants permettent à eux seuls l'expression de plusieurs dizaines de milliers de protéines dans la plupart des eucaryotes supérieurs étudiés. Si les lincRNAs sont traduits, ils représentent un immense réservoir de sPEPs, avec à la clef une multitude de nouvelles fonctions, de nouveaux mécanismes de régulation de l'expression des génomes.

		Mean size	Function
Long ncRNA	Ribosomal RNA	~1.9 kb	Essential for protein synthesis
	XIST RNA	~17 kb	Chromosome X inactivation
	Other lncRNA	> 200 nt	Involved in epigenetic modification, post-transcriptional processing, modulation of chromatin structure, etc.
Small ncRNA	miRNAs	18-21 nt	Gene regulation
	siRNA	~21 nt	Gene regulation; defense against viruses and transposon activity
	rasiRNA	24-27 nt	Orientation of heterochromatin in the formation of centromeres
	snoRNA	60-300 nt	Methylation and pseudo uridylation of other RNAs
	snRNA	100-300 nt	Involved in spliceosome complex
	piRNA	26-30 nt	Regulation of transposon activity and chromatin state

Tableau 5 : Différents types, principales caractéristiques et fonctions des ARN non-codants (Dogini et al., 2014).

2. La fonction de certains lincRNAs passerait-elle par l'expression de sPEPs?

Diverses études ont pu montrer qu'au même titre que les ARNm, les lincRNAs présentent un patron d'expression spécifique et régulé, que ce soit chez la drosophile (Inagaki et al., 2005) ou chez la souris (Mercer et al., 2008). Ceci suggère que ces lincRNAs aient un rôle développemental. Pour certains d'entre eux, leur fonction a été élucidée. Ils jouent un rôle dans la régulation de l'expression génique, via plusieurs mécanismes (Rinn and Chang, 2012). Ils vont par exemple 1/ **séquestrer** un facteur de transcription en jouant le rôle de leurre, ou des facteurs d'épissage en leurrant l'ARNm, 2/ permettre la bonne **architecture** d'un complexe protéique en jouant le rôle d'échafaudage, 3/ **Guider** un facteur protéique sur sa

cible (par exemple un facteur de transcription sur un promoteur), 4/ jouer le rôle d'**enhancer**, dans un triplex ADN/ARN/ADN, où le lincRNA en interaction avec l'ADN, va permettre de guider un facteur sur un promoteur éloigné. Ces 4 modes d'action ainsi que quelques exemples sont répertoriés dans le tableau 6.

De plus, ils peuvent être épissés (Ota et al., 2004) et polyadénylés (Tupy et al., 2005). Ces deux caractéristiques étant synonymes de traductibilité, il était tout naturel de se demander si la fonction de ces lincRNAs ne passerait pas par la traduction des smORFs qu'ils contiennent souvent. C'est le cas de du lincRNA *polished-rice* (*pri*), qui fut l'objet d'étude de ma thèse, et que je vais vous présenter maintenant.

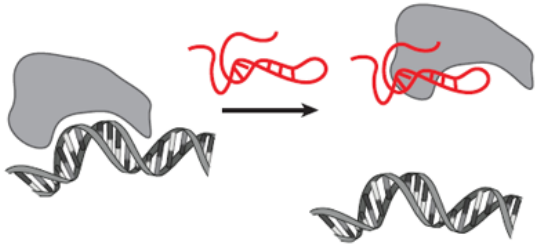
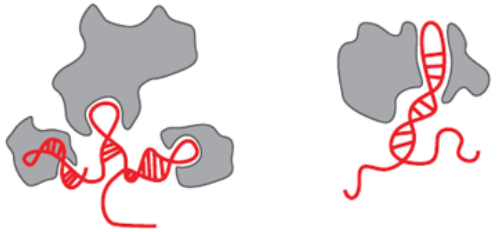
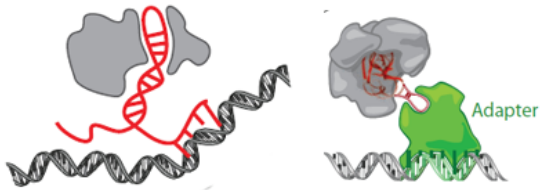
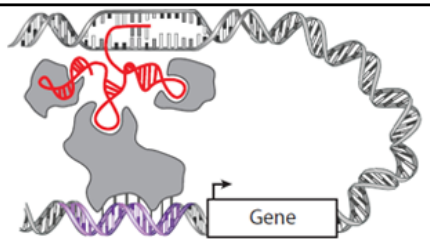
		Long noncoding RNAs	Protéines	Fonction des lncRNAs	Références
	Séquestration/leurre	*Gas5*	Glucocorticoid receptor	Binds to glucocorticoid receptor as a decoy and titrates GR away from target genes	Kino *et al*, 2010
		lincRNA-p21	hnRNP-K	Targets hnRNP-K in *trans* to mediate p53-dependent gene repression	Huarte *et al*, 2010
		PANDA	NF-YA	p53 inducible and titrates away NF-YA to favor survival over cell death during DNA damage	Hung *et al*, 2011
		MALAT1	Serine/arginine-rich splicing factors	Sequesters serine/arginine splicing factors to regulate alternative splicing	Tripathi *et al*, 2010
		DHFR minor	TFIIB	Titrates away TFIIB during cell quiescence to decrease *DHFR* transcription	Martianov *et al*, 2007
	Architecture	*Kcnq1ot1, Air*	G9a	Targets H3K9 methylase G9a in *cis* for imprinting	Nagano *et al*, 2008 Pandey *et al*, 2008
		HOTAIR, many others	LSD1-CoREST	Targets the LSD1 complex to demethylate H3K4me2 to enforce gene silencing	Khalil *et al*, 2009 Tsai *et al*, 2010
		ANRIL, Xist	PRC1	Targets PRC1 in *cis* for gene silencing. ANRIL influences p16INK4a expression and cell senescence	Yap *et al*, 2010 Bernstein *et al*, 2006
		Kcnq1ot1	PRC2	Targets PRC2 either in *cis* or *trans* to mediate H3K27 methylation and gene silencing for dosage compensation, imprinting, and developmental gene expression	Pandey *et al*, 2008
	Guide	*Xist, HOTAIR, Gtl2,*	PRC2	Targets PRC2 either in *cis* or *trans* to mediate H3K27 methylation and gene silencing for dosage compensation, imprinting, and developmental gene expression	Khalil *et al*, 2009 Zhao *et al*, 2008 Rinn *et al*, 2007 Zhao *et al*, 2010
	Enhanceur	*SRA*	CTCF	Enhances insulator function of CTCF	Yao *et al*, 2010
		pRNA	DNMT3b	Targets DNMT3b in *cis* to the rRNA locus via an RNA:DNA:DNA triplex for cytosine methylation and gene silencing	Schmitz *et al*, 2010
		HOTTIP, some eRNAs?	MLL-WDR5	Binds to and localizes the MLL complex and H3K4me3 via chromosomal looping for gene activation	Ørom *et al*, 2010 Wang *et al*, 2011
		Xist	YY1	YY1 binding nucleates Xist on the inactive X chromosome	Jeon *et al*, 2011

Tableau 6: Modèles de modes d'action de longs ARN non-codants, exemples et fonctions (Rinn and Chang, 2012).

Partie 2 : *polished-rice* (*pri*) : la métamorphose d'un lincRNA en 4 peptides

I. La découverte du lincRNA *pri*

pri a été identifié à partir de la banque d'ADNc établie par le Berkeley Drosophila Genome Project (BDGP, www.fruitfly.org, (Rubin et al., 2000) comme lincRNA par deux groupes. L'équipe de Gerald M. Rubin s'est intéressée à l'identification d'ARN non-codants polyadénylés et a trouvé *pri*, correspondant à l'ADNc LD11162 (Tupy et al., 2005). L'équipe de Yugi Kageyama qui s'intéressait elle aussi à l'identification et l'expression d'ARNs non-codant semblables à des ARNm (mRNA-like ncRNAs ou MRE) au cours de l'embryogenèse de la drosophile, a identifié *pri* comme le MRE29, exprimé dans les trachées de la drosophile au stade 9 du développement embryonnaire (Fig. 11) (Inagaki et al., 2005).

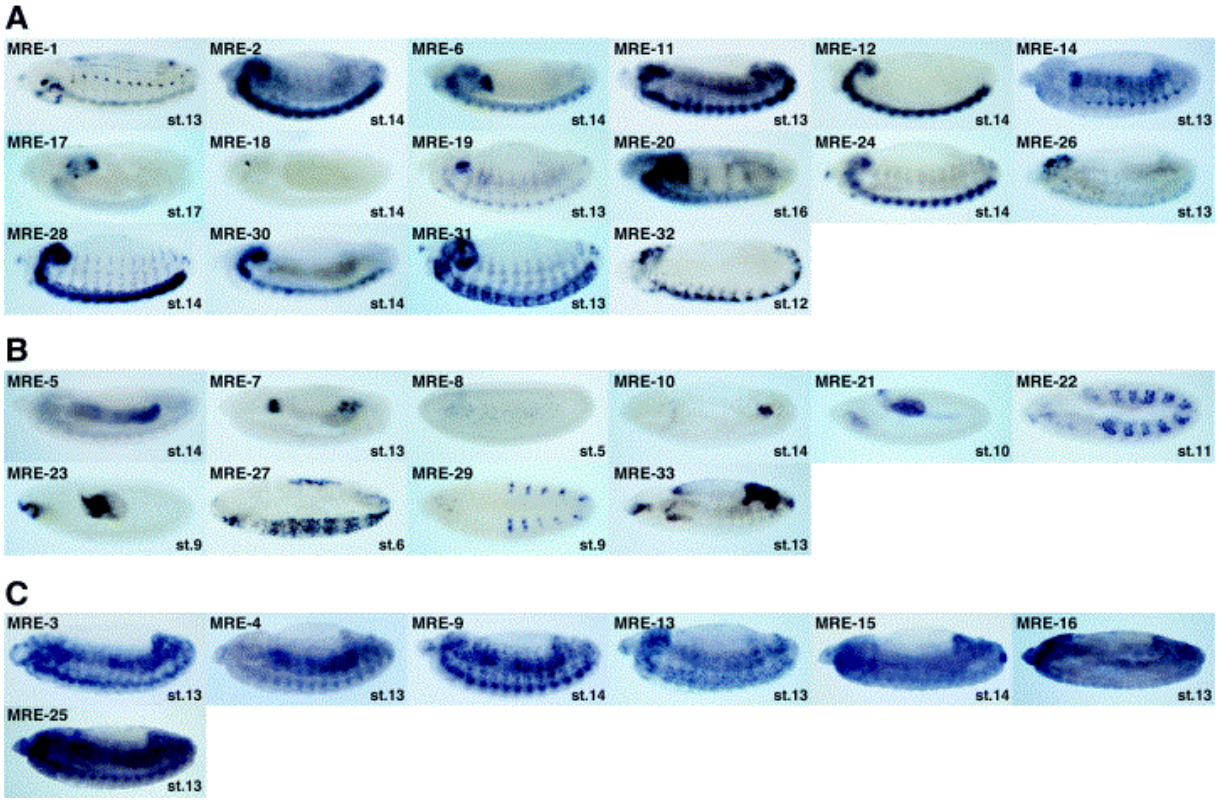


Figure 11 : 33 MRE sont exprimés au cours de l'embryogenèse. 26 des 33 MRE sont exprimés de manière tissu-spécifique. Par hybridation *in situ*, 16 ont été détectés dans le système nerveux central ou périphérique (**A**), et 10 dans d'autres organes (dont le MRE29 dans les trachées) (**B**), alors que les 7 autres sont exprimés ubiquitairement (**C**) (Inagaki et al., 2005).

II. Des patrons d'expression et des phénotypes chez les insectes

A. Chez l'embryon de drosophile

Parmi les 35 MRE identifiés dans leur étude, le groupe de Yugi Kageyama décide d'étudier plus précisément le MRE29, *pri*, car celui-ci présente un patron d'expression spécifique et dynamique au cours de l'embryogenèse de la drosophile (Fig.12).

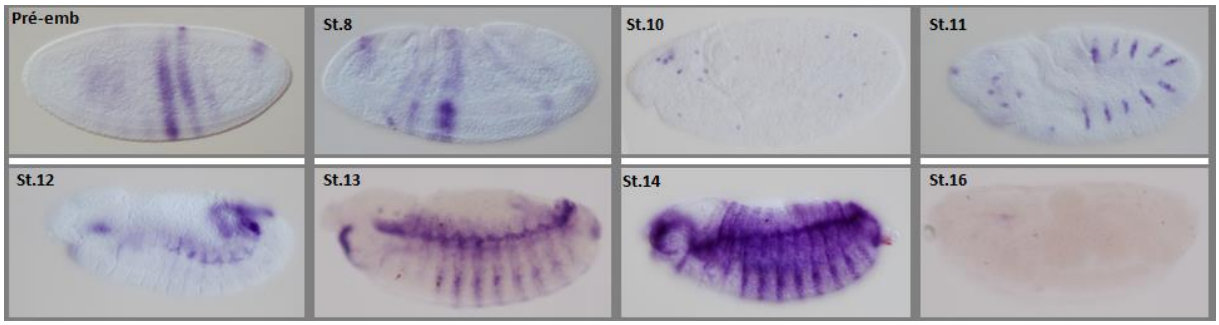


Figure 12 : *pri* a un patron d'expression dynamique au cours de l'embryogenèse. Hybridation *in situ* de *pri*. L'expression de *pri* est précoce (débute au stade pré-embryonnaire) et perdure jusqu'au stade 16 de développement. Pri est exprimé dans les trachées durant toute leur formation, et dans l'épiderme avec un profil en bandes.

Cette caractéristique suggérant un rôle important pour le développement, des mutants par délétion de la région génomique contenant *pri* sont générés. En accord avec ce rôle développemental présumé essentiel, ils meurent en fin d'embryogenèse. Par ailleurs, ils présentent deux phénotypes flagrants : 1/ une perte de l'intégrité du réseau trachéal, qui est l'appareil respiratoire de l'embryon, qui va présenter de nombreuses ruptures ainsi qu'un diamètre très irrégulier (Fig. 14) et 2/ une altération de la cuticule (l'exosquelette des insectes) qui est entièrement lisse (Fig. 13). Nous le verrons un peu plus loin dans l'introduction, la surface des embryons sauvages de drosophile est caractérisée par une alternance très spécifique de segments de cuticule lisse et de cuticule présentant des structures en forme de petits crochets (ou de poils) : les trichomes (Fig. 22). C'est l'absence de formation des trichomes qui donnera à MRE29 son nouveau nom : *polished-rice* (*pri*) puisqu'un embryon de drosophile a la forme et la couleur d'un petit grain de riz, qui dans ce cas est entièrement poli ou lissé (Kondo et al., 2007). En parallèle, deux autres groupes ont aussi identifié *pri*, mais cette fois par des approches génétiques (Savard et al., 2006; Galindo et al., 2007).

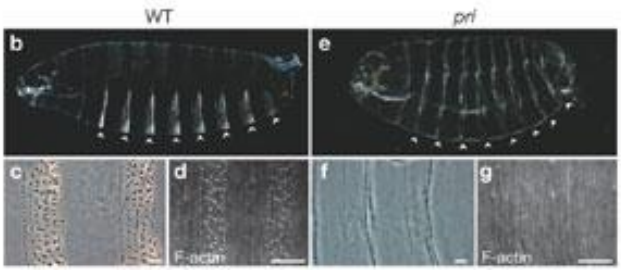


Figure 13 : Les embryons mutants *pri* présentent une altération de la cuticule. Cuticules (**b**, **c, e, f**) et marquage phalloïdine (**d, g**), d'embryons sauvages (gauche) ou mutants *pri* (droite) (Kondo et al., 2007).

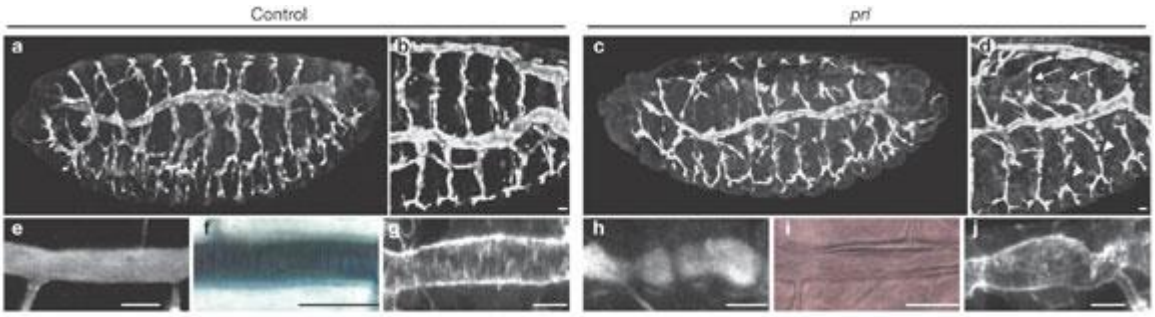


Figure 14 : Les embryons mutants *pri* présentent une altération du réseau trachéal. Les embryons mutants (gauche) présentent un réseau au diamètre régulier (**e, f, g**) et sans rupture (**a, b**), contrairement à des embryons mutants *pri* qui ont des trachées irrégulières (**h, i, j**) avec de nombreuses ruptures (**c, d**) (Kondo et al., 2007).

B. Chez la drosophile adulte

Le groupe de Juan Pablo Couso s'attache à comprendre les mécanismes sous-tendant le développement de la patte de la drosophile. Celle-ci se développe à partir d'une structure larvaire appelée le disque imaginal. Les disques imaginaux correspondent à des îlots de cellules à partir desquelles se développent les appendices de l'adulte (patte, ailes, antennes, yeux…) au cours de la métamorphose. La morphogenèse de ce disque de patte conduit à l'acquisition d'un groupe de repliements qui vont préfigurer la forme de la patte adulte. Cette morphogenèse est hautement régulée. Lors de leurs travaux, ils ont identifié un mutant qui présentait des pattes anormales. Celles-ci, au lieu de se développer en une articulation segmentée en 5 tarses (comme un doigt a 3 phalanges), vont présenter une fusion des tarses (Fig. 15). La suite de leurs travaux leur a permis d'attribuer à *pri* la responsabilité de ce phénotype, et c'est pourquoi dans cette étude il fut nommé *tarsal-less* (*tal*) (Galindo et al., 2007).

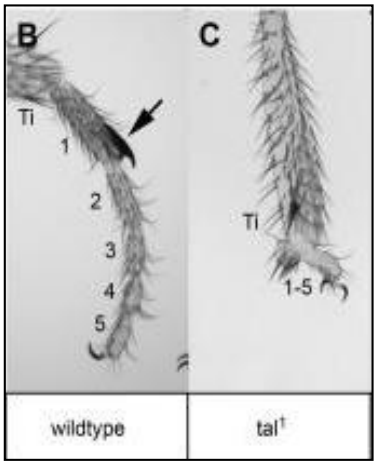


Figure 15 : Les mutants *tal* présentent des pattes anormales. B. Une patte de drosophile sauvage présente un tibia (Ti) et 5 tarses. **C.** Une patte d'un mutant *tal* la région des tarses est non-segmentée, les 5 tarses sont fusionnées (Galindo et al., 2007).

C. Chez *Tribolium castaneum*

Dans le groupe de Diethard Tautz, c'est chez un autre insecte qu'a été identifié *pri* : *Tribolium castaneum* (un petit coléoptère de farine), un organisme modèle très prisé dans la génétique évolutive du développement, notamment parce que ses mécanismes de segmentation embryonnaire sont distincts de ceux de la drosophile (ce qui fait de ces deux insectes des modèles complémentaires dans ce domaine d'étude, (Bonneton, 2010). Joël Savard, lors de sa thèse dans le groupe de Tautz, a réalisé un crible d'expression d'ESTs dans l'embryon de *Tribolium*. Il en a identifié un exprimé dans les segments abdominaux où aucun gène de segmentation n'avait jusque-là été identifié. Ce gène, lorsqu'il est muté, va induire un phénotype combiné de perte de segments abdominaux et de transformation homéotique de segments abdominaux (ne formant pas de patte) en segments thoraciques (formant des pattes). Ainsi, les adultes obtenus auront un phénotype de plus petite taille et d'un surnombre de pattes (Fig. 16), d'où le nom donné au gène *pri* dans cette étude : *mille-pattes* (*mlpt*) (Savard et al., 2006).

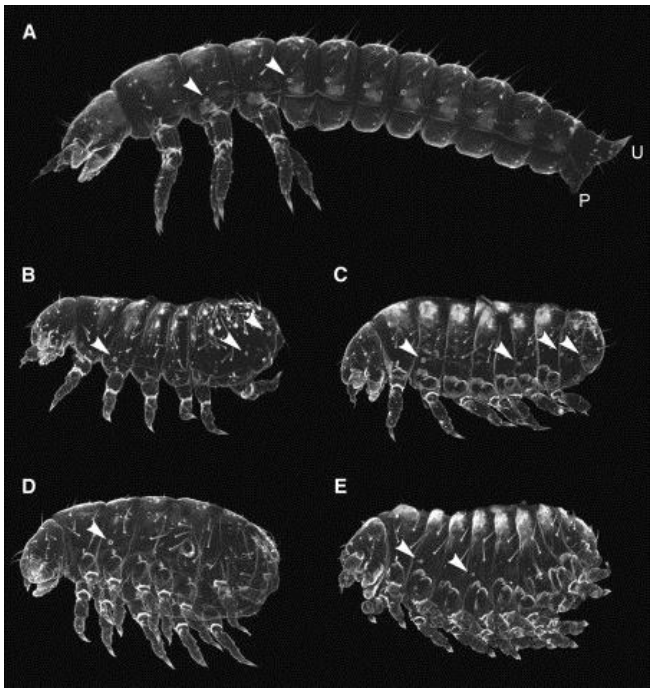


Figure 16 : Préparations de cuticules de larves de *Tribolium castaneum* sauvages ou mutantes *mlpt.* A. Larve sauvage (qui n'a que trois segments thoraciques formant des pattes). **B-E.** Série de phénotypes de plus en plus forts de mutants *mlpt*. Les larves sont plus petites et présentent des transformations homéotiques de segments abdominaux en segments thoraciques, augmentant le nombre de pattes (Savard et al., 2006).

III. Des preuves expérimentales de la traduction de ses smORFs

Ces trois découvertes quasi-simultanées démontrent l'importance du gène *pri* pour le développement des insectes, que ce soit chez la drosophile ou *Tribolium castaneum*. De plus, il a été montré que cette famille de gène *pri* est vieille d'au moins 440 millions d'années puisqu'on le retrouve aussi dans le génome de *Daphnia*, un micro-crustacé (et que les règnes des insectes et des crustacés se sont séparés à cette date) (Fig. 17).

A. *pri* est un ARN polycistronique

Très rapidement, au vu des phénotypes étonnants induits par la mutation de ce gène, les 3 groupes (Yugi Kageyama, Juan Pablo Couso et Diethard Tautz) vont se pencher de plus près sur sa séquence. Son ARN de 1,5kb était, je le rappelle, classé lincRNA sur l'absence d'ORF > 300nt, mais mRNA-like puisqu'il est polyadénylé. Et tous vont se rendre compte qu'il est polycistronique : il contient plusieurs smORFs. En effet, chez la drosophile il contient 5 smORFs de 11 (smORFs 1, 2 et 3), 32 (smORF 4) et 49 codons (smORF 5) (Fig. 17), chez *Tribolium* il en contient 4 de 10, 12, 15 et 23 codons.

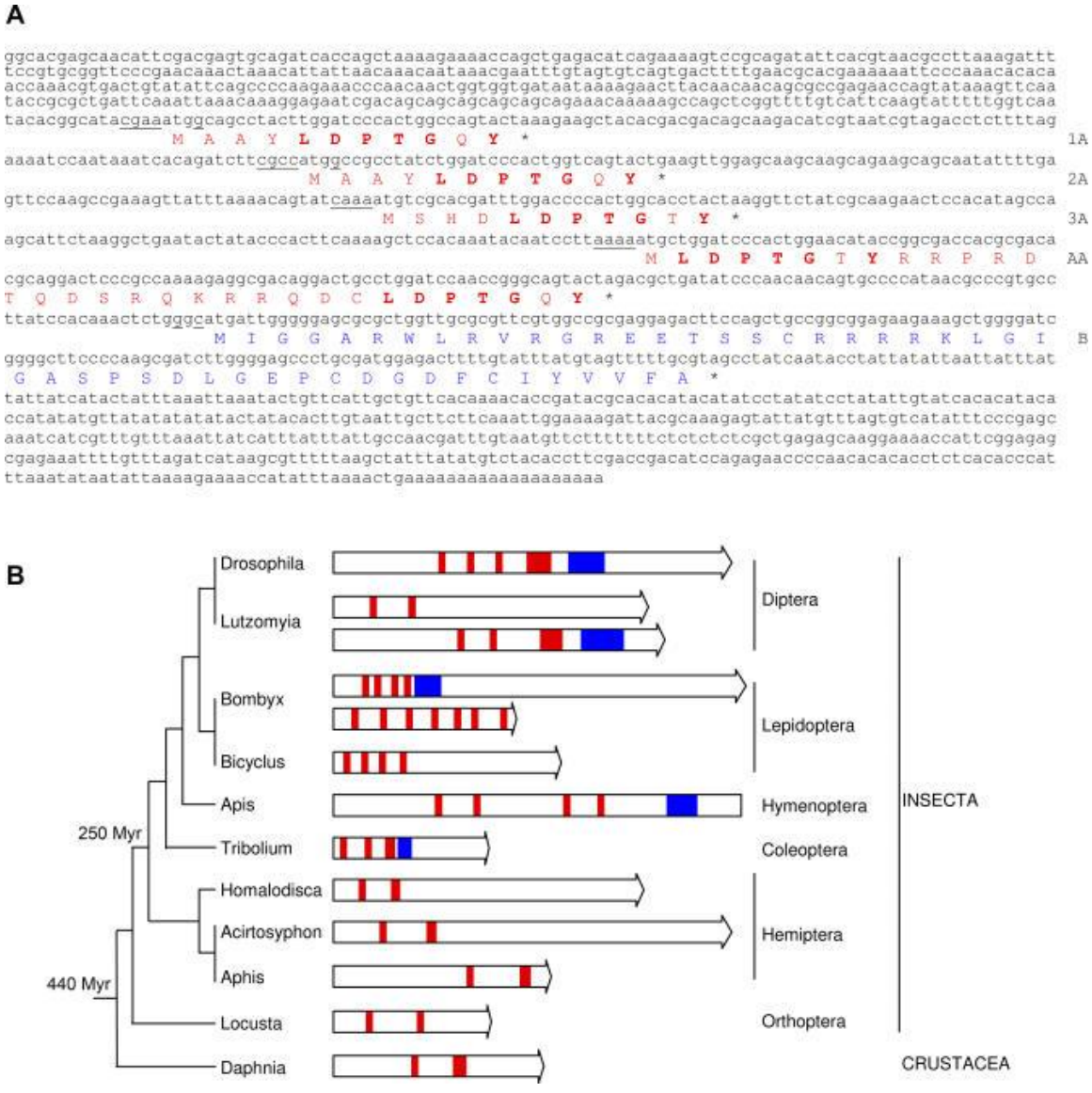


Figure 17 : Le transcrit *tal* chez la drosophile et d'autres espèces. **A.** Séquence de l'ADNc de *pri* avec la traduction conceptuelle de ses ORFs (1A, 2A, 3A, AA et B correspondent aux smORFs 1, 2, 3, 4 et 5). La séquence Kozak autour des codons d'initiation est soulignée. Les résidus conservés sont en gras. **B.** Représentation graphique de *tal* et ses homologues dans d'autres espèces, représentés par des ADNc (flèches) ou des régions génomiques (rectangles). Les boîtes rouges correspondent aux smORFs 1, 2, 3 et 4 et leurs « homologues », les boîtes bleues au smORF 5 et ses « homologues ». La famille de gènes *tal* est vieille d'au moins 440 millions d'années (Galindo et al., 2007).

Trois éléments concernant ces smORFs vont retenir leur attention : 1/ Tous ces smORFs, que ce soit chez la drosophile ou *Tribolium*, présentent un consensus Kozak (plus ou moins clair) autour de leur codon d'initiation (Fig. 17), soutenant l'idée qu'ils puissent supporter tout à fait normalement une initiation de la traduction, 2/ la séquence protéique théorique correspondant à ces smORFs présente un motif LDPTGXY (Fig. 17) retrouvé chez la drosophile une fois dans les smORFs 1, 2 et 3 et deux fois dans le smORF 4 (ici X = Q ou T, et on notera que la séquence correspondant aux smORFs 1 et 2 est identique) et chez *Tribolium* une fois dans chacun des smORFs 1, 2 et 3 (et ici X = Q ou L), 3/ la pression de conservation de ce gène est plus importante au niveau des smORFs que du reste de l'ARN, préservant ainsi la séquence protéique correspondant aux smORFs et ce surtout au niveau du motif LDPTGXY (Fig. 18). Ils vont donc naturellement se demander si la fonction de *pri* ne passerait pas par la traduction de ses smORFs et donc l'expression des sPEPs.

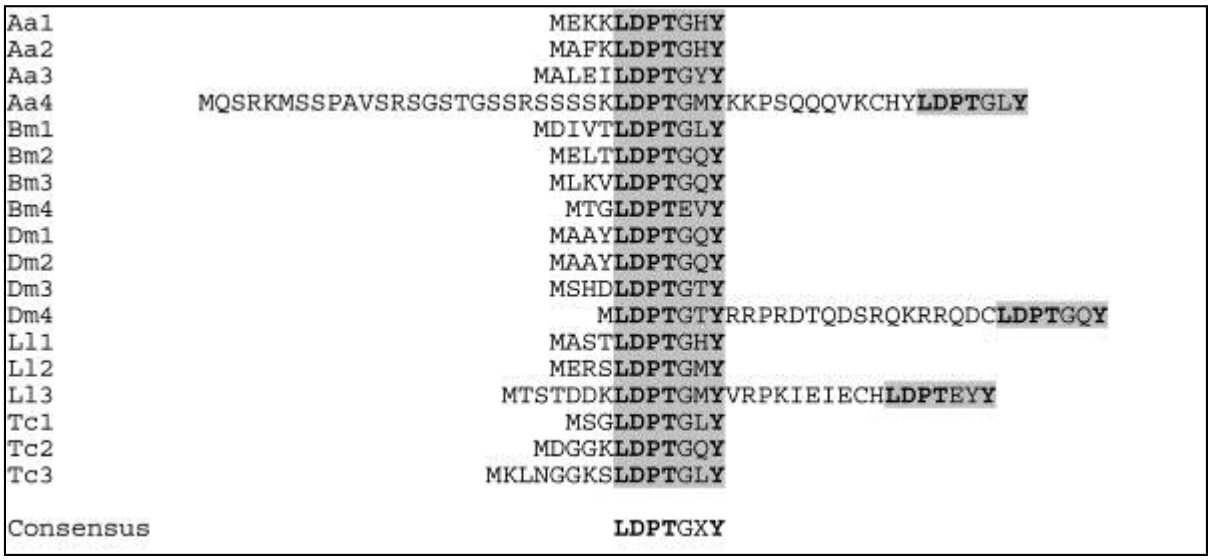


Figure 18 : Alignement des peptides potentiels *mlpt* avec ceux de plusieurs insectes. Aa, *Aedes aegypti*; Bm, *Bombyx mori*; Dm, *Drosophila melanogaster*; Ll, *Lutzomyia longipalpis*; Tc, *Tribolium castaneum.* Les numéros correspondent aux smORFs présents sur l'ARN, de gauche à droite (Savard et al., 2006).

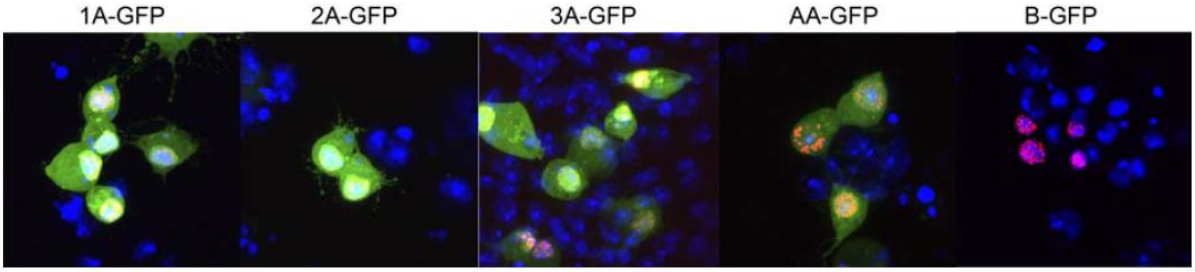
B. La fonction de *pri* requiert la traduction de ses smORFs

Les groupes de Juan Pablo Couso et de Yugi Kageyama, travaillant tous deux sur la drosophile, ont voulu répondre à cette question. Le problème est que les potentiels sPEPs (que nous appellerons désormais les peptides Pri) étant très petits, les preuves directes de leur présence feront longtemps défaut. En effet, leur purification et la production d'anticorps dirigés contre eux se sont restées, à ce jour, infructueuses.

1. Traduction des smORFs 1 à 4 en fusion avec la GFP

Mais les preuves indirectes ne manquent pas. Les deux équipes ont réalisé des transgènes de *pri* dans lesquels la séquence codant la Green Fluorescent Protein (GFP), sans son codon initiateur, est fusionnée alternativement en 3' de chacun des 5 smORFs. Leurs résultats sont similaires : après transfection en cellules S2 (lignée cellulaire de drosophile), un signal GFP est observable pour les fusions avec les smORFs 1, 2, 3 et 4, tandis qu'aucun signal n'est détectable avec le smORF 5. Ceci démontre que les 4 premiers smORFs (contenant le motif LDPTGXY) sont traduits, et que le cinquième (dépourvu de ce motif) ne l'est pas (Fig.19).

Figure 19 : expression de différentes constructions où les smORFs 1, 2, 3, 4 (AA) et 5 (B) sont en phase avec la séquence codant la GFP dans des cellules S2R+. Vert : GFP, Bleu : marquage nucléaire au DAPI, Rouge : DsRed nucléaire, marqueur de transfection (Galindo et al., 2007).



2. L'intégrité d'au moins un des smORFs 1 à 4 est requise pour la fonction génétique de *pri*

Par ailleurs, ils vont tester l'hypothèse de l'importance des smORFs de *pri* pour sa fonction en génétique, par des expériences de sauvetage de phénotype mutant de *pri*. L'équipe de Yugi Kageyama s'est intéressée au sauvetage des phénotypes embryonnaires (trachéaux et épidermiques). Ils ont pu démontrer que la réexpression seule des smORFs 1 (et 2 puisque ce sont les mêmes), 3 ou 4 est suffisante pour restaurer un phénotype sauvage (au même titre que la réexpression de l'ADNc de *pri*). En revanche ils montrent que l'ADNc complet de *pri* dans lequel des décalages de phase de lecture (frame-shift) ont été introduits dans chacun des 4 smORFs (Pri1-4FS, laissant le smORF 5 intègre), une version qui produira des peptides aux séquences différentes, n'est pas capable de sauver les phénotypes (Fig. 20).

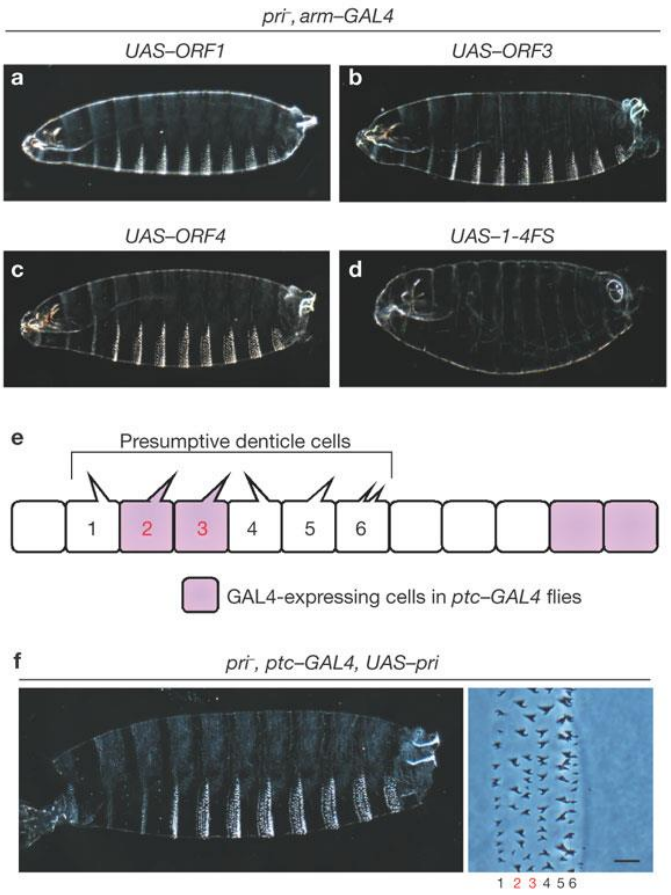


Figure 20 : Sauvetage du phénotype *polished-rice* de mutants *pri* par réexpression des smORFs 1, 3 et 4. Le phénotype de perte de trichome de mutants *pri* est sauvé par les *UAS-ORF1* (**a**), *UAS-ORF3* (**b**), ou *UAS-ORF4* (**c**) dont l'expression est dirigée par *arm-Gal4* (*arm* est un driver ubiquitaire), mais pas par *UAS-1-4FS*, une version dans laquelle ont été introduits des décalages de phase de lecture dans les 4 premiers smORFs (**d**). L'expression ectopique de *pri* dans les cellules exprimant *ptc* (**e**) sauve complètement le phénotype d'absence de trichome de mutants *pri* (suggérant une fonction non-cellulaire autonome des peptides Pri) (Kondo et al., 2007).

Le groupe de Juan Pablo Couso a réalisé le même type d'approche, et regardé le sauvetage à la fois des phénotypes embryonnaires et adultes (pattes sans tarse). De leur côté ils ont aussi montré que l'intégrité des smORFs 1, 2, 3 ou 4 (et non du smORF 5) est nécessaire pour restaurer les phénotypes embryonnaires et adultes. De plus, ils ont montré que ces phénotypes peuvent être sauvés par la réexpression de l'ADNc de *pri* issu d'une autre espèce (*Bombyx mori*) (Fig. 21).

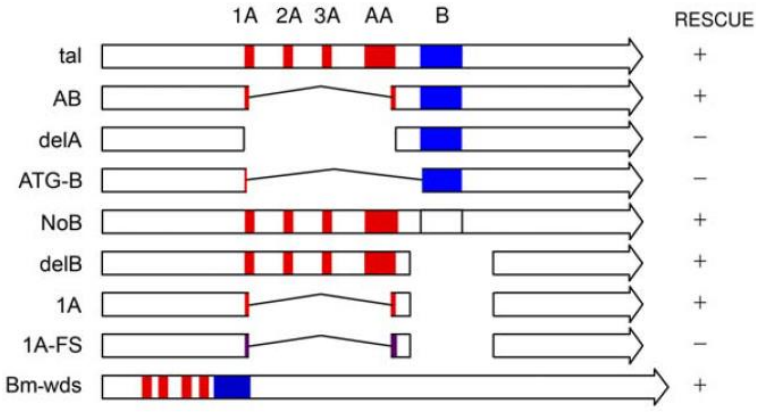


Figure 21 : Schéma des constructions utilisées pour les sauvetages phénotypiques embryonnaires et adultes de mutants *tal*. L'ADNc de *tal* sauve les phénotypes, comme les constructions contenant au moins un des 4 premiers smORFs intègre. Les smORF 5 et 1A-FS (smORF 1 avec un décalage de phase de lecture) ne sont pas capables de sauver les phénotypes. En revanche, l'ADNc de *tal* de *Bombyx mori* (Bm-wds) reproduit les résultats obtenus avec celui de la drosophile (Galindo et al., 2007).

C. *pri* exprime des peptides Pri

Ces données démontrent : 1/ que la fonction du gène *pri* passe par la traduction en peptides Pri à partir des 4 smORFs 1, 2, 3 et 4, dont la taille varie de 11 à 32 codons, 2/ que ces 4 peptides présentent une redondance fonctionnelle puisque l'expression d'un seul des 4 est suffisante pour restaurer un phénotype sauvage chez un mutant *pri*, 3/ qu'au moins pour les phénotypes considérés, l'ARNm de *pri* ne joue aucun rôle puisque l'ADNc de *pri* d'une autre espèce (produisant des peptides similaires mais dont les régions non-codantes ne sont pas conservées chez la drosophile) est capable de sauver les phénotypes, tandis que la réexpression de la version 1-4FS (qui ne diffère de l'ADNc sauvage que par 8 nt sur ses 1500nt) n'en est pas capable.

IV. L'absence de trichome embryonnaire : un phénotype bien connu de l'équipe

Dans cette partie, je vais expliquer pourquoi notre équipe s'est intéressée à ce gène, et finalement comment mon projet de thèse (débuté en master) était en adéquation avec « l'ère-peptidique », puisqu'il vise à décrypter les mécanismes d'action de sPEPs dont le rôle essentiel au cours du développement embryonnaire était déjà établi. Cette histoire a commencé grâce au phénotype « *polished-rice* » : les embryons mutants pour *pri* ne présentent pas de trichome épidermique.

A. La formation des trichomes épidermiques : modèle d'étude de la morphogenèse cellulaire

L'épiderme embryonnaire est composé d'une monocouche de cellules post-mitotiques entourant l'embryon. Ce sont ces cellules qui vont sécréter les nombreux composants de la cuticule. Cette structure, aussi appelée exosquelette, va constituer pour la larve qui va éclore une véritable armure. En effet, la cuticule joue le rôle de protection mécanique, de défense contre le milieu extérieur (attaque de micro-organismes), de lutte contre la déshydratation et participe aussi à la locomotion de la larve.

Comme décrit plus tôt, une caractéristique de l'embryon (en fin d'embryogenèse) de drosophile est qu'il présente un patron segmenté très stéréotypé d'alternance de cuticule lisse et de cuticule présentant des trichomes (Fig. 22). Ces petites structures en forme de poils retranscrivent la forme des cellules qui les ont sécrétées. En effet, jusqu'au 13ème stade de développement, l'épiderme des embryons est homogène et lisse. Mais en fin d'embryogenèse,

certaines cellules épidermiques vont subir une étape de différenciation terminale qui correspond à un important réarrangement du cytosquelette d'actine à leur pôle apical, qui va donner lieu à une extension en forme de crochet. Lorsque la cuticule sera sécrétée par ces cellules, elle prendra cette forme caractéristique, alors que la cuticule sécrétée par les cellules restées lisses sera lisse.

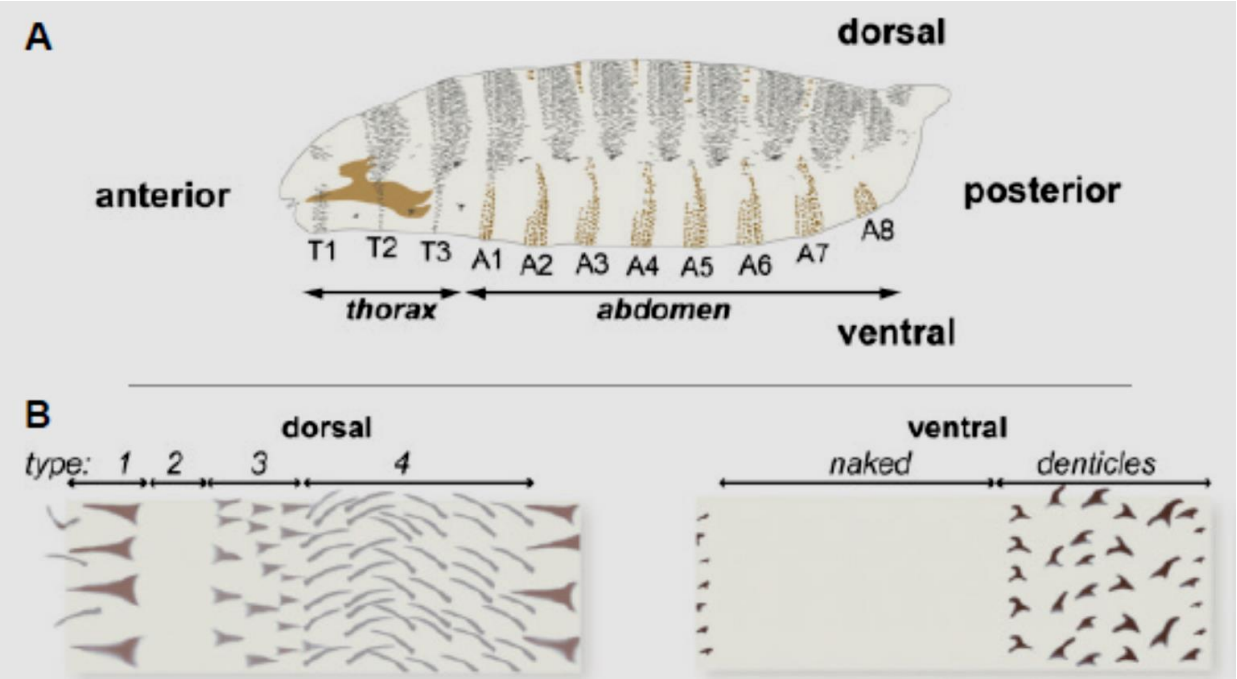


Figure 22 : Patron des extensions de la cuticule de la larve de drosophile. A. Dessin schématisant l'organisation des extensions observables sur la cuticule d'une jeune larve (1er stade larvaire) de drosophile. **B.** Détails des cuticules dorsales et ventrales correspondant au quatrième segment abdominal (A4). La région dorsale se différencie en 4 types de cuticules (1-4), la région ventrale en deux types : la cuticule lisse et la cuticule portant des trichomes (ou denticules) (Payre, 2004).

Notre équipe s'intéresse aux mécanismes de régulation de la morphogenèse cellulaire au cours du développement, et notre modèle d'étude est ce changement de forme des cellules épidermiques. Lors de leur crible génétique, Christiane Nüsslein-Volhard et Eric Wieschaüs ont identifié une mutation pour laquelle un phénotype de cuticule totalement lisse était observable : ils ont nommé cette mutation *shavenbaby* (*svb*, bébé rasé). J'en profite pour noter que c'est une parfaite illustration de la difficulté d'identifier des smORFs par une approche de crible génétique en mutagenèse aléatoire : lors de ce crible, pour un même phénotype, *pri* n'a pas été identifié. D'une part la probabilité d'obtenir une mutation sur une région de 33nt est très faible, d'autre part les 4 smORFs de *pri* présentant une redondance fonctionnelle, même si une mutation avait eu lieu dans un des 4 smORFs, aucun phénotype n'aurait été observé.

B. Shavenbaby : le facteur de transcription à la tête de ce programme de différenciation terminale

Shavenbaby est un facteur de transcription à doigts de zinc (Mével-Ninio et al., 1995), et un des travaux fondateurs de l'équipe a été de montrer que son expression est nécessaire et suffisante à la formation des trichomes (Payre et al., 1999) (Fig. 23). Ainsi, le patron d'expression de Svb préfigure l'implantation des trichomes (Delon et al., 2003; Payre, 2004). Ces dernières années, divers travaux effectués dans l'équipe ont cherché à comprendre comment Svb contrôle le remodelage localisé de la forme des cellules épidermiques et ont permis d'identifier les gènes cibles de Svb (Chanut-Delalande et al., 2006; Fernandes et al., 2010; Menoret et al., 2013). Par analyses transcriptomiques, Delphine Ménoret, lors de sa thèse, a pu définir qu'ils sont environ 150. La réalisation d'immuno-précipitation de chromatine avec un anticorps dirigé contre Svb, ainsi que la validation des séquences fixées comme séquence « enhancer » par transgénèse, a permis de définir que la plupart des gènes régulés sont des cibles directes. Même si pour beaucoup d'entre eux le rôle dans la morphogenèse épidermique reste inconnu, on retrouve entre autres de nombreux acteurs de l'organisation de l'actine, de la mise en place de la matrice extracellulaire (MEC), de la formation et pigmentation de la cuticule (plus épaisse et plus pigmentée au niveau des trichomes), du trafic intracellulaire, qui sont autant de processus impliqués dans la différenciation terminale des trichomes. Ainsi, Svb contrôle de multiples effecteurs aux fonctions variées.

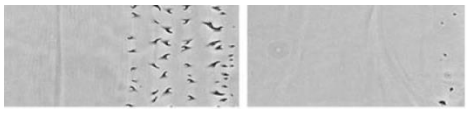


Figure 23 : Svb est requis pour la morphogenèse épidermique embryonnaire. Cuticule d'embryon sauvage (gauche) ou mutant *svb* (droite) (Salles et al., 2002).

Le phénotype d'absence de trichome observé chez les mutants *pri* nous est donc très familier, puisque c'est le cas pour les embryons mutants pour *svb*. Par ailleurs, Isabelle Delon a démontré que Svb est aussi responsable de la morphogenèse épidermique de tissus adultes comme le thorax (Fig. 24) ou l'aile (Fig. 27) (Delon et al., 2003), qui de la même manière que les embryons ne formeront pas d'extension en absence de Svb.

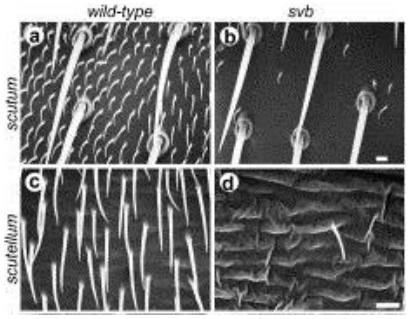


Figure 24 : Svb est requis pour la morphogenèse épidermique adulte. Micrographes électroniques de l'épiderme de drosophiles adultes sauvages (**a**, **c**), présentant de nombreux trichomes, ou mutantes pour *svb* (**b**, **d**), où de grandes zones de cuticule lisse sont observables, dans le scutum (**a**, **b**) et le scutellum (**c**, **d**) (deux régions du thorax) (Delon et al., 2003).

L'étude de la fonction de certains effecteurs (Chanut-Delalande et al., 2006; Fernandes et al., 2010) a révélé que l'inactivation individuelle de cibles de Svb n'affecte que la forme du trichome, pas sa présence (Fig. 25), confortant un rôle instructeur de Svb, comme chef d'orchestre de la morphogenèse, coordonnant la co-expression d'une batterie de gènes. Ainsi, jusqu'à la découverte de *pri*, *svb* était le seul gène connu pour lequel un tel phénotype était observable et c'est pourquoi on le considérait comme le « *master-gene* » de ce processus.

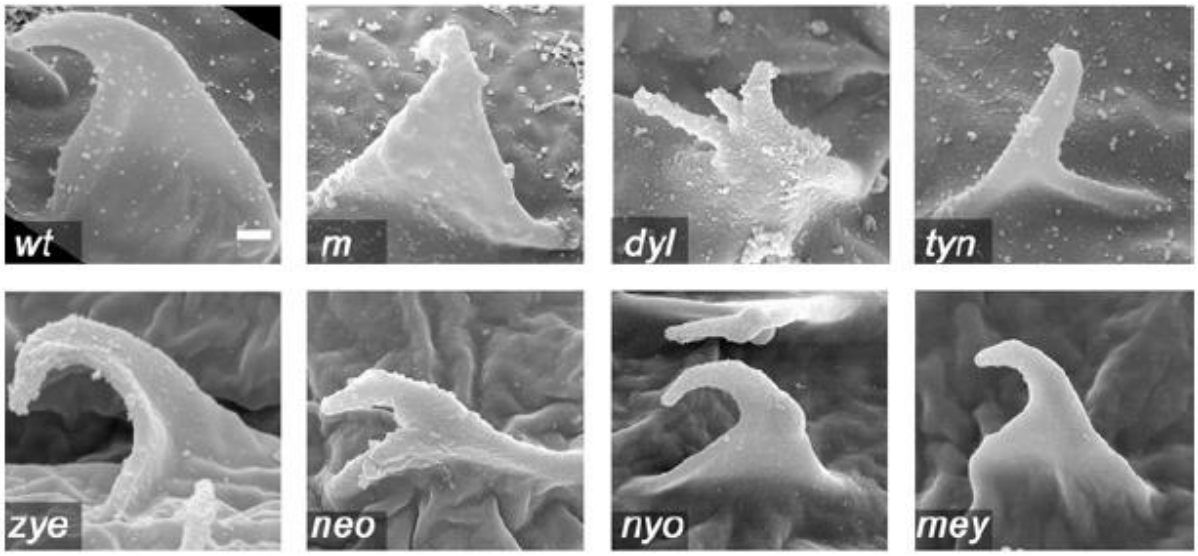


Figure 25 : Les mutants de cibles de Svb présentent des défauts spécifiques de forme du trichome. Micrographes électroniques de denticules d'embryon sauvage (*wt*), et d'embryons mutants pour des cibles de Svb (*m*, *dyl*, *tyn*, *zye*, *neo*, *nyo*, *mey*), présentant des altérations spécifiques de la forme des denticules (jamais leur présence) (Fernandes et al., 2010).

C. OvoA et OvoB : deux isoformes naturelles et une boîte à outils génétiques

Dans cette partie, je vais vous présenter les deux isoformes naturelles de Svb qui par le passé se sont avérées être de précieux outils pour décrypter la fonction de Svb, et qui vous le verrez par la suite, dans le cadre de ce projet nous ont aussi été d'une grande aide.

Svb est exprimé à partir du locus *ovo/svb* (Fig. 26). Ce locus a aussi été identifié comme dirigeant la différenciation de la lignée germinale chez les femelles adultes (Mével-Ninio et al., 1991). En effet, deux promoteurs alternatifs germinaux (Mével-Ninio et al., 1995) dirigent l'expression de deux isoformes protéiques de Svb : OvoA et OvoB (Fig. 26). Les caractérisations moléculaire et fonctionnelle de ces trois isoformes ont permis de définir 1/ qu'elles sont largement recouvrantes, et partagent la région C-terminale, contenant le domaine de liaison à l'ADN et un domaine activateur et diffèrent par l'étendue de leur région N-terminale 2/ qu'OvoB, le plus petit des 3 est un activateur transcriptionnel, 3/ qu'OvoA contient un domaine N-terminal supplémentaire, transformant l'activité de ce facteur en

répresseur de transcription (permettant ainsi de définir un domaine de répression) (Mével-Ninio et al., 1995; Andrews et al., 2000; Delon et al., 2003) (Fig. 26).

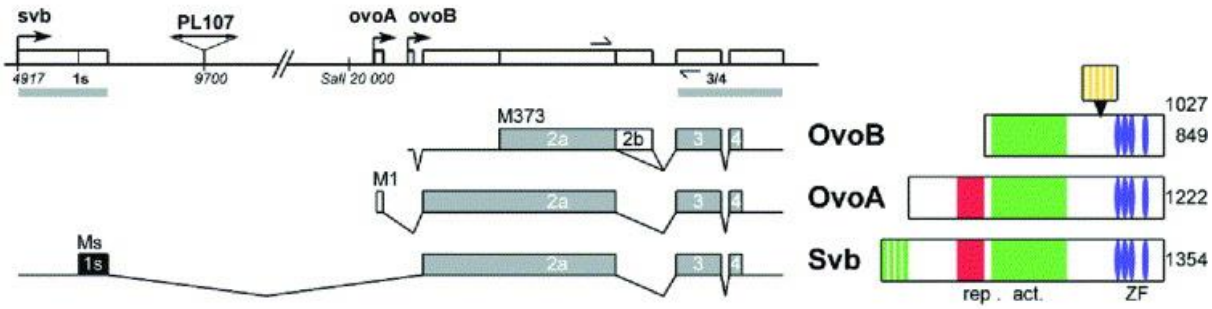


Figure 26 : Caractérisation du locus *ovo/svb* et des trois isoformes produites OvoB, OvoA et Svb. Organisation moléculaire du locus avec les promoteurs somatiques (*svb*) et germinaux (*ovo*). Les transcrits *ovo*/*svb* sont dessinés avec les exons communs en gris, et les exons spécifiques de *svb* et *d'ovo* en noir et blanc respectivement. Les régions protéiques contenant les domaines d'activation ou de répression sont représentés en vert et rouge respectivement. Le domaine de liaison à l'ADN est composé de quatre motifs à doigts de zinc (Zinc Finger, ZF, en bleu). Les tailles des protéines en acides-aminés sont indiquées à droite (Delon et al., 2003).

L'expression ectopique d'OvoA dans un domaine où Svb induit la formation de trichomes engendre une perte de ceux-ci (Fig.27) (Delon et al., 2003), le définissant comme dominant négatif de Svb. Par ailleurs, l'expression ectopique d'OvoB est capable de sauver le phénotype de mutants *svb* (Salles et al., 2002) et l'expression d'OvoB ou de Svb dans un domaine normalement lisse induit la formation de trichomes (Fig.27) (Delon et al., 2003). Ces données montrent que c'est la fonction d'activateur transcriptionnel de Svb qui est requise pour l'expression de ses gènes cibles et la formation des trichomes (Delon et al., 2003). Pourtant, Svb possède un domaine N-terminal supplémentaire par rapport à OvoA, exprimé à partir de l'exon1S (Delon et al., 2003). Il comporte donc, à l'instar d'OvoA, les domaines de répression et d'activation, et devrait selon toute logique se comporter en répresseur de la transcription. Pourtant, les données génétiques sont claires : c'est la fonction d'activateur transcriptionnel de Svb qui est requise pour la formation des trichomes. Une interprétation de ces données était que la partie N-terminale de Svb devait masquer la fonction du domaine répresseur.

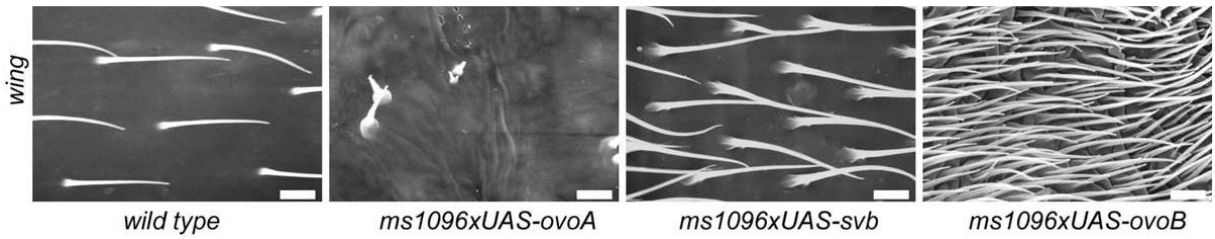


Figure 27 : Expression ectopique des isoformes Ovo et Svb dans l'aile de la drosophile. Micrographes électroniques d'aile sauvage (*wild-type*), ou d'ailes dans lesquelles sont exprimés ectopiquement (avec le driver *ms1096-GAL4*) *ovoA*, *svb* ou *ovoB*. L'expression *d'ovoA* induit une perte des trichomes, tandis que les expressions de *svb* et *d'ovoB* induisent la formation de trichomes surnuméraires (phénotype plus marqué avec *ovoB*) (Delon et al., 2003).

D. Quelle-est la place des peptides Pri dans ce scénario ?

Revenons au gène *pri*. Au vu du phénotype *polished-rice*, il était naturel d'envisager que l'expression de Svb puisse être altérée dans un contexte mutant pour *pri*. C'est ce qu'a vérifié le groupe de Yugi Kageyama : par hybridation *in situ* ils ont analysé l'expression de *svb* dans un embryon mutant pour *pri*, et constaté que celle-ci est tout à fait normale. Réciproquement, la mutation de *svb* n'engendre pas de différence de niveau d'expression de *pri*. (Fig. 28). Ces résultats situent Pri et Svb au même niveau : ils sont tous les deux requis pour la fonction d'activateur transcriptionnel de Svb dans la formation du trichome épidermique (Fig. 29).

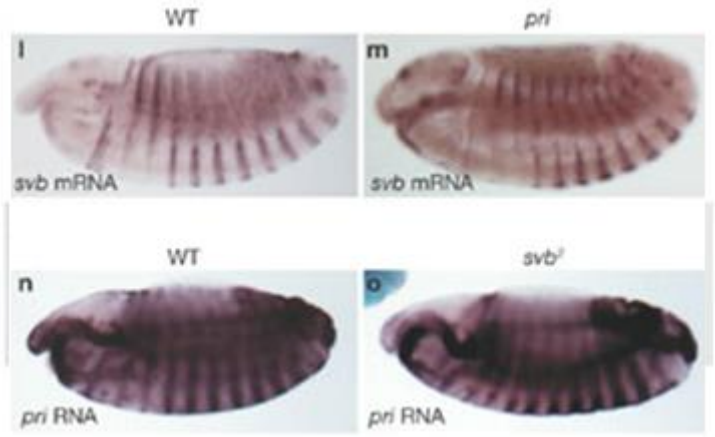


Figure 28: Les expressions de *pri* et *svb* ne dépendent pas l'une de l'autre. Hybridations *in situ* pour révéler soit l'ARN de *svb* dans des embryons sauvage (**l**) ou mutant pour *pri* (**m**), soit l'ARN de *pri* dans des embryons sauvage (**n**) ou mutant pour *svb* (**o**). Dans les deux cas, aucune différence d'expression des transcrits n'est observable entre les génotypes sauvages ou mutants (Kondo et al., 2007).

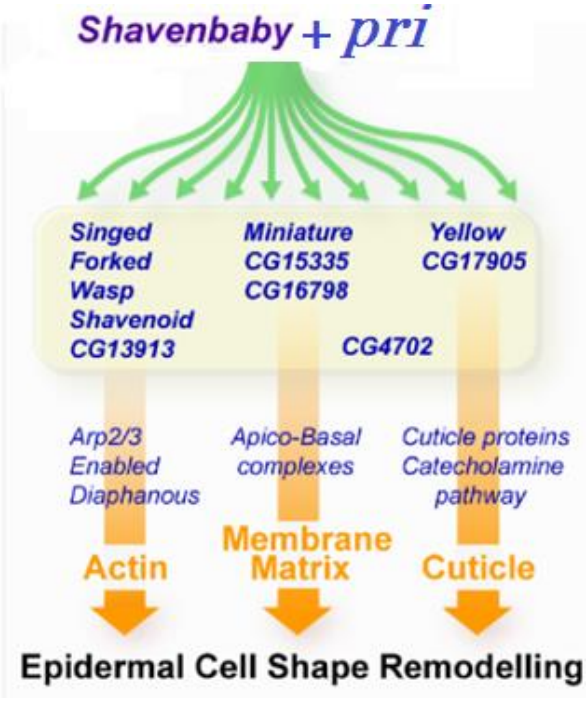


Figure 29 : Schéma du rôle instructeur de Svb dans le remodelage des cellules épidermiques, dépendant de *pri*. Ce schéma, sans *pri*, a été établi en 2006 (Chanut-Delalande et al., 2006). Svb, en présence de *pri*, active l'expression de gènes cibles impliqués dans le remodelage du cytosquelette d'actine, la liaison entre les membrane cellulaire et matrice extracellulaire, et les sécrétion et pigmentation de la cuticule. L'action concertée de tous ces effecteurs conduit au remodelage apical de la cellule épidermique dans laquelle Svb est exprimé.

Mes travaux de master et de thèse ont eu pour objectif de comprendre comment les peptides Pri régulent l'activité transcriptionnelle de Svb afin que celui-ci puisse instruire son programme génétique de différenciation des cellules épidermiques. Les résultats obtenus sont présentés en trois parties.

La première partie reprend les travaux qui ont été faits dans l'équipe, en collaboration avec le groupe de Yugi Kageyama, durant mon année de master. Les résultats obtenus ont été publiés et l'article est inséré dans ce manuscrit. La partie expérimentale que j'ai développée pour ce projet est détaillée à la suite. Nous avons pu montrer que les peptides Pri induisent une maturation post-traductionnelle de Svb induisant un changement de son activité transcriptionnelle. Les différentes hypothèses alors envisagées sur la nature de ce mécanisme sont référencées en discussion de cette première section.

La deuxième partie est présentée sous forme d'article en fin de préparation, qui reprend la quasi-totalité de mes travaux de thèse réalisés en collaboration avec Jennifer Zanet, en post-doctorat dans l'équipe. Ils concernent l'identification des séquences *cis*-régulatrices et des facteurs *trans*-régulateurs de la maturation de Svb en réponse aux peptides Pri.

La troisième partie regroupe, sous forme de discussion, des résultats additionnels qui ont été réalisés pour tester diverses hypothèses.

RESULTATS

Partie 1 : Les peptides Pri : un interrupteur de l'activité transcriptionnelle de Svb.

I. Résumé

Dans ces travaux, auxquels j'ai participé pendant mon stage de master, nous avons montré que les peptides Pri sont requis pour l'expression des cibles de Svb. Nous montrons de plus qu'il existe un parallèle fort entre Svb en absence et en présence de Pri, et ses isoformes OvoA et OvoB, respectivement, à la fois en cellules en culture S2 et *in vivo*.

En effet, des tests d'activité transcriptionnelle en cellules S2 montrent que Svb se comporte en répresseur (comme OvoA), et qu'en réponse à Pri il devient activateur (comme OvoB). Des analyses en Western blot montrent que ce changement d'activité est accompagné d'un changement de taille de Svb, d'une forme longue plus grande qu'OvoA (comme présenté en introduction, Fig. 26) à une forme courte dont le poids moléculaire est proche de celui d'OvoB. En revanche, ces deux approches montrent que les isoformes OvoA et B sont insensibles à Pri. Nous avons par ailleurs observé qu'il existe une corrélation entre l'activité transcriptionnelle de ces facteurs et leur localisation sub-nucléaire : les répresseurs (OvoA et Svb sans Pri) sont situés dans des foyers nucléaires, les activateurs (OvoB et Svb avec Pri) présentent une localisation diffuse et homogène dans le nucléoplasme.

Nous avons pu démontrer que la forme courte et activatrice de Svb est issue de la maturation Pri-dépendante de la forme longue et répresseur : 1/ Le séquençage N-terminal de la forme courte de Svb nous a permis de la caractériser, et montre que son premier résidu ne correspond pas à un codon d'initiation (même non-canonique). Cette forme tronquée en N-terminal ne possède plus le domaine de répression transcriptionnelle commun avec OvoA, 2/ Le suivi de la localisation de la population de Svb exprimée AVANT l'expression des peptides Pri montre que c'est la même protéine qui se relocalise dans le noyau, excluant l'hypothèse d'une néo-synthèse protéique. C'est avec cette analyse que j'ai contribué à ces travaux, et la stratégie mise en œuvre est explicitée à la suite de l'article dans lequel a été publiée cette étude.

En conclusion, ces travaux montrent que les peptides Pri induisent un changement de l'activité transcriptionnelle de Svb : celui-ci, exprimé sous sa forme longue en tant que

répresseur, subit en réponse à Pri l'élimination post-traductionnelle de sa région N-terminale, le convertissant ainsi en activateur de la transcription. Svb est alors capable d'induire l'expression de ses gènes cibles, dirigeant ainsi le programme de différenciation terminale des cellules de l'épiderme embryonnaire : la formation des trichomes.

II. Article

A. Small Peptides Switch the Transcriptional Activity of Shavenbaby During Drosophila Embryogenesis

Small Peptides Switch the Transcriptional Activity of Shavenbaby During Drosophila Embryogenesis
T. Kondo, *et al.*
Science **329**, 336 (2010);
DOI: 10.1126/science.1188158

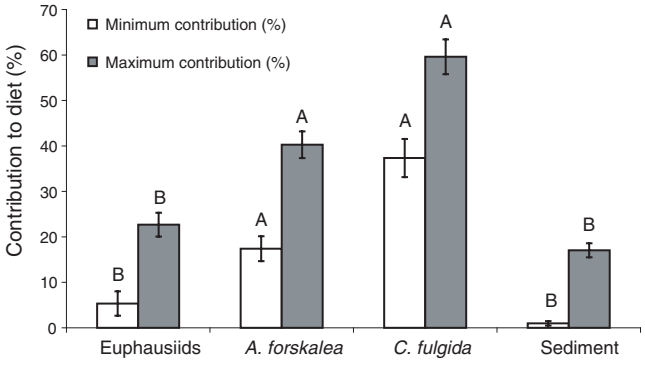


Fig. 3. Isotope analyses showing maximum and minimum contribution of euphausiids, jellyfish, and sediment to the diet of *S. bibarbatus*, as determined by a four-endpoint Isosource model (based on carbon and nitrogen) (*19*). Bars denote mean ± SEM. Bars with identical letters were not significantly different from each other at an α level of 5% (Tukey test, $P < 0.05$; *A. forskalea*, $n = 11$; euphausiids, $n = 6$; *C. fulgida* $n = 25$; mud, $n = 5$ samples; *S. bibarbatus*, $n = 41$).

and ecological adaptations, is already playing a critical role in the ecosystem off Namibia, and it is likely to continue to do so into the future.

References and Notes

1. M. E. Carr, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **49**, 59 (2001).
2. C. P. Lynam *et al.*, *Curr. Biol.* **16**, 1976 (2006).
3. P. Cury, L. Shannon, *Prog. Oceanogr.* **60**, 223 (2004).
4. A. Bakun, *Science* **247**, 198 (1990).
5. C. D. van der Lingen *et al.*, in *Benguela: Predicting a Large Marine Ecosystem. Large Marine Ecosystems 14*, L. V. Shannon, G. Hempel, P. Malanotte-Rizzoli, C. L. Moloney, J. Woods, Eds. (Elsevier, Amsterdam, 2006), pp. 147–184.
6. A. J. Richardson, A. Bakun, G. C. Hays, M. J. Gibbons, *Trends Ecol. Evol.* **24**, 312 (2009).
7. J. Yamamoto *et al.*, *Mar. Biol.* **153**, 311 (2008).
8. R. J. M. Crawford, L. V. Shannon, D. E. Pollock, *Oceanogr. Mar. Biol. Ann. Rev.* **25**, 353 (1987).
9. D. J. Boyer, I. Hampton, *S. Afr. J. Mar. Sci.* **23**, 5 (2001).
10. V. Brüchert, B. Currie, K. R. Peard, *Prog. Oceanogr.* **83**, 169 (2009).
11. V. Brüchert *et al.*, in *Past and Present Water Column Anoxia*, L. N. Neretin, Ed. (Springer, Netherlands, 2006), pp. 161–194.
12. G. Lavik *et al.*, *Nature* **457**, 581 (2009).
13. H. N. Schulz *et al.*, *Science* **284**, 493 (1999).
14. S. J. Weeks, B. Currie, A. Bakun, *Nature* **415**, 493 (2002).
15. S. J. Weeks, B. Currie, A. Bakun, K. R. Peard, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **51**, 153 (2004).
16. J. Rogers, J. M. Bremner, *Oceanogr. Mar. Biol.* **29**, 1 (1991).
17. A. Bakun, S. J. Weeks, *Ecol. Lett.* **7**, 1015 (2004).
18. P. Fréon, M. Barange, J. Arístegui, *Prog. Oceanogr.* **83**, 1 (2009).
19. Materials and methods are available as supporting material on *Science* Online.
20. M. J. Gibbons *et al.*, *S. Afr. J. Mar. Sci.* **22**, 1 (2000).
21. T. Bagarinao, *Aquat. Toxicol.* **24**, 21 (1992).
22. F. J. Millero, T. Plese, M. Fernandez, *Limnol. Oceanogr.* **33**, 269 (1988).
23. T. Bagarinao, R. D. Vetter, *J. Comp. Physiol. B* **160**, 519 (1990).
24. R. Vaquer-Sunyer, C. M. Duarte, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 15452 (2008).
25. P. Domenici, C. Lefrançois, A. Shingles, *Philos. Trans. R. Soc. London Ser. B* **362**, 2105 (2007).
26. N. Matikainen, M. Vornanen, *J. Exp. Biol.* **167**, 203 (1992).
27. S. Kaartvedt, A. Røstad, T. A. Klevjer, *Mar. Ecol. Prog. Ser.* **390**, 237 (2009).
28. E. S. Todd, *Copeia* **2**, 374 (1976).
29. H. Pang, K. N. Bitar, *Am. J. Physiol. Cell Physiol.* **289**, C982 (2005).
30. D. Pauly, V. Christensen, V. J. Dalsgaard, R. Froese, F. Torres Jr., *Science* **279**, 860 (1998).
31. J. J. Heymans, L. J. Shannon, A. Jarre, *Ecol. Model.* **172**, 175 (2004).
32. P. M. S. Monteiro, A. K. van der Plas, J.-L. Melice, P. Florenchie, *Deep Sea Res. Part I Oceanogr. Res. Pap.* **55**, 435 (2008).
33. H. Hamukuaya, M. J. O'Toole, P. M. J. Woodhead, *S. Afr. J. Mar. Sci.* **19**, 57 (1998).
34. R. J. Diaz, R. Rosenberg, *Science* **321**, 926 (2008).
35. C. L. Prosser, F. A. Brown, *Comparative Animal Physiology* (Saunders, Philadelphia, 1961).

36. We thank the crew of the *G. O. Sars*; F. Midtøy for assistance; and P. Ellitson, M. Hordnes, R. Jones, R. Amundsen and the rest of the scientific crew. We thank the National Research Foundation of South Africa, the Research Council of Norway, and our home institutions for funding and support. We thank BENEFIT (Benguela Environment Fisheries Interaction and Training), S. Sundby, D. C. Boyer, J. Otto Krakstad, and the crew of the research vessel *Dr. Fridtjof Nansen* for support with earlier goby cruises, laying the basis for the present study. We thank K. Helge Jensen for statistical support. We appreciate the comments on this manuscript by J. Giske, C. Jørgensen, M. P. Heino, and the anonymous reviewers. Care and handling of experimental animals were performed in accordance with institutional guidelines. J.A.W.S. was a postdoctoral researcher funded by the Natural Sciences and Engineering Reserach Council of Canada at the time when the research was conducted.

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5989/333/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References

9 April 2010; accepted 4 June 2010
10.1126/science.1190708

Small Peptides Switch the Transcriptional Activity of Shavenbaby During *Drosophila* Embryogenesis

T. Kondo,[1,2]*† S. Plaza,[3,4]* J. Zanet,[3,4] E. Benrabah,[3,4] P. Valenti,[3,4] Y. Hashimoto,[1,6]† S. Kobayashi,[1,5] F. Payre,[3,4]‡ Y. Kageyama[1,6]‡

A substantial proportion of eukaryotic transcripts are considered to be noncoding RNAs because they contain only short open reading frames (sORFs). Recent findings suggest, however, that some sORFs encode small bioactive peptides. Here, we show that peptides of 11 to 32 amino acids encoded by the *polished rice* (*pri*) sORF gene control epidermal differentiation in *Drosophila* by modifying the transcription factor Shavenbaby (Svb). Pri peptides trigger the amino-terminal truncation of the Svb protein, which converts Svb from a repressor to an activator. Our results demonstrate that during *Drosophila* embryogenesis, Pri sORF peptides provide a strict temporal control to the transcriptional program of epidermal morphogenesis.

Studies of eukaryotic genomes have revealed that a large proportion of genomic DNA produces atypical long transcripts, the functions of which are controversial (*1*–*4*). These transcripts contain only short open reading frames (sORFs, <100 codons) and thus are generally considered to be non-protein-coding RNAs (ncRNAs). However, there is growing evidence that the sORFs present in some ncRNAs are actually translated into small peptides, the abundance of which is probably greatly underestimated (*5*–*7*). Whereas sORF-encoded peptides may represent an overlooked repertoire of bioactive molecules (*8*), their functions and the mechanisms by which they operate are largely unknown.

We and others recently identified an evolutionarily conserved sORF gene, referred to as *polished rice* (*pri*) or *tarsal-less* (*tal*) in *Drosophila,* and *mille-pattes* (*mlpt*) in *Tribolium* (*9*–*11*). *pri* mRNA is a polycistronic transcript that encodes four similar peptides, 11 to 32 amino acids in length, that play a redundant role in *Drosophila* embryogenesis (*9*, *10*). Embryos that lack *pri* display prominent defects, including the absence of trichomes and aberrant tracheal architecture (*9*, *10*). Reduced *pri* activity in imaginal development results in abnormal leg morphogenesis (*10*, *12*). Similarly, *mlpt* knockdown in *Tribolium* leads to appendage defects and the transformation of segmental identity (*11*).

To gain insight into the molecular function of Pri peptides, we focused on their role in trichome formation during *Drosophila* embryogenesis.

Epidermis differentiation results in a pattern of smooth cells and cells that form apical extensions, called trichomes (ventral denticles and dorsal hairs) (Fig. 1A) (*13*). Modifications of the trichome pattern that have been examined in insects (resulting from laboratory-induced mutations or evolutionary diversification) are so far all attributable to changes in expression of *shavenbaby* (*svb*) (*14*–*16*). Indeed, *svb* encodes a transcription factor that directly regulates the expression of target effectors, which are collectively responsible for trichome formation (*17*, *18*). Although the absence of *pri* results in trichome loss, the expression of *svb* is not altered in *pri* mutants (*9*). Reciprocally, *pri* is expressed normally in *svb* mutants (*9*), showing that *svb* and *pri* act in parallel in trichome formation (fig. S1). Expression of Svb target genes, such as *miniature* and *shavenoid* (*17*), is lost in *pri* mutants, whereas the expression of other epidermal genes is unaffected (Fig. 1B and S2). The activity of isolated Svb-responsive enhancers was also strongly reduced in *pri* mutants (fig. S3). Therefore, *pri* is specifically required for the transcription of Svb downstream targets in trichome cells.

How can Pri peptides regulate the expression of Svb target genes without affecting *svb* expression? The *svb* locus encodes three overlapping protein isoforms: Svb and the germline-specific proteins OvoA and OvoB (Fig. 2A) (*19*, *20*). Ovo/Svb proteins all share the same DNA-recognition and transcriptional-activation domains but differ in their N-termini (Fig. 2A). The shortest isoform, OvoB, is a transcriptional activator and induces trichomes when artificially expressed in the epidermis (Fig. 2F) (*19*). OvoA contains an extended N-terminal region, which switches its function toward active transcriptional repression and thus dominantly inhibits trichome formation (Fig. 2C) (*19*). Svb contains a further N-terminal extension, compared to OvoA, and promotes the formation of ectopic trichomes like OvoB (Fig. 2D) (*20*). To evaluate the specificity of Pri/Svb interaction, we examined the influence of *pri* on the different Ovo/Svb isoforms with respect to trichome formation. In wild-type embryos, seven rows of ventral cells per segment express *svb* and form trichomes (Fig. 2B) (*13*, *15*). Upon its ectopic expression in smooth cells, Svb [or Svb:green fluorescent protein (GFP)] induced supernumerary trichomes in control embryos (Fig. 2D) but not in *pri* mutants (Fig. 2E). In contrast, OvoB (or OvoB:GFP) was insensitive to *pri*, with ectopic trichomes forming in both control and *pri* mutant embryos (Fig. 2, F

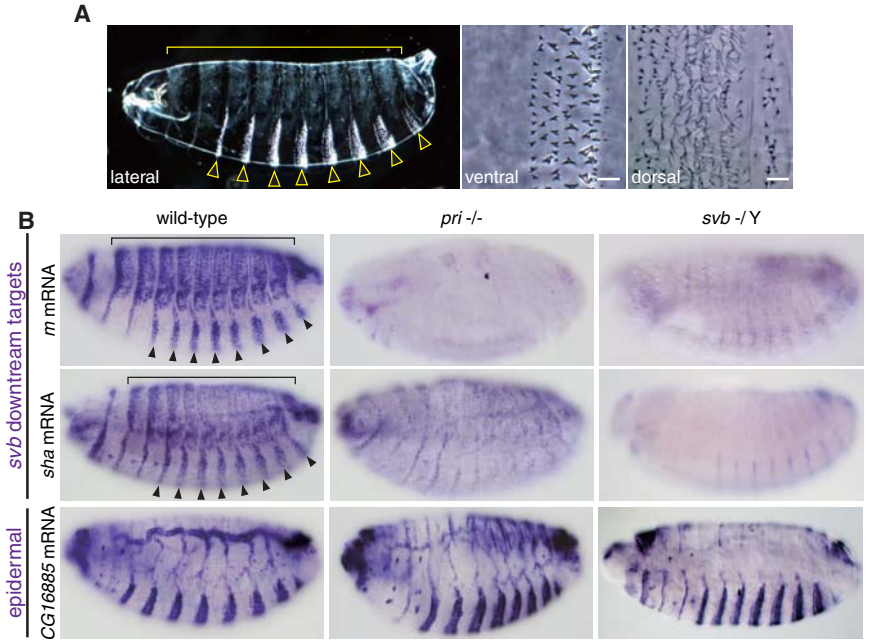


Fig. 1. *pri* is required for the expression of Svb target genes. (**A**) Embryonic cuticle specimens, showing the two types of trichomes, ventral denticles (arrowheads and middle close-up) and dorsal hairs (bracket and right close-up). (**B**) Two Svb downstream targets, *miniature* (*m*) and *shavenoid* (*sha*) are expressed in trichome cells at stages 15 and 16 in wild type, but not in *pri* and *svb* mutants. As a control of *pri* specificity for Svb function, the epidermal expression of *CG16885* (independent of *svb*) was not affected in *pri* mutants. Anterior is to the left and dorsal is to the top, except for the close-ups in (A). Scale bar, 10 μm.

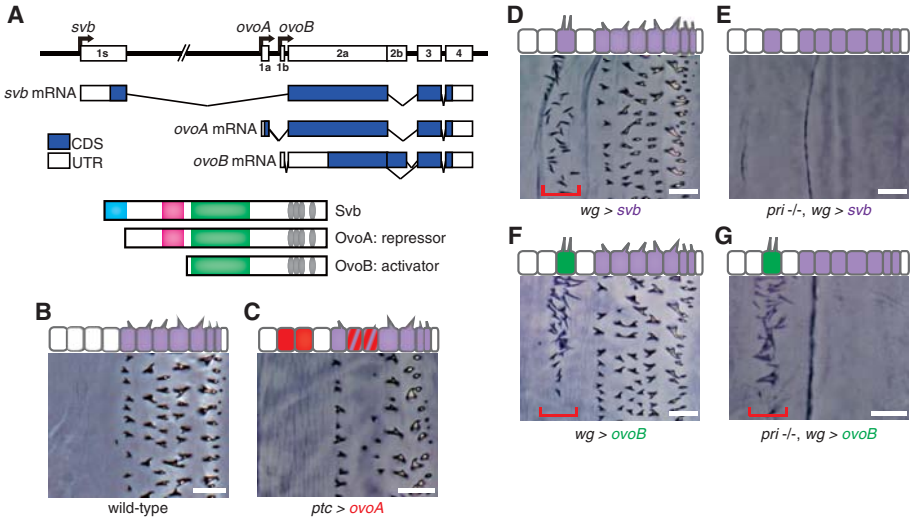


Fig. 2. The ability of Svb to induce trichomes depends on *pri*. (**A**) Scheme of the *ovo*/*svb* locus and protein isoforms. Coding (CDS) and untranslated regions (UTR) of mRNAs are represented by blue and white boxes, respectively. The Svb-specific protein region is in turquoise; the repression, activation, and DNA-binding domains are in red, green, and gray, respectively. (**B** to **G**) Micrographs of ventral trichomes (A4 segment) and illustrations of epidermal cells expressing *svb* (purple), *ovoA* (red), and *ovoB* (green) in embryos of different backgrounds. *pri* mRNA is widely expressed in epidermis and reinforced in trichome cells (fig. S1F). (B) Wild-type embryo. (C) Embryo expressing *UAS-ovoA:GFP* under the control of *ptc-GAL4.* (D) and (E) Embryos expressing *UAS-svb:GFP* under the control of *wg-Gal4* in the presence (D) or absence (E) of *pri*. (F) and (G) Embryos expressing *UAS-ovoB:GFP* under the control of *wg-Gal4* in the presence (F) or absence (G) of *pri.* Red brackets indicate ectopic trichomes. Scale bar, 10 μm.

[1]Okazaki Institute for Integrative Bioscience, National Institute for Basic Biology (NIBB), National Institutes of Natural Sciences, 5-1 Myodaiji-Higashiyama, Okazaki 444-8787, Japan. [2]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan. [3]Université de Toulouse, Université Paul Sabatier, Centre de Biologie du Développement, Bâtiment 4R3, 118 route de Narbonne, F-31062 Toulouse, France. [4]CNRS, UMR 5547, Centre de Biologie du Développement (CBD), F-31062 Toulouse, France. [5]Department of Basic Biology, School of Life Science, Graduate University for Advanced Studies (SOKENDAI), 38 Myodaiji-Nishigonaka, Okazaki 444-8585, Japan. [6]Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.

*These authors contributed equally to this work.
†Present address: RIKEN Center for Developmental Biology, 2-2-3 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan.
‡To whom correspondence should be addressed. E-mail: payre@cict.fr (F.P.); kageyama@nibb.ac.jp (Y.K.)

and G). In the latter case, we observed only OvoB-induced ectopic trichomes and no Svb-dependent endogenous trichomes (Fig. 2G). These results show that whereas *pri* has no effect on the shorter OvoB isoform, *pri* peptides specifically control the ability of Svb to induce trichomes.

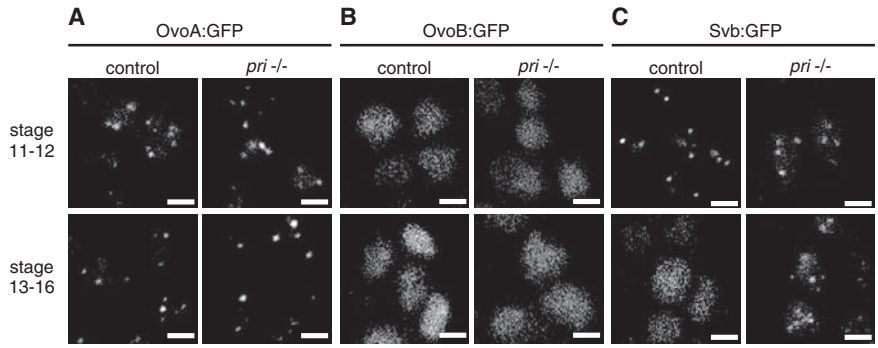


Fig. 3. *pri* regulates the subnuclear localization of Svb in living embryos. Distribution of (**A**) OvoA:GFP, (**B**) OvoB:GFP, and (**C**) Svb:GFP driven by *wg-GAL4* in control (*pri* +/–) or *pri* mutant embryos at stages 11 and 12 and stages 13 to 16. In all cases, the GFP signal was restricted to nuclei (fig. S4). Although the distribution of OvoA (focal) and OvoB (diffuse) was insensitive to *pri*, Svb switched from foci to diffuse in a *pri*-dependent manner. Scale bar, 10 μm.

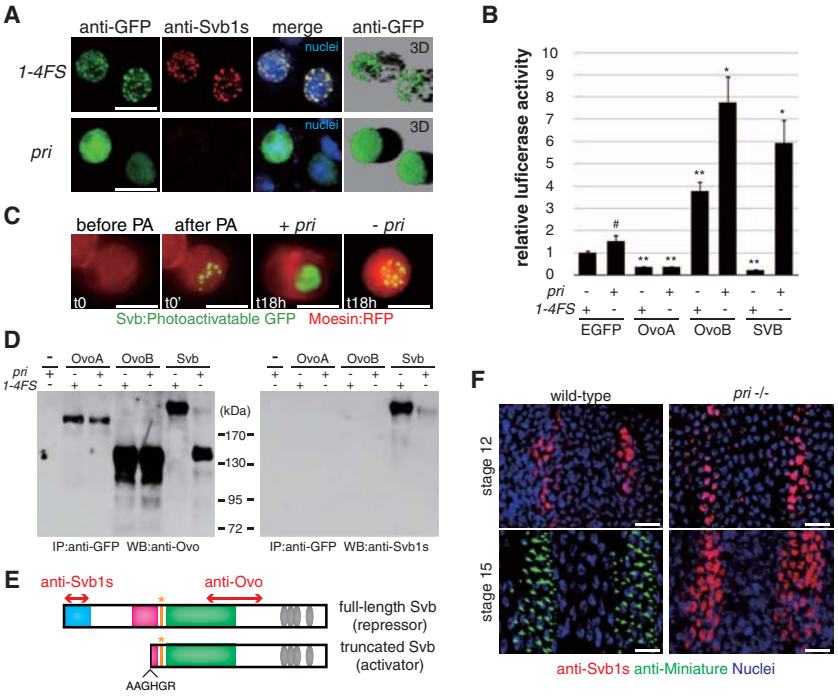


Fig. 4. Pri converts the Svb protein from a transcriptional repressor to an activator by N-terminal truncation. (**A**) Subnuclear localization of Svb:GFP in S2 cells when *pri* is co-expressed. *1-4FS*, a full-length *pri* mRNA with frame-shift mutations in ORF1-4 (*9*), was used as control. Cells were stained with antibody to GFP (green) and anti-Svb1s (red). Nuclei are in blue. The rightmost panel is a three-dimensional representation of Svb nuclear distribution. (**B**) Transcriptional activity of OvoA, OvoB, and Svb in S2 cells. Luciferase activity was used as a reporter for the transcriptional activity of the Enh-*m* enhancer. Error bars represent SE, and significance against GFP/*1-4FS*–transfected cells was evaluated with *t* tests (*$P < 0.05$, **$P < 0.01$, #$P > 0.05$). (**C**) Distribution of Svb:PA-GFP (green) in S2 cells, before photoactivation (t0), after photoactivation (t0′), and after the induction of *pri* expression (t18h). Without *pri* induction, Svb:PA-GFP was retained in foci (*–pri,* t18h). Red is Moesin:red fluorescent protein (RFP) used as a transfection control. (**D**) Western blots analysis of S2 cells expressing OvoA:GFP, OvoB:GFP, and Svb:GFP. Protein extracts were immunoprecipitated with antibody to GFP and probed with antibody to Ovo or anti-Svb1s. (**E**) Schematic representation of predicted form of truncated Svb. The N terminus of truncated Svb matches the sequence AAGHGR, which is located 56 amino acids upstream of the OvoB-initiating methionine (asterisk). Red arrows indicate the regions used to generate antibodies. (**F**) Ventral views of wild-type and *pri* mutant embryos stained with anti-Svb1s (red), and antibody to Miniature (green) that underlies nascent trichomes. Nuclei are in blue. Scale bar, 10 μm.

We next examined whether Pri peptides affect the synthesis or trafficking of Ovo/Svb proteins. Using transgenic C-terminal GFP-fusions (proven functional as described above), we observed that *pri* does not influence the production of Ovo/Svb proteins or their import to the nucleus (Fig. 3 and fig. S4). However, we noticed different patterns of their intranuclear distribution. Regardless of *pri* activity, throughout embryogenesis OvoA accumulated in discrete foci (Fig. 3A), and OvoB was distributed diffusely in the nucleoplasm (Fig. 3B). During stages 11 and 12, before *pri* is expressed in the epidermis (fig. S1), Svb formed intranuclear foci, like OvoA (Fig. 3C). At the onset of *pri* epidermal expression (stage 13 onwards) (fig. S1), the nuclear distribution of Svb became diffuse (Fig. 3C). Therefore, Svb distribution changes from a pattern similar to the OvoA repressor to that of the OvoB activator, and the timing of this conversion correlates with the expression of *pri*. This redistribution of Svb was abolished in *pri* mutants, in which Svb remained in nuclear foci throughout embryogenesis (Fig. 3C). Thus, *pri* participates in the conversion of nuclear distribution of Svb from punctate to diffuse.

The nonpunctuated, diffuse nuclear distribution of Svb in epidermal cells correlates with its ability to induce trichomes, suggesting that Svb redistribution coincides with active transcription of its targets. We explored this hypothesis using assays in *Drosophila* Schneider cells (S2 cells), which are of embryonic origin. Similarly to observations in embryos, the nuclear pattern of Svb was converted from punctate to diffuse in a *pri*-dependent manner in S2 cells (Fig. 4A). We quantified the transcriptional activity of Svb/Ovo using the Enh-*m* enhancer, which is directly activated by Svb in vivo (*17*) (fig. S3). OvoB strongly stimulated the transcription driven by Enh-*m,* and OvoA repressed transcription, both with or without *pri* (Fig. 4B). In contrast, Svb behaved like OvoA in the absence of *pri*, but similar to OvoB, activated Enh-*m* in the presence of Pri peptides (Fig. 4B). Inactivation of the Svb-binding site of Enh-*m* (*17*) suppressed this activation (fig. S5), indicating that *pri* is required for the direct activation of Enh-*m* by Svb. These results demonstrate that Pri peptides switch the transcriptional activity of Svb from that of a repressor accumulated in nuclear foci to a nucleoplasmic activator.

To explore the mechanisms by which Pri peptides trigger this switch in Svb intranuclear distribution, we examined whether Pri requires de novo synthesis of the Svb protein. Using a photoactivatable-GFP (PA-GFP), we observed that photoactivated Svb:PA-GFP switched from foci to diffuse distribution after the induction of *pri* expression (Fig. 4C). Therefore, the same Svb molecules are relocated within the nucleus, suggesting that the action of Pri peptides relies on posttranslational modifications of Svb. Accordingly, Western blot analysis showed that whereas the size of OvoA and OvoB proteins (including that of their minor species) were not affected by *pri*, Svb exhibited a differential electrophoretic mobility in a *pri*-dependent manner (Fig.

4D). In the absence of *pri*, Svb appeared slightly larger than OvoA, as deduced from the cDNA sequences (Fig. 2A). Upon *pri* expression, the Svb protein displayed a faster mobility, corresponding to a truncation of approximately 50 kD, without apparent modification in the size of *svb* mRNA (fig. S6). An antibody raised against the N-terminal Svb-specific region (anti-Svb1s) recognized only the larger Svb protein but not the truncated product formed upon *pri* expression. This truncated Svb protein was detected by antibodies to Ovo and GFP (Fig. 4, A and D, and fig. S7A), showing that it lacks the N-terminal region but retains an intact C terminus. To further characterize Svb truncation, we purified the truncated Svb protein and microsequenced its N-terminal end (fig. S8A). The N terminus of truncated Svb matches the sequence AAGHGR, which is located 56 amino acids upstream of the OvoB-initiating methionine and within a protein region that shows strong evolutionary conservation in insects (Fig. 4E and fig. S8B). The corresponding DNA sequence displays synonymous nucleotide substitutions across species and lacks canonical or alternative initiation codons (fig. S8B), further supporting the view that Svb truncation results from a posttranslational cleavage. Hence, the *pri*-induced truncated form of Svb contains the DNA-binding and activation domains but not the repression domain, explaining why it acts as a transcriptional activator.

Consistent with this idea, we observed a *pri*-dependent truncation of the endogenous Svb protein during embryogenesis. In wild-type embryos, anti-Svb1s detected a transient nuclear signal in trichome cells, at stages 11 and 12, that disappeared at later stages (Fig. 4F and fig. S7B). The loss of the Svb N-terminal region coincided with the onset of *pri* expression in the epidermis (fig. S1). Indeed, *pri* is required for Svb truncation in vivo—as revealed by the persistence of anti-Svb1s signal in *pri* mutants—throughout embryogenesis (Fig. 4F). We conclude that Pri peptides convert Svb from a transcriptional repressor to an activator via the truncation of its N-terminal region.

This study demonstrates that 11– to 32–amino acid peptides encoded by sORFs orchestrate epidermal differentiation through the control of Svb transcriptional activity. At stages 11 and 12, *svb* is already expressed and restricted to presumptive trichome cells, in which the full-length Svb repressor probably prevents the premature expression of cellular effectors. At stages 13 and 14, the expression of *pri* in epidermal cells then triggers N-terminal truncation of the Svb protein, probably through a proteolytic release of the repressor domain, causing a rapid conversion of Svb function toward activation. Thus, although *svb* expression defines the spatial pattern of trichomes, the action of Pri peptides defines the temporality of trichome formation.

Besides the mechanisms of epidermal differentiation, our studies suggest broader functions for Pri peptides. Although *pri* is also required for tracheal morphogenesis (*9*), we observed normal trachea in *svb* mutant embryos (fig. S9), indicating that Pri peptides probably regulate additional developmental factors. Recent large-scale analyses indicate that thousands of unexplored transcripts are also probably encoding polypeptides of less than 100 amino acids in mice and humans (*1*, *21*, *7*). Future functional analyses should elucidate how small peptides encoded by transcripts improperly termed ncRNAs contribute to various biological processes including development and differentiation.

References and Notes

1. P. Carninci *et al.*, RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group), *Science* **309**, 1559 (2005).
2. S. Inagaki *et al.*, *Genes Cells* **10**, 1163 (2005).
3. T. Ota *et al.*, *Nat. Genet.* **36**, 40 (2004).
4. J. L. Tupy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5495 (2005).
5. K. Hanada, X. Zhang, J. O. Borevitz, W. H. Li, S. H. Shiu, *Genome Res.* **17**, 632 (2007).
6. J. P. Kastenmayer *et al.*, *Genome Res.* **16**, 365 (2006).
7. M. C. Frith *et al.*, *PLoS Genet.* **2**, e52 (2006).
8. Y. Hashimoto, T. Kondo, Y. Kageyama, *Dev. Growth Differ.* **50** (suppl. 1), S269 (2008).
9. T. Kondo *et al.*, *Nat. Cell Biol.* **9**, 660 (2007).
10. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, J. P. Couso, *PLoS Biol.* **5**, e106 (2007).
11. J. Savard, H. Marques-Souza, M. Aranda, D. Tautz, *Cell* **126**, 559 (2006).
12. J. I. Pueyo, J. P. Couso, *Dev. Biol.* **324**, 192 (2008).
13. F. Payre, *Int. J. Dev. Biol.* **48**, 207 (2004).
14. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
15. F. Payre, A. Vincent, S. Carreno, *Nature* **400**, 271 (1999).
16. E. Sucena, I. Delon, I. Jones, F. Payre, D. L. Stern, *Nature* **424**, 935 (2003).
17. H. Chanut-Delalande, I. Fernandes, F. Roch, F. Payre, S. Plaza, *PLoS Biol.* **4**, e290 (2006).
18. I. Fernandes *et al.*, *Dev. Cell* **18**, 64 (2010).
19. J. Andrews *et al.*, *Development* **127**, 881 (2000).
20. I. Delon, H. Chanut-Delalande, F. Payre, *Mech. Dev.* **120**, 747 (2003).
21. P. P. Amaral, M. E. Dinger, T. R. Mercer, J. S. Mattick, *Science* **319**, 1787 (2008).

22. We thank the Kyoto Drosophila Genetic Resource Center (DGRC); Bloomington *Drosophila* Stock Center for fly strains; the Indiana DGRC; G. Patterson, J. Lippincott-Schwartz, and C. Hill for plasmids; S. Takada and T. Okubo (NIBB) for technical advice; Y. Latapie and B. Ronsin (CBD) and J. D'Alayer (Institut Pasteur) for excellent technical assistance; and members of the Kobayashi laboratory for helpful comments and discussion. This work was supported by a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science; the JST PRESTO program; the Ministry of Education, Culture, Sports, Science and Technology KAKENHI (20370091); Agence Nationale de la Recherche (Programme Blanc "Netoshape"); Fondation pour la Recherche Médicale (Equipe 2005); and Association pour la Recherche sur le Cancer (1111).

Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5989/336/DC1
Materials and Methods
Figs. S1 to S9
References

10 February 2010; accepted 20 May 2010
10.1126/science.1188158

Hedgehog Signaling Regulates Segment Formation in the Annelid *Platynereis*

Nicolas Dray,[1]*† Kristin Tessmar-Raible,[2,3]* Martine Le Gouar,[1]* Laura Vibert,[1] Foteini Christodoulou,[3] Katharina Schipany,[2] Aurélien Guillou,[4] Juliane Zantke,[2] Heidi Snyman,[3] Julien Béhague,[1,4] Michel Vervoort,[1,4] Detlev Arendt,[3] Guillaume Balavoine[1,4]‡

Annelids and arthropods share a similar segmented organization of the body whose evolutionary origin remains unclear. The Hedgehog signaling pathway, prominent in arthropod embryonic segment patterning, has not been shown to have a similar function outside arthropods. We show that the ligand Hedgehog, the receptor Patched, and the transcription factor Gli are all expressed in striped patterns before the morphological appearance of segments in the annelid *Platynereis dumerilii*. Treatments with small molecules antagonistic to Hedgehog signaling disrupt segment formation. *Platynereis* Hedgehog is not necessary to establish early segment patterns but is required to maintain them. The molecular similarity of segment patterning functions of the Hedgehog pathway in an annelid and in arthropods supports a common origin of segmentation in protostomes.

In the fruit fly, the axial patterning of each individual segment is controlled and maintained by two signaling pathways operating across the segment: Wnt/β-catenin and Hedgehog (*1*, *2*). The function of the Hedgehog (Hh) pathway in segment formation is conserved in other holometabolous insects (*3*) and probably also in noninsect arthropods (*4*, *5*). By contrast, mesodermal somites in vertebrates are patterned by nonhomologous segment polarity genes (*6*). Here, we investigated the role of the Hh pathway during segment formation in an annelid representative of the third great branch of Bilaterians, Spiralians (fig. S1). The nereidid *Platynereis dumerilii* presents two phases of segment formation: larval metamorphosis and juvenile posterior growth (fig. S2). Using degenerate polymerase chain reaction (PCR), we cloned four genes of the Hh pathway in *Platynereis* coding, respectively, for orthologs of the ligand Hedgehog (*Pdu-hh*), its receptor Patched *(Pdu-ptc)*, the transmembrane activator Smoothened (*Pdu-smo*),

B. Informations supplémentaires

www.sciencemag.org/cgi/content/full/329/5989/336/DC1

Supporting Online Material for

Small Peptides Switch the Transcriptional Activity of Shavenbaby During *Drosophila* Embryogenesis

T. Kondo, S. Plaza, J. Zanet, E. Benrabah, P. Valenti, Y. Hashimoto, S. Kobayashi, F. Payre,* Y. Kageyama*

*To whom correspondence should be addressed. E-mail: payre@cict.fr (F. P.); kageyama@nibb.ac.jp (Y.K.)

Published 16 July 2010, *Science* **329**, 336 (2010)
DOI: 10.1126/science.1188158

This PDF file includes:

Supporting online materials:

Materials and methods

Fly strains and germline transformation. Flies were reared at 25 ºC on standard cornmeal/yeast/glucose/agar food. We used the ovo^{svb1}, ovo^{svb2} (*S1*), svb^{R9} (*S2*), pri^{1}, pri^{2} and pri^{3} (*S3*) mutant alleles maintained over YFP/GFP-expressing balancer chromosomes. Unless otherwise noted, all other stocks were obtained from the Bloomington *Drosophila* Stock Center, Kyoto DGRC and National Institute of Genetics (Mishima, Japan). The Oregon-R strain was used as wild-type flies. The *lacZ* reporter constructs Enh-*m* (chrX:11,650,731-11,651,129, corresponding to Emin400 in (*S4*)) and Enh-*sha* (genomic positions chr2R:6,844,033-6,844,696, this work) were introduced into *svb* and *pri* mutant backgrounds by appropriate crosses. Transgenic strains carrying *UAS-svb:GFP*, *UAS-ovoB:GFP* or *UAS-ovoA:GFP* were generated by standard P element-mediated germline transformation (*S5*) using pUAST-svb:GFP, pUAST-ovoB:GFP and pUAST-ovoA:GFP, respectively. Transgenic lines *pr-svb>>Svb:GFP,* expressing a full-length Svb cDNA fused with EGFP and under the control of endogenous *svb cis*-regulatory elements (*S6*), were generated by PhiC31-mediated procedures (*S7*).

Plasmid construction. Expression vectors encoding Svb/Ovo proteins fused with either EGFP or 3xMyc tags were constructed by inserting the corresponding tag sequence immediately upstream of the stop codon of the Svb/Ovo coding sequence (CDS). DNA fragments encoding Svb/Ovo CDS fused to EGFP or 3xMyc tags were amplified by PCR with specific primers containing attB1 and attB2 recombination sites (at the 5'- and 3'-ends, respectively). Resulting DNA fragments are referred to svb/ovo:GFP and svb/ovo:Myc, respectively. These fragments were then subcloned into pDONR211 (Invitrogen) with BP clonase (Invitrogen) and subsequently transferred into pUAST-rfA (*S8*) or pMT-DEST48 (Invitrogen) with LR clonase (Invitrogen), giving rise to the pUAST-svb/ovo:GFP or the pMT-svb/ovo:GFP (or svb/ovo:Myc), respectively. For pMT-DsRed:nls, DsRed-Express

CDS in pCMV-DsRed-Express (Takara) was fused with the nuclear localization signal taken from pRed H-Stinger (obtained from Bloomington DGRC). The resulting fragment was similarly linked to the attB1 site at the 5'-end and the attB2 site at the 3'-end by PCR, subcloned into pDONR221 with BP clonase, and then transferred into pMT-DEST48 with LR clonase. The expression vector pSvb:PA-GFP encoding a full-length Svb protein fused to photoactivatable GFP (*S9*) was obtained by inserting *svb* sequences in frame with PA-GFP ORF under the control of a constitutive actin5C promoter. To construct vectors expressing the full-length *pri* (pMT-pri), 1-4FS *pri* (pMT-1-4FS) and EGFP control (pMT-EGFP), each fragment in pDONR221 (*S3*) was transferred into pMT-DEST48 with LR clonase. To construct the luciferase reporter plasmids pGL4.11-Enh-*m* and pGL4.11-Enh-*m*KO, the wild-type and mutated enhancers fused with a minimal promoter fragment were subcloned into pGL4.11 (Promega) using the *Xho* I/*Bgl* II sites. The *Renilla* luciferase CDS was amplified by PCR from pGL4.74 (Promega) with the attB1 site at the 5'-end and attB2 at the 3'-end and cloned into pDONR221 with BP clonase, then transferred into pMT-DEST48 with LR clonase, giving rise to the pMT-Rluc vector, which was used to standardize transfection efficiency.

Antibody production. DNA fragments encoding the Svb-specific N-terminal extension (Svb1s, positions 1-136 aa) and a region common to all Ovo/Svb protein isoforms (Ovo, positions 982-1228) were subcloned into the pGEX (GE Healthcare) or pKLD116 (*S10*) vectors. GST or MBP fusion proteins were produced and purified according to the supplier's instructions and (*S11*), respectively. The purified proteins were used to immunize rabbits (Agrobio) and for subsequent purification of the immune sera by affinity chromatography.

Embryo staining and observation. Cuticle preparations were performed as previously described (*S2, S3*). Digoxigenin-labeled RNA antisense probes derived from cDNAs corresponding to *miniature* (*CG9369*), *shavenoid* (*CG13209*), *CG16885*, *dusky-like* (*CG15013*), *forked* (*CG5424*), *trynity* (*CG17131*), *singed* (*CG32858*), *ovo*/*shavenbaby* (*CG6824*) and *polished rice* (*CR33327*), were synthesized *in vitro* and processed for *in situ*

hybridization following standard procedures, as described in (*S2, S3*). To identify mutant embryos, we used either a *svb* mutant chromosome carrying the btd^{1} mutation (leading to head defects; a gift from E. Wieschaus), or embryos sorted using GFP-expressing derivatives of *FM7c*, *TM3* or *TM6B* balancers (http://flystocks.bio.indiana.edu/). Immunostaining was performed according to (*S2*), with anti-Miniature at 1/400 (*S4*), anti-ß-galactosidase at 1/1000 (Cappel), Alexa Fluor 488-, 594- or 647-labeled secondary antibodies (Molecular Probes), and TO-PRO-3 (Molecular Probes) or DAPI for nuclear counterstaining (Sigma-Aldrich). We used affinity-purified anti-Svb1s (1/20) to detect the Svb-specific N-terminal region, and anti-GFP (Roche, 1/300) to detect its C-terminal region in *pr-svb>>Svb:GFP* embryos; the staining was revealed using a TSA amplification kit according to the manufacturer's specifications (Molecular Probes). Embryos were mounted in Vectashield (Vector Laboratories) and photographed with a Leica TSC SP2 confocal or Nikon Eclipse 90i microscope. Living embryos expressing GFP and DsRed were dechorionated using 50% bleach and mounted on glass slides with silicone oil (Shin-Etsu Silicones). The GFP and DsRed signals were detected by confocal microscopy (LSM5 Exciter, Carl Zeiss).

Luciferase assays and immunostaining in S2 cells. S2 cells were grown at 25 °C in Schneider's *Drosophila* medium (Invitrogen) containing 10% fetal calf serum, 100 units/ml penicillin and 100 µg/ml streptomycin. For luciferase assays, S2 cells ($0.5x10^{6}$ cells/500 µl) were plated in 24-well plates. 24 hours after plating, the S2 cells were transfected with a mixture of 100 ng pMT-svb/ovo:Myc (or pMT-EGFP), 50 ng pMT-pri (or pMT-1-4FS), 50 ng pGL4.11-Emin400/400KO and 0.5 ng pMT-Rluc in 25 µl, with 1 µl of siLentFect (Bio-Rad) in 25 µl (total 50 µl). To induce the expression of recombinant proteins using pMT-based constructs, $CuSO_4$ was added to the medium at a final concentration of 0.5 mM, 24 hours after transfection. 18 hours later, firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Reporter Assay System (Promega). The values reported in the graphs represent the average of three independent experiments, with error bars showing standard errors. After analysis of variance by F-test, the statistical significance was evaluated by Student's t-tests (for two samples with similar variances) or

Welch's t-test (for two samples having possibly unequal variances). For immunostaining, pMT-svb:GFP was co-transfected either with pMT-pri or pMT-1-4FS in S2 cells using Fugene (Roche). 24 hours after transfection, $CuSO_4$ was added to the medium at a final concentration of 0.5 mM. 48 hours after transfection, cells were fixed in 4% paraformaldehyde in PBS. Cells were further permeabilized briefly in PBS containing 0.3% TritonX-100 and were incubated with anti-Svb1s (1/100) or mouse anti-GFP (Roche. 1/1000) in PAT buffer (PBS containing 5% BSA and 1% TritonX-100) overnight at 4°C. After several washes, anti-rabbit Alexa Fluor 546 (Molecular Probes, 1/200) and anti-mouse Alexa Fluor 488 (1/200) in PAT buffer were added to the cells for two hours at 22°C. After several washes, nuclei were revealed by TO-PRO-3 staining. Cells were mounted in Vectashield (Vector Laboratories) and imaged using a confocal microscope (TCS SP2, Leica). Image stacks were subjected to processing and 3D reconstruction using the Imaris 4.5.2 software (Bitplane).

Photoactivation experiments. S2 cells were co-transfected by a mixture of pAC-Svb:PA-GFP, pMT-pri and pAC-Moesin:RFP plasmids. 24 hours after transfection, transfected cells identified from Moesin:RFP fluorescence were photographed (t0) and then subjected to photoactivation, using the DAPI illumination channel for 30 seconds. After photoactivation the same cells were photographed again (t0') and eventually transferred into fresh medium containing 0.5 mM $CuSO_4$ to induce *pri* expression, from the inducible pMT-pri vector. 18 hours after the induction of *pri*, cells were finally imaged in the same conditions to observe the nuclear distribution of photoactivated Svb:PA-GFP. For control, the same procedures were followed, except that photoactivated cells were transferred into fresh medium without $CuSO_4$. We are grateful to B. Roncin (CBD and Plate-Forme IBISA d'Imagerie Cellulaire de Toulouse, http://tri.ups-tlse.fr/) for his precious help in these studies.

Western blots. For immunoprecipitation of Svb/Ovo:GFP or Svb/Ovo:Myc fusion proteins, S2 cells ($2.5x10^6$ cells/2.5 ml) were plated in 6-well plates. 24 hours after plating, S2 cells were transfected with a mixture of 1 µg pMT-svb/ovo:GFP and 500 ng pMT-pri (or

pMT-1-4FS) in 100 µl of Opti-MEM (Invitrogen) with 3 µl of Fugene, or with 500 ng pMT-svb/ovo:Myc, 250 ng pMT-pri (or pMT-1-4FS) and 50 ng pMT-DsRed:nls in 125 µl with 5 µl of siLentFect in 125 µl (total 250 µl). 24 hours after transfection, expression was induced by $CuSO_4$ at a final concentration of 0.5 mM. 48 hours after induction, a half of the cells was lyzed in RIPA buffer and the rest was subjected to total RNA preparation as described below. GFP-tagged proteins were immunoprecipitated with an anti-GFP antibody (Chromotek) according to the manufacturer's protocol. The Myc-tagged proteins were immunoprecipitated with EZview-Red Anti-c-Myc Affinity Gel (Sigma-Aldrich) according to the manufacturer's protocol. Immunoprecipitated samples were analyzed by western blot, using the NuPAGE system (Invitrogen). Proteins were detected using anti-Ovo (1/500), anti-Svb1s (1/100), anti-GFP (TP401, Acris Antibodies, 1/2500) or mouse anti-Myc (9E10, Santa Cruz, 1/200) antibodies. Secondary antibodies were anti-mouse or anti-rabbit IgG-HRP conjugates (Jackson Laboratory, 1/10,000) and detected using ECL plus (GE Healthcare) and LAS-1000 (Fujifilm).

Pulification and N-terminal sequencing of the truncated Svb. S2 Cells were co-transfected with pMT-svb:GFP, pMT-pri (or pMT-1-4FS) and a hygromycine-resistance pHygro vector (a gift from S.Carreno). One week after transfection, stable transformants were selected by adding 300 µg/ml of Hygromycine B (Invitrogen) to the medium. 10 days later, individual resistant clones were selected and tested for the Svb:GFP expression and *pri* responsiveness. A stable Svb/Pri-expressing line was amplified. 10^8 cells were treated with 1mM CuSO4 and, after overnight induction, cells were harvested, lysed in RIPA buffer and subjected to immunoprecipitation as described above using 20 µl of anti-GFP antibody-conjugated beads (Chromotek). The immunoprecipitate was then subjected to SDS-PAGE and transfer to PVDF membrane (Hybond-P, GE Healthcare). After transfer, truncated Svb:GFP visualized by coomassie staining was subjected to Edman degradation for microsequencing. Microsequencing was performed by J. D'Alayer (Laboratory of protein microsequencing, Institut Pasteur, Paris).

Northern blots and RT-PCR. Total RNA samples were prepared from S2 cells ($2.5x10^6$

cells/2.5 ml) transfected with the svb/ovo:Myc construct as described above. 48 hours after $CuSO_4$ induction, cells were homogenized with RNeasy (Qiagen) and RNA extracts were used for further characterization. For northern blots, RNA was separated by formaldehyde agarose gel electrophoresis and transferred to a nylon membrane (GeneScreen Plus, NEN). Each loaded sample contained 1.2 μg of total RNA. The amount of ribosomal RNA, visualized by methylene blue staining, served as a loading and transfer control. The membranes were hybridized with a digoxigenin (DIG)-labeled antisense RNA probe for *svb*/*ovo* in hybridization buffer (ULTRAhyb, Ambion) at 68˚C overnight, washed according to the DIG Application Manual (Roche), and incubated with an alkaline phosphatase-conjugated anti-DIG antibody (Roche). Labeled probes were detected with chemiluminescence using CSPD (Tropix) and LAS-1000 (Fujifilm). The *svb*/*ovo* probe was amplified from a *svb* cDNA using primers (5'-ATGTACGTGTGCGAGGAGTG-3' / 5'-GCTGGCCGAGCAGATTGTTG-3'), and cloned into pCR-Blunt II-TOPO (Invitrogen). DIG-labeled RNA probe was generated by *in vitro* transcription using T7 RNA Polymerase (Roche). RT-PCR analyses were performed using PrimeScript RTase and PrimeStar (Takara) with the specific primers (5'-CGCTCTCCGAATGGTTTTCCTTG-3' for *pri* cDNA synthesis and 5'-GAGTTCCAAGCCGAAAGTTA-3' / 5'-CACCGAACATTAC AAATCGTTGGC-3' for *pri* PCR) (5'-TCATCGGAGTCATAAGAGGGG-3' for *svb*/*ovo* cDNA synthesis and 5'-GGAGGTATGCCGAAGATTTTCC-3' / 5'-GCCGGTATTATAGAATCGTGGG-3' for *svb* PCR; 5'-TCGCAA ATCAACCGATAATCC-3' / 5'-GCCGGTATTATAGAATCGTGGG-3' for endogenous *svb* PCR; 5'-TTGCCCGTTTTGTCTTTGC-3' / 5'-GCCGGTATTATA GAATCGTGGG-3' for *ovoA* PCR; and 5'-GACAAATTTCTGAGAATCG CACTTC-3' / 5'-GCCGGTATTATAGAATCGTGGG-3' for *ovoB* PCR).

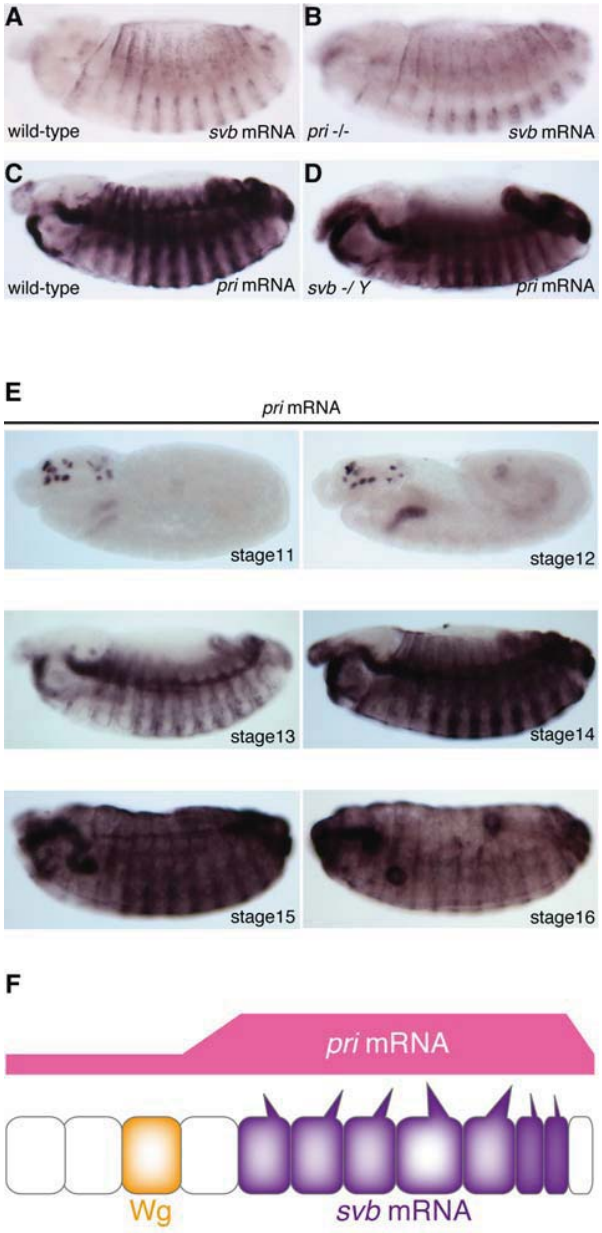


Figure S1. Expression of *pri* and *svb* mRNA during late embryogenesis. (A, B**)** *svb* mRNA expression in wild-type and *pri* mutant embryos at stage 14. (C, D**)** *pri* mRNA expression in wild-type and *svb* mutant embryos at stage 14. The expression of *svb* mRNA was not affected by the lack of *pri*, and reciprocally the expression of *pri* was not affected by the lack of *svb*. (E) *pri* mRNA expression in wild-type embryos from stage 11 to 16. *pri* mRNA was expressed in epidermal cells from stage 13 onward. (F) Schematic representation of the expression of *pri* and *svb* mRNA in the ventral epidermis. *svb* mRNA is strictly restricted to presumptive trichome cells, whereas *pri* mRNA is more broadly expressed in epidermis with a reinforcement in trichome cells, which are located posterior to Wingless (Wg)-positive cells (*S3*).

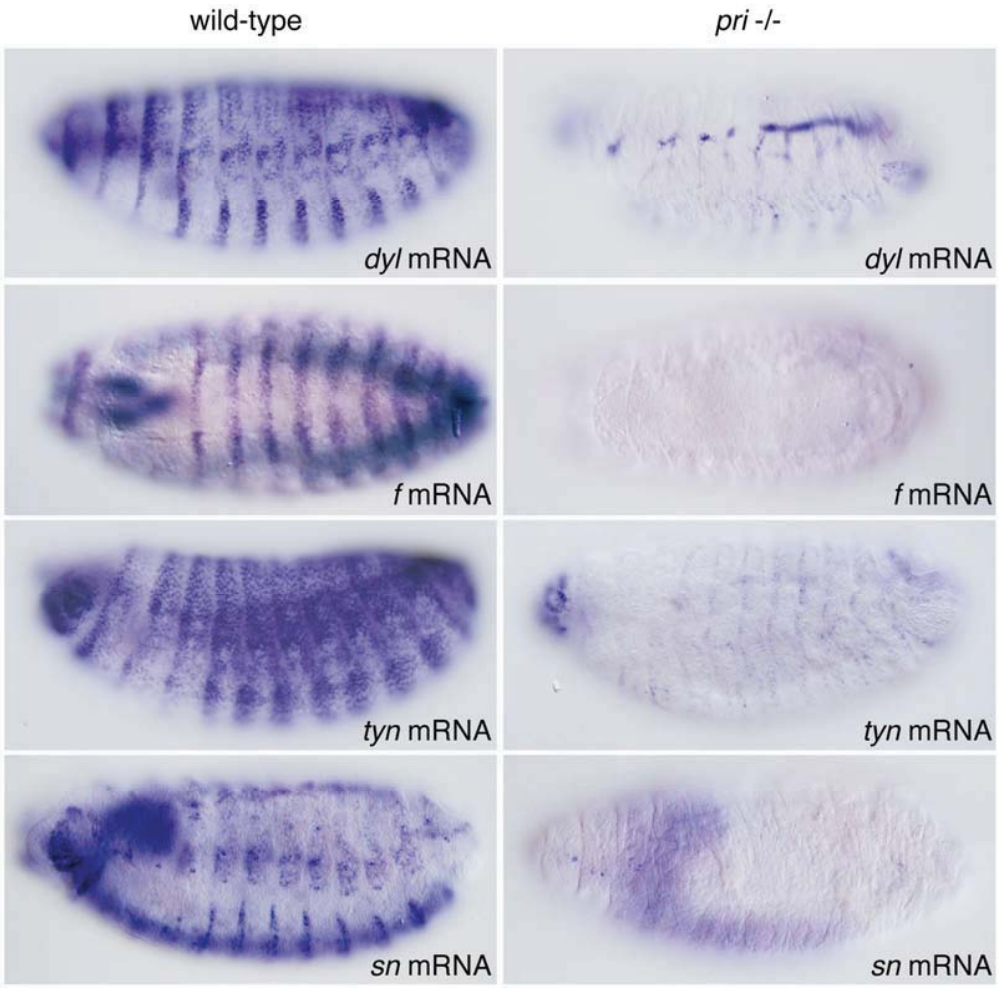


Figure S2. mRNA expression of various Svb target genes (*S4, S11*) in wild-type and *pri* mutant embryos. The epidermal expression of the Svb target genes *dusky-like (dyl)*, *forked (f)*, *trinity (tyn)*, and *singed (sn*) was strongly reduced in *pri* mutants.

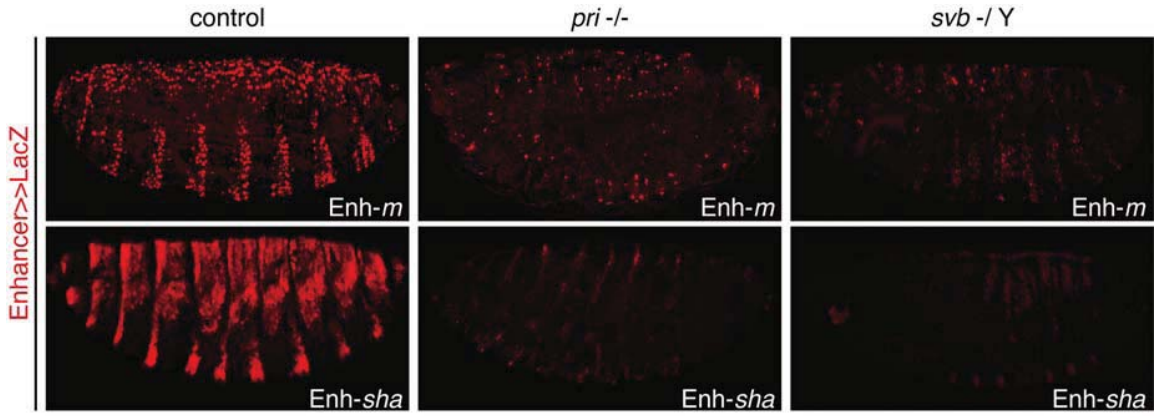


Figure S3. Expression of *lacZ* reporters under the control of Svb-responsive enhancers, taken from the Svb downstream targets *miniature* (Enh-*m*) and *shavenoid* (Enh-*sha*). The *in vivo* activity of the Enh-*m* and Enh-*sha* enhancers requires both *svb* and *pri* functions.

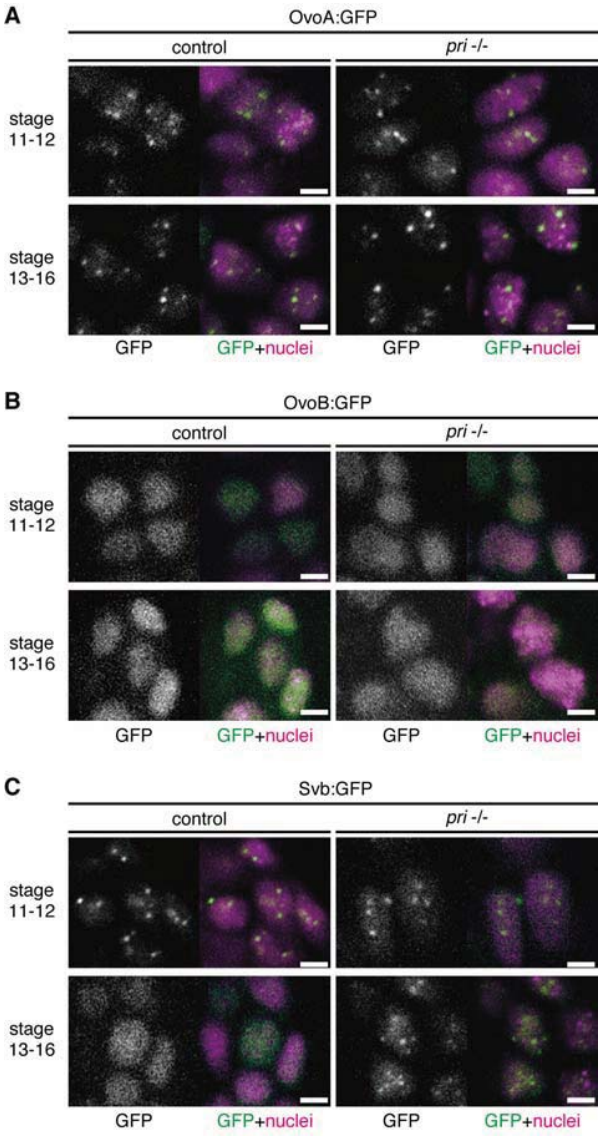


Figure S4. Svb/Ovo proteins are localized into nucleus, regardless of *pri* activity. Distribution of OvoA:GFP (A), OvoB:GFP (B) and Svb:GFP (C) driven by *wg-GAL4* in control (*pri* +/-) or *pri* mutant embryos at stage 11-12 and 13-16. Color images merge GFP (green) and nuclear DsRed (magenta). Scale bar, 10 μm.

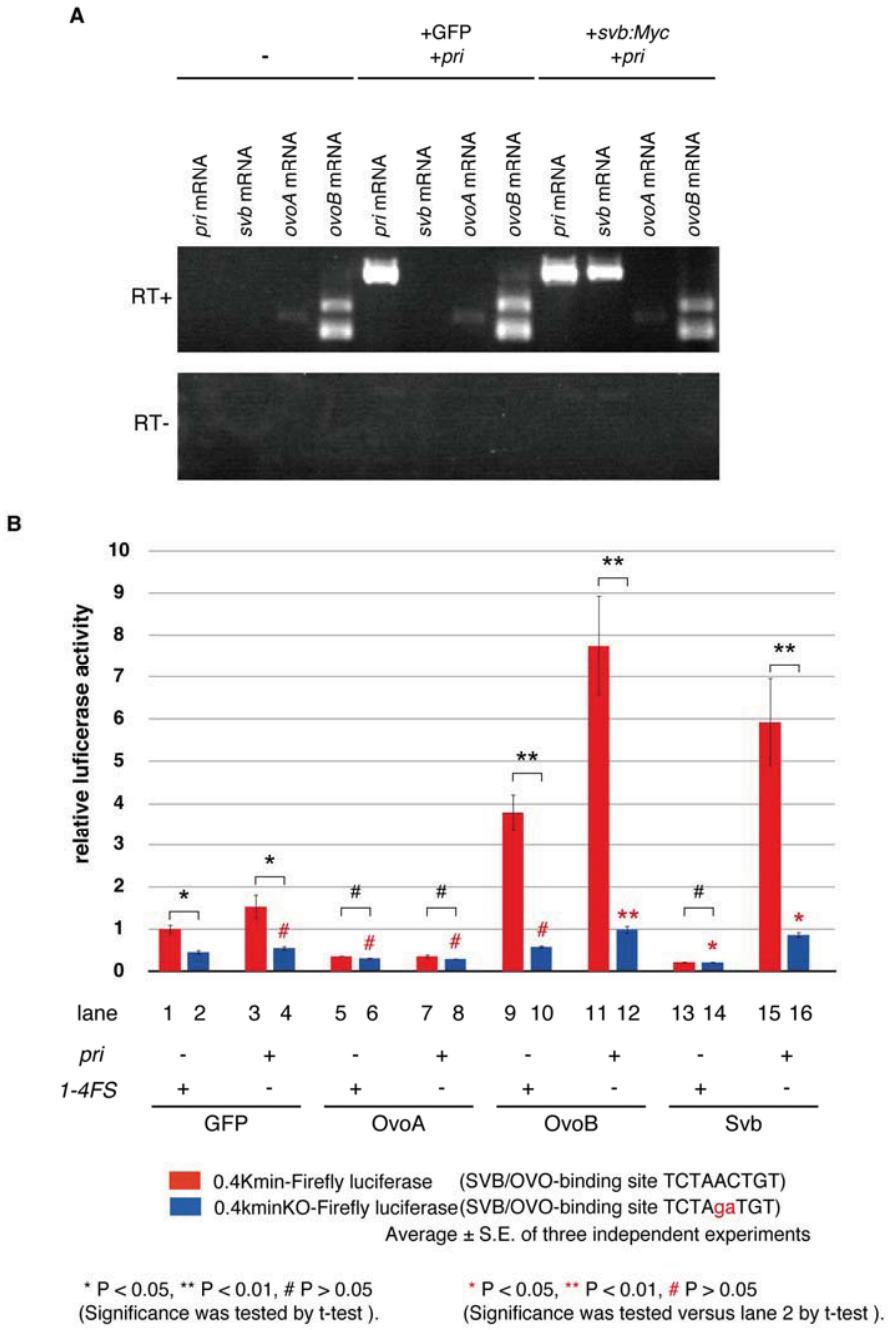


Figure S5. Svb directly regulates the transcriptional activity of the Enh-*m* enhancer, in a *pri*-dependent manner in S2 cells. (A) Endogenous expression of *pri*, *svb*, *ovoA* and *ovoB* mRNA in S2 cells, examined by RT-PCR. Neither *pri* nor *svb* mRNA was found expressed at detectable levels in S2 cells, while we observed endogenous expression of *ovoB* mRNA, as well as residual amounts of *ovoA* mRNA. (B) Transcriptional

assays for Ovo/Svb activity in S2 cells. The transcriptional properties of Svb/Ovo:Myc proteins were evaluated by firefly luciferase reporter activity driven by the Enh-*m* (red bars) or Enh-*m*KO (blue bars) enhancers. Compared to Enh-*m*, the two point mutations in the Svb/Ovo-binding site of Enh-*m*KO suppressed the transcriptional activation mediated by Svb in the presence of Pri peptides (lane 16) (as well as that mediated by OvoB with or without Pri; lane 10, 12), indicating that this activation requires the direct binding of Svb/Ovo. Since Enh-*m*KO showed weaker activity than Enh-*m* even without Svb/Ovo (lanes 1-4), it is likely that the endogenous OvoB detected in S2 cells contributed to basal reporter activity. OvoA (lanes 5-8) and Svb without Pri (lanes 13-14) decreased the reporter expression to the lowest transcription level observed for Enh-*m*KO under control conditions (lane 2). Firefly luciferase activities were normalized to that of *Renilla* luciferase, used as a transfection control. Significance was evaluated by Student's t-test or Welch's t-test after analysis of variance by F-test. Error bars represent standard errors.

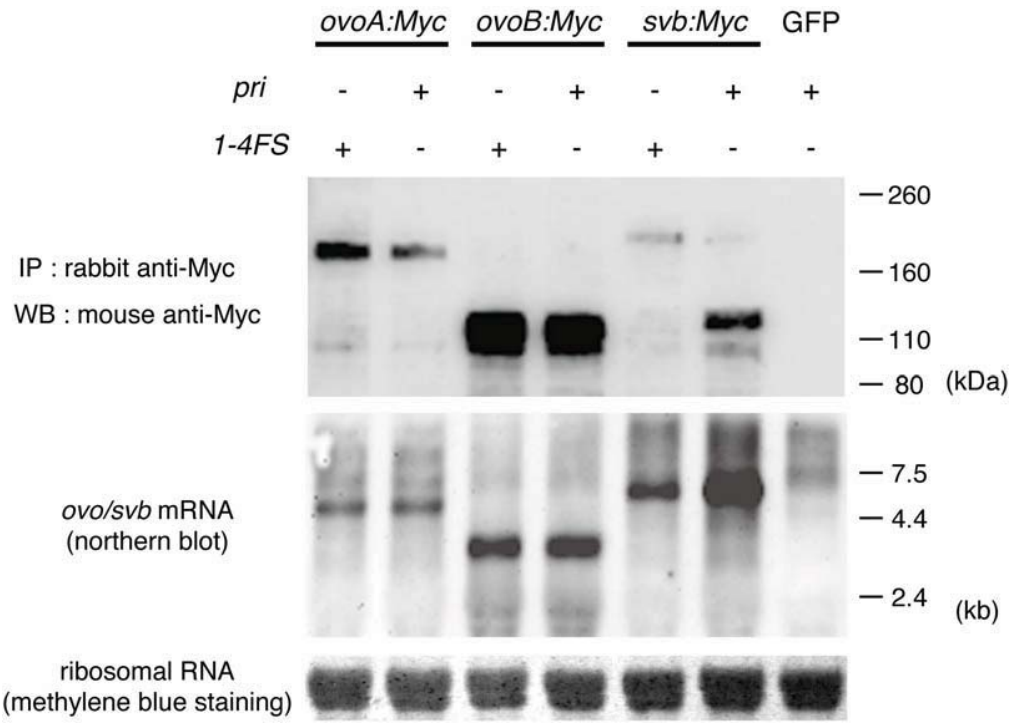


Figure S6. Analysis of the respective modifications of *ovo/svb* protein and mRNA products, following the expression of Pri peptides in S2 cells. The apparent size of *ovo/svb* mRNAs was not modified following expression of *pri,* or the *1-4FS* construct used a negative control. The electrophoretic mobility of OvoA:Myc and OvoB:Myc proteins, were not either affected by the status of *pri* expression. In contrast, the mobility of the Svb:Myc protein became much faster upon *pri* expression, indicating a truncation of Svb. In some experiments, this truncated product appears more abundant than the full-length form, a feature possibly resulting from an increased stability of the protein and/or mRNA products. It is also possible that the repressor activity of full-length Svb affects cell viability or growth of transfected cells. Methylene blue staining of ribosomal RNA was used as a loading control.

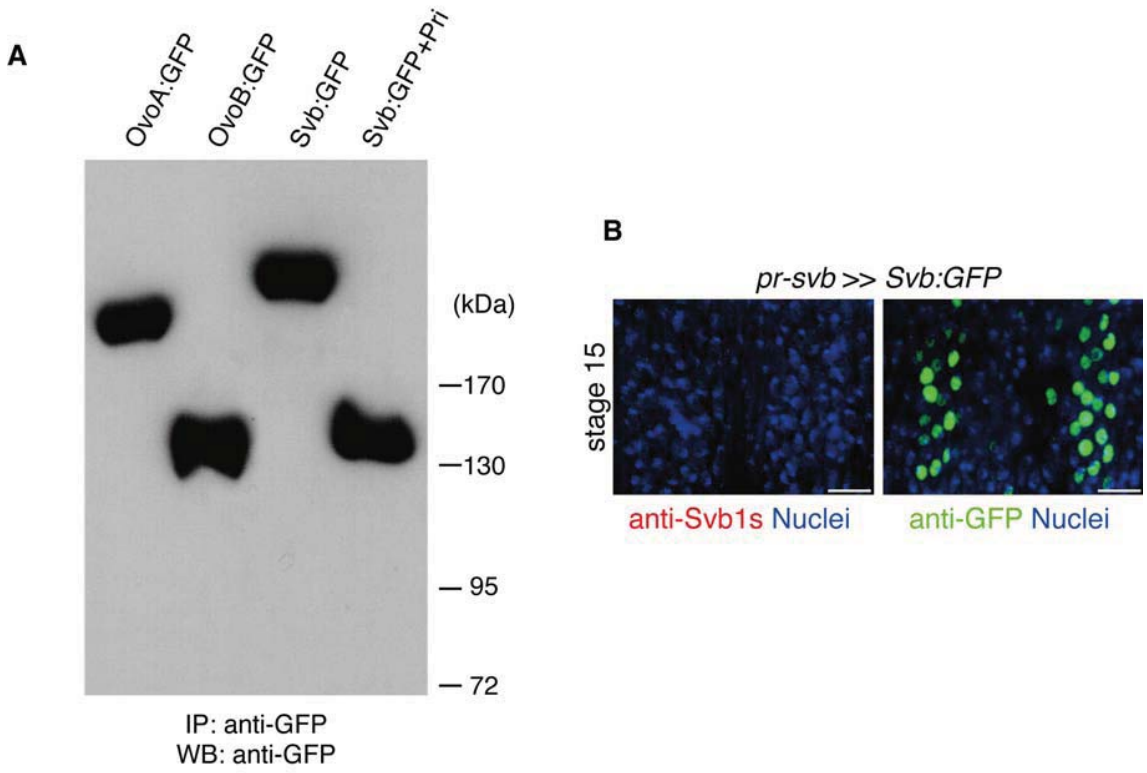


Figure S7. (A) Western blots of protein extracts from S2 cells expressing OvoA:GFP, OvoB:GFP, Svb:GFP and Svb:GFP + Pri. Cell extracts were immunoprecipitated with an anti-GFP antibody, then separated by SDS-PAGE and analyzed by western blotting using an anti-GFP antibody. Detection of the truncated Svb isoform by anti-GFP showed that it contained an intact C-terminal region. (B) Ventral views of transgenic embryos that express Svb:GFP under control of the *svb* promoter, stained with anti-Svb1s (red) or anti-GFP (green). Svb:GFP was recognized by anti-GFP, but not by anti-Svb1s at stage 15, supporting that Svb:GFP was truncated at its N-terminus but retained an intact C-terminal region after onset of *pri* expression in embryos. Nuclei are in blue. Scale bar, 10 μm.

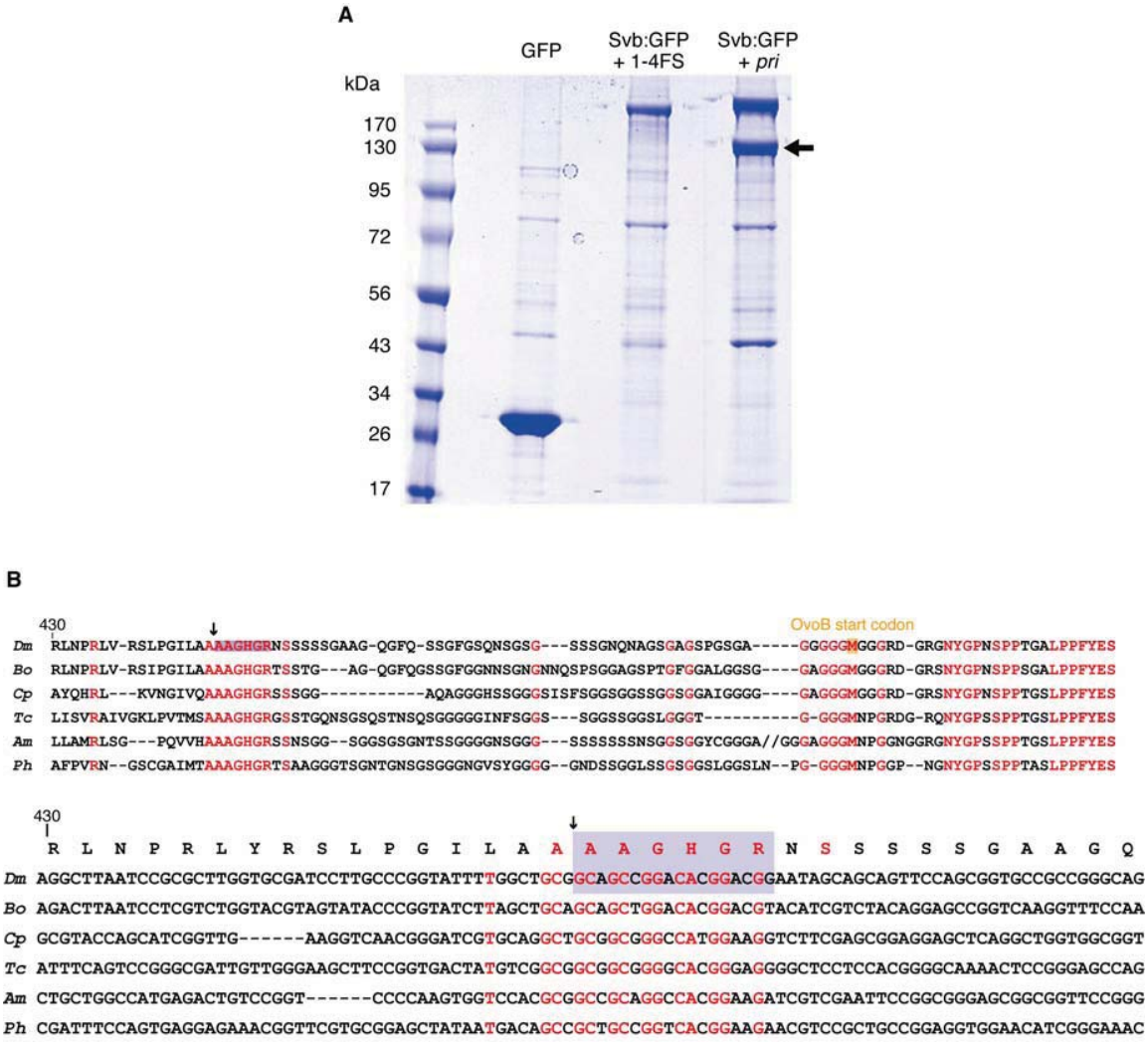


Figure S8. (A) A coomasie stained electrophoretic gel showing results from the biochemical purification of the truncated Svb protein. S2 cell lines stably expressing GFP alone, Svb:GFP and 1-4FS, or Svb:GFP and Pri were immuno-purified with an anti-GFP antibody. The truncated Svb:GFP protein (arrow) was transferred to a PDVF membrane and subjected to Edman sequencing for determination of its N-terminal extremity. (B) Amino acid (upper) and nucleotide (lower) alignments of Ovo/Svb sequences corresponding to the N-terminal extremity of the truncated Svb protein. The AAAGHGRXS protein motif presents strong evolutionary conservation. On the other hand, the corresponding DNA sequence shows synonymous nucleotide substitutions among insect species, supporting that the selective pressure has applied to the protein rather than on the RNA sequence. Although CUG, GUG or ACG have been documented as alternative initiation codons (*S12*), there is no reported evidence of the use of GGC, the codon for this N-terminal Alanine (and also that of the immediately upstream residue). These results strongly suggest that Svb truncation does not result from

alternative translation, but rather on a proteolytic maturation. Sequences are: *Dm, Drosophila melanogaster; Bo, Bactrocera oleae; Cp, Culex pipiens; Tc, Tribolium castaneum; Am, Apis melifera; Ph, Pediculus humanus*. Invariant residues are in red, purple shadowing underlines the N-terminus of truncated Svb determined by microsequencing. Orange shadowing indicates the initiating Methionine of OvoB.

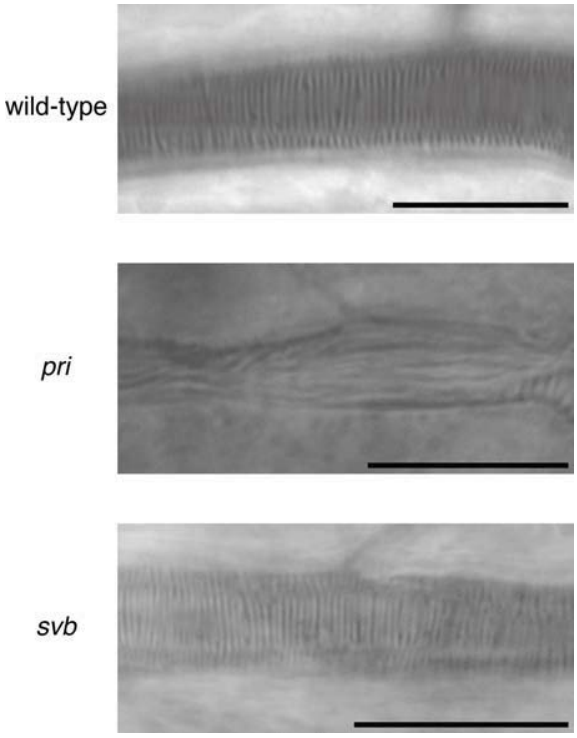


Figure S9. *svb* is not required for tracheal taenidium formation. Differential interference contrast microscopy of the tracheal taenidiae in wild-type embryos showed the characteristic pattern of folds perpendicular to the longitudinal axis of the tracheal dorsal trunk (top). These regular taenidiae were absent in embryos lacking *pri* (*pri*^{2}/*pri*^{3}, middle panel), but were not affected in *svb* mutant embryos (*svb*^{R9}/*Y*; bottom). Scale bar, 10 μm.

Supporting references

S1. E. Wieschaus, C. Nusslein-Volhard, G. Jurgens, *Rouxs Arch. Dev. Biol.* **193**, 296 (1984).

S2. I. Delon, H. Chanut-Delalande, F. Payre, *Mech Dev* **120**, 747 (2003).

S3. T. Kondo *et al.*, *Nat Cell Biol* **9**, 660 (2007).

S4. H. Chanut-Delalande, I. Fernandes, F. Roch, F. Payre, S. Plaza, *PLoS Biol* **4**, e290 (2006).

S5. A. C. Spradling, G. M. Rubin, *Science* **218**, 341 (1982).

S6. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).

S7. J. Bischof, R. K. Maeda, M. Hediger, F. Karch, K. Basler, *Proc Natl Acad Sci U S A* **104**, 3312 (2007).

S8. T. Kondo, S. Inagaki, K. Yasuda, Y. Kageyama, *Genes Genet Syst* **81**, 129 (2006).

S9. G. H. Patterson, J. Lippincott-Schwartz, *Science* **297**, 1873 (2002).

S10. C. J. Rocco, K. L. Dennison, V. A. Klenchin, I. Rayment, J. C. Escalante-Semerena, *Plasmid* **59**, 231 (2008).

S11. I. Fernandes *et al.*, *Dev Cell* **18**, 64 (2010).

S12. C. Touriol *et al.*, *Biol Cell* **95**, 169 (2003).

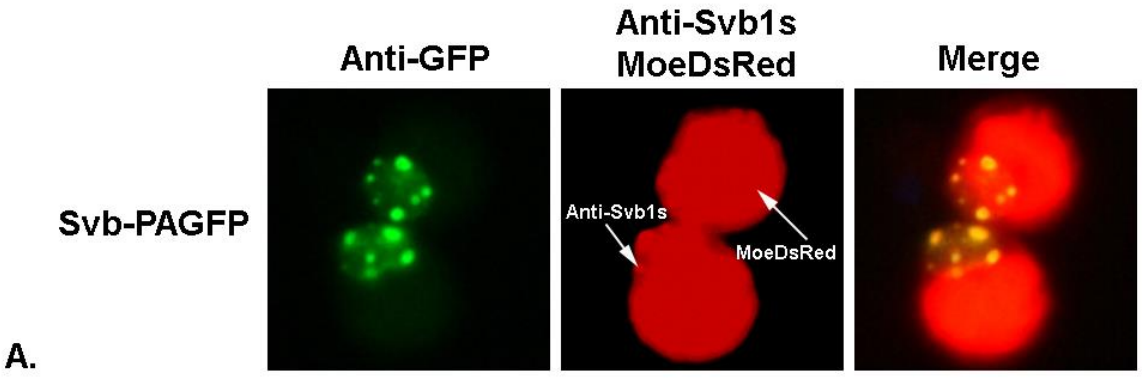
C. Détails sur l'expérience de la GFP photo-activable

Afin de déterminer si la forme courte de Svb résulte d'une néo-synthèse protéique ou d'une maturation post-traductionnelle de la forme longue localisée en foyers nucléaires, j'ai choisi d'utiliser une stratégie où Svb est fusionné à une GFP photo-activable (PAGFP) (Patterson and Lippincott-Schwartz, 2002). Elle a la particularité de n'émettre une fluorescence qu'après exposition à 410nm. Une fois activée, elle se comporte comme une GFP classique, et on peut la suivre au cours du temps. Il est donc possible de suivre les Svb-PAGFP pleine taille présentes dans les foyers, suite à l'expression de *pri*.

J'ai construit un plasmide à partir d'un vecteur exprimant Svb-GFP sous le contrôle d'un promoteur constitutif (actine), auquel j'ai substitué la GFP par la PAGFP. Après transfection des cellules S2, ce plasmide produit bien la protéine Svb-PAGFP pleine taille, comme montré par immuno-marquage avec les anticorps Anti-Svb1s et Anti-GFP (PAGFP et GFP ne diffèrent que par 4 résidus). En absence de *pri*, Svb-PAGFP présente une localisation ponctuée dans le compartiment nucléaire (Fig. 30A).

Cette construction étant fonctionnelle, j'ai réalisé des tests de photo-activation. Une co-transfection avec un vecteur exprimant constitutivement la Moesin-DsRed (MoeDsRed) a permis de repérer les cellules transfectées (Fig. 30A). 24h après transfection, les cellules exprimant la MoeDsRed (et donc sûrement aussi Svb-PAGFP) ont été soumises aux UV pendant différents temps. Quelques secondes se sont avérées suffisantes pour activer la PAGFP. J'ai suivi le maintien de cette fluorescence au cours du temps. La protéine photoactivée est suffisamment stable pour être détectable 24 heures plus tard. Pour tester l'influence de *pri* sur le devenir de Svb-PAGFP, j'ai co-transfecté les cellules avec les vecteurs Moe-DsRed, Svb-PAGFP et pMT-Pri, permettant l'expression de *pri* sous le contrôle d'un promoteur métallothionéine (pMT), inductible aux métaux lourds. 24 heures après transfection, j'ai procédé à la photo-activation de Svb-PAGFP des cellules transfectées, puis induit (ou non) l'expression de *pri* par ajout $CuSO_4$ dans le milieu de culture. 18h après, les cellules ont été fixées et observées. Dans les cellules où *pri* n'a pas été induit, Svb-PAGFP est resté en foyers nucléaires. Dans les cellules où *pri* est exprimé, le pool de Svb-PAGFP, en foyers juste après photo-activation, est diffus dans le noyau (Fig. 30B). **Ces résultats démontrent que les mêmes molécules de Svb présentes dans les foyers sont redistribuées**

dans le nucléoplasme, sans besoin de néo-synthèse. L'activité de *pri* semble donc s'exercer par l'intermédiaire du contrôle de la maturation post-traductionnelle de Svb.



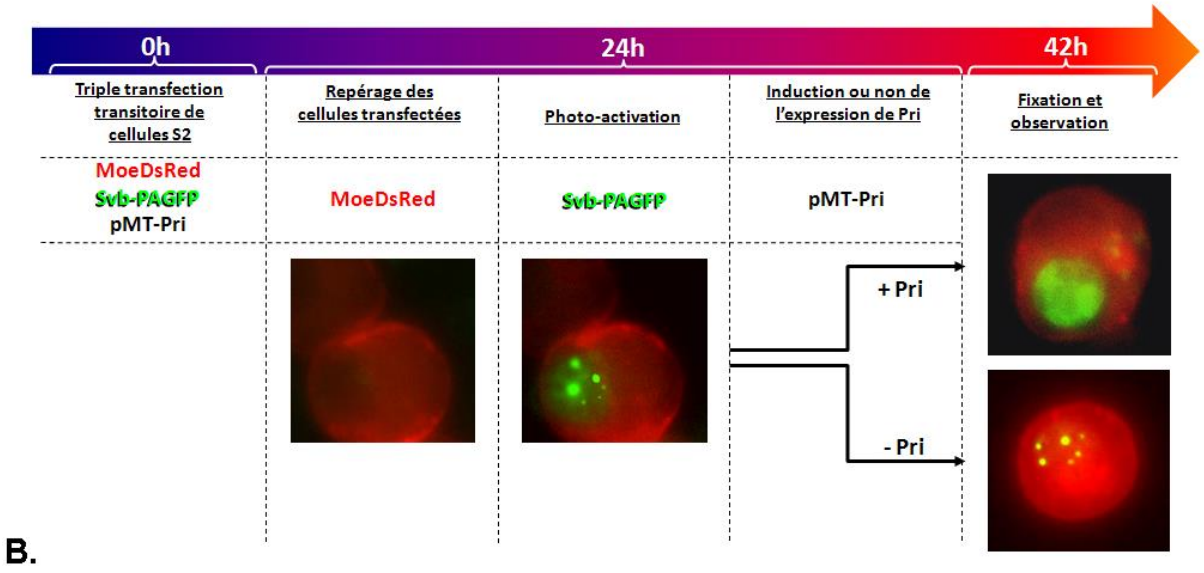


Figure 30 : La forme diffuse de Svb provient de la relocalisation des molécules Svb distribuées en foyers nucléaires avant l'expression de Pri. A. Transfection transitoire de cellules S2 par MoeDsRed et Svb-PAGFP. MoeDsRed est visualisée en rouge dans le cytoplasme. Svb-PAGFP est visualisé en rouge dans les foyers nucléaires avec l'Anti-Svb1s et en vert avec l'Anti-GFP. **B.** Schéma de la stratégie expérimentale utilisée lors des expériences de photo-activation. Les constructions impliquées à chaque étape sont indiquées. MoeDsRed est visualisée dans le cytoplasme en rouge, Svb-PAGFP est visualisé en vert après photo-activation.

III. Discussion et hypothèses

Les peptides Pri induisent la maturation de Svb, conduisant à l'élimination de sa partie N-terminale et ainsi un changement de son activité transcriptionnelle. Pour comprendre le mode d'action des peptides, il était nécessaire de comprendre la mécanistique de cette modification post-traductionnelle. En effet, au commencement de ma thèse, plusieurs pistes étaient permises.

A. Svb est-il clivé par une endoprotéase ?

Le séquençage d'Edman de la forme courte de Svb nous a permis d'identifier son extrémité N-terminale. Il s'agit d'une région qui présente une forte conservation évolutive, que j'appellerai le site AAGH (Fig. 31). Ceci suggérait que cette séquence puisse être requise pour la reconnaissance par une endoprotéase qui viendrait couper Svb à ce niveau là, générant ainsi la forme courte. Dans ce modèle, les peptides Pri pourraient par exemple être requis pour l'activation de cette protéase.

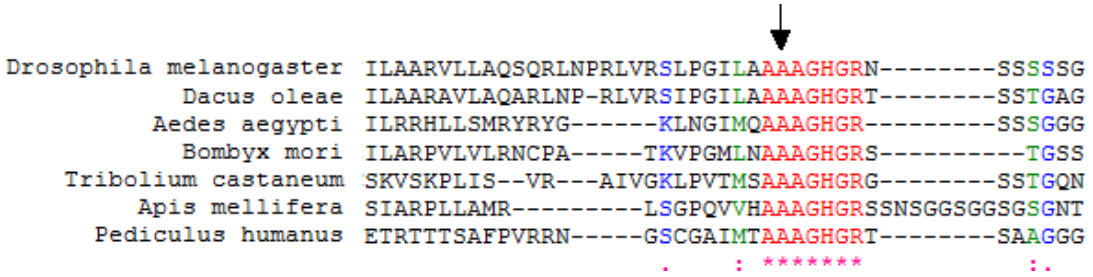


Figure 31 : La région de Svb correspondant au N-terminal de sa forme courte est évolutivement conservée. L'alignement des séquences des homologues de Svb dans d'autres espèces d'insectes montre que la région correspondant à l'extrémité N-terminale de la forme courte générée en réponse aux peptides Pri présente une forte conservation évolutive. La flèche indique le premier résidu de la forme courte, déterminé en séquençage d'Edman (la 2ème Ala rouge).

B. La voie des caspases est-elle impliquée?

La voie des caspases gouverne des fonctions apoptotiques et non-apoptotiques impliquées dans le développement animal (Miura, 2012). Les caspases sont des protéases qui reconnaissent et clivent un site consensus DEVD sur les protéines cibles (Cohen, 1997). Ce clivage peut parfois conduire à un changement d'activité. Par exemple, le clivage par la caspase-3 de la kinase Rho-associated coiled-coil protein kinase 1 (ROCK-1) (Chang et al., 2006) élimine un domaine d'inhibition en C-terminal, et génère ainsi une forme plus courte constitutivement active.

Svb possède un site DEVD en amont du N-terminal de la forme courte. Ce site est trouvé dans une région PEST, segment riche en proline (P), aspartate (D), glutamate (E), sérine (S) et thréonine (T) (Fig. 32). Ce motif PEST, après clivage potentiel par une caspase, constituerait pour le fragment N-terminal une extrémité non-structurée qui pourrait diriger l'entrée au - et la dégradation par le - protéasome (selon le modèle proposé par Belizario (Belizario et al., 2008)). La dégradation par le protéasome sera abordée dans le paragraphe suivant. Cependant, le site DEVD est situé 200 résidus en amont du site AAGH. Cette hypothèse n'était donc pas incompatible avec nos résultats, mais ne suffirait pas à elle seule à expliquer la forme courte de Svb.

MPKIFLIKNRLHQQQQRLLESQNLLQHKNQDDERLVPPLSPSGSGSGPSPTPTSQPPPEPQGQGQQVLGQVPDSDQQPLSLTR
KRFHHRRHYFGQSRHSLDHLNQNQSPNPNANANPNQIQNPAELEVECATGQVQNENFAAELLQRLTPNTATTAQNNIVNNLVN
NSRAATSVLATKDCTIENSPISIPKNQRAEDEEEQEDQEKEKPAERERERKSDERTEQVEKEERVEREEEEDDEVDVGVEAPRP
RFYNTGVVLTQAQRKEYPQEPKDLSLTIAKSSPASPHIHSDSESDSDSDGGCKLIVDEKPPLPVIKPLSLRLRSTPPPADQRP
SPPPPRDPAPAVRCSVIQRAPQSQLPTSRAGFLLPPLDQLGPEQQEPIDYHVPKRRSPSYDSDEELNARRLERARQVREARRR
STILAARVLLAQSQRLNPRLVRSLPGILAAAAGHGRNSSSSSGAAGQGFQSSGFGSQNSGSGSSSGNQNAGSGAGSPGSGAGG
GGGMGGGRDGRGNYGPNSPPTGALPPFYESLKSGQQSTASNNTGQSPGANHSHFNANPANFLQNAAAAAYIMSAGSGGGGCTG
NGGGGASGPGGGPSANSGGGGGGGGGNGYINCGGVGGPNNSLDGNNLLNFASVSNYNESNSKFHNHHHHHQHNNNNNNNGGQT
SMMGHPFYGGNPSAYGIILKDEPDIEYDEAKIDIGTFAQNIIQATMGSSGQFNASAYEDAIMSDLASSGQCPNGAVDPLQFTA
TLMLSSQTDHLLEQLSDAVDLSSFLQRSCVDDEESTSPRQDFELVSTPSLTPDSVTPVEQHNTNTTQLDVLHENLLTQLTHNI
VRGGSNQQQQHHQQHGVQQQQQQQHSVQQQQQHNVQQQHGVQQQHVQQQPPPSYQHATRGLMMQQQPQHGGYQQQAAIMSQQQ
QQLLSQQQQQSHHQQQQQQQHAAAYQQHNIYAQQQQQQQQQHHQQQQQQQHHHFHHQQQQQPQPQSHHSHHHGHGHDNSNMSL
PSPTAAAAAAAAAAAAAAAAAAAHLQRPMSSSSSSGGTNSSNSSGGSSNSPLLDANAAAAAAAALLDTKPLIQSLGLPPDLQLE
FVNGGHGIKNPLAVENAHGGHHRIRNIDCIDDLSKHGHHSQHQQQQGSPQQQNMQQSVQQQSVQQQQSLQQQQQQQHHQHHSN
SSASSNASSHGSAEALCMGSSGGANEDSSSGNNKFVCRVCMKTFSLQRLLNRHMKCHSDIKRYLCTFCGKGFNDTFDLKRHTR
THTGVRPYKCNLCEKSFTQRCSLESHCQKVHSVQHQYAYKERRAKMYVCEECGHTTCEPEVHYLHLKNNHPFSPALLKFYDKR
HFKFTNSQFANNLLGQLPMPVHN

Figure 32: Svb contient un site DEVD et un motif PEST. Séquence protéique de la forme longue de Svb. La séquence correspondant à la forme courte est grisée, le site AAGH est souligné, le site DEVD est en rouge, les résidus définissant le motif PEST sont en jaune.

C. S'agit-il d'une maturation dépendante du système ubiquitine-protéasome?

Il existe des cas ressemblant à Svb, de changement d'activité transcriptionnelle d'un facteur de transcription à la suite d'une dégradation partielle par le protéasome, et ce en réponse à une polyubiquitination (Ubiquitin-Proteasome dependent Processing, RUP). Le premier exemple est la dégradation de p105, précurseur cytosolique de p50 (sous-unité du facteur de transcription NF-κB). p105 subit la dégradation de sa partie C-terminale, la partie N-terminale reste stable. C'est p50, qui est transloqué dans le noyau et joue son rôle d'activateur de la transcription (Hoppe et al., 2001). Ce modèle s'applique aussi à Cubitus interruptus (Ci) qui sous sa forme longue est un activateur, et lorsqu'il est partiellement dégradé devient un répresseur (Tian et al., 2005). Je vais décrire ici la voie de l'ubiquitine-protéasome, et expliquer comment ces facteurs de transcription sont partiellement dégradés.

1. Le système de l'ubiquitine-proteasome (UPS)

a. La voie de la poly-ubiquitination

i. Les acteurs enzymatiques

Une cascade d'enzymes connues sous le nom de E1 ubiquitin activating enzyme, E2 ubiquitin-conjugating enzyme et E3 ubiquitin-ligase permet la conjugaison de molécules d'ubiquitine sur une (des) lysine(s) de la protéine cible (Neutzner and Neutzner, 2012) (Fig. 33). L'UPS est le système majeur de protéolyse chez les eucaryotes, avec des fonctions critiques dans le contrôle du cycle cellulaire, l'apoptose, l'inflammation, la transcription, la transduction de signaux, le contrôle qualité des protéines, etc. La E1 (il n'en existe qu'une

chez la drosophile, Uba1) active une molécule d'ubiquitine : c'est une étape qui nécessite l'hydrolyse d'une molécule d'ATP et conduit à la formation d'un pont thioester entre le C-terminal de l'ubiquitine et une cystéine de la E1. L'ubiquitine activée est ensuite transférée sur une E2 (il en existe 32 chez la drosophile), avec laquelle elle va aussi former un pont thioester. L'ampleur des rôles et des cibles du protéasome s'est faite par la multiplicité des E3, les facteurs spécifiques de l'ubiquitination. Elles sont les plus nombreuses (environ 300 chez la drosophile), et sont divisées en sous-groupes, dont les deux plus représentés sont les E3 de type RING (Really Interesting New Gene) et les E3 de type HECT (Homologous to E6-AP Carboxy Terminus). Les E3 vont être liées à leur substrat et soit catalyser directement le transfert de l'ubiquitine de l'E2 à une lysine du substrat (RING-E3), soit prendre en charge l'ubiquitine et ensuite la transférer sur le substrat (HECT-E3) (Metzger et al., 2014) (Fig. 33). La liaison et l'ubiquitination du substrat par une E3 se fait sur une séquence appelée **dégron** (séquence qui engendre la dégradation).

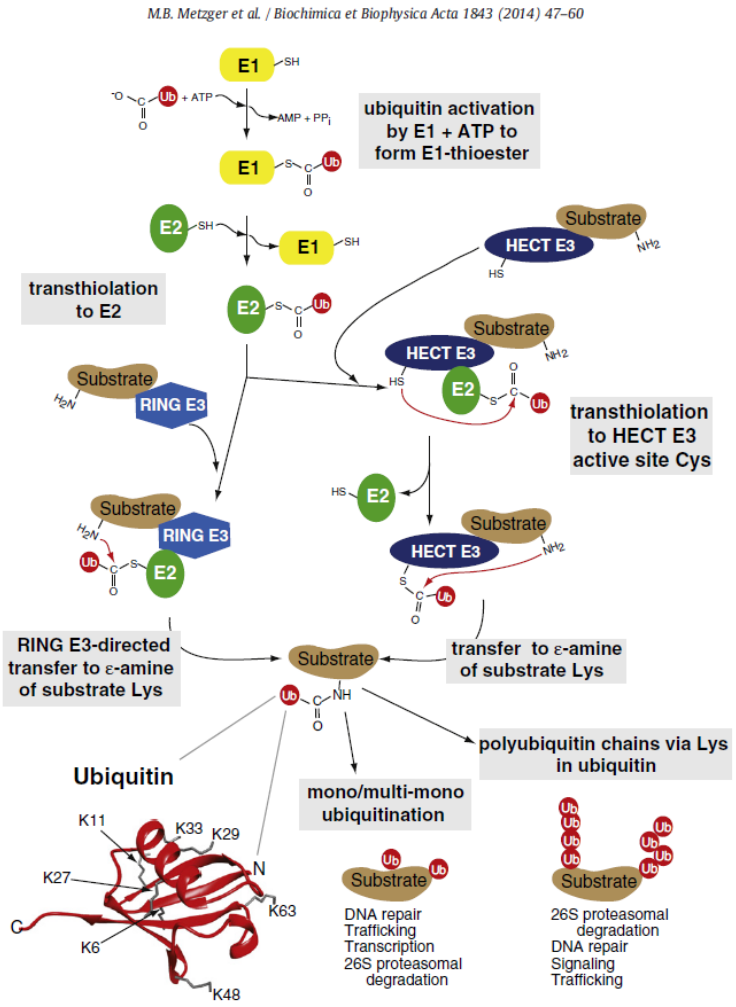


Figure 33: Schéma de la voie de l'ubiquitination d'une protéine.

ii. Le code de l'ubiquitine

Le transfert de l'ubiquitine sur le substrat se fait soit directement sur une lysine, soit sur une lysine d'une ubiquitine déjà liée, formant à terme une chaine de poly-ubiquitines. L'ubiquitine (76 résidus) possède 7 lysines : K6, K11, K27, K29, K33, K48 et K63 (Fig. 33). Il existe un code de branchement des chaînes d'ubiquitines, qui vont donner lieu à des formes et donc des rôles différents (Komander, 2009; Xu et al., 2009; Trempe, 2011) (Fig. 34). Cependant la plus représentée pour l'instant est la chaîne liée en K48 comme signal de dégradation par le protéasome 26S (Hershko and Ciechanover, 1998).

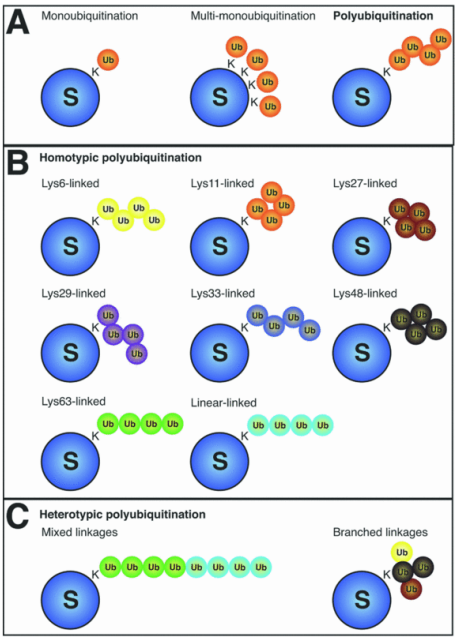


Figure 34: Différentes formes d'ubiquitination. A. Une protéine peut être mono-, multi-, poly- (ou multi-poly-) ubiquitinylée. **B.** Formes de chaînes d'ubiquitines homotypiques : chaque chaîne ne contient qu'un seul type de liaison, et a une forme particulière. **C.** Formes de chaînes hétérotypiques, avec différents types de liaison dans une même chaîne d'ubiquitines (Komander, 2009).

b. La structure du protéasome 26S

Le protéasome 26S est un complexe protéique subdivisé en 2 sous-domaines (Fig. 35), généralement impliqué dans la dégradation des protéines polyubiquitinylées (Coux et al., 1996; Pickart, 2004).

i. La particule cœur 20S (CP20S)

La CP20S (Core Particle 20S) est un complexe de 28 sous-unités en forme de tunnel, arrangées en 4 anneaux hétéroheptamériques. Les anneaux externes, composés chacun de 7 sous-unités α (α1-7) sont situés de part et d'autre des deux anneaux internes composés de 7 sous-unités β (β1-7). Ce sont les sous-unités β1, β2 et β5 qui contiennent des sites d'activité protéolytique nécessaires à la dégradation des protéines adressées au protéasome (Fig. 35). Chaque site peut cliver de multiples séquences peptidiques, avec β1 qui clive en C-terminal

des résidus acides (activité «caspase-like»), β2 qui coupe après les lysines et arginines (activité «trypsine-like») et β5 qui clive après les résidus hydrophobes (activité «chymotrypsine-like») (Arendt and Hochstrasser, 1997; Heinemeyer et al., 1997).

Les anneaux α, les plus externes de cette structure en forme de tunnel, sont là pour réguler l'ouverture de son entrée (Fig. 35). Les sous-unités α présentent une queue N-terminale convergeant en un réseau entremêlé à l'entrée du tunnel, en bloquant l'entrée (Groll et al., 1997; Groll et al., 2000). Ainsi en absence de structure additionnelle, le cœur catalytique du CP20S ne sera pas accessible aux protéines, évitant une dégradation anarchique. La CP20S devra être activée (ouverte) suite à la liaison de particules régulatrices, dont je vais n'en détailler qu'une ici.

ii. La particule régulatrice 19S (RP19S)

Dans le cas du protéasome 26S, les deux particules régulatrices qui se lient de part et d'autre du CP20S sont les mêmes : les RP19S (Regulatory Particle 19S). C'est un complexe de 19 protéines, subdivisé en deux sous-domaines (Fig. 35).

- Le couvercle (Lid)

Le lid est composé de 9 sous-unités Sem1, Rpn3, 5-9,11 et 12 (Fig. 35). Seule l'activité d'Rpn11 a été identifiée : il s'agit d'une enzyme de déubiquitination (DUB) dont le rôle est critique dans l'activité du protéasome 26S (Verma et al., 2002).

- La base

La base est constituée d'un anneau de 6 protéines à activité ATPase (Rpt1-6) (Fig. 35). Rpt 2, 3 et 5 possèdent en C-terminal un motif HbYX, qui en venant se brancher dans des poches formées à l'interface de deux sous-unités α induisent un changement de leur conformation et ainsi l'ouverture de l'entrée dans le CP20S. C'est ce qu'on appelle le *gate-opening* (Smith et al., 2007; Gillette et al., 2008; Rabl et al., 2008). Deux autres sous-unités sont des grosses protéines d'échafaudage : Rpn1 et Rpn2, qui vont être liées à deux récepteurs à l'ubiquitine Rpn10 et Rpn13 (Schreiner et al., 2008) (Fig. 35). Par ailleurs, Rpn1 et Rpn13 sont aussi liés à deux DUBs, Ubp6 (Stone et al., 2004) et Uch37 (Yao et al., 2006) respectivement. Ainsi, une protéine ubiquitinylée est reconnue et maintenue au protéasome grâce à Rpn10 et 13, son étiquette d'ubiquitines est enlevée par Ubp6 et Uch37. Puis l'hydrolyse d'ATP par les Rtp1-6 va générer un cycle de basses et hautes affinités entre le substrat et le protéasome, engendrant ses déstabilisation et déstructuration des conformations

tertiaire et secondaire (Rubin et al., 1998; Liu et al., 2006). Ainsi, petit à petit dépliée, la protéine va pouvoir entrer à l'intérieur du CP20S (Smith et al., 2005; Finley, 2009) et être dégradée par les sous-unités β1, β2 et β5.

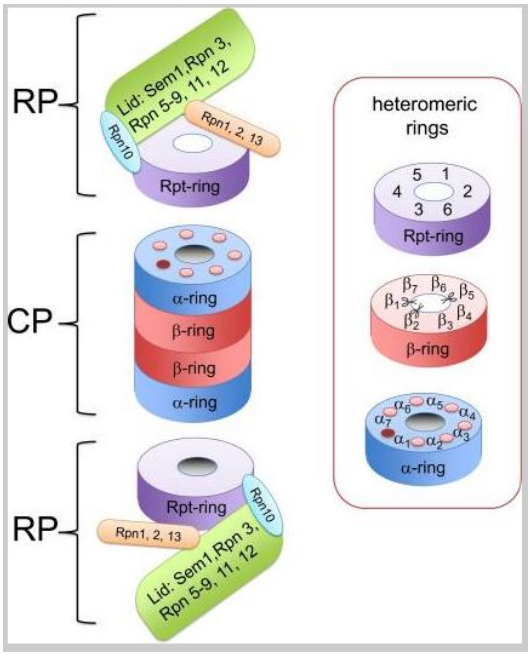


Figure 35 : composition du protéasome 26S (Bedford et al., 2010). RP : RP19S. CP : CP20S.

2. Le signal d'initiation de la dégradation

Il a été décrit que pour qu'une protéine puisse être dégradée par le protéasome, il faut qu'elle présente un site d'initiation, qui est une région non structurée et libre d'accéder à l'intérieur du CP20S (Prakash et al., 2004). Ce site constitue la région par laquelle la translocation dans le CP20S, et donc l'hydrolyse de la protéine, est initiée (Piwko and Jentsch, 2006) (Fig. 36). Ainsi, selon la structure du substrat, la dégradation peut démarrer par une extrémité ou par l'intérieur.

3. La dégradation partielle

Les substrats subissant une dégradation partielle dans le protéasome présentent des points communs au niveau de leur composition : un domaine très structuré adjacent à un domaine de faible complexité (Tian et al., 2005). Dans le cas de p105 il s'agit d'une région riche en glycines (Glycine Rich Region, GRR) (Hoppe et al., 2001). Ces éléments sont généralement situés au milieu de la protéine. La région de faible complexité sert à initier la dégradation, et la protéine rentre dans la CP20S en formant une boucle (Piwko and Jentsch, 2006) (Fig. 36). La région structurée bloque la dégradation, et donc celle-ci ne se fait que du côté de la protéine qui ne contient pas cette région structurée. Ainsi une forme tronquée de la

protéine est générée (Fig. 36). Par contre, lorsque cette région est absente ou instable, la dégradation de la protéine sera totale (Tian et al., 2005).

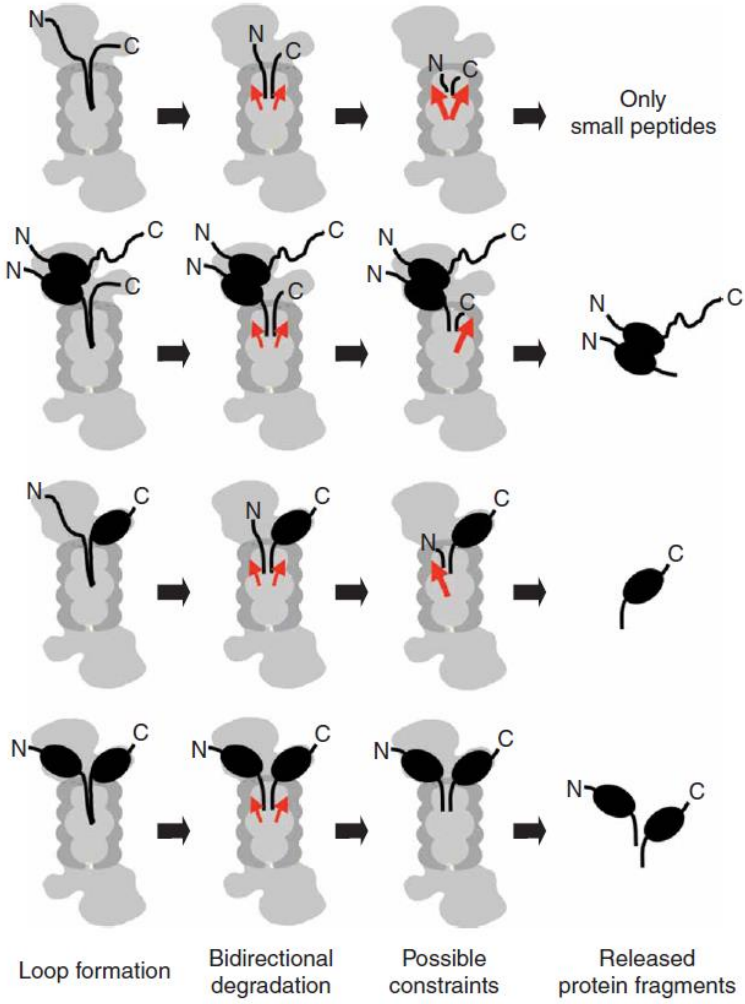


Figure36 : modèles hypothétiques de dégradation et maturation de protéines par le protéasome. La dégradation totale d'une protéine (en haut) peut démarrer par l'insertion d'une boucle interne dans le CP20S. La dégradation continue ensuite bi-directionnellement (flèches rouges). Si aucune contrainte stérique n'est rencontrée, la dégradation est totale. La maturation (dégradation partielle) de protéines par le protéasome a lieu si la protéine contient un ou plusieurs domaines très structurés (boules noires) (qui peuvent être des domaines de dimérisation). Ces domaines agissant comme des blocs structuraux, la dégradation ne se propagera que jusqu'à eux, ou sur la région n'en contenant pas. Ces dégradations partielles engendrent des protéines plus courtes qui peuvent avoir des propriétés différentes de la protéine de départ, dépendant du domaine qui a été éliminé (Piwko and Jentsch, 2006).

D. La maturation de Svb dépend-elle d'ubiquitin-like modifiers (UBLs)?

1. Les peptides Pri : une étiquette ?

On sait qu'il existe de nombreuses protéines ubiquitin-like (Taherbhoy et al., 2012), qui ont été découvertes plus tard et sont globalement moins bien caractérisées. Il n'était pas impossible d'en avoir découvert une nouvelle (qui ne serait pas exactement ubiquitin-like au niveau de sa structure, mais dans le sens où ces peptides constitueraient par leur ajout en tant qu'étiquette de manière covalente sur Svb, un signal de dégradation). De plus, Pri et l'ubiquitine ont un point commun : leur petite taille (76 résidus pour l'ubiquitine). Notons tout de même que Pri ne contient pas de lysines, ce qui constituerait un point de divergence sur le mécanisme d'étiquetage sur la protéine cible (Svb).

2. Svb est-il sumoylé?

La protéine SUMO (Small Ubiquitin-like MOdifier) est quant à elle une UBL relativement bien décrite (Wilkinson and Henley, 2010). La sumoylation de Svb en réponse à Pri était une piste attractive puisque c'est une modification post-traductionnelle qui est souvent associée avec des changements de localisation sub-nucléaire, mais aussi d'activité transcriptionnelle de certains facteurs de transcription (Gill, 2004). Par exemple, LRH-1est diffus en absence de sumoylation, mais localisé dans des foyers nucléaires quand il est sumoylé (Yang et al., 2009). Dans le cas de FLASH, la sumoylation contrôle son activité de co-activateur du facteur de transcription c-Myb, sans influencer sa localisation en foyers nucléaires (Alm-Kristiansen et al., 2009).

Partie 2 : Les peptides Pri sont requis pour l'interaction entre Svb et Ubr3, dirigeant ainsi son ubiquitination et sa dégradation partielle par le protéasome.

L'ensemble de mes travaux de thèse a eu pour objectif de répondre à cette question : Comment les peptides Pri induisent-ils l'élimination de la partie N-terminale de Svb ? Pour y répondre, il était nécessaire d'identifier les facteurs et les séquences de Svb associés à ce processus. Je vais présenter ici les méthodologies choisies pour répondre à ces deux sous-questions.

I. Méthodologie

A. Identification des séquences *cis*-régulatrices de Svb

Pour identifier les séquences *cis*-régulatrices de Svb nécessaires à sa sensibilité aux peptides Pri, j'ai choisi de générer des mutants (délétions internes ou externes, ou mutations ponctuelles) fusionnés en C-terminal à la GFP, et de tester leur capacité à répondre à Pri. Pour cela, les plasmides exprimant constitutivement les versions mutées de Svb sont co-transfectés en cellules en culture S2 avec un plasmide permettant l'expression inductible de *pri* (grâce au promoteur de la métallothionéine). Ainsi pour chaque mutant, l'expression de *pri* est induite ou non par ajout de $CuSO_4$. On peut évaluer la réponse du mutant à Pri en comparant les deux conditions, soit par analyse de taille des protéines en Western blot, soit par immuno-coloration des cellules avec un anticorps spécifique de la forme longue de Svb (l'anti-Svb1s). Cet anticorps permet d'observer en absence de Pri un signal qui co-localise avec le signal GFP, et qui disparait en présence de Pri. Par ailleurs, la localisation sub-nucléaire des mutants est un autre critère d'évaluation de réponse aux peptides, que nous n'avons que très peu exploité. En effet, cette analyse de relocalisation de Svb de foyers nucléaires vers une localisation homogène dans le noyau est délicate et subjective. Par exemple, une cellule qui présentera une localisation diffuse du signal GFP avec un seul foyer nucléaire sera difficilement classable.

B. Identification des facteurs *trans*-régulateurs de la réponse de Svb à Pri

Pour identifier les facteurs associés à la régulation de Svb par les peptides Pri, nous avons décidé de tirer profit de la lignée cellulaire stable co-exprimant de manière inductible Pri et Svb-GFP (Fig. 37B), originellement établie pour l'analyse de la forme courte de Svb en séquençage d'Edman. Grâce à cette lignée et à l'anti-Svb1s, il nous était possible d'envisager une approche pan-génomique et non biaisée de crible par ARN interférence. J'ai réalisé et analysé ce crible sur la plateforme de criblage Drosophila RNAi Screening Center (DRSC), à Harvard Medical School (Boston, MA), dans le laboratoire de Norbert Perrimon. Le principe est le suivant : la plateforme propose une collection de plaques à 396 puits contenant des dsRNA (double strand RNA) dirigés contre chacun des gènes annotés de la drosophile. L'induction au $CuSO_4$ de l'expression de Pri et Svb-GFP est faite 3 jours après ensemencement des cellules dans les puits. Le lendemain, les cellules sont fixées et immuno-colorées avec l'anti-Svb1s (le signal GFP conféré par l'expression de Svb-GFP est interprétable et n'a pas besoin d'être amplifié). Ainsi, si les 3 jours d'incubation avec le dsRNA ont conduit à l'absence d'un facteur requis pour la régulation de Svb par Pri, Svb-GFP sera resté sous sa forme longue, détectable avec l'anti-Svb1s, donnant de nombreuses cellules qui seront positives à la fois pour l'anti-Svb1s et la GFP (Fig. 37D). En revanche, si le facteur éliminé de la cellule n'est pas impliqué dans ce processus, Svb sera bien clivé en réponse à Pri et seul un signal GFP sera observable (Fig. 37C). Ainsi, par simple comptage du nombre de cellules positives pour l'anti-Svb1s parmi les cellules exprimant Svb (positives à la GFP), on peut estimer l'implication du facteur (dont la dégradation est ciblée par le dsRNA) dans la réponse de Svb à Pri.

L'ensemble des gènes est représenté dans 66 plaques. Chacun des gènes est ciblé par un ou plusieurs amplicons. Un amplicon est un dsRNA de 500pb environ, ciblant une région de l'ARNm dont il induit la dégradation. Ainsi les grands gènes, permettant l'expression de grosses protéines, sont souvent représentés par plusieurs amplicons. Le crible a été réalisé en duplicat. Les 132 plaques correspondantes, après immuno-coloration, ont été analysées en microscopie confocale, permettant d'obtenir pour chacun des amplicons 7 champs, dans les 2 canaux correspondant à l'anti-Svb1s et la GFP. Au final cela représente plus de 500000 images à traiter.

Brice Ronsin, responsable de la plateforme d'imagerie du Centre de Biologie du Développement, a développé une macro pour ce projet, permettant de compter dans le logiciel ImageJ, de manière automatique, le nombre de cellules positives pour chacun des 2 signaux séparément, ou des 2 signaux ensembles. Les données obtenues sont exploitables avec Excel. Ainsi ce crible nous a permis d'attribuer à chacun des gènes de la drosophile un score reflétant le pourcentage d'inhibition de la réponse de Svb à Pri. Plus le score est haut, plus le facteur est impliqué dans ce processus.

L'ensemble des résultats obtenus par ces deux approches nous a permis une véritable avancée dans la compréhension du mécanisme de régulation de Svb par les peptides Pri, et sont présentés à la suite, sous forme d'article qui sera soumis à publication très bientôt.

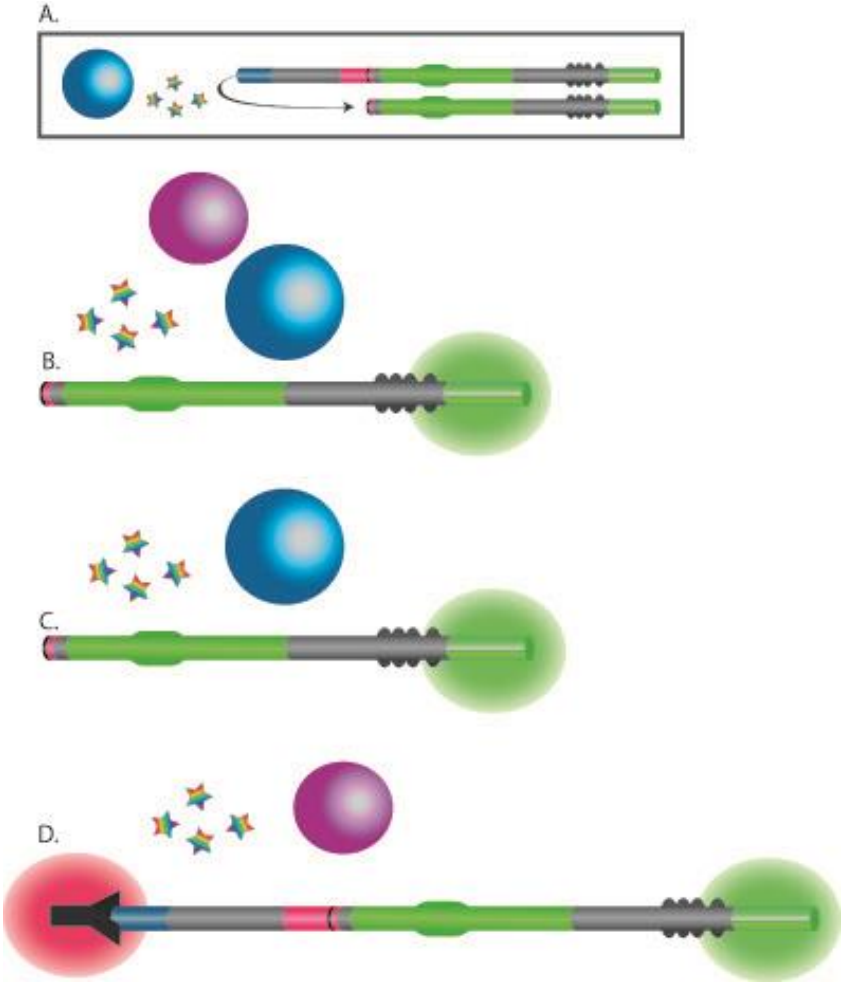


Figure 37 : Schéma de la stratégie du crible RNAi. A. Modèle dans lequel un facteur bleu serait nécessaire au processing de Svb, de sa forme longue à sa forme courte, en réponse à Pri (étoiles). **B.** Dans notre lignée cellulaire exprimant *svb*, *pri*, le facteur bleu et un facteur rose, Svb est clivé en réponse à Pri, et on peut observer un signal GFP (fusionnée en C-terminal). **C.** Après traitement avec un dsRNA dirigé contre le facteur rose, le facteur bleu est présent et Svb est clivé, on voit toujours un signal GFP. **D.** Après traitement avec un dsRNA contre le facteur bleu, les peptides Pri ne sont plus capables d'induire le clivage de Svb. Celui-ci sera resté sous sa forme longue, détectable avec l'anti-Svb1s. Ainsi la quantification des cellules positives pour l'anti-Svb1s parmi les cellules positives pour la GFP nous donnera accès au pourcentage d'inhibition de la réponse de Svb à Pri en fonction du facteur dont l'expression a été inhibée par dsRNA.

II. Résumé

Ces travaux réalisés sur l'étude du mécanisme de la régulation de Svb par les peptides Pri nous ont permis de définir qu'il s'agit d'un RUP (Ubiquitin-Proteasome dependent Processing), comme décrit précédemment pour p105. Nous avons par ailleurs identifié la E3 ubiquitin-ligase (Ubr3) impliquée dans ce processus, qui je le rappelle, est l'enzyme de la voie de l'ubiquitination qui confère la spécificité au système (Fig. 33).

L'expression des peptides Pri induit la formation d'un complexe entre Svb et Ubr3, et donc l'ubiquitination de Svb. Les analyses des séquences *cis*-régulatrices de Svb nous ont permis de définir son dégron, et les lysines cibles de l'ubiquitination. Svb est alors envoyé au protéasome 26S où il subit une dégradation partielle (i.e. élimination de sa région N-terminale contenant le domaine de répression transcriptionnelle). Nous avons identifié les régions dans sa partie C-terminale qui pourraient (en accord avec des prédictions de désordre protéique) jouer le rôle de blocs structuraux, selon le modèle proposé par Piwko (Fig. 36) (Piwko and Jentsch, 2006).

Un résultat très intéressant qui met fin au débat sur l'existence réelle des peptides Pri (que certains n'admettaient pas) est que nous avons pu reproduire *in vitro* l'induction du complexe Ubr3/Svb (obtenu par expression de *pri*), mais en ajoutant des peptides Pri que nous avons fait synthétisés. Ainsi, il n'y a plus l'ARNm *pri* et le résultat observé est clairement dû aux peptides eux-mêmes.

Ce modèle établi à partir d'expériences en cellules en culture S2 a été validé *in vivo*, pour la formation des trichomes épidermiques de la drosophile adulte, où des clones mutants pour *ubr3* ne sont pas capables de former ces extensions en dépit de l'expression de *pri* et *svb*.

III. Article (en préparation)

A. The UBR3 Ubiquitin Ligase Mediates the Role of Pri Small Peptides in Shavenbaby Processing

Title: The UBR3 Ubiquitin Ligase Mediates the Role of Pri Small Peptides in Shavenbaby Processing

Authors: E. Benrabah[1,2]†, J. Zanet[1,2]†, T. Li[3], A. Pélissier-Monier[1,2], B. Ronsin[1,2], H.J. Bellen[3], F. Payre[1,2]*, S. Plaza [1,2]*.

Affiliations:

[1] Université de Toulouse, UPS, Centre de Biologie du Développement, Bâtiment 4R3, 118 route de Narbonne, F-31062 Toulouse, France.

[2] CNRS, UMR5547, Centre de Biologie du Développement, F-31062 Toulouse, France.

[3] Program in Developmental Biology, Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Neurological Research Institute, Baylor College of Medicine, Houston, Texas 77030, USA.

*Correspondence to: francois.payre@univ-tlse3.fr, serge.plaza@univ-tlse3.fr

† these authors contributed equally to this work

Abstract

A wide variety of RNAs contain small Open-Reading-Frames (smORF), but the existence of smORF-encoded peptides and their putative mode of action remain elusive in the absence of functional evidence. The *polished rice (pri)* RNA, which only encodes smORFs, is required during *Drosophila* development to trigger the conversion of a transcription factor, Shavenbaby, from a repressor to an activator by an unknown mechanism. Here we show that Pri smORF peptides act to address Svb at the proteasome, where a degradation limited to the N-term region ensures Svb maturation. Genome-wide screening further identified the critical role of an E3 ligase, UBR3, that ubiquitinates Svb N-terminus and thereby elicits proteasome processing. We find that Ubr3 binds to Svb, in a Pri-dependent manner, as demonstrated by using synthetic Pri peptides. These results provide mechanistic insights into the activity of Pri peptides and open new ways to explore the functions of similar smORF-encoded molecules.

High-throughput genomics has shown the prevalence of atypical RNAs such as long non-coding (lncRNAs) that lack the classical hallmarks of protein-coding genes (1). However, experimental evidence indicates that some contain functional small ORFs (smORF) (2), from yeast or plants to mammals (3-6). *polished rice/tarsal-less (pri),* initially thought to be a lncRNA (7), encodes four similar smORF peptides (11-32aa) that are required at several steps of development across insect species (8, 9). In flies, the Shavenbaby (Svb) TF was identified to be a main target of Pri peptides, underscoring their role in epidermal differentiation as they are required for trichome formation (10, 11). Pri peptides induce a post-translational cleavage of the Svb protein, converting a full-length transcriptional repressor into a shorter activator (Fig. 1A) (10). Once activated by this processing, Svb turns on the transcriptional program of trichomes by the direct activation of a battery of effectors (12, 13). Here we identify the molecular mechanisms by which Pri peptides exert their activity.

To identify factors required for Svb cleavage in response to Pri peptides we performed a genome-wide RNAi screen. We used a stable line of *Drosophila* cells that faithfully reproduces the proper cleavage of Svb, as monitored with an antibody specific to the Svb repressor (anti-Svb1s) (Fig. 1A, S1). In control conditions, Svb is matured and the anti-Svb1s signal is near background levels, while GFP (Svb::GFP) provides an internal control for Svb expression (Fig. 1B). We reasoned that RNAi depletion of factors required for Svb processing would increase the anti-Svb1s signal, therefore providing a positive screening of all *Drosophila* genes. We set up an automated assay quantifying Svb maturation, with an inhibitory score reflecting the proportion of cells unable to release the Svb N-terminus (see Sup Mat). *pri* RNAi displayed the highest score, confirming the suitability of our approach to identify players required for Svb processing (Fig. 1B and Table S2). Different methods used to evaluate genome-wide results all converged towards a key role for the proteasome. For example, COMPLEAT, a recently developed framework based on protein complex analysis (14), identified the proteasome in 55 out of the 62 top complexes (p-value 1,4E-14 to 7,4E-4; see Table S1). A survey of individual proteasome subunits further shows that their depletion impairs Svb maturation, including members of the 19S regulatory particle and the 20S catalytic core (Fig. 1B, C, Table S2). Chemical inhibitors provided an independent way to confirm the role of the proteasome. Indeed, treatment with either MG132 or epoxomicin inhibited Pri-induced Svb maturation (Fig. 1D). These data therefore provide compelling

evidence that Pri peptides require proteasomal activity to trigger the post-translational processing of Svb.

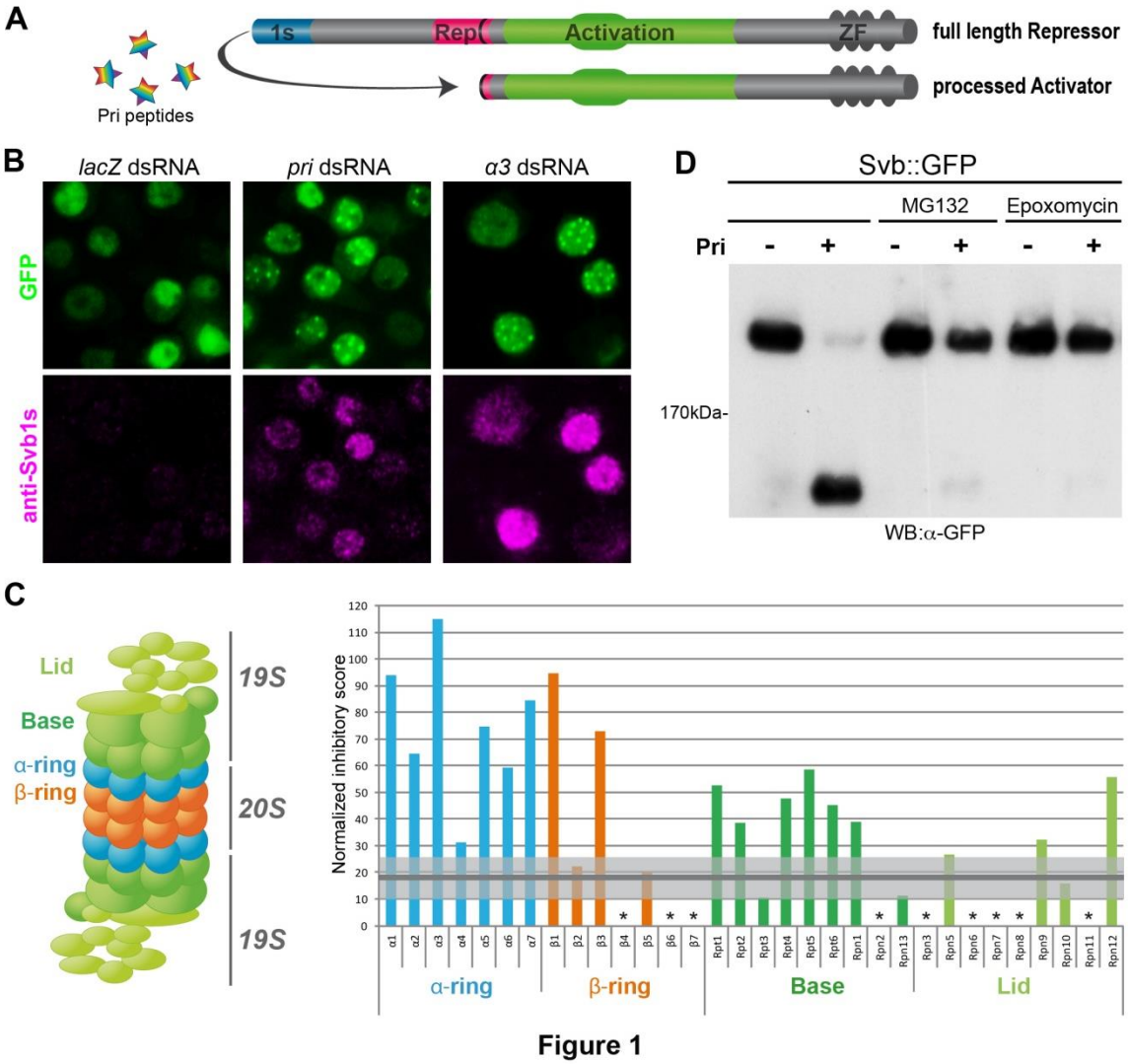


Figure 1

Fig. 1: Pri-dependent processing of Svb relies on the proteasome activity. A) The Svb transcription factor is made as a large repressor protein. Pri peptides induce a post-translational processing, releasing the N-term region that includes a repressor domain (red), thereby switching Svb into a shorter transcriptional activator. **B**) A stable S2 cell line allowing controlled expression of Svb::GFP and Pri was used for genome-wide dsRNA screening. Cells were stained for anti-Svb1s (purple) antibody specific of Svb N-terminus, GFP did not need signal amplification.. **C**) Scheme of the 26S proteasome complex, with different particles being color-coded. The chart plots inhibition scores for each proteasome subunit (see Table S2), with the mean score of all *Drosophila* genes (black) and two standard intervals (grey). * cell death prevented analysis of Svb processing. **D**) The chemical proteasome inhibitors MG132 & Epoxomicin inhibit Pri-dependent processing of Svb, as visualized by western blot using anti-GFP.

The proteasome being involved in a myriad of cell functions (15), as the proteasome may be directly or indirectly involved in the activity of Pri peptides, we sought to discriminate between these possibilities by delineating the region(s) that implement the response to Pri within the Svb protein: a degron. The Svb OvoA isoform that lacks the 137 N-term residues of Svb (Fig. 2A) is insensitive to Pri (10), suggesting a key role for that domain. Systematic deletions further demonstrate the importance of the first 31 N-term aa for Pri response, since the N31 mutant that retains this motif is properly processed (Fig. 2A). Reciprocally, mutants that truncate this motif (N11, N21) display impaired Svb processing and a mutant only lacking the 31aa (Δ31) is refractory to Pri (Fig. 2A). We therefore assessed whether this region can transform an unrelated factor into a Pri target and found that fusing the Svb N-terminus to GFP is sufficient to provide sensitivity to Pri peptides (Fig. 2B). However, unlike Svb, 1S::GFP became fully degraded upon *pri* expression (Fig. 2B, S1B), suggesting that additional domains are required for proper Svb processing. Accordingly, large C-term truncations (1-532, 1-500, 1-468 or 1-445) caused a Pri- (and proteasome-) dependent full degradation (Fig. 2C, S2A) whereas smaller C-term truncations (1-1069, 1-869 and 1-701) retained significant processing in response to Pri, albeit with an altered pattern when compared to wild type (Fig. 2C, S2A). We noticed that the N-terminus of Svb bears features of intrinsically disordered regions whereas the C-term moiety (including the ZF domain) contains regions (Fig. S2B). We therefore postulated that a similar mechanism might prevent the full degradation of Svb once targeted to the proteasome. Adding back Svb ZFs to the 1-445 mutant reverted its full degradation, rescuing the production of shorter product in the presence of Pri (Fig. 2C, S3). We obtained similar results with the DNA-binding domain of Gal4 (Fig. 2C,S3), showing that a heterologous structured domain is able to stop proteasome degradation and restore processing. In sum, these data identify separate protein regions that are indispensable for the post-translational processing of Svb in response to Pri peptides: i) the 31 N-term residues that are primarily responsible for proteasome targeting, ii) C-term regions that prevent full degradation and enhance processing.

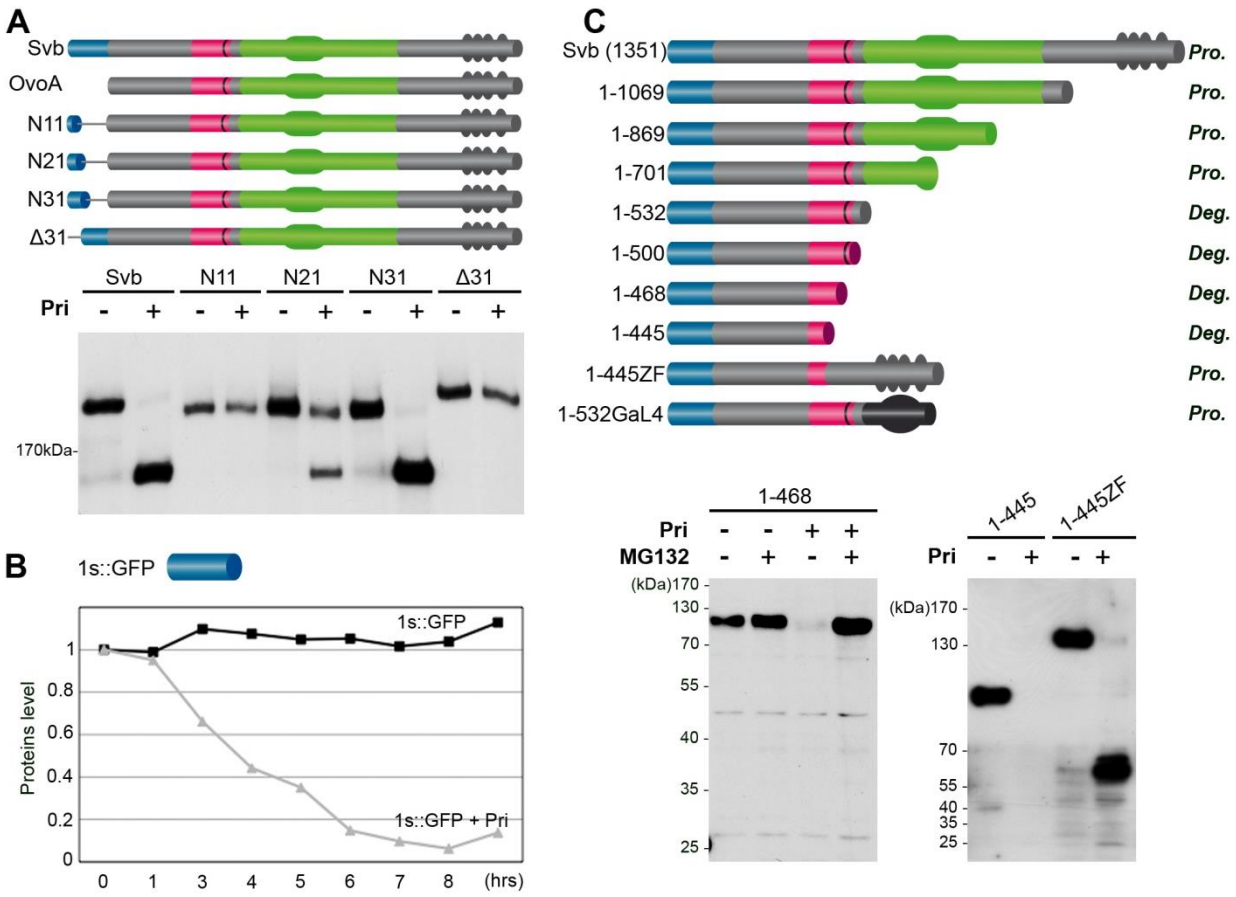


Figure 2

Fig. 2: Separate regions of the Svb protein instruct proper processing

A) Scheme of Svb mutants and the germline-specific OvoA isoform that lacks the Svb N-term region (1s, blue). Processing of Svb mutants engineered within the 1s region was analyzed by western blot. **B**) Time course of 1s::GFP protein level following the onset of *pri* expression, when compared to control conditions. **C**) In response to Pri, Svb mutants bearing C-terminal truncations display either processing (*Pro*) or full degradation (*Deg*), both being inhibited by MG132 (see Fig. S2, S3). Fusion of the Zinc Finger (ZF) domain to the 1-445 mutant prevents full degradation and restores processing.

The canonical proteasome pathway degrades proteins bearing ubiquitinated lysine residues (15). A remarkable feature of the Svb N-terminus is its strong conservation, from arthropods to human (Fig. 3A). Of note, there is the presence of two invariant lysines (K3 and K8) and a third one at a less constrained position (K28 in flies). We therefore examined whether they participate in Svb processing. Although individual lysine substitutions have weak or no effect, simultaneous mutations of K3-K8, or all three together (3Kmut), abolish Svb processing (Fig. 3A). Furthermore, we detected strong ubiquitination of Svb in the presence of Pri and a proteasome inhibitor (Fig. 3B). Importantly, Pri-induced ubiquitination is no longer detected in the 3Kmut mutant, revealing a key role for these lysines in Svb ubiquitination in response to Pri (Fig. 3B).

Ubiquitin conjugation requires three enzymes, E1, E2 and E3. Much specificity comes from E3 ligases that directly bind the substrate, and to a lesser extent from E2s (REF). Consistently, there is a single E1, but 31 E2 and hundreds of predicted E3 in *Drosophila*. When we systematically examined all putative E2 and E3 in flies, Ubr3 and UbcD6 are the only two enzymes that show high scores, indicating that Ubcd6 and Ubr3 are the main ubiquitin enzymes required for Svb processing (Fig. S4, Table S3). The Ubr3/UbcD6 mammalian counterparts have been shown to interact (16), suggesting that they belong to the same complex. These results therefore suggested that Ubr3 was the specific E3 ligase mediating Svb ubiquitination, a hypothesis we tested in a series of experiments. First, RNAi knockdown of *ubr3* in S2 cells significantly inhibits Svb cleavage, as further shown by western blots (Fig. 3C,D). Second, reciprocal immunoprecipitation assays revealed that Svb interacts with Ubr3 and, importantly, this depends on Pri (Fig. 4A). Third, this interaction is stabilized upon proteasome inhibition, allowing the detection of Ubr3 in complex with ubiquitinated Svb (Fig. 4A). Finally, this assay allowed us to explore the mechanistic role of Pri peptides in the formation of Svb/Ubr3 complex. In protein extracts from cells that do not express *pri*, UBR3 does not interact with Svb (Fig. 4A). We found that complementing these extracts with synthetic Pri peptides added *in vitro* is sufficient to promote Ubr3/Svb interaction, in a dose-dependent manner (Fig. 4B). Providing additional evidence for specificity, a scrambled peptide of same aminoacid composition was devoid of activity (Fig. 4B). We therefore conclude that Pri peptides directly promote the binding of Ubr3 to the Svb N-terminus, explaining how they subsequently trigger Svb ubiquitination.

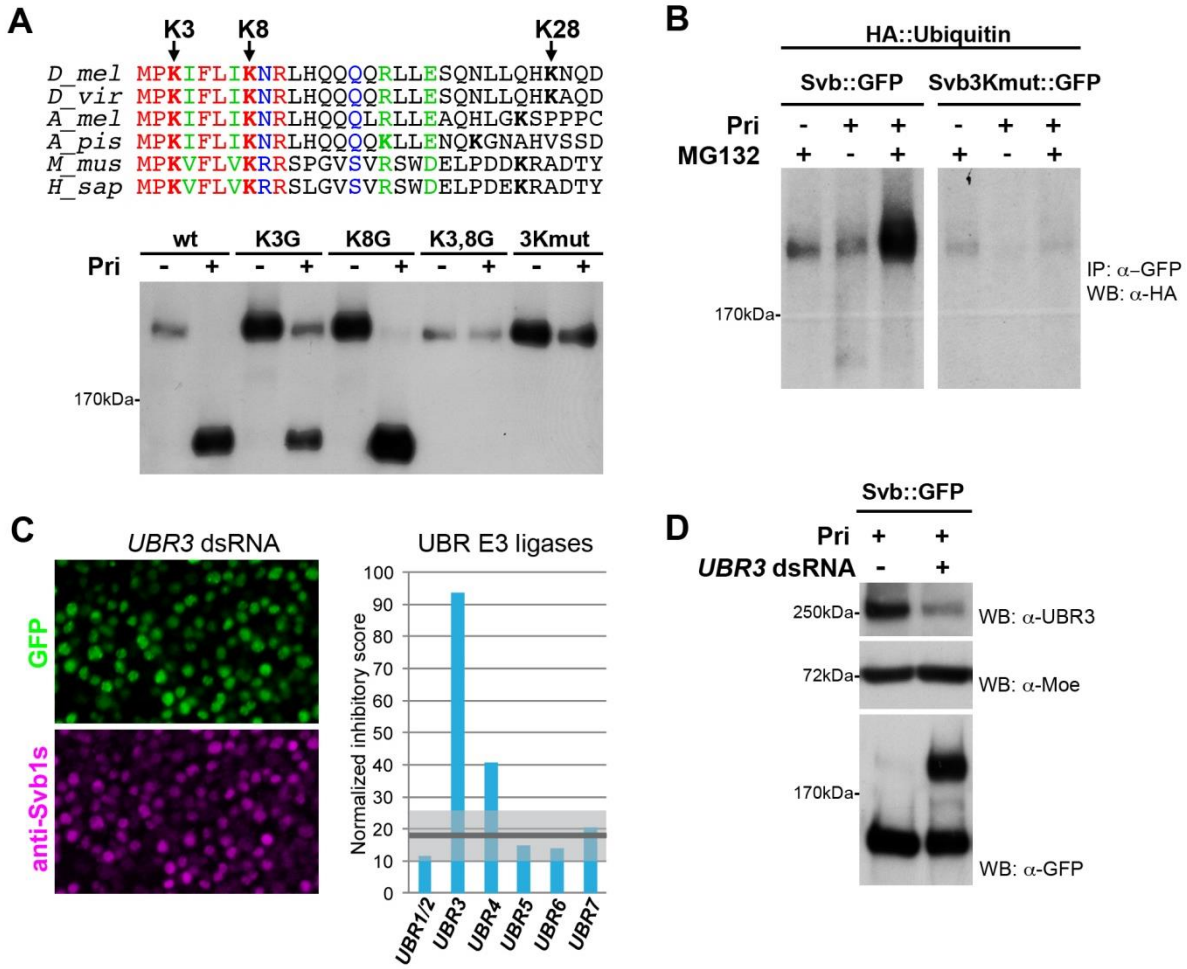


Figure 3

Fig. 3: Svb processing requires UBR3-mediated ubiquitination

A) Alignment of Svb N-term sequences across animals. Point mutations of lysines (3Kmut) inhibit Svb processing in response to Pri. **B**) Conjugation of ubiquitin::HA to the Svb protein is strongly increased upon *pri* expression when combined to proteasome inhibition. The 3Kmut variant is refractory both to processing and to ubiquitination. **C**) Among putative E3 ligases of the Ubr family, Ubr3 displays the highest inhibitory score. **D**) dsRNA knockdown decreases endogenous Ubr3 protein level and impairs Pri-dependent Svb processing.

To investigate the *in vivo* relevance of these mechanisms, we isolated an *ubr3* null allele (17) and assayed whether *ubr3* loss of function affects Svb activity in whole animals (Fig. 4C). Similar to what we observed in the absence of *pri* or *svb* (x, y), we found that *ubr3* mutant cells are unable to form epidermal trichomes (Fig. 4C). Furthermore, the lack of *ubr3* prevents Svb processing, as shown by cell autonomous persistence of the uncleaved repressor form of Svb in *ubr3* mutant cells (Fig. 4C).

In summary, the data show that Ubr3 is an obligate mediator of Pri peptides action for the processing of Svb, leading us to propose the following model (Fig. 4D). The main role of Pri peptides is to enable Ubr3 binding to the N-term region of Svb, a result reminiscent of UBR1 activation by dipeptides in yeast (Du et al, 2002). Ubr3/Ubcd6 then catalyze the ubiquitination of three lysines (K3, 8 & 28), leading to proteasome targeting. Release of the N-term repressor results from a partial degradation, limited by intrinsic properties of C-term regions of the Svb protein likely acting as structural inhibitors. These findings demonstrate that regulation of the Ubr3 protein ligase activity plays a key role in the proteasome-mediated activation of a transcription factor by small peptides. It is possible that additional *pri* targets as of yet to be identified proteins may be regulated through the same mechanisms. This study represents one of the very few cases for which a molecular function has been assigned to smORF-encoded peptides (4, 6). Hence, smORF peptides may provide an unexplored reservoir of peptidic interfaces, well suited to bind to and modify the activity of a wide range of cellular enzymes.

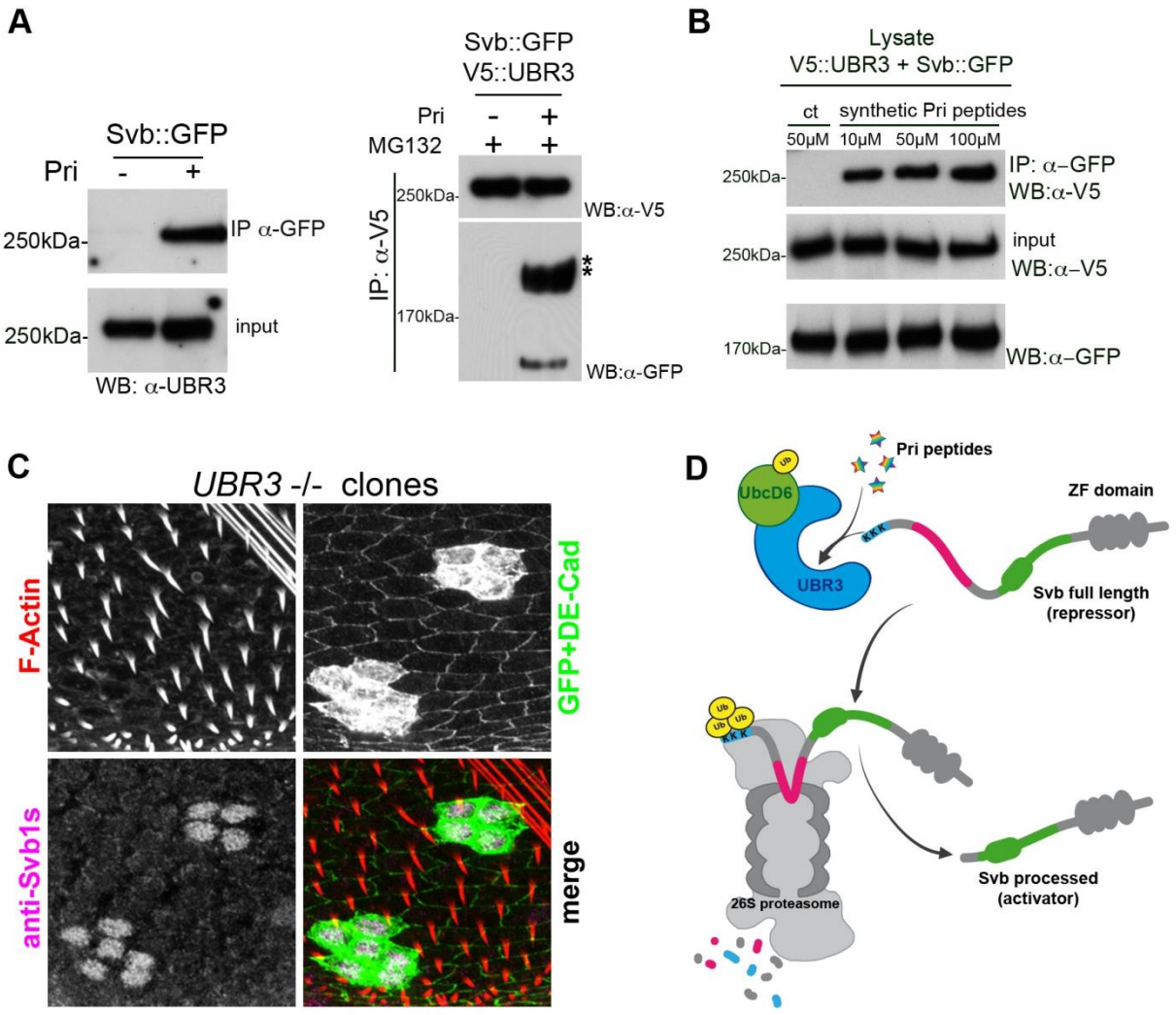


Figure 4

Fig. 4: Pri peptides induce Svb/UBR3 protein interaction

A) Immunoprecipitation of Svb::GFP complexed with endogenous Ubr3 (left), or V5::Ubr3 interacting with Svb::GFP (right), occurs only in the presence of Pri. Note the enrichment of ubiquitinated Svb::GFP(*) bound to Ubr3, following proteasome inhibition. **B**) Protein extracts from cells expressing V5::Ubr3 and Svb::GFP in the absence of *pri*, were subsequently incubated *in vitro* with increasing concentration of synthetic Pri peptides. A scrambled peptide was used for control (ct). **C**) *ubr3-/-* clones (GFP-positive cells, green) in the pupal thorax were stained for anti-Svb1s (purple) and actin (red). Cadherin::GFP (DE-Cad, green) highlights cell contours. **D**) Model of Svb processing. Pri peptides promote Ubr3 binding to the Svb N-terminus, resulting in ubiquitination and ultimately proteasome targeting. The Svb N-terminus, predicted to be intrinsically disordered, is fully degraded. C-term regions, likely acting as structural blocks (green and gray bulges) to prevent complete degradation, allow the release of processed Svb activator.

References and notes

1. M. B. Clark, J. S. Mattick, Semin Cell Dev Biol 22, 366 (2011).

2. S. J. Andrews, J. A. Rothnagel, Nat Rev Genet 15, 193 (2014).

3. K. Hanada et al., Proc Natl Acad Sci U S A 110, 2395 (2013).

4. K. Hanyu-Nakamura, H. Sonobe-Nojima, A. Tanigawa, P. Lasko, A. Nakamura, Nature 451, 730 (2008).

5. J. P. Kastenmayer et al., Genome Res 16, 365 (2006).

6. E. G. Magny et al., Science 341, 1116 (2013).

7. S. Inagaki et al., Genes Cells 10, 1163 (2005).

8. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, J. P. Couso, PLoS Biol 5, e106 (2007).

9. T. Kondo et al., Nat Cell Biol 9, 660 (2007).

10. T. Kondo et al., Science 329, 336 (2010).

11. F. Payre, A. Vincent, S. Carreno, Nature 400, 271 (1999).

12. H. Chanut-Delalande, I. Fernandes, F. Roch, F. Payre, S. Plaza, PLoS Biol 4, e290 (2006).

13. D. Menoret et al., Genome Biol 14, R86 (2013).

14. A. Vinayagam et al., Sci Signal 6, rs5 (2013).

15. A. Varshavsky, Annu Rev Biochem 81, 167 (2012).

16. T. Tasaki et al., J Biol Chem 282, 18510 (2007).

17.Yamamoto S*, Jaiswal M* *et al.* (submitted) *

Acknowledgements:

We thank N. Perrimon and the Drosophila Screening Center; Y. Latapie and P. Valenti for technical assistance; the RIO imaging platform. The authors declare no conflicts of interest. This work was supported by the CNRS, the Agence Nationale de la Recherche (ANR blanc "smORFpep") and the Fondation pour la Recherche Medicale (EB).

B. Matériels et méthodes

Cell culture

Schneider S2 cells were grown in Schneider medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin (Invitrogen). We used S2 stable cell lines co-expressing copper-inducible transgenes pMTpri + pMTSvb::GFP ("1B"), and pMTpri1-4fs + pMTSvb::GFP ("FS") that expresses a frame-shifted variant as a negative (Kondo et al, 2010).

Cloning procedures

All the mutant forms of Svb were generated from pAc-Svb::GFP (Kondo et al, 2010). Exon 1s mutations: a pAc-SvbK7 shuttle vector with unique restriction sites Acc65I and HpaI flanking exon 1s was generated from pAc-Svb::GFP by site-directed mutagenesis (QuickChange kit, Stratagene). Multiple mutations of exon 1s (deletions or punctual mutations) were generated by site-directed mutagenesis by primer extension using 4 primers. External primers 5' CCCCGGATCGGGGTACCTAC and 3' TCGCTGATTTTTTGGTATGC are used for all constucts. Internal specific primers are listed in the table below. Amplified mutated exons 1s are cloned in the pAc-SvbK7 at Acc65I and HpaI sites. C-terminal deletions: sequences corresponding to C-terminal deletant forms of Svb are amplified by PCR using 5' CCCCGGATCGGGGTACCTAC primer and various 3' specific primers listed in table below. PCR fragments were cloned at SmaI sites in pAc-Svb::GFP.1-445ZF and 1-532ZF: zinc fingers region corresponding to exons 3 and 4 were amplified by PCR using 5' CGATCGGACCGCGGACTTGGGCTTGCCGCCGGATCTGC and 3' TCCGATCGCCATGGTGGCGATGGATACTGGCATGGGCAGCTGGCC primers and cloned in 1-445 and 1-532 constructs at SacII and XcmI restriction sites. 1-532Gal4: Gal4 DNA binding region was amplified by PCR using 5'CGATCGGACCGCGGACATGAAGCTACTGTCTTCTATCG and 3' TCCGATCGCCCATGGTGGCGATGGACGATACAGTCAACTGTCTTTG primers and cloned in 1-532 construct at SacII and XcmI restriction sites. All constructs were verified by sequencing.

Exon1S mutations	Punctual mutations	K3G	5'	GATAATCCATCATGCCGGGGATTTTCCTGATC	codon 3 : AAG → GGG
			3'	GATCAGGAAAATCCCCGGCATGATGGATTATC	
		K8G	5'	GAAGATTTTCCTGATCGGAAATCGCCTGCATC	codon 8 : AAA → GGA
			3'	GATGCAGGCGATTTCCGATCAGGAAAATCTTC	
		K3K8G	5'	CCATCATGCCGGGGATTTTCCTGATCGGAAATCGCCTG	codon 3 : AAG → GGG codon 8 : AAA → GGA
			3'	CAGGCGATTTCCGATCAGGAAAATCCCCGGCATGATGG	
		3Kmt	5'	AGCAACAGCGCTTGCTTGAATCGCAAAACTTGCTGCAGCACGGGAATCAGGATGATGAGC	codon 3 :AAG → GGG codon 8 : AAA → GGA codon 28 : AAG → GGG
			3'	TTCAAGCAAGCGCTGTTGCTGCTGATGCAGGCGATTTCCGATCAGGAAAATCCCCGGCATGATGG	
	Deletions	N11	5'	GATTTTCCTGATCAAAAATCGCGAAAACTTTGCCGCCGAGTTG	Red : upstream sequence of deleted region Black : downstream sequence of deleted region NB : codon 1 and 2 are conserved for Δ31
			3'	CAACTCGGCGGCAAAGTTTTCGCGATTTTTGATCAGGAAAATC	
		N21	5'	CAACAGCGCTTGCTTGAATCGGAAAACTTTGCCGCCGAGTTG	
			3'	CAACTCGGCGGCAAAGTTTTCCGATTCAAGCAAGCGCTGTTG	
		N31	5'	CTGCAGCACAAGAATCAGGATGAAAACTTTGCCGCCGAGTTG	
			3'	CAACTCGGCGGCAAAGTTTTCATCCTGATTCTTGTGCTGCAG	
		Δ31	5'	GCAAATCAACCGATAATCCATATGCCGGAGCGCTTGGTGCCGCCCCTC	
			3'	GAGGGGCGGCACCAAGCGCTCCGGCATATGGATTATCGGTTGATTTGC	
C-terminal deletions		1-1069	3'	GCGGTACCCCCGGGGCTTTGGATGAGTGGCTTTGTGTCC	Green : coding sequence corresponding to C-terminal end of deletant forms
		1-869	3'	GCGGTACCCCCGGGATGCTGCTGTTGCACATTGTGCTGC	
		1-701	3'	GCGGTACCCCCGGGAAAGGTGCCAATATCGATCTTGGCC	
		1-532	3'	GCGGTACCCCCGGGGCCGCTCTTGAGGCTCTCATAAAACG	
		1-500	3'	GCGGTACCCCCGGGCATACCACCACCGCCACCGGCACC	
		1-468	3'	GCGGTACCCCCGGGACCGGACGATTGAAAGCCCTGCCCG	
		1-445	3'	GCGGTACCCCCGGGCGCAGCCAAAATACCGGGCAAGGATC	
		1S	3'	GCGGTACCCCCGGGTTCATTTTGAACTTGGCCAGTGGC	

Drosophila cell-based RNAi screen

The RNAi screen has been performed at the Drosophila Screening RNAi Center (DSRC) (http://www.flyrnai.org/, Flockhart et al., 2012) on the 1B cell line according to their recommendations. We used the DSRC dsRNAs collection, directed against the 13900 drosophila genes, distributed in 66 assay 384-well plates (Perkin Elmer Opera compatible). 0.25µg of each dsRNA in 5µL of water was distributed in each well. We added 5µL of dsRNA against *pri* (0,05µg/µL) in the 4 empty wells destined to positive control. 15000 cells are added in 40µL of complete Schneider's medium per well. Pri and Svb::GFP expression was induced after 72h by adding 5µL of 10mM CuSO4 (final concentration of 1mM). 18h later, cells were fixed in 4% paraformaldehyde in phosphate-buffered saline

(PBS) for 30 min, then permeabilized and saturated 1h in PBS with 0.1%Triton X-100 (PBT 0,1%) and 1% bovine serum albumin (BSA). Primary anti-Svb1s (1:3000) and secondary anti-rabbit Alexa Fluor® 555 (1:1000) antibodies were incubated for 2 hours in PBT 0,1% with 0,1% BSA. Cells were washed in PBT 0,1%. Plates were imaged with an automated high-throughput confocal microscope (PerkinElmer Evotec Opera). For each well, seven pictures of different fields were recorded with a 20× objective, sequentially acquired in green (Svb::GFP) and red (anti-Svb1s) channels, and converted from .flex to .tiff with the Evotec Acapella image analysis software. Pictures were further analyzed with ImageJ software. For each position, green and red pictures were thresholded and transformed into binary masks. Then, particles corresponding to positive cells (in term of size and circularity) were counted. We applied a boolean operation "AND" on green and red masks, particles found were counted automatically. For each well, only the 5th positions exhibiting the highest number of GFP-positive cells were kept for further analysis. Medians of GFP-positive cell numbers (medG) and "GFP and anti-Svb1s positive cells" numbers (medG&R), and their ratio (medG&R/medG) were calculated. Wells with medG<50 (not enough surviving cells) were excluded from the analysis. For each plate, we applied a normalization of these ratios, in order to have a mean score of 100 for the 4 positive controls. We obtained scores for each dsRNA (performed in duplicate) corresponding to the 13900 Drosophila genes. The 380 top genes were re-tested in a "cherry-pick" step, in sextuplates, and analyzed with the same method, giving rise to a list of 84 top genes (positive in the six replicates).

UBR3 antibody production

The cDNA sequence encoding 751-1500 amino acids of UBR3 was cloned into pET21 expression construct. Purified inclusion bodies were used to immunize guinea pigs.

UBR3 cDNA cloning

UBR3 cDNA was constructed from exon sequences and cloned into pUASTattB using a GENEART Seamless Cloning and Assembly Kit (Life Technologies). The cDNA sequence was verified by sequencing and identical to 1074..7733 of NM_132200.4.

Immunoprecipitation and Western blotting

S2 cells were plated in 6-well plates ($1{,}5x10^6$ cells/3mL) and transfected in 100 µL of Opti-MEM (Invitrogen) with 3 µL of Fugene^{HD} (Promega) with a mixture of 1µg of indicated

constructs. pMT plasmid expression was induced by $CuSO_4$ at the final concentration of 0.5 mM. The following plasmids were used: pAc-Svb::GFP and its mutated versions, pMT-pri (Kondo et al, 2010), pAc-V5::Ubr3 and pAc-HA::Ubiquitin (kind gift from P Meier and N Tapon labs). Cells were lysed in RIPA buffer (50mM TrisHCl pH 7,5, 150mM NaCl, 1mM EDTA, 1% NP40, 1% deoxycholate, 0,1% SDS, protease inhibitors) for immunoprecipitation, or in NP40 buffer (20nM Tris-HCl pH 7,5, 150mM NaCl, 4mM $MgCl_2$, 0,5% NP40, protease inhibitors) for co-immunoprecipitation assays. To chemically block proteasome activity, cells were treated for 4 hours with MG132 (Calbiochem) at 25µM, or with Epoxomycin (Sigma) at 1µM. Tagged proteins were immunoprecipitated with an anti-GFP antibody (Chromotek) or with an anti-V5 antibody (Invitrogen), according to manufacturer's protocols. Immunoprecipitated samples were analyzed by Western blot, using the NuPAGE system (Invitrogen). Proteins were detected using anti-Svb1s (1/10000), anti-GFP (TP401, Acris Antibodies, 1/10000), anti-HA (Covance, 1/2000), anti-V5 (Invitrogen, 1:2500), anti-Ubr3 (1/1000) and anti-Moesin (Carreno et al, 2008, 1/100000) antibodies. Secondary antibodies were anti- mouse or anti-rabbit IgG-HRP conjugates (Jackson Laboratory, 1/10,000) and detected using ECL pico kit (Pierce) or Lumilight ^{Plus} kit (Roche) and Amersham hyperfilm ECL (GE healthcare). 1S::GFP and Moesin protein levels on estern blot were quantified with ImageJ software.

Fly stocks and clonal analysis

Fly strains used for Flip-out experiments were tub-Gal80, hsFLP, FRT19A ; act-Gal4, UASmCD8GFP and Ubr3b, FRT19A/DP(1;Y). The Ubr3b allele was generated by EMS induced mutagenesis (Yamamoto et al, 2014) and is considered as null as it introduces a stop codon at Leu^{788}. Homozygous clones Ubr3b were induced *via* FLP-FRT recombination, during the second and third instar larvae. Heat shock was performed for 1h at 37°C. White pupae were picked and timed at 25°C. Pupal thorax were dissected 42h after pupal formation, fixed in 4% paraformaldehyde 1x PBS, and stained for F-actin, Svb-1s and DE-Cad. To reveal F-actin, thorax were incubated 30 min in PBT 0,1% + Phalloidin-TRITC 1/100. For Svb-1s and DE-Cad stainings, primary anti-Svb1s was used at 1/2500 and anti DE-Cad (hybridoma bank) at 1/100 in PBT 0,1% with 0,1% BSA (overnight) and revealed using an anti-rabbit 647 and anti rat 488 secondary antibodies (AlexaFluor®, 1/1000, 2h), respectively. Images were acquired on a laser-scanning confocal microscope (LSM710, Zeiss).

C. Figures supplémentaires

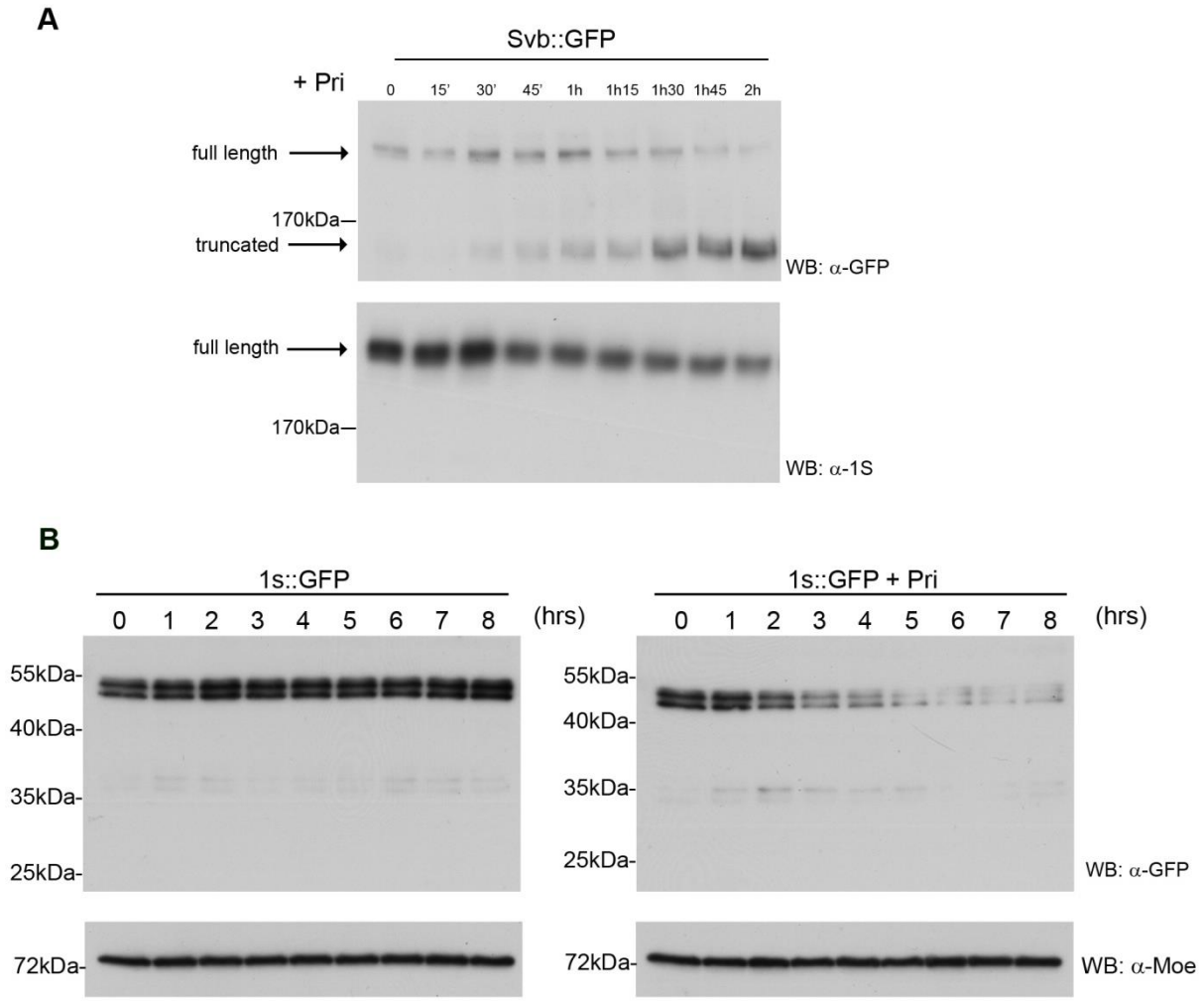


Figure S1

Figure S1: A) Time course of Svb::GFP protein following the onset of *pri* expression. Protein extracts were immunoblotted with anti-GFP and anti-Svb1s antibodies. After 1 hour, the short form of Svb starts to be detected. After 2h30, the total amount of full length Svb proteins have been cleaved. B) Time course of 1s::GFP protein following the onset of *pri* expression. Moesin was used as the loading control. In Fig.2, quantities of proteins were measured with ImageJ software, the ratio 1s::GFP/Moesin were calculated and normalized to 1 at time point 0.

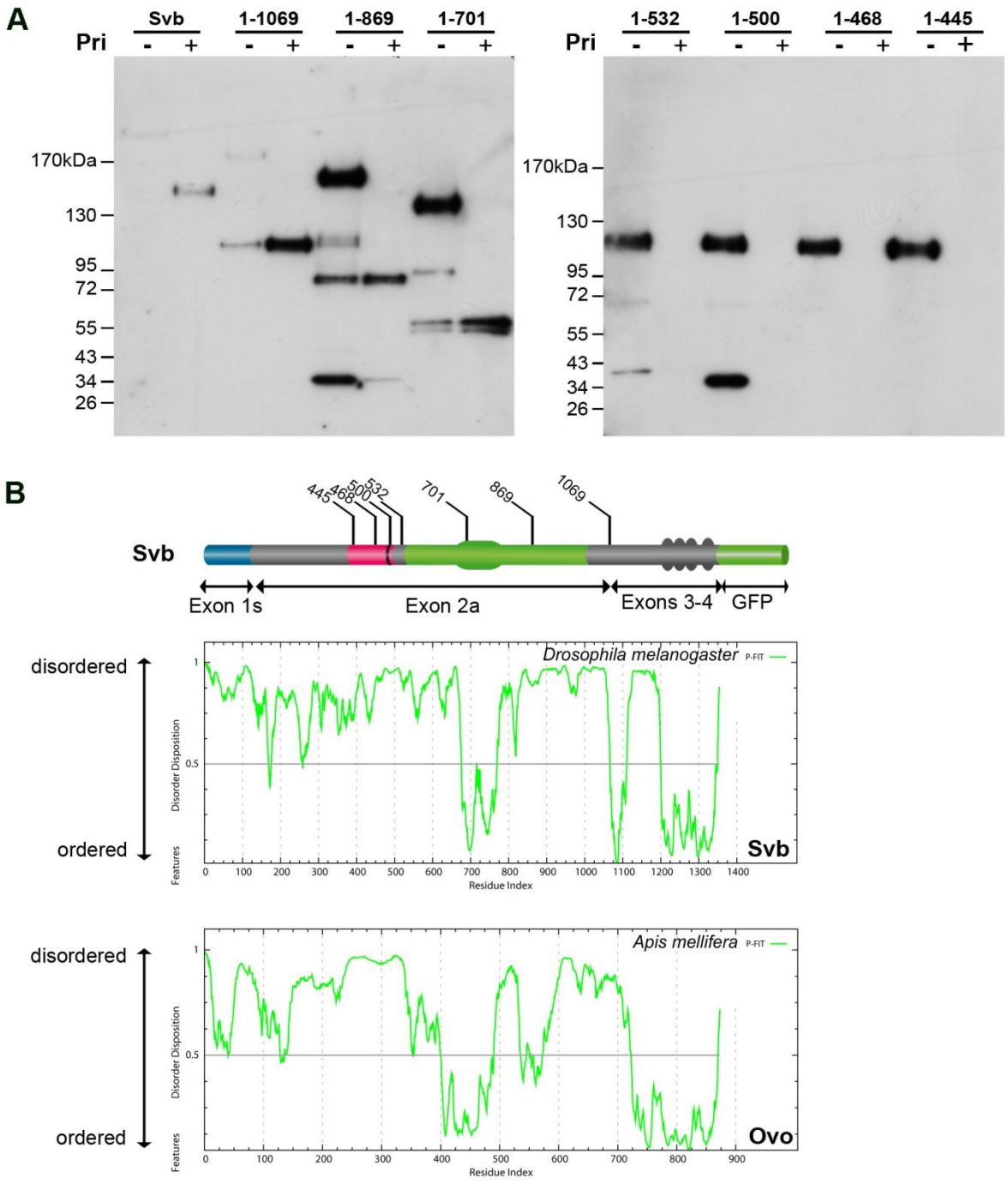


Figure S2

Figure S2: A) Western blot analyses of protein extracts from S2 cells co-transfected with Svb::GFP and its mutant forms with or without Pri and probed with anti-GFP. Progressive C-terminal deletions destabilize the protein in presence of Pri peptides resulting in the complete degradation. Corresponding graphs are displayed in Fig. 2C. B) On top, Svb protein with boundaries of C-terminal deletions of Svb highlighted with the amino acid number. Bottom, Graphs represent the ordered/ disordered amino acids region of Svb protein and its ortholog in *Apis melifera*. Note the most ordered regions are localized in the exon 2A around the amino acid 700 and in the Zinc Finger (grey bulges in Exons 3 and 4).

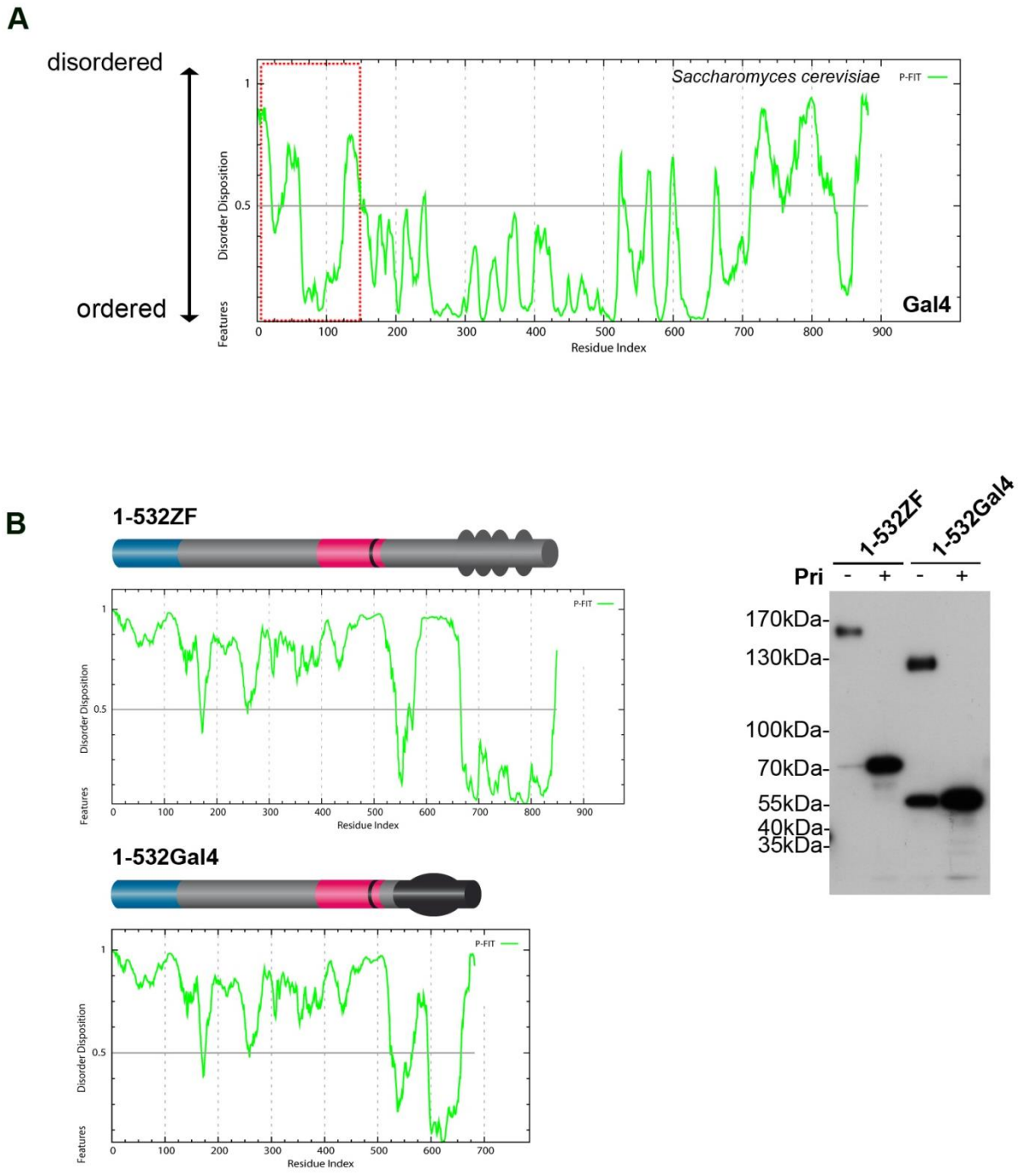


Figure S3

Figure S3: A) Graph represents the predicted ordered/disordered amino acids regions of GAL4 protein of *Saccharomyces cerevisae*. The red region corresponds to the DNA binding domain (DBD) we fused to 1-532 contruct. B) Inserting Zinc Finger (ZF, encoded by exons 3 and 4) or the unrelated Gal4 DBD (black bulge) sequences in the fully degraded 1-532 Svb mutant is able to prevent or limit proteasome degradation.

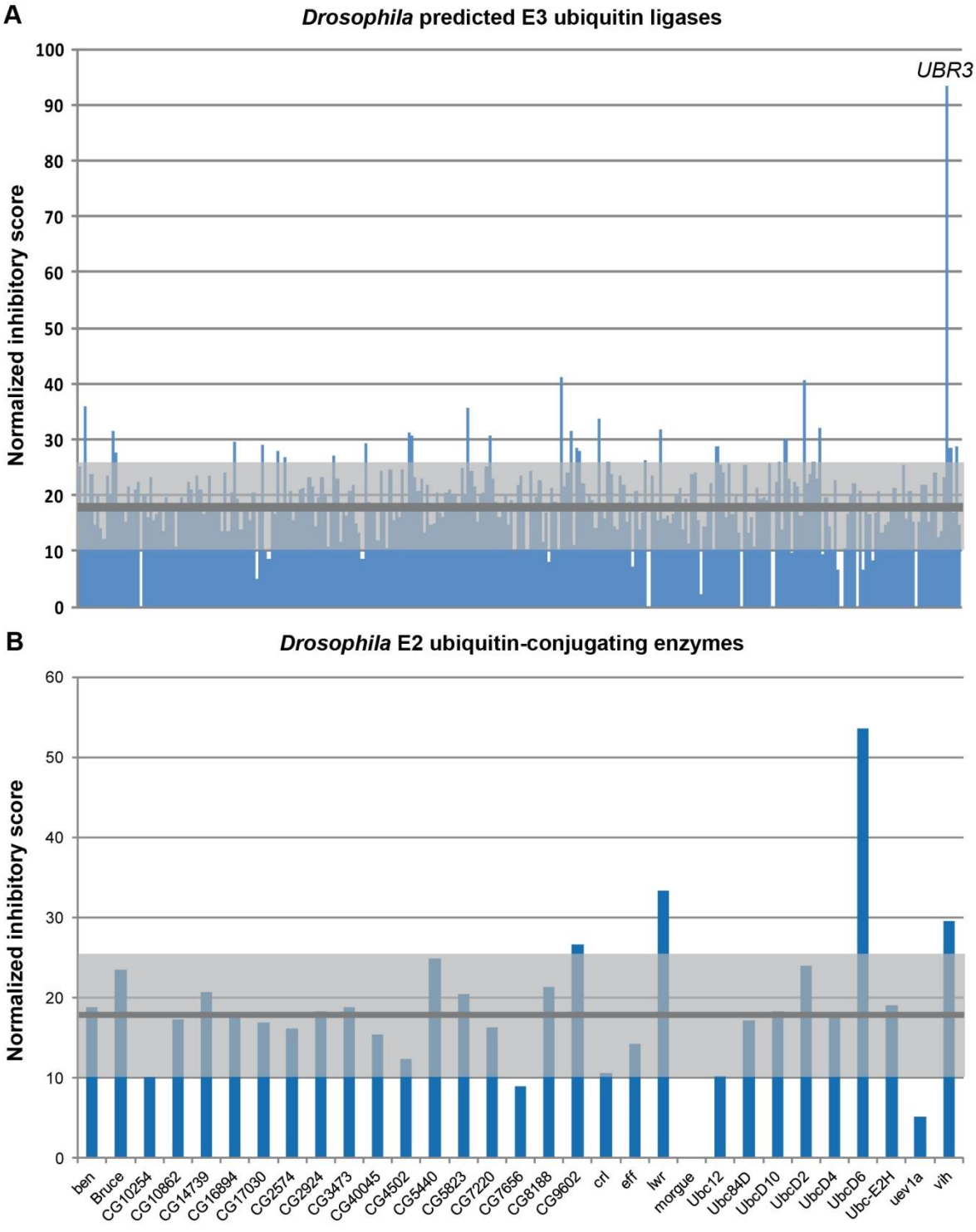


Figure S4

Figure S4: Chart plots levels of inhibitory scores for each predicted E3 ubiquitin ligases (A) and E2 ubiquitin-conjugating enzymes (B) (see Table S3), with the mean score of all *Drosophila* genes (black) and two standard intervals (grey). * massive cell death prevented analysis of Svb processing. Note the E2 UbcD6 and the E3 UBR3 display the highest scores.

Table S1: Protein complex analysis of the screen (46827 hits) using COMPLEAT software (http://www.flyrnai.org/compleat/) and G.O. analysis of the 84 top genes of the screen using DAVID software (http://david.abcc.ncifcrf.gov/home.jsp).

Table S2: Screen scores of the different proteasome components. Scores reported in Figure 1B are in bold. *Pri* and LacZ (negative control) scores are also shown. * indicates when GFP positive cells are < 50, and so the corresponding amplicon was removed from analysis.

Table S3: List of predicted E3 ubiquitin ligases and E2 ubiquitin-conjugating enzymes normalized inhibitory scores corresponding to Fig. S4. * indicates when GFP positive cells are < 50, and so the corresponding amplicon was removed from analysis.

Partie 3 : Discussion et résultats additionnels

I. Ubr3, membre de la famille des protéines à UBRbox

Le système de dégradation par l'UPS fait intervenir des E3 ubiquitin-ligases (Fig. 33). Une E3 reconnait une séquence signal dans le substrat à dégrader, appelé le dégron. Il en existe un sous-groupe : les N-dégrons. Ceux-ci, en N-terminal, sont caractérisés par un résidu N-terminal dit déstabilisant et la présence de lysine(s) interne(s) (Varshavsky, 1996), cibles de la poly-ubiquitination (Chau et al., 1989). L'ensemble de ces résidus N-terminaux déstabilisants a donné lieu à une loi appelée le « **N-end rule** » (Fig. 38). Cette loi met en corrélation le temps de demi-vie d'une protéine et la nature de son résidu N-terminal (Bachmair et al., 1986; Varshavsky, 1996).

A. La voie du N-end rule

Le N-end rule a une structure hiérarchique (Fig. 38). Le résidu N-terminal peut être tertiaire (N, Q et C), secondaire (D, E et C) ou primaire. Dans le cas de résidus tertiaires ou secondaires, la génération du N-dégron se fait par déamidation du résidu tertiaire par Ntan1 ou Ntaq1 (N devient D et Q devient E), puis par conjugaison d'une arginine par Ate1sur le résidu secondaire (R-D et R-E). Les résidus primaires sont classés en 2 sous-groupes : les résidus de type 1, basiques (R, K, H) et de type 2, gros et hydrophobes (L, I, F, Y et W) (Fig. 38) (Tasaki et al., 2012).

Je n'entrerai pas dans les détails, mais il existe chez les eucaryotes une autre branche dans la voie du N-end rule, appelée Ac/N-end rule (Hwang et al., 2010; Shemorry et al., 2013) (la première s'appelant en fait la Arg/N-end rule). Dans cette voie, les résidus déstabilisants ne sont pas les mêmes, et doivent être acétylés pour signaler la dégradation. Par ailleurs, il a aussi récemment été montré que la méthionine (1er résidu) elle-même peut jouer le rôle de signal de dégradation dans les voies des Arg/N-end rule et Ac/N-end rule (Kim et al., 2014).

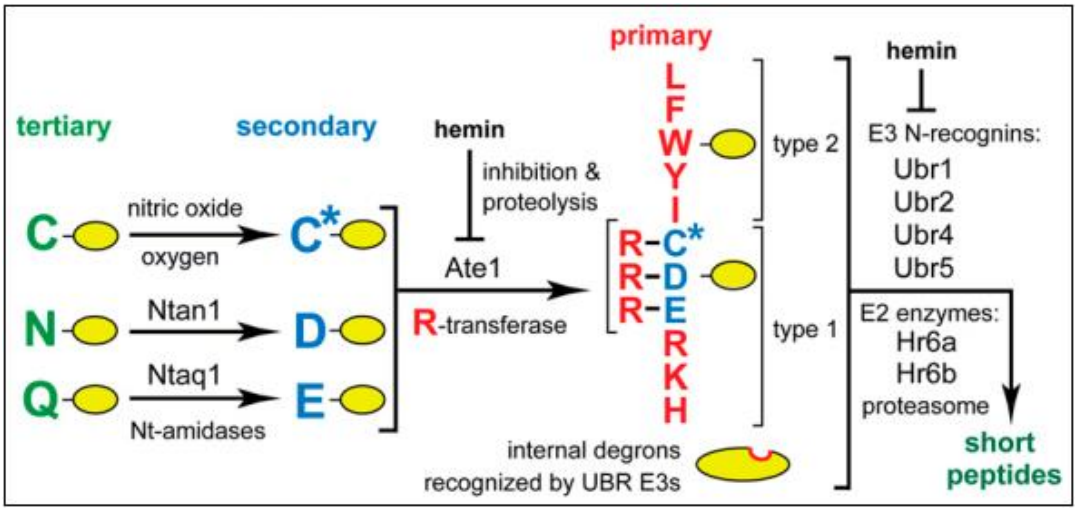


Figure 38 : La voie de l'Arg/N-end rule (mammifère).

B. Les N-recognins

Les E3 ubiquitin-ligases spécialisées dans la reconnaissance des N-dégrons sont appelées les N-recognins (Bartel et al., 1990). Le mécanisme de sélection des substrats a été révélé par la découverte d'un domaine conservé dans les N-recognins, l'**UBRbox**, comme site de liaison au substrat (Tasaki et al., 2005; Xia et al., 2008; Tasaki et al., 2009). Ce domaine a permis d'identifier au moins 7 protéines exprimées par le génome humain : UBR1-7 (Fig. 39).

Cette famille est subdivisée en 2 : 1/UBR1-3 sont dites canoniques, sur la base de leur homologie de séquences, leur taille et leurs domaines conservés [UBR box : site de liaison pour les N-dégrons de type 1, N-domain : site de liaison pour les N-dégrons de type 2, domaine RING finger : domaine d'ubiquitination, et le domaine AI (AutoInhibitory)]. 2/UBR4-7 sont dites non-canoniques car elles présentent une importante divergence évolutive (Fig. 39).

A l'exception d'UBR4, toutes les UBR contiennent un domaine signature d'activité E3 ubiquitin-ligase (RING finger pour UBR1-3, HECT pour UBR5, F-box pour UBR6, PHD (Plant HomeoDomain, ressemblant au domaine RING finger) pour UBR7 (Tasaki et al., 2005) (Fig. 39). Sur la base d'expériences de liaison et de dégradation de substrats contenant un N-dégron, seules les UBR1, 2, 4 et 5 ont pu être caractérisées en tant que N-recognins. Les UBR1, 2 et 4 peuvent dégrader des substrats de type 1 ou 2, UBR5 seulement de type 1. Pour les UBR3, 6 et 7, aucune implication dans la voie du N-end rule n'a pu être montrée, elles sont classées non-N-recognins (Tasaki et al., 2005; Tasaki et al., 2007; Tasaki et al., 2009).

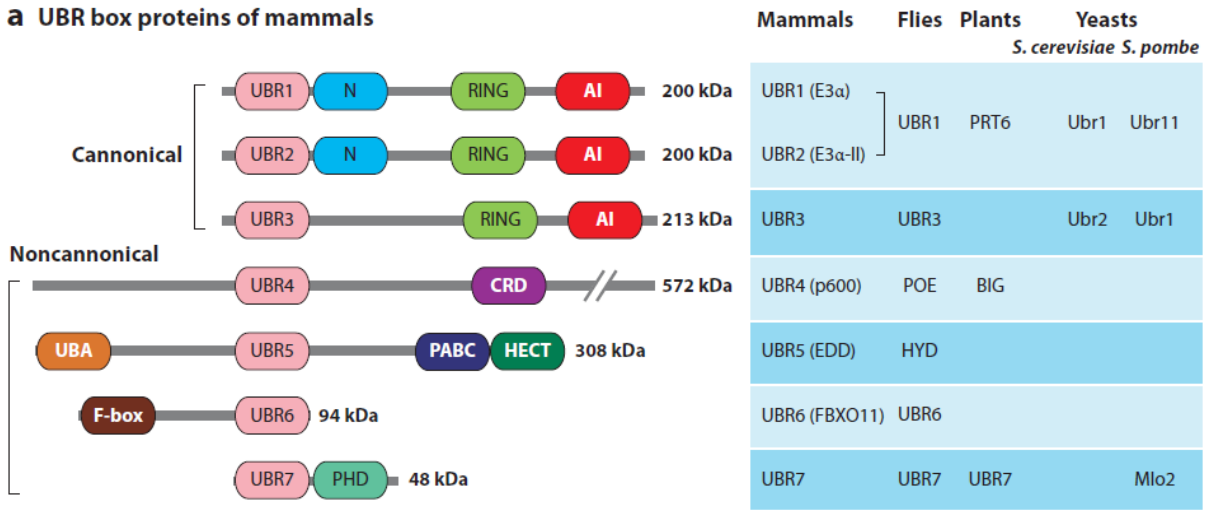


Figure 39 : Famille des protéines à UBRbox (Tasaki et al., 2012).

C. Le domaine d'auto-inhibition

1. L'activité d'Ubr1 est régulée par des dipeptides

Dans nos travaux, nous avons montré que la liaison entre Ubr3 et son substrat Svb est dépendante de l'expression des peptides Pri. Ce modèle est très similaire à ce qui a été décrit pour UBR1 chez *Saccharomyces cerevisiae*. UBR1, en plus de son activité N-recognin, est capable de lier des substrats via un dégron interne. C'est le cas pour Cup9, un répresseur de l'expression de PTR2, qui exprime un transporteur de peptides. Dans ce cas-là, UBR1 est activée allostériquement par la liaison de dipeptides avec des résidus N-terminaux de type 1 et 2, au niveau de l'UBRbox et du N-domain respectivement (Fig. 40). En effet, en absence de dipeptides, UBR1 est sous une conformation fermée. Son domaine d'auto-inhibition (AI), en C-terminal, est replié sur la partie N-terminale, masquant le N-domain. La liaison d'un dipeptide de type1 sur l'UBRbox déstabilise cette liaison et libère le N-domain, qui peut alors se lier à un dipeptide de type 2. Ceci induit un passage d'UBR1 en forme ouverte, dépliée, démasquant un site de liaison pour les substrats contenant un dégron interne (Cup9) (Fig. 40) (Turner et al., 2000; Du et al., 2002). Il s'agit donc d'une boucle d'auto-amplification : PTR2 permet d'importer des peptides, qui induisent le dépliement d'UBR1, qui dégrade Cup9, qui ne réprime plus l'expression de PTR2.

On notera cependant que la caractérisation du rôle du domaine AI pour la fonction d'Ubr11 (homologue d'UBR1 chez *Schizosaccharomyces pombe*) a donné des résultats différents. Dans le cas d'Ubr11, la liaison de Cup9 se fait indépendamment de la liaison de dipeptides, mais Ubr11ΔAI n'est plus capable de dégrader les substrats de type 1 et 2

(Kitamura et al., 2012). Ceci suggère que ce domaine qualifié d'auto-inhibiteur d'après les résultats obtenus avec UBR1 chez *Saccharomyces cerevisiae* joue d'autres rôles dont pour l'instant on ne comprend pas la mécanistique.

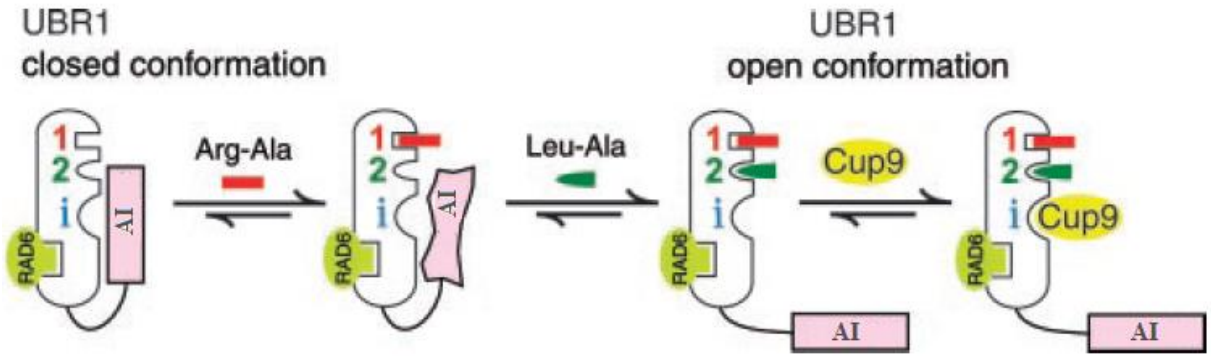


Figure 40 : Régulation de l'auto-inhibition d'UBR1par la liaison de dipeptides. **1** : UBRbox, site de liaison des résidus de type 1, **2** : N-domain, site de liaison des résidus de type 2, **i** : site de liaison des substrats contenant un dégron interne (**Cup9**), **Arg-Ala** : dipeptide de type 1, **Leu-Ala** : dipeptide de type 2, **AI** : domaine d'auto-inhibition, **RAD6** : E2 ubiquitin-conjugating enzyme (homologue de UbcD6 chez la drosophile) (Du et al., 2002).

2. Rôle du domaine d'auto-inhibition d'Ubr3 pour la liaison de Svb

UBR1, 2 et 3 présentent une forte conservation de leur UBRbox et de leur domaine AI. Nous nous sommes donc demandé si les peptides Pri ne pouvaient pas induire un changement conformationnel d'Ubr3 selon le modèle décrit pour UBR1, libérant le site de liaison à Svb. J'ai donc choisi de générer une version d'Ubr3 tronquée en C-terminal : Ubr3-1445. Ainsi, si le rôle des peptides Pri est d'induire un dépliement d'Ubr3, ce mutant devrait être capable de lier Svb en absence des peptides. C'est ce que nous avons testé par co-immuno-précipitation (Fig. 41). Le résultat va à l'encontre de ce modèle : Ubr3-1445 n'est pas lié à Svb en absence des peptides Pri. Ce mutant se comporte de la même manière qu'Ubr3, excluant un rôle de ce domaine AI dans le cadre de la liaison de Svb en réponse à l'expression des peptides Pri. J'ai aussi généré le mutant Ubr3-2, restreint autour de l'UBRbox (car nous pensions que ce domaine pouvait être le site de liaison avec Svb), et son interaction avec Svb dépend aussi de la présence des peptides Pri.

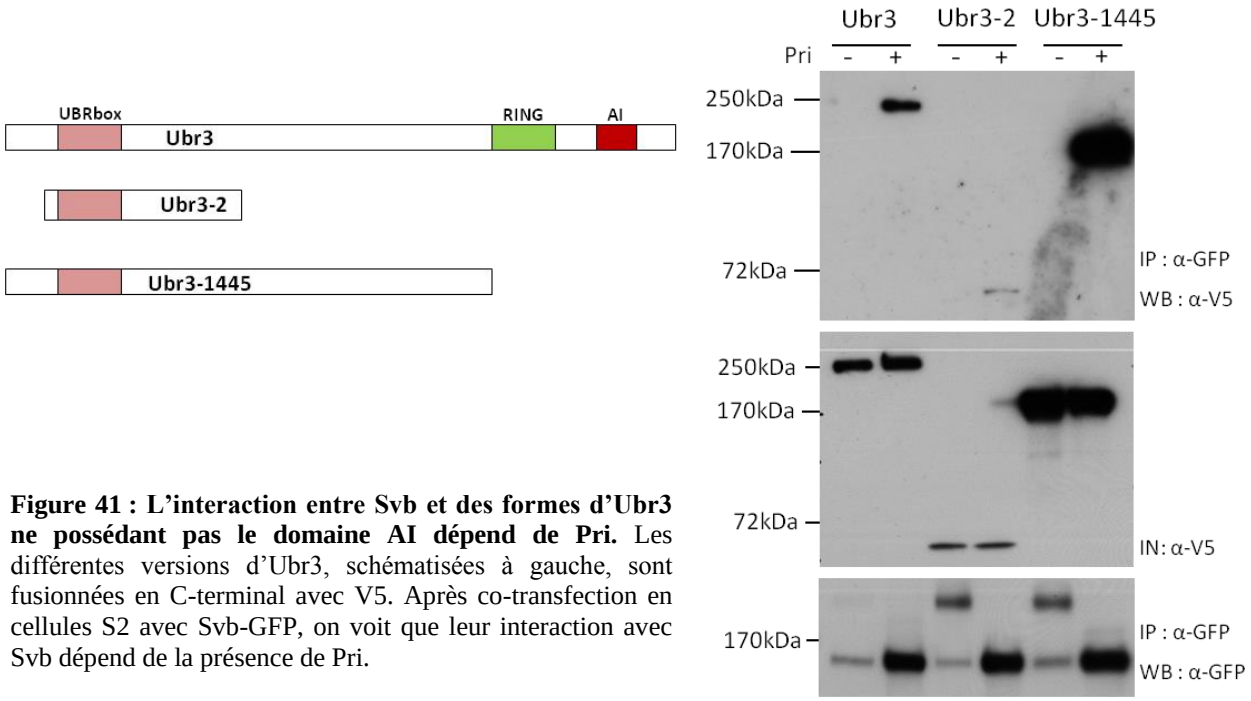


Figure 41 : L'interaction entre Svb et des formes d'Ubr3 ne possédant pas le domaine AI dépend de Pri. Les différentes versions d'Ubr3, schématisées à gauche, sont fusionnées en C-terminal avec V5. Après co-transfection en cellules S2 avec Svb-GFP, on voit que leur interaction avec Svb dépend de la présence de Pri.

D. Le N-dégron de Svb

1. Cible de la voie du N-end rule?

Il a été montré qu'Ubr3 n'est pas capable de dégrader des substrats dans le cadre du N-end rule (Tasaki et al., 2005). Cependant, la dégradation de Svb se fait par la reconnaissance d'un dégron en N-terminal. Le résidu N-terminal de Svb (après la méthionine) est pourtant une Proline, résidu qui n'a pas été décrit comme « déstabilisant ». Les résultats du crible RNAi montrent que le gène *purity of essence* (*poe*), codant l'Ubr4 de la drosophile, avec un score de 42 (*ubr3* a un score de 94), est aussi partiellement impliqué dans la régulation de Svb par Pri. Ceci pourrait suggérer une certaine redondance fonctionnelle entre Ubr3 et Ubr4. Ubr4 est, contrairement à Ubr3, une N-recognin (Tasaki et al., 2005). Nous nous sommes donc demandé si la voie du N-end rule ne pouvait pas, au moins partiellement, participer à la dégradation de Svb. Nous avons envisagé de transformer Svb en substrat de type1, en remplaçant sa Proline N-terminale par une Arginine, un résidu déstabilisant de type 1 (Varshavsky, 2008). Ainsi Svb pourrait devenir substrat d'Ubr1/2, 4 ou 5 et être dégradé sans avoir besoin des peptides Pri.

Nous avons généré ce mutant ponctuel, P2R, et testé sa capacité à être dégradé en absence de Pri. Le résultat obtenu était totalement inattendu : ce mutant n'est pas dégradé, ni en absence, ni en présence des peptides Pri (Fig. 42). Ceci souligne l'importance de cette proline N-terminale, en plus des trois lysines déjà identifiées dans le dégron de Svb, pour sa

maturation en réponse à Pri. Une hypothèse envisageable est que la Proline soit requise pour la conformation de l'interaction entre Svb et Ubr3. En effet, c'est le seul acide aminé qui a une fonction amine secondaire cyclique, créant des « coudes » dans les chaînes polypeptidiques. Mais un autre mutant, P2G (que je justifie dans la section suivante), invalide cette explication, puisque lui est toujours sensible à Pri (Fig. 42). On notera cependant qu'il présente une dégradation en forme courte en absence de Pri plus importante que ce que l'on observe parfois pour Svb (que nous expliquons par le promoteur pMT dirigeant l'expression de *pri* qui peut avoir une activité basale même en absence de $CuSO_4$).

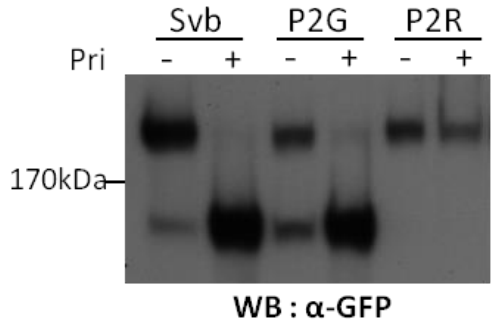


Figure 42 : Le deuxième résidu de Svb est important pour la fonction du dégron. Des mutants ponctuels P2G et P2R de Svb en fusion avec la GFP ont été transfectés en cellules S2 en présence ou en absence de Pri, et ont été analysés en Western Blot avec un anticorps anti-GFP.

De manière intéressante, chez la souris un substrat d'UBR3 a été identifié (Meisenberg et al., 2012). Il s'agit d'APE1 (AP Endonuclease-1). Cette protéine est essentielle à la survie cellulaire (sa délétion est létale embryonnaire pour les souris, et aucune lignée cellulaire ne l'exprimant pas n'a pu être établie). Cependant, le niveau d'APE1 doit être finement régulé par UBR3. Les études sur cette régulation ont permis d'identifier le dégron d'APE1 : celui-ci est en son extrémité N-terminale, et sa séquence est : MPKRGKK…, ce qui n'est pas sans nous rappeler le N-dégron de Svb : MPKIFLI… Deux substrats d'une même E3, appartenant à une famille impliquée dans la voie du N-end rule présentent les 2 mêmes premiers résidus (après la méthionine). Sachant que la fonction non-N-recognin d'UBR3, (en dépit de son homologie forte avec UBR1-2, les acteurs majeurs du N-end rule), est très intrigante, on pourrait se demander s'il ne s'agit pas en fait d'un nouveau type de N-end rule. En effet, comme je le disais en introduction de cette voie, il y a eu des découvertes récentes, notamment sur le rôle de la méthionine elle-même en tant que signal de dégradation (Kim et al., 2014). Il n'est donc pas impossible que l'on découvre un jour que MPK est le signal déstabilisant N-terminal d'une dégradation par UBR3.

2. Domaine SNAG

Le dégron de Svb (31 premiers résidus) est conservé au cours de l'évolution, notamment sur les 10 premiers résidus. Cette séquence correspond à un motif appelé SNAG (SNAil/Gfi). Il est retrouvé dans les membres de la famille des facteurs de transcription de doigts de zinc Snail/Scratch (Nieto, 2002; Barrallo-Gimeno and Nieto, 2009; Kerner et al., 2009). Le rôle de ce domaine a été étudié dans l'orthologue mammifère de Svb : Ovol1, un répresseur transcriptionnel (Dai et al., 1998; Nair et al., 2006) (qui possède un domaine de répression). Il a été montré que sa région SNAG permet d'interagir avec une Histone Déacétylase (HDAC1), donc de la recruter au niveau du promoteur des gènes cibles, pour renforcer passivement cette répression. En effet, HDAC1 va déacétyler des histones H3 et induire une compaction chromatinienne défavorable à l'accès de la machinerie transcriptionnelle (Nair et al., 2007).

Au début de mon travail sur ce projet, en master, j'avais exploré cette piste, en imaginant que le changement de localisation nucléaire et d'activité de Svb puisse avoir un lien avec cette interaction potentielle SNAG/HDAC. D'après les données dont je disposais à ce moment-là, j'avais établi un modèle de la régulation de Svb par Pri que je présente ici (Fig. 43), juste pour illustrer le cheminement qui a été fait depuis.

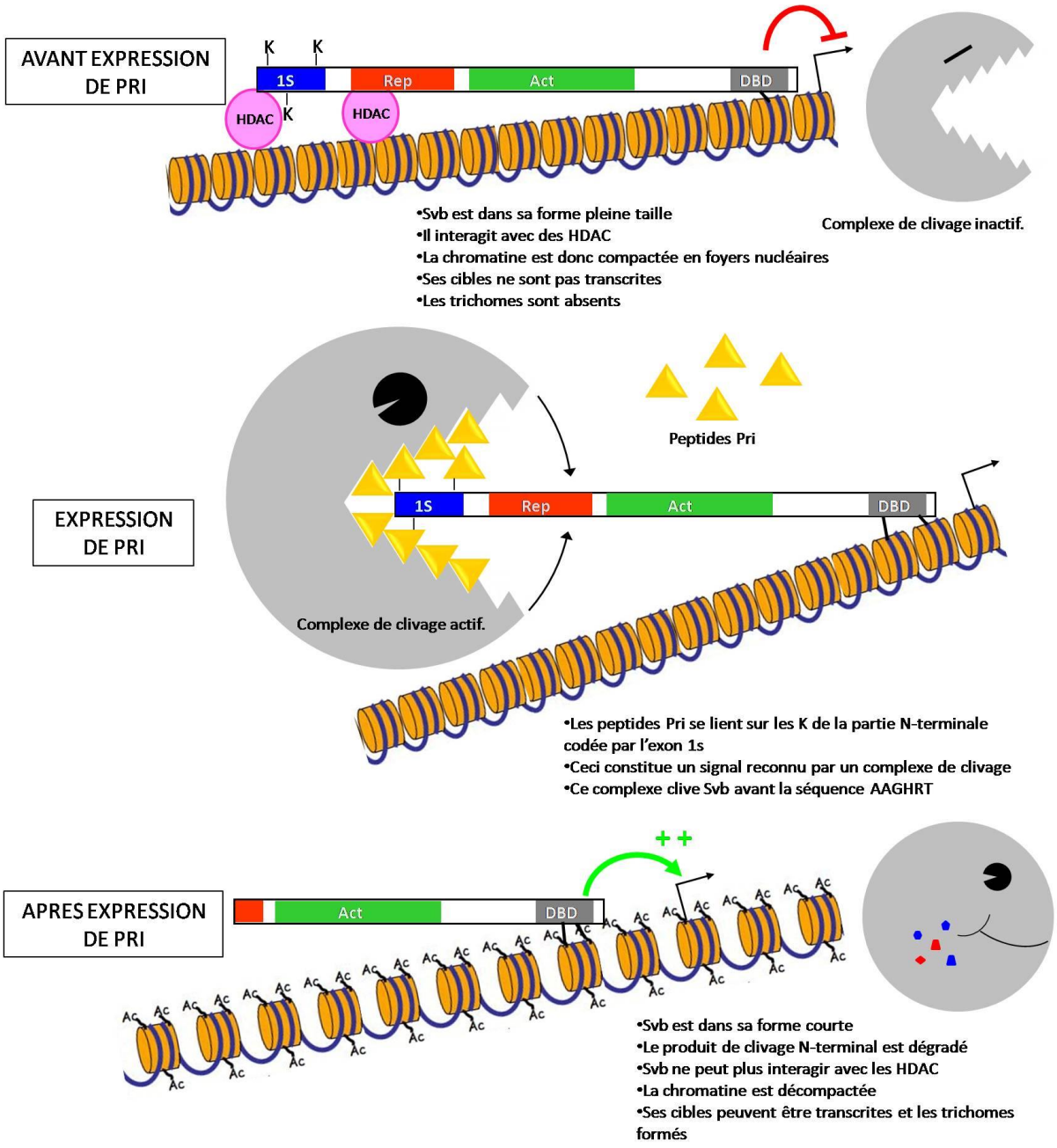


Figure 43: Modèle de la conversion du facteur de transcription Svb par les peptides Pri et effets sur la transcription des gènes-cibles de Svb.

J'avais donc réalisé 3 mutants ponctuels : Ovol1like et Ovol2like, dans lesquels je remplaçais le motif SNAG de Svb par celui de ses orthologues (Ovol1like : K3R, I4A et I7V, Ovol2like : I4V et I7V), et P2G. La Proline 2 présente une très forte conservation dans ce motif SNAG, et les travaux sur Ovol1 ont montré que sa mutation engendre une perte de l'interaction avec la HDAC1 et donc une perte partielle de son activité de répresseur (Nair et al., 2007).

M'intéressant surtout à la relocalisation de Svb en réponse à Pri, j'avais quantifié le nombre de cellules présentant des foyers nucléaires versus une localisation diffuse de chacun

de ces mutants, en absence et en présence de Pri. Seul le mutant Ovol1like est partiellement altéré, avec 39% de relocalisation contre 65% pour Svb. P2G et Ovol2like sont quant à eux normalement relocalisés. Des immuno-colorations avec l'anti-Svb1s montrent qu'en présence de Pri, alors qu'ils sont relocalisés, Ovol1like et Ovol2like sont réactifs à cet anticorps (ce qui n'est pas le cas pour P2G) (Fig. 44). Des analyses de leur taille par Western Blot montrent pourtant que la réponse à Pri, en termes de dégradation, est partielle pour Ovol1like (avec une légère persistance de la forme longue en présence de Pri), et totale pour Ovol2like et P2G.

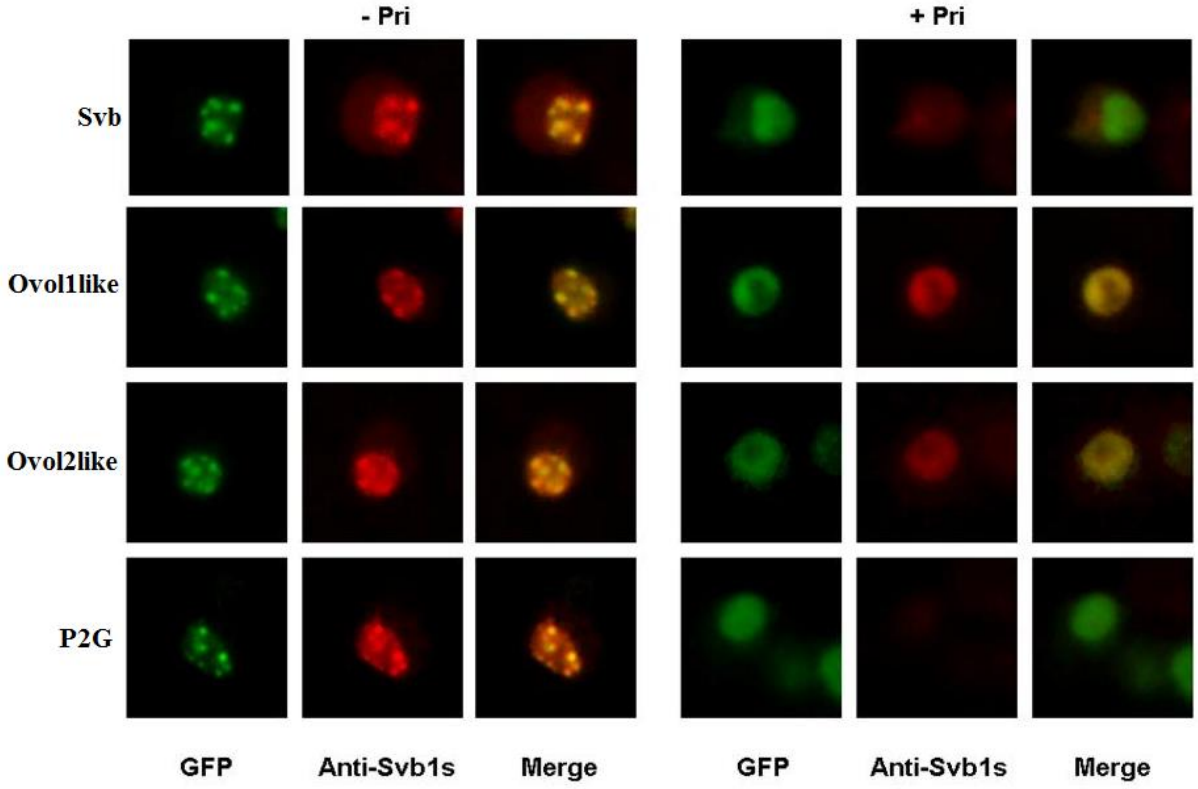


Figure 44 : Localisation nucléaire des protéines Svb-GFP mutants dans le dégron en absence et présence de Pri. Les cellules S2 ont été transfectées avec le vecteur contrôle (Svb) ou un des 3 mutants, avec ou sans induction de Pri. Détection des protéines avec la GFP (en vert) et l'Anti-Svb1s (en rouge).

L'immuno-réactivité à l'anti-Svb1s des formes Ovol1like et Ovol2like diffuses et maturées en présence de Pri est un résultat que nous n'expliquons pas, mais qu'il faut garder à l'esprit. D'autre part, il serait intéressant d'étudier la réponse à Pri d'Ovol1 et Ovol2. En effet ces orthologues mammifères présentant avec Svb une forte conservation N-terminale, il est envisageable qu'il existe ici un même mécanisme de régulation de leur activité transcriptionnelle, dépendante de l'UPS et, pourquoi pas, induite par l'expression d'éventuels peptides Pri-like mammifères.

3. Rôle "non-N-dégron" de cette région

Finalement, au gré des hypothèses envisagées, j'ai généré un total de 11 mutants ponctuels affectant de 1 à 3 résidus dans le dégron de Svb (31 premiers résidus). Leur sensibilité à Pri a été testée par Western blot, mais aussi parfois par analyse de la localisation sub-nucléaire (Tableau 7). Plusieurs phénotypes sont observables. S21A est normalement dégradé en réponse à Pri, mais présente un problème de relocalisation sub-nucléaire, avec seulement 35% de relocalisation contre 65% pour le sauvage, P2G est partiellement dégradé en absence de Pri, P2R n'est pas sensible à Pri, etc. L'ensemble de ces résultats montre que cette région, qui joue le rôle de N-dégron, est aussi impliquée dans la localisation correcte de Svb, et donc sûrement (on ne l'a pas testé) dans son activité transcriptionnelle de répresseur.

			Analyse des tailles en WB	
Nom	Séquence des 31 aa N-ter	% relocalisation	Réponse Pri indépendante	Réponse à Pri
Wt	MPKIFLIKNRLHQQQQRLLESQNLLQHKNQD	65	NON	OUI
P2G	-G-----------------------------	56	un peu	OUI
P2R	-R-----------------------------	-	NON	NON
K3G	--G----------------------------	-	NON	moitié
K8G	-------G---------------------	-	NON	OUI
K3K8G	--G----G---------------------	-	NON	NON
3Kmt	--G----G-----------------G---	22	NON	NON
K3RK8R	--R----R---------------------	-	NON	OUI
3KR	--R----R-----------------R---	-	NON	NON
S21A	------------------A----------	35	NON	OUI
Ovol1like	--RA--V----------------------	39	NON	moitié
Ovol2like	---V--V----------------------	57	NON	OUI

Tableau 7 : Résumé des analyses des mutants ponctuels de Svb. La séquence correspondant au dégron de chacun des mutants est indiquée. Les mutants insensibles à Pri (aucun changement en réponse à Pri) sont en rouge, les mutants intermédiaires (pour lesquels Pri induit une relocalisation et/ou un changement de taille, mais moindre que ce qui est observé pour Svb) sont en orange, les mutants qui se comportent comme Svb sont en vert. Les mutants qui présentent des phénotypes en absence de Pri, ou que l'on ne peut pour l'instant pas relier à une sensibilité à Pri sont en bleu (Ovol1like et Ovol2like sont en bleu à cause de leur immunoréactivité à l'Anti-Svb1s détaillée précédemment).

4. Implication dans le rôle de répresseur de la forme longue de Svb

Des résultats préliminaires obtenus en immuno-précipitation de la chromatine avec un anticorps dirigé contre Svb dans des embryons de drosophile montrent que Svb est déjà sur les promoteurs de ses gènes-cibles avant l'expression de Pri. Il est possible que sa fonction de

répresseur soit requise pour empêcher leur expression basale. Ainsi, son activation en réponse à Pri donne lieu à une expression synchrone et concertée de toutes les cibles. Si on admet que cela puisse être un critère pour le bon déroulement de la formation du trichome, il n'est pas impossible que des mécanismes aient été mis en place pour « protéger » la forme longue de Svb. Ceci va dans le sens des résultats cités précédemment : la partie N-terminale de Svb ne sert pas qu'à la liaison avec Ubr3 en réponse à Pri.

a. Identification de co-facteurs

On pourrait imaginer que cette région est requise pour l'interaction avec des co-facteurs qui stabiliseraient Svb sous sa forme longue. Les résultats du crible RNAi peuvent permettre d'explorer cette piste. En effet, je rappelle que dans ce crible le score d'un gène est corrélé à son implication dans la réponse de Svb à Pri. Pour identifier les facteurs associés, nous avions donc considéré les gènes qui avaient un haut score (significativement supérieurs à la moyenne des scores de tous les gènes du crible). Mais on peut considérer les scores significativement inférieurs à cette moyenne. Il s'agit de gènes qui lorsqu'ils ne sont pas exprimés vont favoriser la réponse de Svb à Pri, donc il peut s'agir de co-facteurs de la forme longue.

3500 gènes ont un score inférieur à 10,3 (la moyenne – l'écart-type des scores du crible). Les analyses de clusters de gènes avec le programme DAVID (**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery, http://david.abcc.ncifcrf.gov/) ne permettent pas de mettre en évidence un mécanisme en particulier (contrairement aux résultats obtenus pour l'UPS à partir de la liste des gènes avec un haut score). Cependant, on trouve par exemple Hdac3 qui a un score de 3,08. Si on compare à ce qui est décrit avec Ovol1, qui interagit avec HDAC1 via son domaine SNAG (Nair et al., 2007), on peut imaginer qu'une interaction existe entre Svb et Hdac3, stabilisant Svb sous sa forme longue répresseur. De plus, Hdac3 pourrait renforcer la répression de Svb via une compaction chromatinienne.

Cette liste de gènes pourrait donc représenter des pistes à explorer dans le cadre de l'étude de la fonction de répresseur de Svb.

b. Identification de déubiquitinases

D'autres gènes défavorisant la dégradation de Svb en réponse à Pri pourraient agir en contrant son ubiquitination par Ubr3. Il existe un groupe d'enzymes appelées les déubiquitinases (DUBs) qui peuvent s'opposer à l'activité d'une E3 ubiquitin-ligase en dépolymérisant la chaîne d'ubiquitines qu'elle génère (Fig. 45) (Clague et al., 2012). Ainsi, s'il existe une DUB qui s'oppose à Ubr3 pour l'ubiquitination de Svb (dont l'action sur Svb pourrait aussi dépendre de la reconnaissance de son N-dégron), elle devrait avoir un faible score dans le crible RNAi. En effet si l'expression de cette DUB est inhibée, l'action d'Ubr3 est favorisée, donc Svb est ubiquitinylé et dégradé plus efficacement. J'ai donc analysé les scores des 41 DUBs référencées chez la drosophile (Tsou et al., 2012) (Tableau 8). 11 d'entre elles ont un score < 10,32, révélant une potentielle implication dans la régulation de la stabilisation de Svb (par opposition à sa dégradation dépendante de l'ubiquitine).

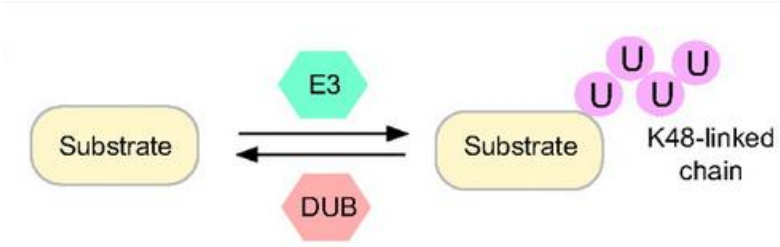


Figure 45: Les DUBs régulent la stabilité des protéines. Une DUB s'oppose à l'action d'une E3 ubiquitin-ligase, stabilisant le substrat si la chaîne d'ubiquitines générée par la E3 est un signal de dégradation (liées en K48).

Flybase ID	Gene symbol	RNAi screen score
FBgn0026738	CG7857	0,00
FBgn0000542	ec	3,96
FBgn0037270	CG9769	4,00
FBgn0050421	CG30421	4,05
FBgn0029763	CG4165	4,09
FBgn0031187	CG14619	5,28
FBgn0032214	CG4968	6,46
FBgn0029819	CG3016	7,13
FBgn0052479	CG32479	7,64
FBgn0036913	CG8334	9,72
FBgn0033352	PAN2	9,89
FBgn0260936	scny	10,93
FBgn0039773	CG2224	11,45
FBgn0027053	CSN5	11,78
FBgn0003023	otu	12,53
FBgn0037734	trbd	12,68
FBgn0029853	CG3781	14,25
FBgn0033738	DUBAI	14,47
FBgn0028476	CG15817	15,29
FBgn0035593	CG4603	15,32
FBgn0033916	CG8494	15,39
FBgn0010288	Uch	15,47
FBgn0016756	Ubp64E	15,47
FBgn0039025	Usp-12-46	15,67
FBgn0005632	faf	17,14
FBgn0030969	CG7288	17,17
FBgn0030366	Usp7	17,24
FBgn0032216	CG5384	17,52
FBgn0035402	CG12082	17,85
FBgn0039214	puf	18,07
FBgn0262166	calypso	18,32
FBgn0011327	Uch-L5	18,33
FBgn0013717	not	18,54
FBgn0036180	CG6091	19,64
FBgn0032210	CYLD	26,22
FBgn0038862	Ubpy	26,29
FBgn0028837	CSN6	28,43
FBgn0032348	CG4751	30,48
FBgn0002787	Mov34	-
FBgn0028694	Rpn11	-
FBgn0033688	Prp8	-

Tableau 8 : Scores obtenus dans le crible RNAi pour les DUBs de la drosophile. Les DUBs qui ont un score inférieur à 10,32 sont en bleu.

II. Implication d'Ubr3 dans les autres phénotypes des mutants *pri*

Comme décrit en introduction, les mutants *pri* présentent deux autres phénotypes majeurs (autres que l'absence de trichome embryonnaire) 1/ un problème de formation de la patte chez l'adulte (Galindo et al., 2007), 2/ un réseau trachéal embryonnaire fortement altéré (Kondo et al., 2007).

A. Phénotype dans la patte adulte

Les travaux de Juan Pablo Couso ont montré que les peptides Pri contrôlent la voie de signalisation Notch pour la formation des articulations qui vont donner lieu aux tarses de la patte. Ils montrent que ce contrôle requiert la présence de Svb (Pueyo and Couso, 2011). Ceci suggère que le mécanisme de régulation de Svb par Ubr3 en réponse à Pri, identifié pour la formation des trichomes, puisse être réédité dans ce tissu. Afin de tester cette hypothèse, Anne Pélissier-Monier, maître de conférences dans l'équipe, est actuellement en train de mettre au point la génération de clones mutants *ubr3* dans les disques imaginaux de pattes. Nous attendons les résultats avec impatience. Si ces clones engendrent une patte présentant le même phénotype que des clones mutant *svb* ou *pri*, cela validera l'hypothèse de la conservation du mécanisme de régulation de Svb par Ubr3 en réponse aux peptides Pri.

B. Phénotype des trachées embryonnaires

En revanche, il en va autrement pour le phénotype de malformation des trachées de l'embryon. Les trachées de mutants pour *svb* sont formées normalement (Kondo et al., 2010). Svb n'est donc pas impliqué dans la fonction de *pri* pour la formation des trachées. Pri permettrait donc de réguler d'autres cibles que Svb. Delphine Ménoret, lors de sa thèse dans l'équipe, a réalisé l'analyse du transcriptome d'embryons mutants pour *svb* et pour *pri*. Ces génotypes permettent tous deux d'identifier les gènes cibles de Svb, qui présentent une perte d'expression. En revanche, de nombreux gènes sont dérégulés seulement chez les embryons mutants pour *pri* (positivement ou négativement). On sait, par comparaison des phénotypes et transcriptomes, que Pri a des fonctions indépendantes de Svb.

Maintenant que le mécanisme de régulation de Svb par Pri a été mis en évidence, qu'il fait intervenir l'E3 ubiquitin-ligase Ubr3, la question qui se pose est : **les fonctions de *pri* indépendantes de *svb* sont-elles toutes dépendantes d'Ubr3 ?** Autrement dit, est-ce que les autres fonctions de *pri* requièrent la dégradation (totale ou partielle) d'autres substrats

d'Ubr3 ? Pour tester cette hypothèse dans la cadre de la mise en place des trachées, il faudrait observer les trachées d'un embryon mutant pour *ubr3*. Si celles-ci présentent les mêmes défauts qu'observés chez les mutants *pri*, cela signifierait que la fonction de Pri dans la formation des trachées et d'induire la dégradation d'un facteur par Ubr3 (facteur qu'il faudrait bien sûr identifier).

Dans le cadre de l'étude de la formation des trichomes, j'avais observé les cuticules d'embryons mutants *ubr3*. Celles-ci étaient normales, mais les embryons mourraient en fin d'embryogenèse (Fig. 46). La contribution maternelle en ARNm *ubr3* pouvait expliquer ce résultat. J'ai réalisé une hybridation *in situ* sur des embryons sauvages pour voir le profil d'expression *d'ubr3*, et constaté qu'il est présent dans tout l'embryon, confortant l'hypothèse de la contribution maternelle (Fig. 48). Ainsi, pour travailler sur les phénotypes embryonnaires de mutants *ubr3*, il est nécessaire de générer des clones germinaux, afin d'obtenir des embryons qui n'auront d'expression ni maternelle ni zygotique d'*ubr3*. C'est ce qu'est en train de réaliser Jennifer Zanet, en post-doctorat dans l'équipe, dans le but de pouvoir comparer les phénotypes embryonnaires de mutants *pri* et de mutants *ubr3*.

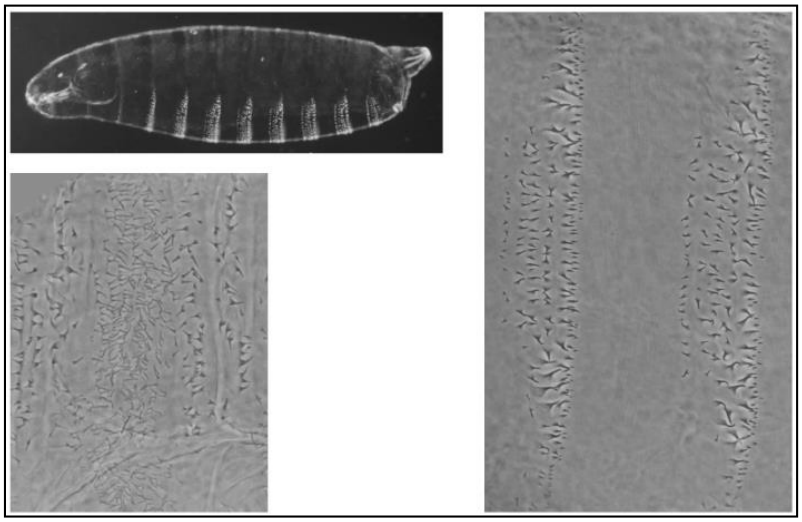


Figure 46 : Cuticule d'une larve mutante *Ubr3* (PL85). Le patron général de l'implantation des trichomes, et leur morphologie à la fois sur le dos (en bas du panel) et le ventre (à droite) sont normaux.

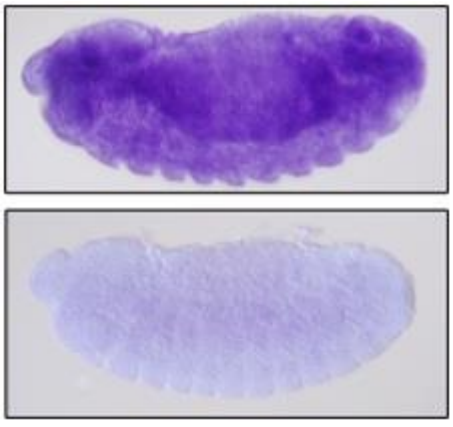


Figure 47 : *ubr3* est exprimé ubiquitairement dans les embryons de drosophile. Hybridation *in situ* sur des embryons sauvages de drosophile, avec une sonde anti-sens au transcrit *ubr3* (en haut) et une sonde contrôle sens (en bas).

C. La dégradation d'autres substrats d'Ubr3 dépend-elle des peptides Pri ?

Les E3 ubiquitin-ligases ont souvent plusieurs substrats dont elles dirigent la dégradation. On est donc en droit de se demander si la dégradation d'autres substrats d'Ubr3 dépend ou non de la présence des peptides Pri. Le groupe d'Hugo Bellen, avec qui nous avons collaboré dans le cadre de ce projet, a identifié un autre substrat d'Ubr3 (les données n'étant pas encore publiées, je ne peux pas révéler son identité). Ignorant tout de l'histoire de Pri et Svb au moment où ont été entrepris ces travaux, toutes leurs analyses ont été réalisées en absence de Pri. Ils sont capables de montrer par des expériences de co-immuno-précipitation, une liaison et ubiquitination de ce substrat par Ubr3. Il faudrait bien sûr vérifier si Pri est capable de modifier ou non cette interaction, mais tout semble cependant indiquer qu'Ubr3 puisse ubiquitinyler des substrats en absence de Pri.

Ceci rappelle le modèle établi pour UBR1 chez *Saccharomyces cerevisiae*, où il y a trois classes de substrats (N-dégron de type1, de type 2, et dégron interne), reconnues par trois régions différentes d'UBR1 (UBRbox, N-domain, et région plus centrale non-définie, respectivement), et où seulement la liaison des substrats possédant un dégron interne (Cup9) dépend de la liaison de didpeptides (Turner et al., 2000; Du et al., 2002). Mais les données expérimentales sur Ubr3 chez la drosophile sont encore trop peu nombreuses pour pouvoir extrapoler ce modèle. De plus, la non-implication du domaine AI dans la liaison de Svb par Ubr3 en réponse à Pri prouve déjà qu'il ne s'agira pas du même type de régulation.

CONCLUSION

Partie 1 : Les sPEPs, des molécules essentielles à la vie.

Pri est l'un des premiers sPEPs à avoir été identifié. Mais d'autres études ont permis de mettre en évidence que la fonction d'ARN présumés non-codants dans le développement est conférée par l'expression de peptides. Je vais en présenter ici trois exemples démontrant l'importance de ces molécules. D'autres exemples sont référencés dans le tableau 9 (Andrews and Rothnagel, 2014).

Species	Gene	sPEP size	Notes	References
Arabidopsis thaliana	*PLS*	36	Required for correct auxin-cytokinin homeostasis to modulate root growth and leaf vascular patterning.	Casson *et al*, 2002
	ROT4	53	Involved in regulation of leaf shape by reducing cell proliferation in lateral organs.	Narita *et al*, 2004
Soybean	*ENOD40-1*	12 and 24	Binds to nodulin100 (which is a subunit of sucrose synthase) and is likely to be involved in the control of sucrose use in nitrogen-fixing nodules.	Rohrig *et al*, 2002
Maize	*brk1*	84	Promotes multi actin-dependent cell polarization events in the developing leaf epidermis (orthologous to HSPC300).	Frank and Smith, 2002
Drosophila melanogaster	*HSPC300*	75	Component of the WAVE-SCAR complex, important in nervous system development for axogenesis and neuromuscular synapse morphogenesis.	Qurashi *et al*, 2007
Zebrafish	*Toddler*	58	Activates Apelin (G protein-couples receptor) to promote cell motility in the early fish embryo, required for heart development.	Chng *et al*, 2013 Pauli *et al*, 2014

Tableau 9 : Exemple de sPEPs identifiés et leur fonction (Andrews and Rothnagel, 2014)

I. Polar granule component (Pgc)

Chez la drosophile, le gène *polar granule component* (*pgc*) avait originairement été rapporté pour fonctionner en tant qu'ARN non-codant. Ce gène est exprimé dans les cellules germinales primordiales de l'embryon, et son rôle est de bloquer transitoirement la transcription en empêchant l'activation de l'ARN polymérase II (PolII) (Martinho et al., 2004). Mais 4 ans plus tard, il a été montré que *pgc* permet l'expression d'un peptide de 71 résidus (Pgc), qui forme un complexe avec CycT et Cdk9 (Hanyu-Nakamura et al., 2008). Ces deux protéines forment le complexe P-TEFb (Positive Transcription Elongation Factor-b), qui est responsable de la phosphorylation de la Sérine 2 de la queue CTD (Carboxy Terminal Domain) de la PolII (Peterlin and Price, 2006). Cette phosphorylation et celle de la Sérine 5 sont nécessaires à l'activation de la transcription (Proudfoot et al., 2002). Pgc, en interagissant avec ce complexe, empêche son recrutement au niveau des sites de transcription. PolII n'est pas phosphorylée sur la Sérine 2, la transcription est inhibée. Cette répression transcriptionnelle est essentielle pour la mise en place de la lignée germinale, en évitant leur

différenciation en cellules somatiques (Pirrotta, 2002). Pgc est donc un peptide essentiel au développement de la drosophile fertile.

II. Sarcolamban (Scl)

Le groupe de Juan Pablo Couso (qui avait identifié *pri* via son phénotype dans la patte de la drosophile adulte) a quant à lui développé (Ladoukakis et al., 2011) et utilisé une méthode bioinformatique basée sur la conservation évolutive pour identifier des smORFs fonctionnels dans les ARNs non-codants et polyadénylés (Tupy et al., 2005) de la drosophile. Ils ont ainsi pu identifier deux smORFs sur un même transcrit (28 et 29 résidus) qui comme les quatre peptides Pri présentent entre eux une homologie de séquence (Magny et al., 2013). Ces peptides sont exprimés dans les muscles et dans le cœur des embryons, et leur mutation entraîne une importante arythmie cardiaque due à un problème de transport de Ca^{2+}. Ces peptides présentent une structure prédite en hélice. Des analyses d'homologie de structure leur ont permis de trouver les orthologues humains de ces peptides, *sarcopilin* (*sln*, 31 résidus) (Wawrzynow et al., 1992) et son paralogue *phospholamban* (*pln*, 52 résidus) (Bhupathy et al., 2007). Ils se proposent donc d'appeler leur homologue chez les arthropodes *sarcolamban* (*scl*). Leurs travaux ont permis de démontrer que le mécanisme d'action de ces peptides dans la régulation de la contraction des muscles est le même chez les mammifères et les drosophiles : le peptide se lie à SERCA (Sarco-Endomplasmique Reticulum Ca^{2+} ATPase, ou Ca-P60A chez la drosophile), réduisant son activité dans le transport de Ca^{2+} intracellulaire (Periasamy and Kalyanasundaram, 2007). De plus, le phénotype d'arythmie d'une drosophile mutante *Scl* peut être sauvé par l'expression de son orthologue humain *Pln*.

Cette étude a donc permis de révéler une famille de sPEPs conservée sur une distance évolutive de plus de 550 millions d'années, qui malgré une divergence de leur séquence protéique primaire ont une structure et un rôle quasi identiques. C'est une parfaite illustration de ce que les progrès en bioinformatique apportent à «l'ère peptidique».

III. Modulator of Retrovirus Infection-2 (MRI-2)

Le groupe d'Alan Saghatelian s'est attaché à l'identification de sPEPs à partir de cellules humaines en utilisant une stratégie qui combine des données de RNA-seq et de peptidomique (Slavoff et al., 2013). Leur protocole de peptidomique est basée sur la technique traditionnelle de protéomique LC-MS/MS (chromatographie en phase liquide couplée à la spectrométrie de masse en tandem), mais optimisée pour préserver les peptides

(Tinoco et al., 2010). Ceci leur a permis de découvrir 90 sPEPs, dont seulement 4 avaient été identifiés auparavant, dans leur lignée cellulaire (Oyama et al., 2004; Oyama et al., 2007). Ces résultats ont pu être validés en recoupant avec des données obtenues en ribosome-profiling sur des cellules souches embryonnaires de souris (Ingolia et al., 2011).

Parmi ces sPEPs, ils se sont intéressés au sPEP MRI-2 (Modulator of Retrovirus Infection-2, 69 résidus). Des expériences de co-immuno-précipitation leur ont permis de définir qu'MRI-2 interagit avec l'hétérodimère Ku, impliqué dans la réparation des cassures double brins (DSB) d'ADN. Leurs travaux ont montré que ce sPEP est requis pour stimuler la réparation des DSB (Slavoff et al., 2014). Ce groupe continue d'optimiser leurs techniques d'identification de sPEPs à partir de lignées cellulaires et tissus humains, et a pu récemment en découvrir 237 nouveaux (Ma et al., 2014).

Ici encore, on voit comment la combinaison de plusieurs techniques à grande échelle nées ces dernières années suite au développement des NGS a permis d'identifier puis d'étudier le rôle d'un sPEP.

Partie 2 : L'ère-peptidique, le nouveau chapitre dans la quête de la compréhension de la vie.

I. Pri et les autres sPEPs témoignent de l'importance de cette nouvelle ère.

Les efforts déployés depuis le milieu des années 2000 pour identifier de la manière la plus exhaustive qu'il soit l'ensemble des acteurs de l'expression des génomes ont permis d'immenses progrès techniques, et donc une quantité d'informations et des connaissances toujours grandissantes. Parmi les changements engendrés cette décennie, on a vu émerger une nouvelle classe de molécules : les sPEPs. De plus en plus de moyens et de volonté sont mobilisés pour l'exploration de ce réservoir potentiel de fonctions biologiques.

Les exemples particuliers tels que Pri et les quelques autres sPEPs décrits précédemment attestent de l'importance de ces nouvelles molécules, et de la nécessité de continuer sur cette lancée. Il pourrait y avoir un très grand nombre de sPEPs exprimés par les génomes, avec pour chacun la possibilité d'être essentiel à la vie, comme l'illustrent les peptides Pri.

II. La vie c'est de la micro-horlogerie.

Un point commun des sPEPs identifiés jusqu'alors est qu'ils n'ont pas d'activité propre, mais ont plutôt une fonction d'orchestration des interactions protéiques. On pourrait comparer une cellule à une montre suisse. Jusqu'à récemment, les techniques un peu « grossières » (dans le sens du manque de résolution) ont permis de trouver les pièces les plus grandes des engrenages, telles que les roues crantées. Mais on ne fait pas une montre suisse de qualité sans micro-visserie. Alors c'est comme cela que je vois les choses : les sPEPs constituent la micro-visserie de la cellule, sans laquelle les évènements ne pourraient pas être si parfaitement interconnectés, à la fois dans l'espace et dans le temps.

La cellule étant la brique de base d'un organisme, la perfection de ses rouages est primordiale. La micro-horlogerie illustre bien que pour atteindre cette perfection, il faut en passer par des micro-pièces (sans quoi les montres se dérèglent d'elles-mêmes). Et sans outil de précision on ne peut pas travailler à cette échelle. C'est le message que je voulais transmettre en introduction : les techniques développées ces dernières années sont les outils de

précision qui nous permettent d'accéder à la micro-visserie des engrenages des cellules, afin d'en comprendre le fonctionnement à une nouvelle échelle, et sans doute pas la dernière.

REFERENCES

Alm-Kristiansen, A. H., Norman, I. L., Matre, V. and Gabrielsen, O. S. (2009) 'SUMO modification regulates the transcriptional activity of FLASH', *Biochem Biophys Res Commun* 387(3): 494-9.

Andrews, J., Garcia-Estefania, D., Delon, I., Lü, J., Mével-Ninio, M., Spierer, A., Payre, F., Pauli, D. and Oliver, B. (2000) 'OVO transcription factors function antagonistically in the Drosophila female germline', *Development* 127(4): 881-92.

Andrews, S. J. and Rothnagel, J. A. (2014) 'Emerging evidence for functional peptides encoded by short open reading frames', *Nat Rev Genet* 15(3): 193-204.

Archetti, M. (2004) 'Selection on codon usage for error minimization at the protein level', *J Mol Evol* 59(3): 400-15.

Arendt, C. S. and Hochstrasser, M. (1997) 'Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation', *Proc Natl Acad Sci U S A* 94(14): 7156-61.

Bachmair, A., Finley, D. and Varshavsky, A. (1986) 'In vivo half-life of a protein is a function of its amino-terminal residue', *Science* 234(4773): 179-86.

Badger, J. H. and Olsen, G. J. (1999) 'CRITICA: coding region identification tool invoking comparative analysis', *Mol Biol Evol* 16(4): 512-24.

Barrallo-Gimeno, A. and Nieto, M. A. (2009) 'Evolutionary history of the Snail/Scratch superfamily', *Trends Genet* 25(6): 248-52.

Bartel, B., Wünning, I. and Varshavsky, A. (1990) 'The recognition component of the N-end rule pathway', *EMBO J* 9(10): 3179-89.

Basrai, M. A., Hieter, P. and Boeke, J. D. (1997) 'Small open reading frames: beautiful needles in the haystack', *Genome Res* 7(8): 768-71.

Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C. et al. (2014) 'Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation', *EMBO J* 33(9): 981-93.

Bedford, L., Paine, S., Sheppard, P. W., Mayer, R. J. and Roelofs, J. (2010) 'Assembly, structure, and function of the 26S proteasome', *Trends Cell Biol* 20(7): 391-401.

Belizario, J. E., Alves, J., Garay-Malpartida, M. and Occhiucci, J. M. (2008) 'Coupling caspase cleavage and proteasomal degradation of proteins carrying PEST motif', *Curr Protein Pept Sci* 9(3): 210-20.

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E. and Allis, C. D. (2006) 'Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin', *Mol Cell Biol* 26(7): 2560-9.

Bhupathy, P., Babu, G. J. and Periasamy, M. (2007) 'Sarcolipin and phospholamban as regulators of cardiac sarcoplasmic reticulum Ca2+ ATPase', *J Mol Cell Cardiol* 42(5): 903-11.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L. and Rubin, E. M. (2003) 'Phylogenetic shadowing of primate sequences to find functional regions of the human genome', *Science* 299(5611): 1391-4.

Boffelli, D., Nobrega, M. A. and Rubin, E. M. (2004a) 'Comparative genomics at the vertebrate extremes', *Nat Rev Genet* 5(6): 456-65.

Boffelli, D., Weer, C. V., Weng, L., Lewis, K. D., Shoukry, M. I., Pachter, L., Keys, D. N. and Rubin, E. M. (2004b) 'Intraspecies sequence comparisons for annotating genomes', *Genome Res* 14(12): 2406-11.

Boguski, M. S., Tolstoshev, C. M. and Bassett, D. E. (1994) 'Gene discovery in dbEST', *Science* 265(5181): 1993-4.

Bonneton, F. (2010) '[When Tribolium complements the genetics of Drosophila]', *Med Sci (Paris)* 26(3): 297-303.

Brent, M. R. and Guigó, R. (2004) 'Recent advances in gene structure prediction', *Curr Opin Struct Biol* 14(3): 264-72.

Brocchieri, L. and Karlin, S. (2005) 'Protein length in eukaryotic and prokaryotic proteomes', *Nucleic Acids Res* 33(10): 3390-400.

Carninci, P. Kasukawa, T. Katayama, S. Gough, J. Frith, M. C. Maeda, N. Oyama, R. Ravasi, T. Lenhard, B. Wells, C. et al. (2005) 'The transcriptional landscape of the mammalian genome', *Science* 309(5740): 1559-63.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M. et al. (1996) 'High-efficiency full-length cDNA cloning by biotinylated CAP trapper', *Genomics* 37(3): 327-36.

Casson, S. A., Chilley, P. M., Topping, J. F., Evans, I. M., Souter, M. A. and Lindsey, K. (2002) 'The POLARIS gene of Arabidopsis encodes a predicted peptide required for correct root growth and leaf vascular patterning', *Plant Cell* 14(8): 1705-21.

Chang, J., Xie, M., Shah, V. R., Schneider, M. D., Entman, M. L., Wei, L. and Schwartz, R. J. (2006) 'Activation of Rho-associated coiled-coil protein kinase 1 (ROCK-1) by caspase-3 cleavage plays an essential role in cardiac myocyte apoptosis', *Proc Natl Acad Sci U S A* 103(39): 14495-500.

Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. and Plaza, S. (2006) 'Shavenbaby couples patterning to epidermal cell shape control', *PLoS Biol* 4(9): e290.

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(4898): 1576-83.

Cheng, H., Chan, W. S., Li, Z., Wang, D., Liu, S. and Zhou, Y. (2011) 'Small open reading frames: current prediction techniques and future prospect', *Curr Protein Pept Sci* 12(6): 503-7.

Chew, G. L., Pauli, A., Rinn, J. L., Regev, A., Schier, A. F. and Valen, E. (2013) 'Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs', *Development* 140(13): 2828-34.

Chng, S. C., Ho, L., Tian, J. and Reversade, B. (2013) 'ELABELA: a hormone essential for heart development signals via the apelin receptor', *Dev Cell* 27(6): 672-80.

Chung, W. Y., Wadhawan, S., Szklarczyk, R., Pond, S. K. and Nekrutenko, A. (2007) 'A first look at ARFome: dual-coding genes in mammalian genomes', *PLoS Comput Biol* 3(5): e91.

Clague, M. J., Liu, H. and Urbé, S. (2012) 'Governance of endocytic trafficking and signaling by reversible ubiquitylation', *Dev Cell* 23(3): 457-67.

Clark, T. A., Sugnet, C. W. and Ares, M. (2002) 'Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays', *Science* 296(5569): 907-10.

Claverie, J. M. (1997) 'Computational methods for the identification of genes in vertebrate genomic sequences', *Hum Mol Genet* 6(10): 1735-44.

Cloonan, N., Forrest, A. R., Kolle, G., Gardiner, B. B., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G. et al. (2008) 'Stem cell transcriptome profiling via massive-scale mRNA sequencing', *Nat Methods* 5(7): 613-9.

Cohen, G. M. (1997) 'Caspases: the executioners of apoptosis', *Biochem J* 326 (Pt 1): 1-16.

Coux, O., Tanaka, K. and Goldberg, A. L. (1996) 'Structure and functions of the 20S and 26S proteasomes', *Annu Rev Biochem* 65: 801-47.

Crappé, J., Van Criekinge, W., Trooskens, G., Hayakawa, E., Luyten, W., Baggerman, G. and Menschaert, G. (2013) 'Combining in silico prediction and ribosome profiling in a genome-wide search for novel putatively coding sORFs', *BMC Genomics* 14: 648.

Crowe, M. L., Wang, X. Q. and Rothnagel, J. A. (2006) 'Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides', *BMC Genomics* 7: 16.

Cvijović, M., Dalevi, D., Bilsland, E., Kemp, G. J. and Sunnerhagen, P. (2007) 'Identification of putative regulatory upstream ORFs in the yeast genome using heuristics and evolutionary conservation', *BMC Bioinformatics* 8: 295.

Dai, X., Schonbaum, C., Degenstein, L., Bai, W., Mahowald, A. and Fuchs, E. (1998) 'The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice', *Genes Dev* 12(21): 3452-63.

Delon, I., Chanut-Delalande, H. and Payre, F. (2003) 'The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in Drosophila', *Mech Dev* 120(7): 747-58.

DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997) 'Exploring the metabolic and genetic control of gene expression on a genomic scale', *Science* 278(5338): 680-6.

Dinger, M. E., Pang, K. C., Mercer, T. R. and Mattick, J. S. (2008) 'Differentiating protein-coding and noncoding RNA: challenges and ambiguities', *PLoS Comput Biol* 4(11): e1000176.

Dogini, D. B., Pascoal, V. D., Avansini, S. H., Vieira, A. S., Pereira, T. C. and Lopes-Cendes, I. (2014) 'The new world of RNAs', *Genet Mol Biol* 37(1 Suppl): 285-293.

Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T. and Varshavsky, A. (2002) 'Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain', *Proc Natl Acad Sci U S A* 99(22): 14110-5.

Dujon, B., Alexandraki, D., André, B., Ansorge, W., Baladron, V., Ballesta, J. P., Banrevi, A., Bolle, P. A., Bolotin-Fukuhara, M. and Bossier, P. (1994) 'Complete DNA sequence of yeast chromosome XI', *Nature* 369(6479): 371-8.

Fernandes, I., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., Payre, F. and Plaza, S. (2010) 'Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes', *Dev Cell* 18(1): 64-76.

Finley, D. (2009) 'Recognition and processing of ubiquitin-protein conjugates by the proteasome', *Annu Rev Biochem* 78: 477-513.

Frank, M. J. and Smith, L. G. (2002) 'A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells', *Curr Biol* 12(10): 849-53.

Fresno, M., Jiménez, A. and Vázquez, D. (1977) 'Inhibition of translation in eukaryotic systems by harringtonine', *Eur J Biochem* 72(2): 323-30.

Frith, M. C., Bailey, T. L., Kasukawa, T., Mignone, F., Kummerfeld, S. K., Madera, M., Sunkara, S., Furuno, M., Bult, C. J., Quackenbush, J. et al. (2006) 'Discrimination of non-protein-coding transcripts from protein-coding mRNA', *RNA Biol* 3(1): 40-8.

Frith, M. C., Pheasant, M. and Mattick, J. S. (2005) 'The amazing complexity of the human transcriptome', *Eur J Hum Genet* 13(8): 894-7.

Fälth, M., Sköld, K., Norrman, M., Svensson, M., Fenyö, D. and Andren, P. E. (2006) 'SwePep, a database designed for endogenous peptides and mass spectrometry', *Mol Cell Proteomics* 5(6): 998-1005.

Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A. and Couso, J. P. (2007) 'Peptides encoded by short ORFs control development and define a new eukaryotic gene family', *PLoS Biol* 5(5): e106.

Gerhard, D. S. Wagner, L. Feingold, E. A. Shenmen, C. M. Grouse, L. H. Schuler, G. Klein, S. L. Old, S. Rasooly, R. Good, P. et al. (2004) 'The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC)', *Genome Res* 14(10B): 2121-7.

Gill, G. (2004) 'SUMO and ubiquitin in the nucleus: different functions, similar mechanisms?', *Genes Dev* 18(17): 2046-59.

Gillette, T. G., Kumar, B., Thompson, D., Slaughter, C. A. and DeMartino, G. N. (2008) 'Differential roles of the COOH termini of AAA subunits of PA700 (19 S regulator) in asymmetric assembly and activation of the 26 S proteasome', *J Biol Chem* 283(46): 31813-22.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. et al. (1996) 'Life with 6000 genes', *Science* 274(5287): 546, 563-7.

Groll, M., Bajorek, M., Köhler, A., Moroder, L., Rubin, D. M., Huber, R., Glickman, M. H. and Finley, D. (2000) 'A gated channel into the proteasome core particle', *Nat Struct Biol* 7(11): 1062-7.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D. and Huber, R. (1997) 'Structure of 20S proteasome from yeast at 2.4 A resolution', *Nature* 386(6624): 463-71.

Guillén, G., Díaz-Camino, C., Loyola-Torres, C. A., Aparicio-Fabre, R., Hernández-López, A., Díaz-Sánchez, M. and Sanchez, F. (2013) 'Detailed analysis of putative genes encoding small proteins in legume genomes', *Front Plant Sci* 4: 208.

Gulcicek, E. E., Colangelo, C. M., McMurray, W., Stone, K., Williams, K., Wu, T., Zhao, H., Spratt, H., Kurosky, A. and Wu, B. (2005) 'Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies', *Curr Protoc Bioinformatics* Chapter 13: Unit 13.1.

Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. and Lander, E. S. (2013) 'Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins', *Cell* 154(1): 240-51.

Hanada, K., Akiyama, K., Sakurai, T., Toyoda, T., Shinozaki, K. and Shiu, S. H. (2010) 'sORF finder: a program package to identify small open reading frames with high coding potential', *Bioinformatics* 26(3): 399-400.

Hanada, K., Higuchi-Takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., Nishi, R., Ohashi, C., Iida, K., Tanaka, M. et al. (2013) 'Small open reading frames associated with morphogenesis are hidden in plant genomes', *Proc Natl Acad Sci U S A* 110(6): 2395-400.

Hanada, K., Zhang, X., Borevitz, J. O., Li, W. H. and Shiu, S. H. (2007) 'A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection', *Genome Res* 17(5): 632-40.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P. and Nakamura, A. (2008) 'Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells', *Nature* 451(7179): 730-3.

Harbers, M. and Carninci, P. (2005) 'Tag-based approaches for transcriptome research and genome annotation', *Nat Methods* 2(7): 495-502.

Hayden, C. A. and Bosco, G. (2008) 'Comparative genomic analysis of novel conserved peptide upstream open reading frames in Drosophila melanogaster and other dipteran species', *BMC Genomics* 9: 61.

Hayden, C. A. and Jorgensen, R. A. (2007) 'Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes', *BMC Biol* 5: 32.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. and Wolf, D. H. (1997) 'The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing', *J Biol Chem* 272(40): 25200-9.

Hershko, A. and Ciechanover, A. (1998) 'The ubiquitin system', *Annu Rev Biochem* 67: 425-79.

Holmes, I. (2003) 'Using guide trees to construct multiple-sequence evolutionary HMMs', *Bioinformatics* 19 Suppl 1: i147-57.

Hoppe, T., Rape, M. and Jentsch, S. (2001) 'Membrane-bound transcription factors: regulated release by RIP or RUP', *Curr Opin Cell Biol* 13(3): 344-8.

Huang, M. T. (1975) 'Harringtonine, an inhibitor of initiation of protein biosynthesis', *Mol Pharmacol* 11(5): 511-9.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., Khalil, A. M., Zuk, O., Amit, I., Rabani, M. et al. (2010) 'A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response', *Cell* 142(3): 409-19.

Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., Horlings, H. M., Shah, N., Umbricht, C., Wang, P. et al. (2011) 'Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters', *Nat Genet* 43(7): 621-9.

Hwang, C. S., Shemorry, A. and Varshavsky, A. (2010) 'N-terminal acetylation of cellular proteins creates specific degradation signals', *Science* 327(5968): 973-7.

Iacono, M., Mignone, F. and Pesole, G. (2005) 'uAUG and uORFs in human and rodent 5'untranslated mRNAs', *Gene* 349: 97-105.

Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A. and Kageyama, Y. (2005) 'Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila', *Genes Cells* 10(12): 1163-73.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. and Weissman, J. S. (2009) 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling', *Science* 324(5924): 218-23.

Ingolia, N. T., Lareau, L. F. and Weissman, J. S. (2011) 'Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes', *Cell* 147(4): 789-802.

James, P. (1997) 'Protein identification in the post-genome era: the rapid rise of proteomics', *Q Rev Biophys* 30(4): 279-331.

Jeon, Y. and Lee, J. T. (2011) 'YY1 tethers Xist RNA to the inactive X nucleation center', *Cell* 146(1): 119-33.

Kageyama, Y., Kondo, T. and Hashimoto, Y. (2011) 'Coding vs non-coding: Translatability of short ORFs found in putative non-coding transcripts', *Biochimie* 93(11): 1981-6.

Kastenmayer, J. P., Ni, L., Chu, A., Kitchen, L. E., Au, W. C., Yang, H., Carter, C. D., Wheeler, D., Davis, R. W., Boeke, J. D. et al. (2006) 'Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae', *Genome Res* 16(3): 365-73.

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H. et al. (2001) 'Functional annotation of a full-length mouse cDNA collection', *Nature* 409(6821): 685-90.

Kerner, P., Hung, J., Béhague, J., Le Gouar, M., Balavoine, G. and Vervoort, M. (2009) 'Insights into the evolution of the snail superfamily from metazoan wide molecular phylogenies and expression data in annelids', *BMC Evol Biol* 9: 94.

Khalil, A. M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A. et al. (2009) 'Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression', *Proc Natl Acad Sci U S A* 106(28): 11667-72.

Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A. and Hwang, C. S. (2014) 'The N-terminal methionine of cellular proteins as a degradation signal', *Cell* 156(1-2): 158-69.

Kino, T., Hurt, D. E., Ichijo, T., Nader, N. and Chrousos, G. P. (2010) 'Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor', *Sci Signal* 3(107): ra8.

Kitamura, K., Nakase, M., Tohda, H. and Takegawa, K. (2012) 'The Ubiquitin ligase Ubr11 is essential for oligopeptide utilization in the fission yeast Schizosaccharomyces pombe', *Eukaryot Cell* 11(3): 302-10.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M. et al. (2006) 'CAGE: cap analysis of gene expression', *Nat Methods* 3(3): 211-22.

Komander, D. (2009) 'The emerging complexity of protein ubiquitination', *Biochem Soc Trans* 37(Pt 5): 937-53.

Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S. and Kageyama, Y. (2007) 'Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA', *Nat Cell Biol* 9(6): 660-5.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F. and Kageyama, Y. (2010) 'Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis', *Science* 329(5989): 336-9.

Kozak, M. (1986) 'Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes', *Cell* 44(2): 283-92.

Kozak, M. (1987) 'An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs', *Nucleic Acids Res* 15(20): 8125-48.

Kozak, M. (1996) 'Interpreting cDNA sequences: some insights from studies on translation', *Mamm Genome* 7(8): 563-74.

Krogh, A., Brown, M., Mian, I. S., Sjölander, K. and Haussler, D. (1994) 'Hidden Markov models in computational biology. Applications to protein modeling', *J Mol Biol* 235(5): 1501-31.

Ladoukakis, E., Pereira, V., Magny, E. G., Eyre-Walker, A. and Couso, J. P. (2011) 'Hundreds of putatively functional small open reading frames in Drosophila', *Genome Biol* 12(11): R118.

Lander, E. S. Linton, L. M. Birren, B. Nusbaum, C. Zody, M. C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. et al. (2001) 'Initial sequencing and analysis of the human genome', *Nature* 409(6822): 860-921.

Lease, K. A. and Walker, J. C. (2006) 'The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics', *Plant Physiol* 142(3): 831-8.

Liu, C. W., Li, X., Thompson, D., Wooding, K., Chang, T. L., Tang, Z., Yu, H., Thomas, P. J. and DeMartino, G. N. (2006) 'ATP binding and ATP hydrolysis play distinct roles in the function of 26S proteasome', *Mol Cell* 24(1): 39-50.

Ma, J., Ward, C. C., Jungreis, I., Slavoff, S. A., Schwaid, A. G., Neveu, J., Budnik, B. A., Kellis, M. and Saghatelian, A. (2014) 'Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue', *J Proteome Res* 13(3): 1757-65.

Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A. and Couso, J. P. (2013) 'Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames', *Science* 341(6150): 1116-20.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. and Akoulitchev, A. (2007) 'Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript', *Nature* 445(7128): 666-70.

Martinho, R. G., Kunwar, P. S., Casanova, J. and Lehmann, R. (2004) 'A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells', *Curr Biol* 14(2): 159-65.

Mathé, C., Sagot, M. F., Schiex, T. and Rouzé, P. (2002) 'Current methods of gene prediction, their strengths and weaknesses', *Nucleic Acids Res* 30(19): 4103-17.

Meisenberg, C., Tait, P. S., Dianova, I. I., Wright, K., Edelmann, M. J., Ternette, N., Tasaki, T., Kessler, B. M., Parsons, J. L., Kwon, Y. T. et al. (2012) 'Ubiquitin ligase UBR3 regulates cellular levels of the essential DNA repair protein APE1 and is required for genome stability', *Nucleic Acids Res* 40(2): 701-11.

Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., Latapie, Y., Ferrer, P., Rouault, H., White, K. P. et al. (2013) 'Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization', *Genome Biol* 14(8): R86.

Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A., Crappé, J., Gevaert, K. and Van Damme, P. (2013) 'Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events', *Mol Cell Proteomics* 12(7): 1780-90.

Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. and Mattick, J. S. (2008) 'Specific expression of long noncoding RNAs in the mouse brain', *Proc Natl Acad Sci U S A* 105(2): 716-21.

Metzger, M. B., Pruneda, J. N., Klevit, R. E. and Weissman, A. M. (2014) 'RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination', *Biochim Biophys Acta* 1843(1): 47-60.

Michel, A. M., Choudhury, K. R., Firth, A. E., Ingolia, N. T., Atkins, J. F. and Baranov, P. V. (2012) 'Observation of dually decoded regions of the human genome using ribosome profiling data', *Genome Res* 22(11): 2219-29.

Minamino, N., Tanaka, J., Kuwahara, H., Kihara, T., Satomi, Y., Matsubae, M. and Takao, T. (2003) 'Determination of endogenous peptides in the porcine brain: possible construction of peptidome, a fact database for endogenous peptides', *J Chromatogr B Analyt Technol Biomed Life Sci* 792(1): 33-48.

Miura, M. (2012) 'Apoptotic and nonapoptotic caspase functions in animal development', *Cold Spring Harb Perspect Biol* 4(10).

Morin, R., Bainbridge, M., Fejes, A., Hirst, M., Krzywinski, M., Pugh, T., McDonald, H., Varhol, R., Jones, S. and Marra, M. (2008) 'Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing', *Biotechniques* 45(1): 81-94.

Morris, D. R. and Geballe, A. P. (2000) 'Upstream open reading frames as regulators of mRNA translation', *Mol Cell Biol* 20(23): 8635-42.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) 'Mapping and quantifying mammalian transcriptomes by RNA-Seq', *Nat Methods* 5(7): 621-8.

Mével-Ninio, M., Terracol, R. and Kafatos, F. C. (1991) 'The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development', *EMBO J* 10(8): 2259-66.

Mével-Ninio, M., Terracol, R., Salles, C., Vincent, A. and Payre, F. (1995) 'ovo, a Drosophila gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning', *Mech Dev* 49(1-2): 83-95.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008) 'The transcriptional landscape of the yeast genome defined by RNA sequencing', *Science* 320(5881): 1344-9.

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R. and Fraser, P. (2008) 'The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin', *Science* 322(5908): 1717-20.

Nair, M., Bilanchone, V., Ortt, K., Sinha, S. and Dai, X. (2007) 'Ovol1 represses its own transcription by competing with transcription activator c-Myb and by recruiting histone deacetylase activity', *Nucleic Acids Res* 35(5): 1687-97.

Nair, M., Teng, A., Bilanchone, V., Agrawal, A., Li, B. and Dai, X. (2006) 'Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription', *J Cell Biol* 173(2): 253-64.

Narita, N. N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J. and Tsukaya, H. (2004) 'Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana', *Plant J* 38(4): 699-713.

Neafsey, D. E. and Galagan, J. E. (2007) 'Dual modes of natural selection on upstream open reading frames', *Mol Biol Evol* 24(8): 1744-51.

Nekrutenko, A., Makova, K. D. and Li, W. H. (2002) 'The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study', *Genome Res* 12(1): 198-202.

Neutzner, M. and Neutzner, A. (2012) 'Enzymes of ubiquitination and deubiquitination', *Essays Biochem* 52: 37-50.

Nieto, M. A. (2002) 'The snail superfamily of zinc-finger transcription factors', *Nat Rev Mol Cell Biol* 3(3): 155-66.

Nüsslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in Drosophila', *Nature* 287(5785): 795-801.

Okazaki, Y. Furuno, M. Kasukawa, T. Adachi, J. Bono, H. Kondo, S. Nikaido, I. Osato, N. Saito, R. Suzuki, H. et al. (2002) 'Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs', *Nature* 420(6915): 563-73.

Okoniewski, M. J. and Miller, C. J. (2006) 'Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations', *BMC Bioinformatics* 7: 276.

Oliver, S. G., van der Aart, Q. J., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J. P. and Benit, P. (1992) 'The complete DNA sequence of yeast chromosome III', *Nature* 357(6373): 38-46.

Ota, T. Suzuki, Y. Nishikawa, T. Otsuki, T. Sugiyama, T. Irie, R. Wakamatsu, A. Hayashi, K. Sato, H. Nagai, K. et al. (2004) 'Complete sequencing and characterization of 21,243 full-length human cDNAs', *Nat Genet* 36(1): 40-5.

Oyama, M., Itagaki, C., Hata, H., Suzuki, Y., Izumi, T., Natsume, T., Isobe, T. and Sugano, S. (2004) 'Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs', *Genome Res* 14(10B): 2048-52.

Oyama, M., Kozuka-Hata, H., Suzuki, Y., Semba, K., Yamamoto, T. and Sugano, S. (2007) 'Diversity of translation start sites may define increased complexity of the human short ORFeome', *Mol Cell Proteomics* 6(6): 1000-6.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M. and Milos, P. M. (2009) 'Direct RNA sequencing', *Nature* 461(7265): 814-8.

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D. and Kanduri, C. (2008) 'Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation', *Mol Cell* 32(2): 232-46.

Patterson, G. H. and Lippincott-Schwartz, J. (2002) 'A photoactivatable GFP for selective photolabeling of proteins and cells', *Science* 297(5588): 1873-7.

Payre, F. (2004) 'Genetic control of epidermis differentiation in Drosophila', *Int J Dev Biol* 48(2-3): 207-15.

Payre, F., Vincent, A. and Carreno, S. (1999) 'ovo/svb integrates Wingless and DER pathways to control epidermis differentiation', *Nature* 400(6741): 271-5.

Pedersen, J. S. and Hein, J. (2003) 'Gene finding with a hidden Markov model of genome structure and evolution', *Bioinformatics* 19(2): 219-27.

Periasamy, M. and Kalyanasundaram, A. (2007) 'SERCA pump isoforms: their role in calcium transport and disease', *Muscle Nerve* 35(4): 430-42.

Peterlin, B. M. and Price, D. H. (2006) 'Controlling the elongation phase of transcription with P-TEFb', *Mol Cell* 23(3): 297-305.

Pickart, C. M. (2004) 'Back to the future with ubiquitin', *Cell* 116(2): 181-90.

Pickrell, W. O., Rees, M. I. and Chung, S. K. (2012) 'Next generation sequencing methodologies--an overview', *Adv Protein Chem Struct Biol* 89: 1-26.

Pirrotta, V. (2002) 'Silence in the germ', *Cell* 110(6): 661-4.

Piwko, W. and Jentsch, S. (2006) 'Proteasome-mediated protein processing by bidirectional degradation initiated from an internal site', *Nat Struct Mol Biol* 13(8): 691-7.

Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E. and Matouschek, A. (2004) 'An unstructured initiation site is required for efficient proteasome-mediated degradation', *Nat Struct Mol Biol* 11(9): 830-7.

Proudfoot, N. J., Furger, A. and Dye, M. J. (2002) 'Integrating mRNA processing with transcription', *Cell* 108(4): 501-12.

Pueyo, J. I. and Couso, J. P. (2011) 'Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor', *Dev Biol* 355(2): 183-93.

Qurashi, A., Sahin, H. B., Carrera, P., Gautreau, A., Schenck, A. and Giangrande, A. (2007) 'HSPC300 and its role in neuronal connectivity', *Neural Dev* 2: 18.

Rabl, J., Smith, D. M., Yu, Y., Chang, S. C., Goldberg, A. L. and Cheng, Y. (2008) 'Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases', *Mol Cell* 30(3): 360-8.

Ribrioux, S., Brüngger, A., Baumgarten, B., Seuwen, K. and John, M. R. (2008) 'Bioinformatics prediction of overlapping frameshifted translation products in mammalian transcripts', *BMC Genomics* 9: 122.

Rinn, J. L. and Chang, H. Y. (2012) 'Genome regulation by long noncoding RNAs', *Annu Rev Biochem* 81: 145-66.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. et al. (2007) 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs', *Cell* 129(7): 1311-23.

Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J. and John, M. (2002) 'Soybean ENOD40 encodes two peptides that bind to sucrose synthase', *Proc Natl Acad Sci U S A* 99(4): 1915-20.

Royce, T. E., Rozowsky, J. S. and Gerstein, M. B. (2007) 'Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification', *Nucleic Acids Res* 35(15): e99.

Rubin, D. M., Glickman, M. H., Larsen, C. N., Dhruvakumar, S. and Finley, D. (1998) 'Active site mutants in the six regulatory particle ATPases reveal multiple roles for ATP in the proteasome', *EMBO J* 17(17): 4909-19.

Rubin, G. M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M. and Harvey, D. A. (2000) 'A Drosophila complementary DNA resource', *Science* 287(5461): 2222-4.

Salles, C., Mével-Ninio, M., Vincent, A. and Payre, F. (2002) 'A germline-specific splicing generates an extended ovo protein isoform required for Drosophila oogenesis', *Dev Biol* 246(2): 366-76.

Savard, J., Marques-Souza, H., Aranda, M. and Tautz, D. (2006) 'A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides', *Cell* 126(3): 559-69.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science* 270(5235): 467-70.

Schmitz, K. M., Mayer, C., Postepska, A. and Grummt, I. (2010) 'Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes', *Genes Dev* 24(20): 2264-9.

Schrader, M. and Schulz-Knappe, P. (2001) 'Peptidomics technologies for human body fluids', *Trends Biotechnol* 19(10 Suppl): S55-60.

Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K. J. and Groll, M. (2008) 'Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction', *Nature* 453(7194): 548-52.

Schulz-Knappe, P., Zucht, H. D., Heine, G., Jürgens, M., Hess, R. and Schrader, M. (2001) 'Peptidomics: the comprehensive analysis of peptides in complex biological mixtures', *Comb Chem High Throughput Screen* 4(2): 207-17.

Shemorry, A., Hwang, C. S. and Varshavsky, A. (2013) 'Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway', *Mol Cell* 50(4): 540-51.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S. et al. (2005) 'Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes', *Genome Res* 15(8): 1034-50.

Siepel, A. C. (2003) 'An algorithm to enumerate sorting reversals for signed permutations', *J Comput Biol* 10(3-4): 575-97.

Siepel, A. and Haussler, D. (2004) 'Phylogenetic estimation of context-dependent substitution rates by maximum likelihood', *Mol Biol Evol* 21(3): 468-88.

Skarshewski, A., Stanton-Cook, M., Huber, T., Al Mansoori, S., Smith, R., Beatson, S. A. and Rothnagel, J. A. (2014) 'uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation', *BMC Bioinformatics* 15: 36.

Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A. and Saghatelian, A. (2014) 'A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining', *J Biol Chem* 289(16): 10950-7.

Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., Karger, A. D., Budnik, B. A., Rinn, J. L. and Saghatelian, A. (2013) 'Peptidomic discovery of short open reading frame-encoded peptides in human cells', *Nat Chem Biol* 9(1): 59-64.

Sleator, R. D. (2010) 'An overview of the current status of eukaryote gene prediction strategies', *Gene* 461(1-2): 1-4.

Smith, D. M., Chang, S. C., Park, S., Finley, D., Cheng, Y. and Goldberg, A. L. (2007) 'Docking of the proteasomal ATPases' carboxyl termini in the 20S proteasome's alpha ring opens the gate for substrate entry', *Mol Cell* 27(5): 731-44.

Smith, D. M., Kafri, G., Cheng, Y., Ng, D., Walz, T. and Goldberg, A. L. (2005) 'ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins', *Mol Cell* 20(5): 687-98.

Steitz, J. A. (1969) 'Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA', *Nature* 224(5223): 957-64.

Stone, M., Hartmann-Petersen, R., Seeger, M., Bech-Otschir, D., Wallace, M. and Gordon, C. (2004) 'Uch2/Uch37 is the major deubiquitinating enzyme associated with the 26S proteasome in fission yeast', *J Mol Biol* 344(3): 697-706.

Svensson, M., Sköld, K., Svenningsson, P. and Andren, P. E. (2003) 'Peptidomics-based discovery of novel neuropeptides', *J Proteome Res* 2(2): 213-9.

Taherbhoy, A. M., Schulman, B. A. and Kaiser, S. E. (2012) 'Ubiquitin-like modifiers', *Essays Biochem* 52: 51-63.

Takahashi, H., Takahashi, A., Naito, S. and Onouchi, H. (2012) 'BAIUCAS: a novel BLAST-based algorithm for the identification of upstream open reading frames with conserved amino acid sequences and its application to the Arabidopsis thaliana genome', *Bioinformatics* 28(17): 2231-41.

Tasaki, T., Mulder, L. C., Iwamatsu, A., Lee, M. J., Davydov, I. V., Varshavsky, A., Muesing, M. and Kwon, Y. T. (2005) 'A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons', *Mol Cell Biol* 25(16): 7120-36.

Tasaki, T., Sohr, R., Xia, Z., Hellweg, R., Hörtnagl, H., Varshavsky, A. and Kwon, Y. T. (2007) 'Biochemical and genetic studies of UBR3, a ubiquitin ligase with a function in olfactory and other sensory systems', *J Biol Chem* 282(25): 18510-20.

Tasaki, T., Sriram, S. M., Park, K. S. and Kwon, Y. T. (2012) 'The N-end rule pathway', *Annu Rev Biochem* 81: 261-89.

Tasaki, T., Zakrzewska, A., Dudgeon, D. D., Jiang, Y., Lazo, J. S. and Kwon, Y. T. (2009) 'The substrate recognition domains of the N-end rule pathway', *J Biol Chem* 284(3): 1884-95.

Tian, L., Holmgren, R. A. and Matouschek, A. (2005) 'A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB', *Nat Struct Mol Biol* 12(12): 1045-53.

Tinoco, A. D., Tagore, D. M. and Saghatelian, A. (2010) 'Expanding the dipeptidyl peptidase 4-regulated peptidome via an optimized peptidomics platform', *J Am Chem Soc* 132(11): 3819-30.

Tran, M. K., Schultz, C. J. and Baumann, U. (2008) 'Conserved upstream open reading frames in higher plants', *BMC Genomics* 9: 361.

Trempe, J. F. (2011) 'Reading the ubiquitin postal code', *Curr Opin Struct Biol* 21(6): 792-801.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A. et al. (2010) 'The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation', *Mol Cell* 39(6): 925-38.

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E. and Chang, H. Y. (2010) 'Long noncoding RNA as modular scaffold of histone modification complexes', *Science* 329(5992): 689-93.

Tscherne, J. S. and Pestka, S. (1975) 'Inhibition of protein synthesis in intact HeLa cells', *Antimicrob Agents Chemother* 8(4): 479-87.

Tsou, W. L., Sheedlo, M. J., Morrow, M. E., Blount, J. R., McGregor, K. M., Das, C. and Todi, S. V. (2012) 'Systematic analysis of the physiological importance of deubiquitinating enzymes', *PLoS One* 7(8): e43112.

Tupy, J. L., Bailey, A. M., Dailey, G., Evans-Holm, M., Siebel, C. W., Misra, S., Celniker, S. E. and Rubin, G. M. (2005) 'Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster', *Proc Natl Acad Sci U S A* 102(15): 5495-500.

Turner, G. C., Du, F. and Varshavsky, A. (2000) 'Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway', *Nature* 405(6786): 579-83.

Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F. M. and Roucou, X. (2013) 'Direct detection of alternative open reading frames translation products in human significantly expands the proteome', *PLoS One* 8(8): e70698.

Vanderperre, B., Lucier, J. F. and Roucou, X. (2012) 'HAltORF: a database of predicted out-of-frame alternative open reading frames in human', *Database (Oxford)* 2012: bas025.

Varshavsky, A. (1996) 'The N-end rule: functions, mysteries, uses', *Proc Natl Acad Sci U S A* 93(22): 12142-9.

Varshavsky, A. (2008) 'The N-end rule at atomic resolution', *Nat Struct Mol Biol* 15(12): 1238-40.

Vaughn, J. N., Ellingson, S. R., Mignone, F. and Arnim, A. (2012) 'Known and novel post-transcriptional regulatory sequences are conserved across plant families', *RNA* 18(3): 368-84.

Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. Holt, R. A. et al. (2001) 'The sequence of the human genome', *Science* 291(5507): 1304-51.

Vera, J. C., Wheat, C. W., Fescemyer, H. W., Frilander, M. J., Crawford, D. L., Hanski, I. and Marden, J. H. (2008) 'Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing', *Mol Ecol* 17(7): 1636-47.

Verma, R., Aravind, L., Oania, R., McDonald, W. H., Yates, J. R., Koonin, E. V. and Deshaies, R. J. (2002) 'Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome', *Science* 298(5593): 611-5.

Wang, K. C. and Chang, H. Y. (2011) 'Molecular mechanisms of long noncoding RNAs', *Mol Cell* 43(6): 904-14.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A. et al. (2011) 'A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression', *Nature* 472(7341): 120-4.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet* 10(1): 57-63.

Wawrzynow, A., Theibert, J. L., Murphy, C., Jona, I., Martonosi, A. and Collins, J. H. (1992) 'Sarcolipin, the "proteolipid" of skeletal muscle sarcoplasmic reticulum, is a unique, amphipathic, 31-residue peptide', *Arch Biochem Biophys* 298(2): 620-3.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J. and Bähler, J. (2008) 'Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution', *Nature* 453(7199): 1239-43.

Wilkinson, K. A. and Henley, J. M. (2010) 'Mechanisms, regulation and consequences of protein SUMOylation', *Biochem J* 428(2): 133-45.

Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M. and Varshavsky, A. (2008) 'Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway', *J Biol Chem* 283(35): 24011-28.

Xu, H., Wang, P., Fu, Y., Zheng, Y., Tang, Q., Si, L., You, J., Zhang, Z., Zhu, Y., Zhou, L. et al. (2010) 'Length of the ORF, position of the first AUG and the Kozak motif are important factors in potential dual-coding transcripts', *Cell Res* 20(4): 445-57.

Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D. and Peng, J. (2009) 'Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation', *Cell* 137(1): 133-45.

Yang, F. M., Pan, C. T., Tsai, H. M., Chiu, T. W., Wu, M. L. and Hu, M. C. (2009) 'Liver receptor homolog-1 localization in the nuclear body is regulated by sumoylation and cAMP signaling in rat granulosa cells', *FEBS J* 276(2): 425-36.

Yang, X., Tschaplinski, T. J., Hurst, G. B., Jawdy, S., Abraham, P. E., Lankford, P. K., Adams, R. M., Shah, M. B., Hettich, R. L., Lindquist, E. et al. (2011) 'Discovery and annotation of small proteins using genomics, proteomics, and computational approaches', *Genome Res* 21(4): 634-41.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D. and Felsenfeld, G. (2010) 'Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA', *Genes Dev* 24(22): 2543-55.

Yao, T., Song, L., Xu, W., DeMartino, G. N., Florens, L., Swanson, S. K., Washburn, M. P., Conaway, R. C., Conaway, J. W. and Cohen, R. E. (2006) 'Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1', *Nat Cell Biol* 8(9): 994-1002.

Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J. and Zhou, M. M. (2010) 'Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a', *Mol Cell* 38(5): 662-74.

Zhang, Z. and Dietrich, F. S. (2005) 'Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in Saccharomyces cerevisiae', *Curr Genet* 48(2): 77-87.

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M. and Lee, J. T. (2010) 'Genome-wide identification of polycomb-associated RNAs by RIP-seq', *Mol Cell* 40(6): 939-53.

Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. and Lee, J. T. (2008) 'Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome', *Science* 322(5902): 750-6.

Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q. et al. (2010) 'Long noncoding RNAs with enhancer-like function in human cells', *Cell* 143(1): 46-58.

At the end of 1990's, genome sequencing projects have marked the beginning of the post-genomic area. First annotation works revealed that the coding proportion of genomes is quite little (only 2-3% is transcribed into messenger RNA, mRNA). Furthermore, we know thanks to transcriptomic progress that almost all the genome is transcribed, underlying the amount of information remaining to be discovered.

Transcriptomic analysis revealed a great number of long non-coding RNAs (lncRNAs), such classified because of the historic 100 codons threshold. Indeed, behind 100 codons, an open reading frame (ORF) was considered as biologically meaningless.

But studies have shown that, as mRNAs, lncRNAs are sometimes capped, spliced and polyadenylated, 3 synonymous step of an RNA translation into protein. Moreover, their expression pattern is highly regulated during invertebrate (drosophila) as well as mammals (mice) development, suggesting their involvement in embryogenesis. That's why mutation of some of them has been performed, leading sometimes to an irrevocable result: an improper development followed by death.

That is what we observed with *polished-rice* (*pri*) lncRNA, identified in insects, especially in drosophila, the model organism we work on. When mutated, *pri* is embryonic lethal. This go with a strong smooth cuticle phenotype, which is unable to form typical hook structure: the trichomes. *Pri* studies have demonstrated that his function actually depend on 4 small peptides Pri (11-32aa) expression, showing as the 100 codons threshold was unsuitable.

We work on the terminal differentiation of embryonic epidermal cell (forming trichome) as a paradigm of cell morphogenesis. Our studies have attributed this shape change to a transcription factor: Shavenbaby (Svb), necessary and sufficient to trichome formation. Indeed, in cells expressing Svb, it induces one hundred target genes expression, all of them involved in an extension formation at the apical pole of the cell.

Svb expression is unaltered in *pri* mutant, and vice versa. Thus, Svb and Pri peptides are at a same level in the trichome differentiation program. Various approaches (as culture cell RNAi screen, molecular deciphering of Svb followed by biochemical analysis or genetic in adult drosophila) allowed us to understand this new mechanism in which small ORF encoded peptides (sPEP) control a genetic program. Our works demonstrate that Pri peptides trigger the interaction between Svb and Ubr3 (an E3 ubiquitin-ligase). Svb is thus polyubiquitynated and targeted to the proteasome, the proteins degradation factory. Here, unlike the majority of proteasomal degraded proteins, Svb undergoes a partial degradation. We show this is due to Svb own domains which may act as structural blocks, preventing the full degradation. The eliminated part of Svb contains a repressor domain and the released shorter form of Svb an activator one. This is why this partial degradation leads to a switch in Svb transcriptional activity, from a repressor to an activator of a morphogenesis program.

This study, among few others, is pionner in describing sPEPs mode of action in morphogenesis process. Many works have now demonstrated that this relatively new class of molecules certainly embodies an unexplored reservoir of biological functions.

Le grand projet des séquençages des génomes a marqué à la fin des 1990 l'entrée dans l'ère post-génomique. Les premiers travaux d'annotation ont révélé que la proportion codante des génomes est infime (seul 2-3% du génome est transcrit en ARN messagers, ARNm). Par ailleurs, les progrès en transcriptomique ont quant à eux démontré que la grande majorité du génome est transcrite, soulignant ainsi la grande part d'informations qu'il reste à découvrir.

Une des découvertes issues des analyses transcriptomiques est qu'il existe un grand nombre de longs ARN non-codants (lncRNAs). Ils ont été classés ainsi sur la base du seuil historique de 100 codons en deçà duquel on considérait que le cadre ouvert de lecture présent n'avait que peu de chance d'être biologiquement significatif.

Pourtant, des études ont montré que ces lncRNAs, à l'instar des ARNm, peuvent être coiffés, épissés et polyadénylés, trois étapes corrélées avec la traduction de l'ARN en protéine. De plus, ils présentent des patrons d'expression hautement régulés au cours du développement d'invertébrés (drosophile) ou de mammifères (souris). Ceci suggérant que ces lncRNAs puissent être impliqués dans l'embryogenèse, des mutations ont été générées et pour certains le résultat est sans appel : un développement incorrect suivi de la mort.

C'est le cas du lncRNA *polished-rice* (*pri*) identifié chez les insectes, et plus particulièrement chez la drosophile, l'organisme sur lequel nous travaillons. Lorsqu'il est muté, *pri* est embryonnaire létal. Ceci est accompagné d'un phénotype notoire : la cuticule des embryons est lisse, elle ne présente pas de structure en forme de poil : les trichomes. Les travaux sur *pri* ont montré que sa fonction dépend de l'expression de 4 peptides Pri (11- 32aa), démontrant l'inadéquation du seuil de 100 résidus.

Nous étudions la différenciation terminale des cellules épidermiques embryonnaires (formant les trichomes) comme paradigme de la morphogenèse cellulaire. Nos études avaient attribué ce remodelage à un facteur de transcription : Shavenbaby (Svb), nécessaire et suffisant à la formation des trichomes. En effet, dans les cellules où il est exprimé, Svb induit l'expression d'une centaine de gènes-cibles impliqués dans la formation d'une extension au pôle apical.

L'expression de Svb n'est pas altérée dans un mutant *pri*, et réciproquement, plaçant Svb et les peptides Pri au même niveau dans le programme de différenciation du trichome. Grâce à diverses approches (crible RNAi en culture cellulaire, découpage moléculaire de Svb suivi d'analyses biochimiques ou encore génétique chez la drosophile adulte), nous avons pu décrypter ce mécanisme nouveau où des peptides exprimés à partir de petits cadres ouverts de lecture (sPEP) contrôlent un programme génétique. En effet, nos travaux montrent que les peptides Pri sont requis dans la formation d'un complexe entre Svb et Ubr3 (une E3 ubiquitine-ligase). Ainsi, Svb est polyubiquitinylé et dirigé vers le protéasome, la machinerie de dégradation des protéines. Là, contrairement à la majorité des cas, Svb va subir une dégradation partielle. Nous montrons que ceci est dû à la structure même de Svb, dont certains domaines pourraient agir comme des blocs, empêchant sa dégradation totale. La région de Svb éliminée contient un domaine de répression, la forme courte de Svb relarguée un domaine d'activation. Cette dégradation partielle a donc pour conséquence un changement d'activité transcriptionnelle de Svb, de répresseur à activateur d'un programme génétique.

Cette étude est, avec de rares autres, pionnière, en ce sens où elle décrit le mécanisme d'action de sPEP dans des processus de morphogenèse, et que de nombreux travaux démontrent que cette catégorie de molécules relativement nouvelle pourrait représenter un réservoir encore inexploré de fonctions biologiques.