U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

The role of *Drosophila melanogaster polo* 3'untranslated region in gene expression

Pedro Alexandre Borges Pinto

Dissertação de Doutoramento em Ciências Biomédicas

2008

The role of *Drosophila melanogaster polo* 3' untranslated region in gene expression

Pedro Alexandre Borges Pinto

Dissertação de Doutoramento em Ciências Biomédicas

2008

Pedro Alexandre Borges Pinto

The role of *Drosophila melanogaster polo* 3' untranslated region in gene expression

Dissertação de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador: Doutora Maria Alexandra Marques Moreira Mourão Carmo

Categoria: Professora Auxiliar Convidada

Afiliação: Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Co-orientador: Doutor Claudio E. Sunkel

Categoria: Professor Catedrático

Afiliação: Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Este trabalho foi financiado pela
Fundação para a Ciência e Tecnologia

O trabalho apresentado nesta dissertação, resultou na redação de um manuscrito, que se encontra submetido para publicação e pendente de aceitação, com o título:

Pinto PAB, Freitas M, Henriques T, Coelho PA, Carmo AM, Sunkel CE, Moreira A. Polyadenylation signal selection unravels a new function for Polo in *Drosophila melanogaster* development. Submitted to *Science*.

Aos meus pais

Contents

Acknowledgments/Agradecimentos i
Summary ii
Resumo iv
Résumé vi

PART I - GENERAL INTRODUCTION

1. mRNA 3'end formation in gene expression: general overview 2

2. *Cis*-acting elements required for mRNA 3'end formation 3
2.1. Polyadenylation signal 3
2.2. Downstream Sequence Elements (DSEs) 4
2.3. Cleavage site 4
2.4. Upstream Sequence Elements (USEs) 5

3. *Trans*-acting factors required for mRNA 3'end processing 6
3.1. Cleavage/Polyadenylation Specificity Factor (CPSF) 6
3.2. Cleavage stimulation Factor (CstF) 8
3.3. Cleavage Factors (CF) 10
3.4. RNA polymerase II (RNA polII) 11
3.5. Symplekin 12
3.6. Poly(A) Polymerase (PAP) 12
3.7. Poly(A) Binding Protein Nuclear 1 (PABPN1) 13

4. mRNA 3'end formation 13

5. Alternative polyadenylation as a regulatory mechanism of gene expression 17
5.1. *Immunoglobulin M heavy chain (μ)* 18
5.2. *suppressor of forked (su(f))* 22
5.3. *enhancer of rudimentary (e(r))* 26

6. The *polo* gene 28

7. Aims of the thesis 29

PART II - EXPERIMENTAL WORK

Chapter I – The longer 2.5 kb polo mRNA is required for abdominal histoblasts proliferation during metamorphosis 31

1. Introduction 32

2. Results 34

2.1. Generating transgenic flies expressing only one *polo* mRNA 34

2.2. Characterization of the *gfp-polo∆pA1* and *gfp-polo∆pA2* transgenic flies 36

2.3. The longer *polo* mRNA is required for proper abdomen formation in *Drosophila* 40

2.4. Both *polo* mRNAs are expressed during the last stages of embryogenesis 42

2.5. *gfp-polo∆pA2; polo*^{9}abdominal histoblasts are absent in pupae with 26-27 hours APF 44

2.6. *gfp-polo∆pA2; polo*^{9} abdominal histoblasts fail to proliferate 47

2.7. *gfp-polo∆pA2; polo*^{9} abdominal histoblasts express lower levels of GFP-Polo 48

3. Discussion 50

Chapter II – Identification and characterization of Upstream Sequence Elements in the vicinity of the polo proximal poly(A) site 53

1. Introduction 54

2. Results 55

2.1. Identification of putative regulatory elements in the *polo* 3'UTR 55

2.2. Mutation of lod 14/ USE 1 inhibit *in vitro* polyadenylation of the proximal poly(A) site 59

2.3. USE 1 is required for the assembly of a protein complex 62

2.4. USE 2 and USE 3 are necessary for efficient *in vitr*o polyadenylation at the proximal poly(A) site 68

2.5. Different protein complexes are assembled in the vicinity of each poly(A) site 71

2.6. Mapping the p35 binding site in the USE 1 73

2.7. Purification of p35 75

2.8. Analysis of the tripartite Upstream Sequence Element 77
2.9. *In vivo* analysis of the Upstream Sequence Elements 82
2.10. *gfp-polo⊿USE 1* and *gfp-polo⊿pA1* transgenic flies present a similar phenotype 87

3. Discussion 90

PART III - GENERAL DISCUSSION

General Discussion 99

PART IV – MATERIALS AND METHODS

A. Materials 109
A1. Enzymes 109
A2. Kits 109
A3. Media 110
A4. General solutions and buffers 111
A5. Radioisotopes 113
A6. Antibodies 114
A7. Oligonucleotides 114

B. Methods 115
B1. Plasmid constructions 115
B2. Subcloning and DNA preparation 116
B2.1. Preparation of competent bacteria 116
B2.2. Restriction Enzyme digestion and DNA modifications 116
B2.3. Ligation of inserts into cloning vectors 116
B2.4. Polymerase Chain Reaction amplification 117
B2.5. Site-directed mutagenesis 117
B2.6. Transformation of competent bacteria 117
B2.7. Plasmid preparation 118
B3. Cell transfections 118
B4. RNA extraction 119
B5. RNA mapping 119
B5.1. DNA 3'end-labelled probes 119
B5.2. S1 Nuclease protection assay 120
B6. *In vitro* transcription 120

B7. *In vitro* cleavage and polyadenylation analysis 121
B8. UV. cross-linking of proteins to RNA 122
B9. Immunoprecipitations 122
B10. Preparation of HeLa cell nuclear extracts 123
B11. Affinity chromatography 123
B12. Protein precipitation with TCA 124
B13. *Drosophila* stocks 124
B14. Quantification of the viability of the transgenic flies 125
B15. Mitotic index 125
B16. *In situ* hybridisation on polytene chromosomes from salivary glands 126
B16.1. Biotin-labelled DNA probe 126
B16.2. Preparation of polytene chromosomes 126
B16.3. *In situ* hybridisation 126
B17. Fluorescent analysis of *Drosophila* embryos 127
B18. Adult abdomen preparations 127
B19. Dissection of pupa epidermis 128
B20. Immunostaining of third instar larvae abdominal histoblasts 128
B21. Quantification of the GFP levels in 26-27h APF pupae epidermis 129
B22. Western blot analysis 129
B23. Northern blot analysis 130

PART V – APPENDIX

Tables 132
Table AI 132
Table AII 133
Table AIII 136
Table AIV 136
Table AV 137
Table AVI 138

List of Abbreviations 139

PART VI – REFERENCES

References 145

Acknowledgments/Agradecimentos

Em primeiro lugar e como não podia deixar de ser, à Professora Alexandra Moreira, pela forma como me recebeu como aluno de doutoramento, pelo seu apoio ao longo de todo este precurso e pela amizade com que me presenteou. O seu entusiasmo e energia foram uma constante motivação.

Ao Professor Claudio Sunkel, pela sua disponibilidade, conselhos e "feedback" de tudo que está relacionado com o *polo*.

A todos os elementos do "Cell Activation and Gene Expression Lab", passado e presente. Em especial, ao Alexandre do Carmo por todos as sugestões e "input", à Mónica Castro, um verdadeiro exemplo de força, coragem e dedicação. To my good friend Martina Bamberger! Your friendship and generosity helped me to overcome the more stressful moments! We will always have Mozart! À Marta Oliveira por todo o carinho e amizade, à minha companheira de bancada Raquel Nunes pela infindável e contagiante boa disposição, à Carine Gonçalves, Telmo Henriques, Marta Freitas, Mafalda Santos e Joana Saraiva, pela vossa amizade e paciência com que me aturaram naqueles dias menos bons. Para aqueles que contribuiram directa ou indirectamente para este projecto e pela paciência ao longo destes anos, um Muito Obrigado por tudo.

À Carla Lopes e ao Paulo Pereira, por toda a ajuda e conhecimento que me transmitiram. Devo-vos tudo que sei sobre *Drosophila* e a paixão que dai nasceu.

À Paula Coelho e Joana Machado, pela vossa amizade e pronta disponibilidade em ajudar.

Ao Torcato Martins por todas as noitadas passadas à frente do microscopio.

À Rita Reis, por todas as conversas profundas e existenciais.....

À Paula Sampaio por toda a ajuda prestada com a microscopia.

À Augusta Monteiro sem a qual não poderia ir de férias sem ter que levar as moscas....

À Madalena Costa, Julio Santos, Cristina Macedo, Elsa Bronze,Tatiana Moutinho e Daniel Perez pela vossa amizade e pelos bons momentos que passamos ao longo destes anos.

À Adelaide Santos e à Maria João Falcão por toda a ajuda e pronta disponibilidade.

Ao Hugo, meu irmão! Quem é que disse que os irmãos mais velhos não podem apreender com os mais novos?!

Finalmente, aos meus pais! Por todo o vosso amor, apoio e sacrifícios que me permitiram chegar ao fim deste longo empreendimento!

Summary

The formation of an mRNA poly(A) tail is a complex process that requires several *cis*- and *trans*-acting elements. The synthesis of different mRNAs from a single gene as a consequence of the presence of several polyadenylation sites throughout a single transcription unit is a contribution to protein diversity as well as to the regulation of gene expression. In opposition to genes where polyadenylation sites are found in internal exons or introns resulting in the synthesis of different mRNAs coding for different protein isoforms, the presence of several polyadenylation sites in the 3'most exon results in the synthesis of mRNAs with the same open reading frame but with different 3'UTRs. The presence or absence in the 3'UTR, of specific *cis*-acting elements involved in mRNA transport, localization, stability or translation, provides further mechanisms to regulate gene expression. Consequently, mRNA 3'end formation is one of the key regulatory steps in eukaryotic gene expression.

The *Drosophila melanogaster polo* gene codes for a cell cycle kinase involved in multiple functions throughout the cell cycle. It has two polyadenylation signals in the 3'most exon which results in the synthesis of a 2.2 and 2.5 kb mRNAs. Both transcripts share the same open reading frame and although much data has been reported concerning the function and regulation of the Polo protein, little is known about the role of each transcript during development.

In this study, we analysed the requirement of each *polo* mRNA in the Polo function during *Drosophila* development. To address this question, transgenic flies expressing only the smaller or the longer *polo* mRNA were generated. Our results showed that there is tissue-specific requirement for the longer mRNA, produced by selection of the distal poly(A) signal, during the adult abdomen formation. In the absence of the longer *polo* mRNA, flies have a reduced viability and an abnormal abdomen formation, whereas flies expressing the longer mRNA are viable and present only a very mild abdominal phenotype. An analysis of the abdominal histoblasts nests, the clusters of cells responsible for the formation of the adult abdomen during the pupae stage, shows that in the absence of the longer *polo* transcript, these cells fail to proliferate upon entrance in metamorphosis, most likely due to low levels of Polo.

To understand the molecular mechanism underlying poly(A) site selection, we performed a search for regulatory *cis*-acting elements in the 3'UTR. By PCR mutagenesis, three Upstream Sequence Elements, USE 1, 2 and 3, positioned upstream the proximal poly(A) signal, were identified. *In vitro* analysis showed that all three elements were required for an efficient polyadenylation reaction through the assembly of a protein complex. An analysis the proximal and the distal poly(A) signals showed that different

protein complexes assembled upstream the polyadenylation signals. These results suggest that USE 1 may be a key regulatory element in polyadenylation site selection.

Finally, we investigated the function of the USE 1 *in vivo* by generating transgenic flies in which USE 1 was deleted. These flies present an identical phenotype to flies expressing only the longer *polo* mRNA. Therefore, these results support our *in vitro* findings, and suggest that USE 1 is likely to be a key regulatory element in alternative polyadenylation of *polo* during development.

Resumo

A síntese da cauda de poli(A) de RNAs mensageiros é um processo complexo que requer vários elementos em *cis* e *trans*. Como consequência, a formação da extremidade 3' de RNAs mensageiros é um dos principais pontos de regulação da expressão genética em eucariontes. A síntese de diferentes mRNAs a partir de um único gene devido à presença de vários locais de poliadenilação ao longo da unidade de transcrição contribui não só para a diversidade proteica como também para a regulação da expressão genética. Contrariamente aos genes em que os locais de poliadenilação se encontram em exões internos ou em intrões, o que resulta na síntese de diferentes mRNAs que codificam para diferentes proteínas, a presença de vários locais de poliadenilação no último exão resulta na síntese de RNAs mensageiros com o mesma região codificante mas com diferentes regiões 3' não traduzidas. A presença ou ausência na região 3' não traduzida, de elementos em *cis* específicos envolvidos em transporte/localização, estabilidade ou tradução de RNAs mensageiros, providencia mecanismos de regulação genética adicionais.

O gene *polo* em *Drosophila melanogaster* codifica uma cinase com múltiplas funções durante o ciclo celular. Possui dois locais de poliadenilação no último exão cujo uso resulta na síntese de dois RNAs mensageiros de 2.2 e 2.5 kb, respectivamente. Ambos os transcritos possuem a mesma região codificante e apesar de haver muita informação sobre a função e regulação da proteína Polo, muito pouco é conhecido sobre a função de cada transcrito durante o desenvolvimento.

Neste estudo , foi analisada a necessidade de cada RNA mensageiro do *polo* para a função do Polo durante o desenvolvimento de *Drosophila*. Para abordar esta questão foram geradas moscas transgénicas capazes de sintetizar apenas ou o RNA mensageiro mais curto ou mais longo do *polo*. Os nossos resultados mostram a existência de uma especificidade ao nível de tecido para o RNA mensageiro mais longo durante a formação do abdómen. Na ausência do mRNA mais longo, as moscas têm uma menor viabilidade e apresentam severas malformações no abdómen, enquanto que moscas que sintetizam apenas o mRNA mais longo são viáveis apresentando apenas um ligeiro fenótipo abdominal. Uma análise dos ninhos de histoblastos, agrupamentos de células responsáveis pela formação do abdómen durante o estadio de pupa, revela que na ausência do transcrito mais longo, estas células não proliferam após a entrada em metamorfose, muito provavelmente devido a baixos níveis de Polo.

Para compreender o mecanismo molecular responsável pela selecção do local de poliadenilação, foi realizada uma procura de elementos em *cis* reguladores na região 3' não traduzida. Por mutagénese, através de PCR, foram identificados três *Upstream*

Sequence Elements, USE 1, 2 e 3, localizados a montante do primeiro local de poliadenilação. Análises *in vitro* revelaram que os três elementos são necessários para uma eficiente reacção de poliadenilação, por intermédio do recrutamento de um complexo proteico. A análise mais detalhada do primeiro e segundo locais de poliadenilação revelou a existência de diferentes complexos proteicos localizados a montante dos locais de poliadenilação.

Finalmente, a função do USE 1 foi investigada *in vivo* através da geração de moscas transgénicas nas quais o USE 1 foi removido. Estas moscas apresentam um fenótipo semelhante ao das moscas que sintetizam apenas o RNA mensageiro mais longo. Assim, estes resultados apoiam as evidências obtidas *in vitro* e sugerem que o USE 1 é provavelmente um elemento regulador chave na poliadenilação alternativa do *polo* durante o desenvolvimento.

Résumé

La formation de la queue polyadénylée des ARN messagers est un processus complexe qui implique plusieurs éléments agissant en *cis* ou en *trans*. La synthèse de messagers différents à partir d'un même gène –grâce à la présence de différents sites de polyadénilation à l'intérieur d'une même unité de transcription- contribue à la diversité des protéines et permet la régulation l'expression des gènes. À l'inverse des sites de polyadénilation situés dans des exons ou introns internes, la présence de plusieurs sites de polyadénylation à l'extrémité 3' des gènes, induit la synthèse d'ARNm avec des séquences codantes identiques mais des extrémités 3' UTR différentes. La présence ou l'absence dans la région 3' UTR d'éléments agissant en *cis* et impliqués dans le transport, la localisation, la stabilité ou la traduction des messagers, permet l'action de différents mécanismes pour contrôler de l'expression des gènes. Par conséquent, la partie 3' des ARNm est un élément clef de l'expression des gènes chez les eucaryotes.

Le gène Polo de *Drosophila melanogaster* code pour une kinase impliquée dans différentes fonctions au cours du cycle cellulaire. Elle possède deux sites de polyadénilation situés en 3', qui induisent la synthèse de deux ARNm de 2,2 et 2,5 kb. Les deux messagers ont la même séquence codante. Bien que la fonction et la régulation de la kinase Polo soient bien décrites, on ignore le rôle que ces deux transcrits jouent au cours du développement.

Dans cette étude, nous avons analysé la contribution de chacun des transcrits pour la fonction de Polo, au cours du développement. Pour répondre à cette question, des mouches transgéniques qui expriment uniquement la forme courte ou la forme longue des messagers, ont été générées. Nos résultats montrent que la forme longue de l'ARNm est spécifiquement requise au cours de la formation de l'abdomen et qu'elle est sélectionnée par le signal distal poly(A). En absence du messager long, la viabilité des mouches est réduite et elles ont un abdomen anormal. À l'inverse, les mouches qui expriment la forme longue, sont viables et présentent un faible phénotype abdominal. Une analyse des histoblastes abdominaux – le groupe de cellules responsables de la formation de l'abdomen au cours du stade pupal- montre qu'en l'absence du transcrit long de *polo*, ces cellules ne prolifèrent pas lors de l'entrée en métamorphose, probablement du fait de la faible présence de Polo.

Afin de comprendre les mécanismes moléculaires impliqués dans la sélection des sites poly(A), nous avons recherché la présence d'éléments agissant en *cis*, dans la

partie 3' UTR. En utilisant la mutagenèse par PCR, nous avons pu identifier 3 éléments « Upstream Sequence Element » appelés USE 1, 2 et 3, et situés en amont de la partie poly(A). Une analyse *in vitro* nous a permis de montrer que ces éléments sont requis pour une polyadénilation efficace, à travers la formation d'un complexe protéique. Une analyse des signaux poly(A) proximaux et distaux montre que différents complexes protéiques s'assemblent en amont des signaux de polydénylation. Les résultats obtenus suggèrent qu'USE 1 soit un élément clef de sélection du site de polydénylation.

Nous avons finalement recherché la fonction de la séquence USE 1 *in vivo*, en produisant des mouches chez lesquelles la séquence USE 1 est supprimée. Ces mouches ont le même phénotype que celles qui expriment uniquement la forme longue du messager du gène Polo. Ces résultats corroborent l'étude faite *in vitro* et suggèrent qu'USE 1 soit un élément clef de la polyadénylation alternative du gène *polo* au cours du développement.

PART I

GENERAL INTRODUCTION

1. mRNA 3'end formation in gene expression: general overview

The genetic expression of protein-coding genes starts with transcription by RNA polymerase II. During this event, the polymerase generates a primary transcript (or pre-mRNA) which then undergoes a series of modifications. These are known as mRNA processing reactions and include the addition of a 7-methyl guanine nucleotide to the 5'end known as the Cap structure, the splicing of introns and the modification of the 3'end through cleavage and addition of a poly(A) tail with approximately 200-300 adenine nucleotides in a process known as mRNA 3'end formation (reviewed in Proudfoot *et al.,* 2002).

The study of the poly(A) tail, its synthesis and role in mRNA biogenesis, started almost half a century ago. Since then, the poly(A) tail has proved to be vital for almost all aspects of gene expression (reviewed in Atwater *et al.,* 1990; Nakielny *et al.,* 1997; Coller and Parker 2004; Kapp and Lorsch 2004). Although initially viewed as a postranscriptional modification, it is now known to occur cotranscriptionally, to influence other mRNA processing reactions and to be necessary for eukaryotic transcription termination (reviewed in Proudfoot *et al.,* 2002). After the synthesis of the mRNA, the poly(A) tail plays several important roles during the lifetime of the transcript. mRNA transport from the nucleus to the cytoplasm and the subcellular localization of specific mRNAs, its stability and translation, are all influenced and dependent of a functional poly(A) tail (reviewed in Atwater *et al.,* 1990; Nakielny *et al.,* 1997; Coller and Parker 2004; Kapp and Lorsch 2004).

In the last three decades, since the discovery of the poly(A) signal (Proudfoot and Brownlee 1976), the identification of the *cis*-acting elements involved in mRNA 3'end formation and the biochemical purification and cloning of the corresponding *trans*-acting factors, showed that this initially apparently simple process is indeed a highly complex mechanism requiring multiple protein complexes (reviewed in Colgan and Manley 1997; Zhao *et al.,* 1999; Edmonds 2002). Such degree of complexity and its relevance in gene expression has shown that mRNA 3'end formation is an important step in gene regulation (reviewed in Edwalds-Gilbert *et al.,* 1997). The analysis of the genomes of different organisms has further reinforced this view with the discovery of a significant number of genes that undergo the process of alternative polyadenylation (Tian *et al.,* 2005).

In the following chapters, we present a comprehensive overview of mRNA 3'end formation and its role in gene expression.

2. *Cis*-acting elements required for mRNA 3'end formation

The 3'end formation of virtually all eukaryotic mRNAs is a two step reaction. The transcribed pre-mRNA first undergoes an endonucleotic cleavage (cleavage reaction) followed by the polymerization of a poly(A) tail to the upstream cleavage product (polyadenylation reaction) (reviewed in Zhao *et al.,* 1999; Edmonds 2002). This process and its efficiency depends on *cis*-acting elements present in the pre-mRNA surrounding the cleavage site (Fig.1). The *cis*-acting elements involved in the 3'end formation have been well characterized for several organisms during the last thirty years. Here we describe the mammalian and *Drosophila melanogaster* polyadenylation signals.

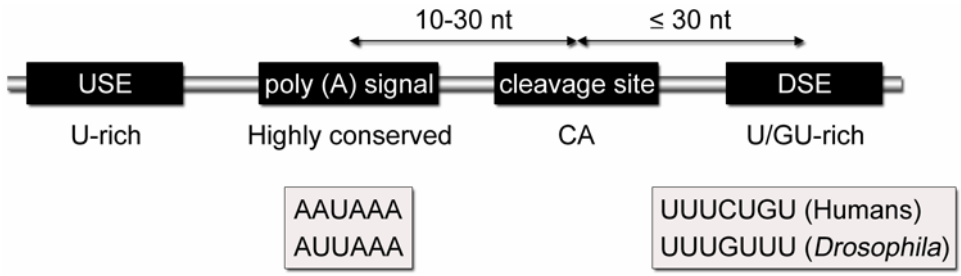

Figure 1. Schematic representation of the *cis*-acting elements involved in 3'end formation in Humans and *Drosophila melanogaster*. The AAUAAA and the AUUAAA represent the first and the second most conserved variants of the poly(A) signal in Humans and *Drosophila*. The cleavage site is usually defined by a CA dinucleotide. The USE (Upstream Sequence Element) and the DSE (Downstream Sequence Element) are poorly conserved elements usually defined as U and U/GU-rich elements, respectively. The similar UUUCUGU and UUUGUUU elements represent the consensus sequence for the DSE in Humans and *Drosophila*, respectively, accordingly to Retelska *et al.* (2006). Adapted from Zhao *et al.* (1999).

2.1. *Polyadenylation signal*

The polyadenylation signal was first described as the hexamer AAUAAA upon comparison of six different mRNAs (Proudfoot and Brownlee 1976). Subsequent analysis of several genes showed that this hexamer was present in the vast majority of the genes, making the poly(A) signal one of the most conserved *cis*-acting elements ever described (reviewed in Colgan and Manley 1997). However, as mentioned, it is not present in all genes. Bioinformatics studies performed using a detailed collection of ESTs show that the AAUAAA hexamer is indeed the most conserved variant of the poly(A) signal, present in 48-69 % of the genes (Graber *et al.,* 1999; Beaudoing *et al.,* 2000; MacDonald and Redondo 2002; Zarudnaya *et al.,* 2003; Retelska *et al.,* 2006). The most common variation of this canonical signal is a single A-U nucleotide substitution on the second

position, AUUAAA. This variant is present in 12-16% of the genes. Other variants are also present in the genome although in a very low frequency (Beaudoing *et al.,* 2000; MacDonald and Redondo 2002; Zarudnaya *et al.,* 2003; Retelska *et al.,* 2006).

Mutagenesis analysis has shown that the poly(A) signal is required for both the endonucleotic cleavage and polyadenylation reactions (Wickens and Stephenson 1984; Wilusz *et al.,* 1989; Sheets *et al.,* 1990). Quantification of the cleavage and polyadenylation reactions efficiency of pre-mRNAs containing variations to the canonical signal has shown that the AAUAAA is the most efficient hexamer. All other non-canonical signals show a strong reduction on cleavage and polyadenylation efficiency. The only variant with a similar efficiency to the canonical poly(A) signal is the AUUAAA (Wilusz *et al.,* 1989; Sheets *et al.,* 1990). Therefore, there is a direct correlation between the efficiency of the different poly(A) signals variants and their frequency in the genome.

2.2. *Downstream Sequence Elements (DSEs)*

The Downstream Sequence Element, positioned approximately within the 30 nucleotides downstream of the cleavage site (Chen *et al.,* 1995; Legendre and Gautheret 2003; Retelska *et al.,* 2006) and present in the vast majority of the human genes (~80% of total pre-mRNAs with AAUAAA or AUUAAA poly(A) signal) (Zarudnaya *et al.,* 2003), is necessary only for the cleavage reaction (reviewed in Edmonds 2002). However, there are genes in which the DSE is absent (Moreira *et al.,* 1995; Zarudnaya *et al.,* 2003) and contrary to the polyadenylation signal, this *cis*-acting element is poorly conserved. This element has been described as a U/GU rich element (reviewed in Colgan and Manley 1997; Zhao *et al.,* 1999). However, bioinformatic analysis of the human genome shows that the DSE is not as much GU-rich as usually stated but rather U-rich (Legendre and Gautheret 2003; Retelska *et al.,* 2006). *In vitro* mRNA 3'end processing reactions have shown that the function of this *cis*-acting element does not depend on its particular sequence but rather on its nucleotide composition (Takagaki and Manley 1997).

2.3. *Cleavage site*

The cleavage site defines the nucleotide where the pre-mRNA is cleaved. It is defined by the position of the poly(A) signal and the DSE (Chen *et al.,* 1995). The distance between these two *cis*-acting elements not only define the major cleavage site but also determines the strength of the poly(A) site (Chou *et al.,* 1994; MacDonald *et al.,* 1994).

A CA dinucleotide is usually present immediately upstream the site where the endonucleotic cleavage occurs. Initial studies showed that although a CA dinucleotide is

preferred, it is not absolutely required (Sheets *et al.,* 1990). However, 59% and 71% of the analysed mRNAs have a cytosine and adenine, respectively, at the -2 and -1 position. Moreover, mutagenesis analysis showed that although substitution of the adenosine does not affect the efficiency of the cleavage reaction, it does lead to small changes on the 3'end of the cleaved pre-mRNA (Sheets *et al.,* 1990; Chen *et al.,* 1995). These studies show that the cleavage site is influenced by the nucleotide preference in the order A>U>C>>G.

However, study of the genetic disorder known as *thrombophilia* revealed a more active role of the cleavage site in mRNA 3'end formation (Gehring *et al.,* 2001; Danckwardt *et al.,* 2006). This condition is characterized among other things by elevated Prothrombin plasma levels and an increased risk of thrombosis (reviewed in Kujovich 2006). The *prothrombin* or *coagulation factor II* (*F2*) gene has a CG dinucleotide that precedes the cleavage site. Molecular analysis of two point mutations found in *thrombophilia* patients at the -2 or -1 position has shown that the dinucleotide preceding the cleavage site has a role in the mRNA 3'end efficiency (Gehring *et al.,* 2001; Danckwardt *et al.,* 2006). Indeed, a change in the CG dinucleotide to either TG (Danckwardt *et al.,* 2006) or CA (Gehring *et al.,* 2001), as seen in these patients, leads to an increase efficiency of mRNA 3'end processing, resulting in increased mRNA and protein levels.

2.4. *Upstream Sequence Elements (USEs)*

Unlike the poly(A) signal and the DSE which are primary *cis*-acting elements necessary for mRNA 3'end formation, USEs are auxiliary elements that can modulate the overall mRNA 3'end efficiency (reviewed in Zhao *et al.,* 1999). Like DSEs, USEs are poorly conserved and have been also described as U-rich elements.

These elements originally identified in viral poly(A) sites (Carswell and Alwine 1989; DeZazzo and Imperiale 1989; Russnak and Ganem 1990; Brown *et al.,* 1991; DeZazzo *et al.,* 1991; Russnak 1991; Sanfacon *et al.,* 1991; Valsamakis *et al.,* 1991; Cherrington and Ganem 1992) have also been identified in a small number of cellular genes like the *C2 complement* (Moreira *et al.,* 1995; Moreira *et al.,* 1998), the *lamin B2* (Brackenridge and Proudfoot 2000), the *cyclooxygenase-2* (Hall-Pogar *et al.,* 2005), the *collagen* (Natalizio *et al.,* 2002) and the *prothrombin* genes (Danckwardt *et al.,* 2004; Danckwardt *et al.,* 2007). Despite the small number of genes where these elements have been identified, analysis of the human genome shows that they are likely to have a broader distribution in the genome since an uracil nucleotide enrichment is observed upstream the poly(A) signals (Legendre and Gautheret 2003).

3. *Trans*-acting factors required for mRNA 3'end processing

The development of *in vitro* systems able to reproduce mRNA 3'end formation (Manley 1983; Moore and Sharp 1984; Moore and Sharp 1985) allowed the biochemical characterization of this reaction. The isolation and purification of activities required for this process *via* biochemical fractionation of protein extracts, allowed the cloning and subsequent characterization of several *trans*-acting factors involved in mRNA 3'end formation. This way it was possible to identify and characterize several multiprotein complexes with specific functions in the endonucleotic cleavage and polyadenylation reactions (reviewed in Zhao *et al.,* 1999; Edmonds 2002). In mammals, eight protein complexes involved in mRNA 3' end formation have been identified and characterized: the *Cleavage/Polyadenylation Specificity Factor* (*CPSF*), *Cleavage stimulation Factor* (*CstF*), *Cleavage Factors* I_m and II_m (*CFI*$_m$, *CFII*$_m$), *RNA Polymerase II (RNA polII)*, *Symplekin*, *Poly (A) Polymerase* (*PAP*), and *Poly (A) Binding Protein Nuclear 1* (*PABPN 1*).

3.1. *Cleavage/Polyadenylation Specificity Factor (CPSF)*

The CPSF is a multiprotein complex first isolated as an AAUAAA-specific activity required for pre-mRNA 3'end processing (Christofori and Keller 1988; Gilmartin and Nevins 1989; Takagaki *et al.,* 1989). This activity was subsequently purified as a complex of four protein subunits of 160, 100, 73, and 30 kDa (CPSF-160, CPSF-100, CPSF-73 and CPSF-30, respectively) (Bienroth *et al.,* 1991; Murthy and Manley 1992). More recently, work from the laboratory of Walter Keller identified the human homologue of Fip1 (hFip1) as an additional component of CPSF (Kaufmann *et al.,* 2004). CPSF binds specifically to AAUAAA containing pre-mRNAs and is required for both the endonucleotic cleavage and polyadenylation reactions (Christofori and Keller 1988; Gilmartin and Nevins 1989; Takagaki *et al.,* 1989; Bienroth *et al.,* 1991; Keller *et al.,* 1991). Therefore, the function of CPSF is directly related with the requirement for the poly(A) signal in mRNA 3'end formation, as both are required for the cleavage and polyadenylation reactions.

The binding of CPSF to the RNA occurs via the 160 kDa subunit (CPSF-160) (Murthy and Manley 1995). CPSF-160 has a bipartite nuclear localization signal (NLS) and two short sequences, RNP1 and RNP2, similar to the RNP (rinonucleoprotein) motif consensus sequence described for a wide range of RNA binding proteins. Although CPSF-160 shows a preferred binding to RNAs containing the poly(A) signal, the entire CPSF complex exhibits a higher specificity towards AAUAAA containing RNAs which suggests that the other subunits are required to enhance the specificity of the complex for the poly(A) signal.

The binding of CPSF to RNA occurs with low affinity. However, it increases in the presence of CstF (Gilmartin and Nevins 1991; Weiss *et al.,* 1991; Murthy and Manley 1992) which indicates that these two protein complexes interact cooperatively enhancing each others binding to the pre-mRNA. This is in agreement with an increased binding of CstF-64 and an enhancement of the cleavage reaction when both CPSF and CstF are present (Murthy and Manley 1992). This cooperative interaction is likely to occur via CPSF-160 since this subunit interacts with CstF-77 (Murthy and Manley 1995) (see bellow).

In vitro studies showed that CPSF and PAP are sufficient to reconstitute the polyadenylation reaction of pre-mRNA substrates (Christofori and Keller 1988; Takagaki *et al.,* 1989). This requirement is likely to reflect the interaction between CPSF-160 and PAP which was shown by immunoprecipitation assays (Murthy and Manley 1995). However, high levels of CPSF-160 inhibit the PAP activity in unspecific assays (Murthy and Manley 1995).

Immunoprecipitation using a CPSF-100 monoclonal or a polyclonal antibody co-immunoprecipitates CPSF-160, CPSF-73 and CPSF-30, supporting the idea that all four proteins are present in a complex (Jenny *et al.,* 1994). Incubation of the 3'end processing reaction with a CPSF-100 antibody inhibits both cleavage and polyadenylation reactions, as expected from the function of CPSF in mRNA 3'end processing. However, unlike CPSF-160, the 100 kDa subunit does not bind to pre-mRNA. Its precise function is unknown.

CPSF-100 is related to CPSF-73 since both protein share 26% identity, 48% similarity over a 375 amino acid fragment which suggests that both subunits evolved from an ancestral gene and therefore might have similar functions in CPSF (Jenny *et al.,* 1994; Jenny *et al.,* 1996). CPSF-73 is a member of the β-CASP subfamily of the metalo-β-lactamase superfamily (Callebaut *et al.,* 2002). Proteins from this subfamily comprise proteins with nucleic acid substrates. Recent studies have shown that this protein is the actual endonuclease activity responsible for the cleavage reaction during mRNA 3'end formation (Ryan *et al.,* 2004; Mandel *et al.,* 2006). In agreement with the similarity between the CPSF-100 and CPSF-73, CPSF-100 is also a member of the β-CASP subfamily (Callebaut *et al.,* 2002). However, unlike CPSF-73, the β-CASP domain of CPSF-100 is predicted to be inactive. Thus it has been suggested that CPSF-100 may play a modulator role of the enzymatic activity (Callebaut *et al.,* 2002).

Although the yeast Fip1 protein (Factor interacting with PAP 1) was first identified in 1995 (Preker *et al.,* 1995) as a mRNA 3'end processing factor involved in polyadenylation, its human homologue, hFip1, was only recently cloned and characterized (Kaufmann *et al.,* 2004). hFip1 interacts with several mRNA 3'end processing factors

(Kaufmann *et al.,* 2004). Pull down assays showed that the N-terminal of hFip1 is able to interact with CPSF-160, CPSF-30, CstF-77 and PAP. The C-terminal of hFip1 is necessary for RNA binding as well for further interactions with CPSF-160. Recently, it was suggested that hFip1 interacts directly with CFI_m (Venkataraman *et al.,* 2005). *In vitro* studies showed that hFip1 is able to stimulate PAP activity regardless of poly(A) signal specificity (Kaufmann *et al.,* 2004). hFip1 binds preferentially to U-rich sequences and has been shown to bind to the Ad-L3 USE and to stimulate PAP activity in a concentration dependent manner (Kaufmann *et al.,* 2004). Indeed, hFip1 is able to stimulate polyadenylation of other USE containing pre-mRNAs such as the SV40 late and the human *C2 complement* pre-mRNAs (Kaufmann *et al.,* 2004). Therefore, since USEs stabilize or/and enhance the CPSF-dependent stimulation, it has been suggested that hFip1 may mediate the USE-dependent stimulation of polyadenylation (Kaufmann *et al.,* 2004).

The precise function of CPSF-30 is unclear. Although some studies have suggested that this subunit is not necessary for mRNA 3'end processing (Murthy and Manley 1992; Gilmartin *et al.,* 1995), the *Danio rerio* and *Saccharomyces cerevisiae* homologues are essential for viability (Gaiano *et al.,* 1996; Barabino *et al.,* 1997). Moreover, antibodies against CPSF-30 co-immunoprecipitates the other CPSF subunits, 160, 100 and 73 kDA subunits (Jenny *et al.,* 1994; Barabino *et al.,* 1997) and it was shown that CPSF-30 also directly interacts with hFip1 (Kaufmann *et al.,* 2004). Immunodepletion of CPSF-30 inhibits the cleavage and polyadenylation reactions (Barabino *et al.,* 1997).

Analysis of the CPSF-30 primary amino acid structure revealed five putative CCCH-type zinc-binding motifs and a CCHC-type carboxy-terminal "zinc knuckle", both motifs implicated in binding nucleic acids (Barabino *et al.,* 1997). Indeed, together with CPSF-160, CPSF-30 was shown to bind preferentially to U-rich sequences in the pre-mRNA, through the zinc knuckle motif (Jenny *et al.,* 1994; Barabino *et al.,* 1997). Thus it may have a function in further stabilizing the CPSF– RNA complex.

3.2. *Cleavage stimulation Factor (CstF)*

This protein complex is required for the cleavage reaction but not for polyadenylation (Gilmartin and Nevins 1989; Takagaki *et al.,* 1989; Takagaki *et al.,* 1990). It is composed of three protein subunits of 77, 64 and 50 kDa (Cstf-77, CstF-64 and CstF-50), respectively (Takagaki *et al.,* 1990).

The 64 kDa subunit (CstF-64) was one of the first identified RNA-binding factors involved in the mRNA 3'end processing (Moore *et al.,* 1988; Wilusz and Shenk 1988;

Wilusz *et al.,* 1990). Cloning of CstF-64 has revealed an N-terminal RNA-binding domain (RBD) responsible for the binding to RNA (Takagaki *et al.,* 1992). The RBD is connected through a hinge region to a 211 amino acid long stretch enriched in proline (19.3%) and glycine (18.9%). Immediately after this structure is a domain of twelve repeats of MEARA/G consensus that is predicted to form a long α-helix stabilized by salt bridges with neighbouring acidic residues, which is followed by another proline (19.3%) and glycine (24.6%) rich motif (56 amino acid residues long) (Takagaki *et al.,* 1996). Although CstF-64 is highly conserved between human, mouse and chicken, the two most conserved domains are the RBD (100% conservation between human and chicken) and the hinge domain (92% conservation between human and chicken) (Takagaki *et al.,* 1996). Whereas the conservation of the RBD is related to its RNA-binding ability, the highly conserved hinge domain is required for interaction with CstF-77 and the 3'end processing factor Symplekin (see bellow) (Takagaki and Manley 2000).

CstF-64 is the only RNA-binding protein in the CstF complex. It is through this protein that the entire complex binds to the DSE (MacDonald *et al.,* 1994). As seen with SELEX studies using the N-terminal fragment of CstF-64 (that contains the RNA-binding domain), this protein specifically selects GU-rich sequences (Takagaki and Manley 1997). Although with no consensus sequence (as is the case of DSE) the selected sequences are characterized by GU dinucleotides, frequently repeated, and/or by U repeats (up to 4 consecutives Us). Moreover, in agreement with reports that five consecutive Us are able to replace the natural DSE of SV40L and SV40E RNAs, these SELEX-selected elements are able to function as DSEs in *in vitro* cleavage reactions by replacing the DSE of Ad-L3 (Wilusz and Shenk 1990; Takagaki and Manley 1997).

The CstF-77 is the subunit that holds together the CstF complex (Takagaki and Manley 1994). The C-terminal domain, that includes a proline rich motif, interacts with CstF-64 and CstF-50 (Takagaki and Manley 2000). These two subunits do not interact with each other. Instead, CstF-77 bridges the two proteins (Takagaki and Manley 1994; Takagaki and Manley 2000). The importance of CstF-77 in the assembly of CstF is further supported by its putative bipartite nuclear localization signal. Despite the fact that all three CstF subunits have a nuclear localization, CstF-77 is the only subunit with a NLS, which suggests that CstF is probably assembled in the cytoplasm prior to transport to the nucleus, with CstF-77 playing a central role in the complex assembly (Takagaki and Manley 2000). CstF-77 also has eleven HAT (half a TPR) motifs that lacks the highly conserved alanine and glycine residues of the TPR motifs (Preker and Keller 1998). This motif is likely to mediate protein-protein interactions. Consistent with this, interaction with CstF-50 requires a C-terminal fragment containing the two last HAT motifs and the proline rich motif (Takagaki and Manley 2000). It is also likely that the HAT domain may be

required for interaction with CPSF-160 through which it may stabilize the assembly of CPSF and CstF to the pre-mRNA. In addition to its interaction with CstF-64 and CstF-50, CstF-77 also shows a strong self-association through its C-terminus (Takagaki and Manley 2000).

The CstF-50 has a significant homology with the β-subunits of the human and bovine G protein/transducin (Takagaki and Manley 1992). It has seven WD-40/transducin repeats that have been suggested to be involved in protein-protein interactions. Interestingly these motifs were found to be specifically necessary for interaction with CstF-77 (Takagaki and Manley 2000). Like CstF-77, CstF-50 interacts with itself, with this association occurring independently of the WD-40 repeats, through its N-terminal fragment (Takagaki and Manley 2000).

Although CstF-77 and CstF-50 self-associate, the molecular weight of the CstF complex as well the stoichiometry of its subunits (1:1:1) suggest that the purified CstF is a monomer composed of only three subunits (Takagaki *et al.,* 1990). However, despite the fact that the self-association of these subunits may compete with the actual CstF-50/CstF-77 interaction, it has been suggested that it may allow CstF complexes to interact with each other during the course of mRNA 3'end processing (Takagaki and Manley 2000).

3.3. *Cleavage Factors (CF)*

The Cleavage Factors I_m (CFI_m) and II_m ($CFII_m$) were first described as necessary for the endonucleotic cleavage of the pre-mRNA (Takagaki *et al.,* 1989).

The CFI_m was purified and characterized as composed of three subunits of 68, 59 and 25 kDa (Ruegsegger *et al.,* 1996). All three proteins are able to bind RNA. The high similarity between CFI_m68 and CFI_m59 has suggested that different dimers may be formed with CFI_m25 (Ruegsegger *et al.,* 1998). In fact it has been shown that $CFI_m^{68/25}$ and $CFI_m^{59/25}$ do exist and are required for the mRNA 3'end processing (Ruegsegger *et al.,* 1998; Millevoi *et al.,* 2006). Although initially described as necessary for the cleavage reaction, recent reports showed that CFI_m is also involved in the polyadenylation reaction (Brown and Gilmartin 2003; Venkataraman *et al.,* 2005; Millevoi *et al.,* 2006). It is involved in the early steps of mRNA 3'end processing, interacting with CPSF and bringing it to the RNA, stabilizing the assembly of the mRNA 3'end machinery (Ruegsegger *et al.,* 1996; Venkataraman *et al.,* 2005).

The $CFII_m$ has been purified into two fractions: $CFIIA_m$ necessary for the cleavage reaction and $CFIIB_m$ that seems to have a stimulatory role (de Vries *et al.,* 2000). The $CFIIA_m$ has been purified and hClp1 (human Cleavage/Polyadenylation Protein 1) and

hPCF11 (human Protein 1 of CFI) have been suggested to be two subunits of $CFIIA_m$ (de Vries *et al.,* 2000). $CFIIB_m$ has not yet been purified to homogeneity.

3.4. *RNA polymerase II (RNA polII)*

In the last ten years the role of RNA polII in mRNA processing has been extensively studied. Apart from transcription, RNA polII influences all steps in mRNA processing: capping, splicing and 3'end formation (reviewed in Hirose and Manley 2000; Proudfoot 2004; Bentley 2005). The initial observations that showed that RNA polII had a role in mRNA processing came from transcription of protein coding genes using promoters from T7 RNA polymerase and RNA polymerase I and III (Smale and Tjian 1985; Sisodia *et al.,* 1987; Mifflin and Kellems 1991). These transcripts did not undergo efficient 3'end processing, while those synthesized by RNA polII did.

The major difference between RNA polymerase II and these other polymerases lies on the unusual C-terminal domain (CTD) of its largest subunit. This domain comprises tandem heptads (52 in mammals and 26 in yeast) with the consensus sequence YSPTSPS (reviewed in Lewin 1997). The CTD was shown to be necessary for mRNA processing *in vivo* (McCracken *et al.,* 1997). Truncation of the CTD not only inhibits splicing but also mRNA 3'end formation (McCracken *et al.,* 1997). Its function in 3'end processing was further characterized *in vitro*, showing that CTD is necessary for the endonucleotic cleavage of the pre-mRNA (Hirose and Manley 1998). In agreement with its role in the cleavage reaction, CTD associates with CPSF and CstF (Dantonel *et al.,* 1997; McCracken *et al.,* 1997). In fact, the RNA polII CTD directly interacts with CstF through CstF-50 (via the N-terminal 211 amino acid residues) and CstF-77 (McCracken *et al.,* 1997).

Moreover, purification of the transcription factor TFIID shows that CSPF is present in these purified fractions (Dantonel *et al.,* 1997). Indeed CPSF-160 directly interacts with several TFIID subunits. Analysis of the TFIID-CPSF interaction showed that CPSF is recruited to the pre-initiation complex via interaction with TFIID and once RNA polII enters elongation, CPSF associates with the RNA polII elongation complex (Dantonel *et al.,* 1997). Furthermore, an interaction with a transcription co-activator has also been described for CstF-64 (Calvo and Manley 2001). It was shown that the CstF 64 kDa subunit interacts both *in vitro* and *in vivo* with PC4.

Therefore, the present data supports that mRNA 3'end processing factors are brought to the transcription pre-initiation complex and loaded on the RNA polII. As it enters the elongation phase, the CTD helps in the recruitment of mRNA 3'processing factors. As soon as poly(A) signals are transcribed by polII and exposed to the processing

factors, mRNA 3'end takes place. Thus, through interaction with other processing factors, the CTD of RNA polII participates in the formation and/or function of a stable and catalytically active processing complex (reviewed in Hirose and Manley 2000).

3.5. *Symplekin*

Symplekin is a 141 kDa protein that appears to act as an assembly/scaffolding protein in mRNA 3'end formation (Takagaki and Manley 2000). This notion arose from the fact that Symplekin can be isolated together with CPSF and CstF and from the observation of a direct interaction between CstF-64 and Symplekin (Takagaki and Manley 2000). It was suggested that Symplekin may establish additional interactions with CPSF. Indeed, in *Xenopus laevis*, Symplekin was also shown to interact with CPSF (Hofmann *et al.,* 2002).

The observation that Symplekin is specifically modified with addition of SUMO and that this modification is required for the assembly of the 3'end processing complex and for the cleavage and polyadenylation reactions, further supports the role of Symplekin as an assembly/scaffolding protein (Vethantham *et al.,* 2007).

3.6. *Poly(A) Polymerase (PAP)*

The Poly(A) Polymerase is a monomer (Wahle 1991b) required both for the cleavage and polyadenylation reactions (Christofori and Keller 1988; Takagaki *et al.,* 1988; Christofori and Keller 1989; Takagaki *et al.,* 1989). Several alternative splicing isoforms have been identified in several organisms (reviewed in Colgan and Manley 1997; Zhao *et al.,* 1999). Despite being required for the cleavage reaction, the enzymatic activity of PAP is only necessary for the polyadenylation reaction.

Although it catalyses the polymerization of the poly(A) tail, PAP has a very low affinity for any RNA substrate *in vitro* (Wahle 1991b). An increased affinity to the substrate is achieved when PAP is incubated in the presence of Mn^{2+}. Additionally, under these conditions, PAP presents a higher activity *in vitro* (Takagaki *et al.,* 1988; Wahle 1991b). The specificity towards pre-mRNAs shown under physiological conditions is probably due to interactions with CPSF via the CPSF-160 and hFip1 (Murthy and Manley 1995; Kaufmann *et al.,* 2004). Recognition of the poly(A) signal by CPSF, brings PAP closer to the pre-mRNA stabilizing the CPSF-RNA complex and specifying polyadenylation of pre-mRNAs (Murthy and Manley 1992; Bienroth *et al.,* 1993).

3.7. *Poly(A) Binding Protein Nuclear 1 (PABPN1)*

Poly(A) Binding Protein Nuclear 1 is as a polyadenylation stimulation factor (reviewed in Kuhn and Wahle 2004). Although both CPSF and PAP are sufficient to carryout polyadenylation (Christofori and Keller 1988; Christofori and Keller 1989; Bienroth *et al.,* 1991; Murthy and Manley 1992), PABPN1 enhances the efficiency of the reaction as well, as together with CPSF and PAP, determines the length of the poly(A) tails (Wahle 1991a; Bienroth *et al.,* 1993).

Polymerization of the poly(A) tail by PAP alone occurs in a distributive mode, *i.e*, with addition of a single AMP (adenosine monophosphate) per binding event (Wahle 1991b). In the presence of CPSF, PAP presents a limited processivity (addition of more than one AMP per binding event) and as a result the length of the poly(A) tails are smaller than the length observed for mRNAs *in vivo* (Wahle 1991a; Bienroth *et al.,* 1993). Only in the presence of CPSF and PABPN1 is a higher processivity obtained (Wahle 1991a; Bienroth *et al.,* 1993). PABPN1 functions by binding to the growing poly(A) tail and further stabilizing the CPSF-PAP complex (Wahle 1991a; Bienroth *et al.,* 1993), probably through direct interaction between PABPN1 and CPSF-30 (Chen *et al.,* 1999). As a result, a higher polyadenylation efficiency is obtained.

In vivo, the maximal length of the poly(A) tail is 200-300 adenine nucleotides. PABPN1, together with CPSF and PAP determines the poly(A) tail length, with the PABPN1 coating the growing poly(A) tail (with a stoichiometry of 23 nt/PABPN1) through the binding to a twelve adenine nucleotide binding site (Wahle *et al.,* 1993; Wahle 1995). It is believed that the upper limit of a poly(A) tail is achieved through disassembly of the CPSF-PAP-PABPN1 complex (Wahle 1995).

4. mRNA 3’end formation

mRNA 3’end formation is, in very simple terms, two coupled reactions in which the pre-mRNA is first cleaved (cleavage reaction) followed by the addition of an adenylate tail to the upstream cleavage product (polyadenylation reaction). However, this process is a far more complex event. The requirement of multiple protein complexes and the intricate nature of the several interactions between the different subunits of these complexes makes mRNA 3’end formation a major step in the regulation of gene expression (reviewed in Edmonds 2002).

The importance of mRNA 3’end formation in gene expression is further reinforced when the interactions between the different reactions that constitute mRNA processing

are considered: capping of the 5'end, the splicing of introns and 3'end formation (reviewed in Proudfoot *et al.,* 2002). Although these reactions are biochemically distinct, they influence each other (Fig. 2). Furthermore, each of these reactions is coupled with transcription, via protein interactions with the CTD of RNA polII, which plays a central role in the coupling of all these reactions (reviewed in Proudfoot *et al.,* 2002). Therefore mRNA processing occurs cotranscriptionally.

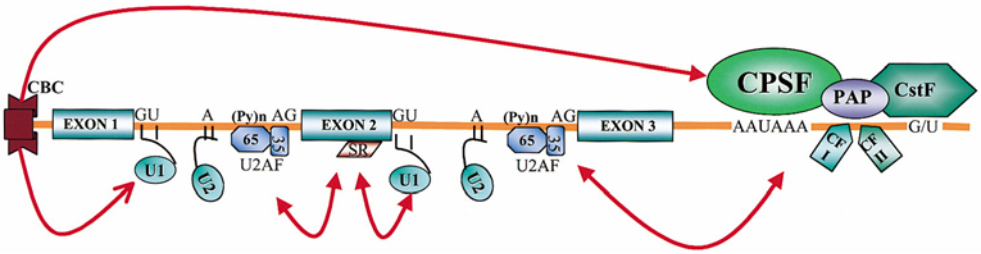

Figure 2. Molecular interactions between mRNA processing reactions. Although biochemically distinct, the capping, splicing and 3'end formation reactions influence each other. CBC: Cap Binding Complex; U1: U1 snRNP; U2: U2 snRNP; SR: SR proteins; U2AF: U2 Auxiliary Factor; (Py)n: Pyrymidine tract; GU and AG: represent the 5' and 3' splice sites respectively; A: branch point; CPSF: Cleavage/Polyadenylation Specificity Factor; CstF: Cleavage stimulation Factor; PAP: Poly(A) Polymerase; CFI: Cleavage Factor I; CFII: Cleavage Factor II; AAUAAA: poly(A) signal; G/U: Downstream Sequence Element. The red arrows represent interactions between the different reactions. Adapted from Proudfoot *et al.* (2002).

The link between transcription and mRNA 3'end formation starts at the promoter. Through interaction with the general transcription factor TFIID, CPSF is recruited to the pre-initiation complex (Dantonel *et al.,* 1997). Upon phosphorylation of Ser-2, RNA polymerase II starts transcription and CPSF is transferred from TFIID to the CTD of RNA polymerase II (Dantonel *et al.,* 1997; McCracken *et al.,* 1997). It is likely that together with CPSF, both CstF and CFI_m are also associated with RNA polII since the early steps of transcription (McCracken *et al.,* 1997; Venkataraman *et al.,* 2005). In fact, CstF interacts with both hypo- and hyperphosporylated forms of RNA polII (McCracken *et al.,* 1997). Moreover, chromatin immunoprecipitation (ChIp) assays have shown that CPSF, CstF and CFI_m25 are detected throughout the coding region of the human housekeeping gene *g6pd* (*glucose-6-phosphate 1-dehydrogenase*) in a similar manner as the elongating RNA polII (Venkataraman *et al.,* 2005).

One of the early steps in the assembly of the 3'end processing is the recognition of the poly(A) signal and of the Downstream Sequence Element (DSE) by CPSF and CstF, respectively. Once transcribed by RNA polII, the several *cis*-acting elements that constitute the poly(A) site are recognized by the mRNA 3'end processing factors that rapidly bind them (Fig. 3).

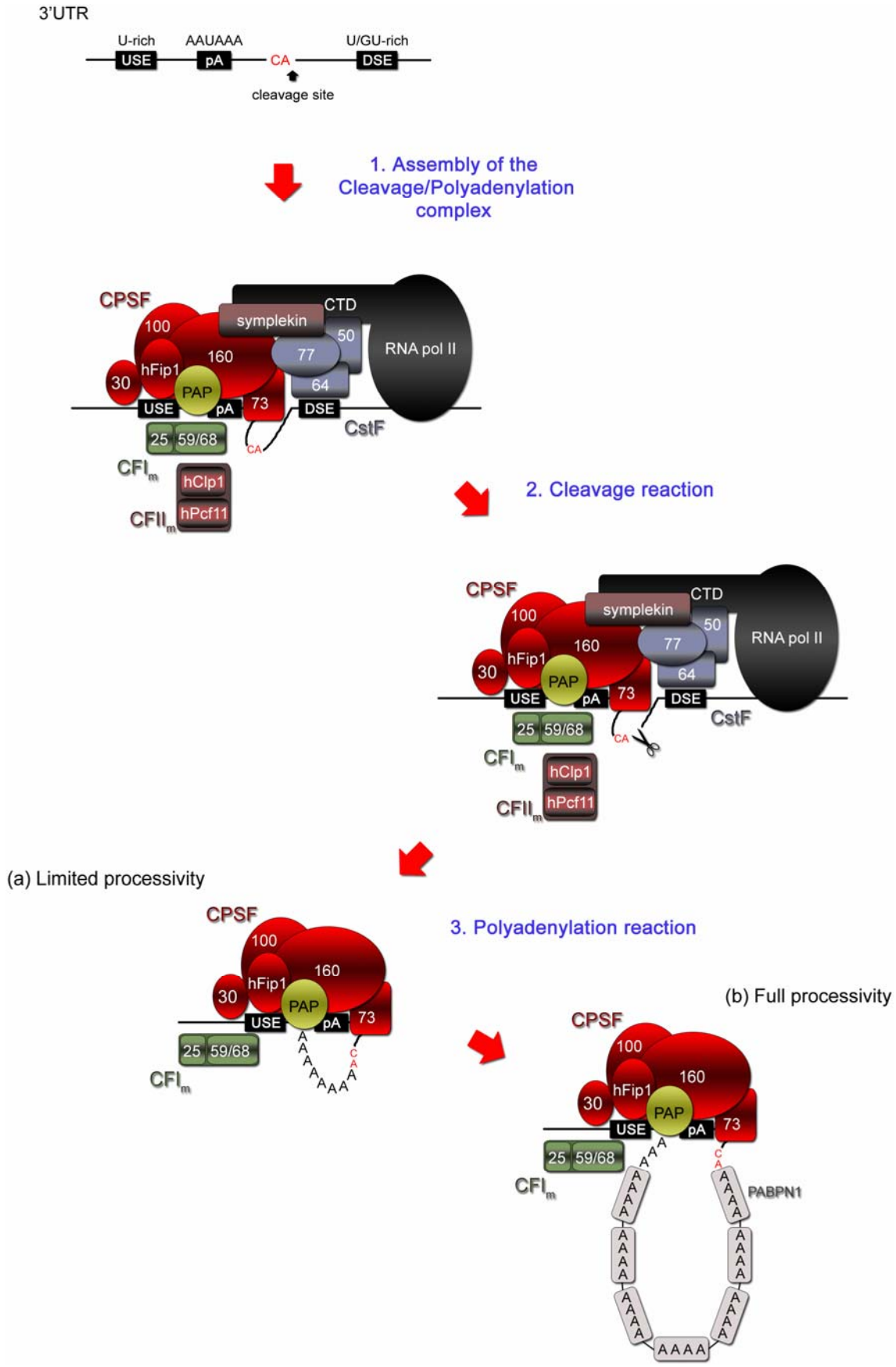

Figure 3. (see legend on next page).

Figure 3. Schematic representation of the mRNA 3'end processing complex. Upon recognition of the poly(A) site by the 3'end processing machinery (1), the pre-mRNA undergoes an endonucleotic cleavage at the cleavage site (2) followed by polymerization of the poly(A) tail to the upstream cleavage product (3). The polyadenylation reaction starts with a limited processivity by PAP (a). Upon the synthesis of a short poly(A) tail with 12 adenine nucleotides, PABPN1 binds to it and further stabilizes the CPSF-PAP complex, inducing full processivity (b). USE: Upstream Sequence Element; pA: poly(A) signal; DSE: Downstream Sequence Element; CA (red): dinucleotide that precedes the cleavage site (indicated by a black arrow). The proteins complexes are represented with different colours; Cleavage/Polyadenylation Specificity Factor (CPSF):red; Cleavage stimulation Factor (CstF): blue; Poly(A) Polymerase (PAP): yellow; Cleavage Factor I_m (CFI_m): green; Cleavage Factor II_m ($CFII_m$): purple; Symplekin: brown; RNA polymerase II (RNA polII): black; Poly(A) Binding Protein Nuclear 1 (PABPN1): grey. The scissor represents the endonucleotic cleavage. The size of the different proteins is not at scale. The scheme of the 3'end processing complex is simplified and does not intent to reflect the complete set of described interactions between the different complexes.

The binding of CPSF to the poly(A) signal is one of the first events and is probably stabilized through interaction with CFI_m, which also displays RNA-binding specificity (Ruegsegger *et al.,* 1996; Brown and Gilmartin 2003; Venkataraman *et al.,* 2005). The CPSF-RNA complex is further stabilized through cooperative interaction with CstF, via CPSF-160 and CstF-77 (Murthy and Manley 1992; Murthy and Manley 1995). Through this interaction the binding of CstF to the DSE is also stabilized. Although the recognition of the poly(A) signal occurs via CPSF-160, CPSF-30 and hFip1 also bind the pre-mRNA with hFip1 most likely being involved in the recognition of Upstream Sequence Elements (USEs) (Jenny *et al.,* 1994; Murthy and Manley 1995; Barabino *et al.,* 1997; Kaufmann *et al.,* 2004). The assembly of CPSF-CstF complex is likely to be further promoted and stabilized by Symplekin and the CTD of RNA polII (Hirose and Manley 1998; Takagaki and Manley 2000). In fact, the purification of Symplekin with CPSF and CstF raises the question whether or not Symplekin, CPSF and CstF may already be assembled prior to pre-mRNA recognition with binding to the RNA occurring in one single interaction instead of a sequential recognition of the different *cis*-acting elements (Takagaki and Manley 2000). The observation that Symplekin is brought to the promoter of *hsp70* through interaction with HSF1 supports this view (Xing *et al.,* 2004).

Once the CPSF-CstF-CFI_m-RNA complex is assembled, it is likely to recruit additional mRNA 3'end processing factors: the Poly(A) Polymerase (via interactions with CPSF-160 and hFip1) and $CFII_m$ (Murthy and Manley 1995; Kaufmann *et al.,* 2004). With exception of Symplekin, all mRNA 3'end processing factors are essential for the endonucleotic cleavage of the pre-mRNA (reviewed in Zhao *et al.,* 1999; Edmonds 2002). As soon as the whole mRNA 3'end processing complex is assembled, the pre-mRNA is cleaved at the cleavage site by the 73 kDa subunit of CPSF (Ryan *et al.,* 2004; Mandel *et al.,* 2006).

After the cleavage reaction, polymerization of the poly(A) tail takes place. Through interactions with CPSF, via CPSF-160, PAP polymerizes with limited processivity short poly(A) tails with 12 adenine nucleotides (Sheets and Wickens 1989; Wahle 1991a; Bienroth *et al.,* 1993; Murthy and Manley 1995). These short tails serve as binding site for PABPN1 that binds to it and promotes the stabilization of the CPSF-PAP complex (Bienroth *et al.,* 1993; Wahle *et al.,* 1993). Through interactions between CPSF, PAP and PABPN1, the polymerase undergoes polymerization of the poly(A) tail in a full processive mode, adding ~200 adenine nucleotides to the 3'end (Wahle 1991a; Bienroth *et al.,* 1993).

In the end, a fully processed mRNA is obtained, with a Cap structure on the 5'end, without the respective introns and with a 3'end poly(A) tail, which is then transported to the cytoplasm and translated into a protein (reviewed in Bruce Alberts 2002).

5. Alternative polyadenylation as a regulatory mechanism of gene expression

3'end formation is a key regulatory step in gene expression (reviewed in Edwalds-Gilbert *et al.,* 1997; Zhao *et al.,* 1999). The overall efficiency of 3'end processing and, therefore, the levels of mRNA produced, are directly dependent on the strength of the poly(A) site (Wickens and Stephenson 1984; Wilusz *et al.,* 1989; Sheets *et al.,* 1990; Chou *et al.,* 1994; Chen *et al.,* 1995). Pre-mRNAs with stronger poly(A) sites are processed more efficiently and therefore result in the production of higher mRNA levels. Moreover, since transcription termination requires a functional poly(A) site, the release of the RNA polymerase II from the DNA template and the reload of the promoter are directly dependent of an efficient mRNA 3'end processing (reviewed in Proudfoot *et al.,* 2002). Therefore, the presence of a poly(A) signal is a way through which gene expression can be regulated. Moreover, the presence of *cis*-regulatory elements and specific *trans*-acting factors provides a way to further regulate gene expression via up- or downregulation of mRNA 3'end processing of polyadenylation sites.

However, mRNA 3'end formation has an even more complex role in gene expression, as several genes have more than one polyadenylation site (reviewed in Edwalds-Gilbert *et al.,* 1997). The synthesis of different transcripts from a single transcription unit that may or may not code for different proteins and that may or may not have different *cis*-regulatory elements involved in processes such as mRNA transport, localization and translation, further reinforces the importance of this mechanism in gene expression regulation. The study of the *cis*- and *trans*-acting elements involved in mRNA 3'end formation, has allowed an increased understanding of how this process can regulate gene expression. During the course of the years, several genes that undergo

alternative polyadenylation have been studied (reviewed in Edwalds-Gilbert *et al.,* 1997). To better describe how alternative polyadenylation can regulate gene expression, a few examples are presented.

5.1. *Immunoglobulin M heavy chain (μ)*

Just as alternative splicing, alternative polyadenylation can also generate different protein isoforms when several poly(A) sites are positioned along the different exons of a gene. One of the most well studied examples of this type of alternative polyadenylation is the *immunoglobulin M heavy chain* (*μ*) gene (reviewed in Edwalds-Gilbert *et al.,* 1997; Edmonds 2002). Like all five types of immunoglobulins, there are two IgM isoforms: a membrane and a secreted-form. This is due to the synthesis of a μ_m and a μ_s mRNA, coding for the membrane and secreted-forms of the heavy chain, respectively (Fig. 4). The μ_m mRNA results from the use of a distal poly(A) site (μ_m pA) and includes the M1 and M2 exons and the 3'end of the last constant region exon, Cμ4. This transcript includes the C-terminal sequences that allows the IgM membrane-form to be inserted into the cytoplasmic membrane. The use of a proximal poly(A) site present in the Cμ4 exon (μ_s pA) and the corresponding synthesis of the μ_s mRNA leads to the exclusion of these C-terminal sequences resulting in the synthesis of the secreted-form of the IgM.

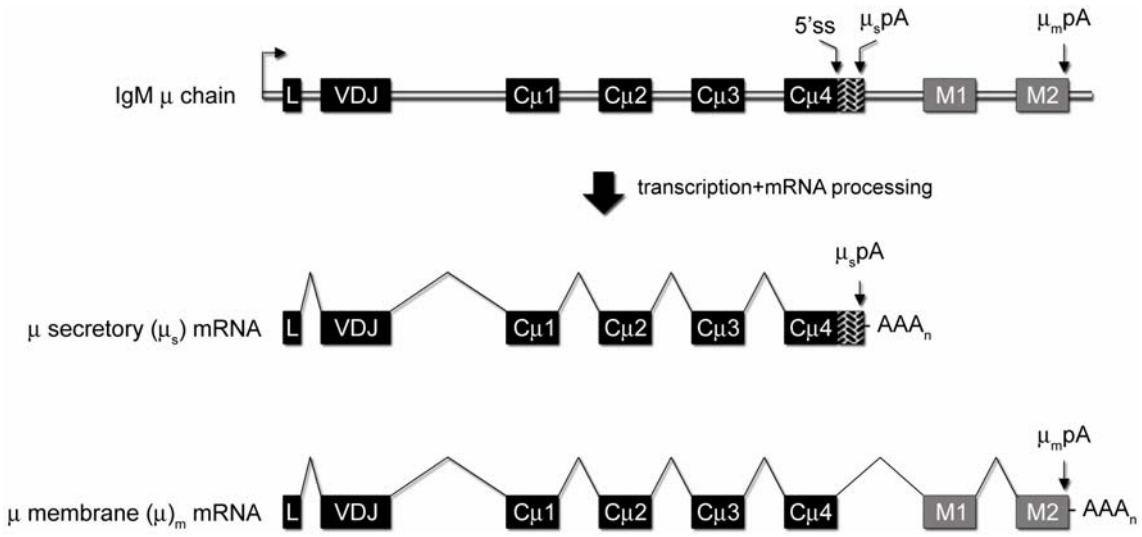

Figure 4. mRNA processing of the *immunoglobulin M heavy chain (μ)* pre-mRNA. The synthesis of the μ secretory (μ_s) and μ membrane (μ_m) mRNAs results from a competition between splicing and 3'end formation. When the use of μ_s poly(A) site at the Cμ4 exon is favoured, the μ_s mRNA is synthesized resulting in the production of the IgM secreted-form. When splicing between the Cμ4-M1 exons is favoured the synthesis of the μ_m mRNA occurs with inclusion of the M1 and M2 exons, allowing insertion of the IgM in the cytoplasmic membrane (IgM membrane-form). The

grey boxes represent the M1 and M2 exons while the hatched box represents the 3'end of the Cμ4 exon present in the μ_s mRNA and excluded from the μ_m mRNA. The black boxes represent all remaining exons. The grey bar between the different exons represents the introns. 5'ss: 5'splice site; μ_spA: μ_s poly(A) signal; μ_mpA: μ_m poly(A) signal. The vertical arrow at the beginning of the IgM μ chain represents the promoter.

The production of the membrane and the secreted-forms of IgM is regulated during B cell development (reviewed in Edwalds-Gilbert *et al.,* 1997; Edmonds 2002). In resting B cells, approximately equal amounts of membrane and secreted-forms of the IgM are produced. However, when resting B cells are stimulated by the presence of an antigen, they proliferate and differentiate into plasma cells. During differentiation, there is a change in gene expression with the synthesis of the μ_s mRNA being highly favoured resulting in the production of high levels of the secreted-form of IgM. A central feature in this regulation is a competition between the usage of the 5' splice site (5'ss) and the μ_s poly(A) site in the Cμ4 exon (Galli *et al.,* 1987; Peterson and Perry 1989). Whereas the use of the 5'ss results in the synthesis of the μ_m mRNA with the use of the distal μ_m poly(A) site, the use of the proximal μ_s poly(A) site leads to the synthesis of the μ_s mRNA. Therefore, these two reactions, splicing of the 5'ss and the use of the μ_s poly(A) site, are mutually exclusive.

Contrary to what is observed in other developmentally regulated genes, the regulation of the IgM does not require any IgM-specific *cis*-acting elements (Galli *et al.,* 1987; Peterson and Perry 1989; Peterson 1994). Instead, it relies on a delicate equilibrium of strength between the 5'ss, the μ_s and the μ_m poly(A) sites (Peterson and Perry 1989). The 5'ss in the Cμ4 exon, is a sub-optimal 5'ss. Point mutations introduced to restore it to a optimal 5'ss, disrupt the μ_s/μ_m regulation with the μ_m mRNA being favoured in plasma cells. Likewise, the strength of each μ poly(A) site is also important in this regulation (Galli *et al.,* 1987; Peterson and Perry 1989). The μ_s poly(A) site is weaker than the μ_m poly(A) site (Galli *et al.,* 1987). This difference in strength is required for the expression of μ_s and μ_m mRNAs. Swapping both poly(A) sites (Galli *et al.,* 1987), or substituting the μ_s poly(A) site for a stronger one (Peterson and Perry 1989) changes the μ_s/μ_m mRNA ratio between B cells and plasma cells, with the synthesis of the μ_s mRNA being favoured. Therefore, the developmental regulation of the IgM heavy chain relies on a specific organization of RNA processing signals. In fact, heterologous RNAs containing this specific arrangement are able to be processed in the same manner as the μ mRNAs in B cells and plasma cells (Peterson 1994).

The non-requirement of *igM μ*-specific *cis*-acting elements implies that there are no-specific *trans*-acting factors involved. All the mentioned evidences support the

hypothesis that regulation of the IgM μ mRNA is achieved through modulation of the levels of basal mRNA processing factors. Indeed it was initially shown that stimulation of mouse resting B cells was accompanied by increased levels of CstF-64 (Takagaki *et al.,* 1996). Overexpression of CstF-64 in the DT40 chicken B cell line induces approximately an 8-fold increase in the μ_s/μ_m ratio, similar to the 7-fold increase observed for mouse resting B cells stimulated with LPS (Takagaki *et al.,* 1996). These results are further supported by the analysis of the effects of decreased levels of CstF-64 (Takagaki and Manley 1998). Downregulation of CstF-64 specifically decreases the μ_s/μ_m ratio, showing that these effects are not due to a general decrease in 3'end processing. Moreover, both CstF-64 and the purified CstF complex bind μ_m poly(A) site with a higher affinity compared to the μ_s poly(A) site which correlates with a higher cleavage efficiency for the μ_m poly(A) site (Takagaki *et al.,* 1996). The observation that CstF-64 is a limiting factor for CstF complex formation during B cell development, further supports that modulation of the levels of CstF-64 are an important aspect in the μ mRNA regulation (Takagaki *et al.,* 1996). Therefore, based on these evidences, Takagaki *et al.* 1996 proposed the following model (Fig. 5): in mature resting B cells, the low levels of CstF-64 limit the amount of active CstF complex available. Under these conditions and due to a higher affinity of CstF towards the μ_m poly(A) site, CstF binds to it, despite the fact that the μ_s poly(A) site is transcribed first and, therefore, the first to be presented to the 3'end processing machinery. As a result, approximately equal amounts of μ_s and μ_m mRNAs are produced. However, upon stimulation with an antigen, B cells proliferate and differentiate concomitantly with a 10-20-fold increase in the CstF-64 levels. Under these circumstances, there is an increased usage of the μ_s poly(A) site that out-competes the Cμ4-M1 splicing event, the μ_m 3'end formation or both. This model is supported by a recent study using microarray analysis of RAW macrophages, showing that the levels of CstF-64 can influence gene expression and induce alternative poly(A) site selection (Shell *et al.,* 2005).

Despite the fact that these evidences support a model in which CstF-64 levels are the principal mechanism for the μ_s/μ_m mRNA regulation, other studies suggest a more complex mechanism (Martincic *et al.,* 1998). The switch from the membrane to the secreted-form of IgM occurs with resting B cells differentiation into plasma cells. Exposure of B cells to the CD40 ligand promotes proliferation. However, only when exposed to CD40 ligand and Interleukin-10 (IL-10) together, do B cells undergo proliferation and differentiation with the increased production of the IgM secreted-form (Martincic *et al.,* 1998).

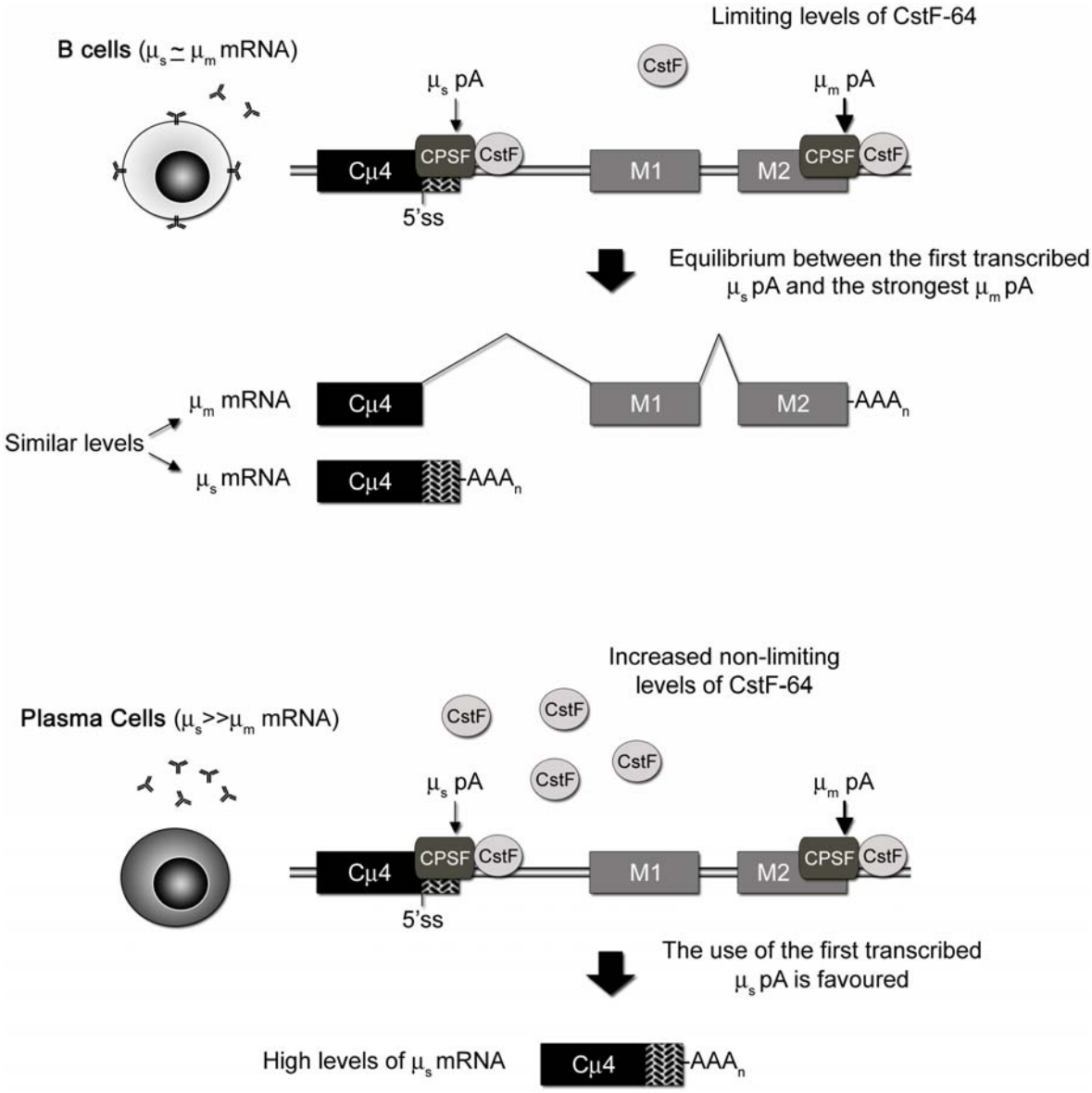

Figure 5. Model for mRNA processing of the *immunoglobulin heavy chain* (*μ*) pre-mRNA during B-lymphocyte development. In resting B cells, CstF-64 is present in limiting amounts resulting in low levels of active CstF. Under these conditions the stronger μ_m poly(A) site is preferentially used resulting in equilibrium between Cμ4-M1 splicing and 3'end formation at the μ_s poly(A) site with the synthesis of similar amounts of μ_s and μ_m mRNA. Upon activation, B cells differentiate into plasma cells and there is an increase in the levels of CstF-64. CstF-64 is no longer present in limiting amounts. The increased levels of active CstF favours the use of the first-transcribed μ_s poly(A) site instead of the μ_m poly(A) site. As a result plasma cells express higher levels of μ_s mRNA. The thickness of the poly(A) site arrows are proportional to the poly(A) site strength: μ_m pA is stronger than μ_s pA. The grey boxes represent the M1 and M2 exons while the hatched box represents the 3'end of Cμ4 present in the μ_s mRNA and excluded form the μ_m mRNA. The black boxes represent all remaining exons. The grey bar between the different exons represents the introns. 5'ss: 5'splice site; μ_spA: μ_s poly(A) site; μ_mpA: μ_m poly(A) site.

Analysis of CstF-64 expression shows that stimulation of B cells with CD40 ligand alone is sufficient to promote a 5-fold increase in CstF-64 but insufficient to promote the IgM switch (Martincic *et al.,* 1998). Exposure to both CD40 ligand and IL-10 however, induces the production of the IgM secreted-form although the increase in CstF-64 levels is the same 5-fold. Similar results were obtained with the SK 6.4 cell line (Martincic *et al.,* 1998). Upon addition of IL-6, over 90% of these cells switch to the IgM secreted-form. However, the levels of CstF-64 remained unchanged after addition of IL-6.

Moreover, although the initial studies suggested that the switch in production from the membrane to secreted-form of IgM is due to an increase in the general cleavage-polyadenylation activity in the plasma cells that enhanced the μ_s poly(A) site recognition (Galli *et al.,* 1987; Galli *et al.,* 1988; Peterson and Perry 1989; Peterson *et al.,* 1991), observation that overexpression of splicing factors can influence the μ_s/μ_m regulation in B cells and plasma cells and that in fact, the splicing environment changes during B-lymphocyte development, has challenged this view (Bruce *et al.,* 2003).

Therefore, the results suggest that CstF-64 may play a role in the developmental regulation of μ mRNA but does not exclude that other factors may also be determinant for this regulation. The observation of an activity in B-cell extracts that specifically promotes instability of the CPSF-CstF-μ_s pA complex indicates a more complex mechanism in μ mRNA regulation with the involvement of other factors (Yan *et al.,* 1995). More recently, U1A and hnRNP F were also suggested to be *trans*-acting factors involved in μ mRNA regulation during B-lymphocyte development (Phillips *et al.,* 2001; Veraldi *et al.,* 2001; Phillips *et al.,* 2004; Ma *et al.,* 2006; Peterson *et al.,* 2006). However, their precise function in this regulation remains unclear.

5.2. *suppressor of forked (su(f))*

The *suppressor of forked*, *su(f)*, provides a good example of how alternative polyadenylation can be used to regulate gene expression in a developmental and tissue-specific manner.

suppressor of forked is the *Drosophila* homologue of the human CstF-77 (Mitchelson *et al.,* 1993; Takagaki and Manley 1994). Although it codes for a 84 kDa protein, the *su(f)* transcription unit produces three mRNAs: a 2.9, 2.6 and 1.3 kb mRNAs (Mitchelson *et al.,* 1993) (Fig. 6). While the 2.6 and 2.9 kb transcripts result from the use of two polyadenylation sites present in the 3'most exon (Fig. 6, pA1 and pA2 respectively) and therefore have the same ORF that codes for the 84 kDa protein, the use of a poly(A) site present in intron 4 (Fig. 6, see pA3) results in the synthesis of the 1.3 kb mRNA. This

transcript has an ORF that lacks a stop codon and genetic studies showed that it is not required for Su(f) activity (Williams and O'Hare 1996). It is likely that this transcript is degraded by a cellular surveillance mechanism such as non-stop mRNA decay (Frischmeyer *et al.,* 2002).

Analysis of temperature sensitive mutants revealed that Su(f) is required for cell proliferation (Audibert and Simonelig 1999). Consisting with this, Su(f) has a tissue-specific expression pattern (Audibert *et al.,* 1998; Audibert and Simonelig 1999). It is highly expressed in mitotic active cells, whereas in differentiated cells, Su(f) is present in very low levels. Interestingly, the use of temperature sensitive mutants at the restrictive temperature showed that Su(f) activity is necessary for its tissue-specific expression (Juge *et al.,* 2000). In the absence of Su(f) activity, the protein is present in mitotic and pos-mitotic tissues.

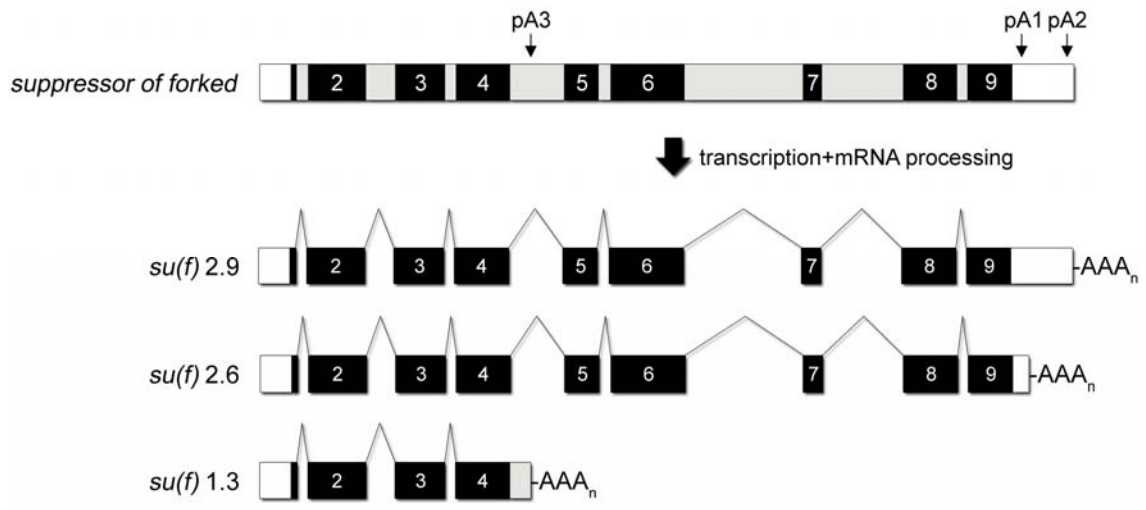

Figure 6. Schematic representation of the *suppressor of forked* (*su(f)*) pre-mRNA processing. The *su(f)* gene produces three mRNAs due to the presence of three polyadenylation sites. The use of pA1 and pA2 present in the 3'most exon leads to the synthesis of a 2.6 (*su(f)* 2.6) and a 2.9 kb (*su(f)* 2.9) transcripts, respectively, both coding for the 84 kDa Su(f) protein. The third transcript of 1.3 kb long is dispensable for Su(f) function and results from the use of pA3 present in intron 4. The black boxes represent the exons with the white boxes representing the 5'UTR and 3'UTR of exons 1 and 9, respectively. The grey boxes represent the introns. The (-AAA$_n$) represents the poly(A) tail.

Therefore, Su(f) seems to be involved in a autoregulatory mechanism necessary for its downregulation in differentiated tissues. The analysis of *su(f)* mRNA expression in these mutants further supports this idea: the synthesis of the 1.3 kb mRNA is dependent upon Su(f) activity (Audibert and Simonelig 1998). Since Su(f) is the *Drosophila* homologue of the human CstF-77 (Mitchelson *et al.,* 1993; Takagaki and Manley 1994), it was suggested that the Su(f), as part of the *Drosophila* CstF complex, may regulate its

own levels in non-proliferating tissues by favouring the use of the intronic polyadenylation site (Juge *et al.,* 2000). However, this mechanism is not yet fully understood, in particular how this autoregulation is tissue specific. It has been suggested that the expression of an yet unidentified protein, specifically expressed in differentiated tissues, may interact with Su(f) and enhance the recognition of the intronic non-canonical poly(A) signal (Juge *et al.,* 2000) (Fig. 7). Alternatively, it is also possible the specific expression in proliferative tissues of a specific repressor that inhibits the recognition and use of the intronic poly(A) site, therefore resulting in the accumulation of Su(f).

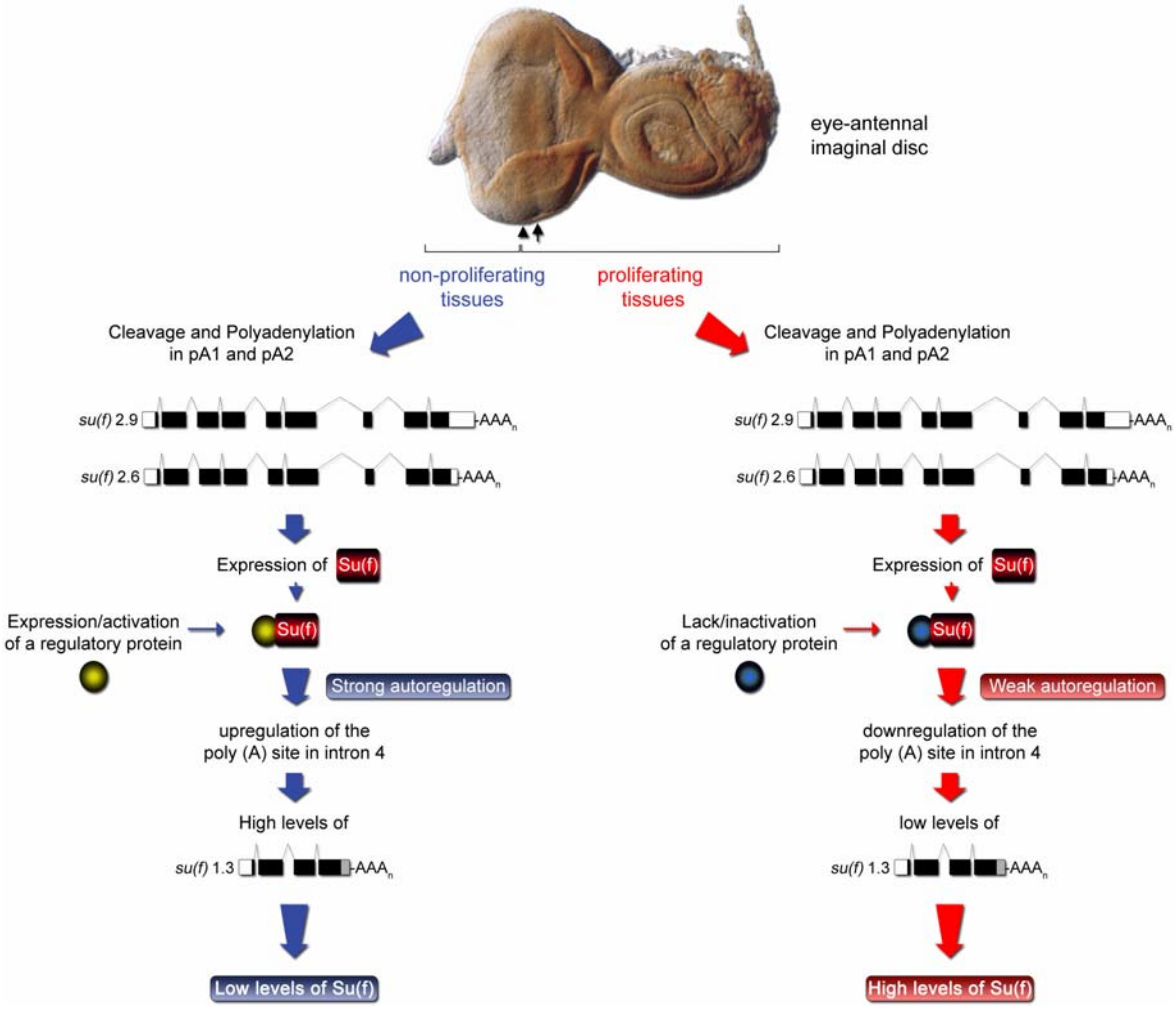

Figure 7. Model for the tissue-specific *su(f)* autoregulation. In non-proliferating tissues, Su(f) recognizes the intronic pA site promoting the synthesis of the 1.3 kb *su(f)* mRNA. As a results, non-proliferating tissues accumulate low levels of Su(f). The autoregulation of *su(f)* in non-proliferating tissues is dependent on the expression or activation of a specific regulatory protein that allows the recognition of the intronic poly(A) site by Su(f). In proliferating tissues, the absence or inactivation of this regulatory protein, does not allow the use of the intronic poly(A) site leading to synthesis of the 2.6 and 2.9 kb mRNAs and accumulation of Su(f). The third-instar larvae eye-antennal imaginal disc is positioned from posterior (left) to anterior (right). The accumulation of Su(f) (brown) in the eye-antennal imaginal disc occurs in the proliferating tissues. The morphogenetic furrow (black

arrow) and the second mitotic wave (black arrowhead) signal the advancing differentiation wave that starts from posterior towards anterior. The black boxes represent the exons present in the 2.9, 2.6 and 1.3 kb *su(f)* mRNAs. The white boxes represent the 5'UTR and 3'UTR of exons 1 and 9, respectively. The grey box in the 1.3 kb *su(f)* mRNA represents part of intron 4 present in the transcript. The (-AAA_n) represents the poly(A) tail. Adapted from Audibert *et al.* (1998) and Juge *et al.* (2000).

Why do mitotic cells accumulate *suppressor of forked* is not fully understood. Since Su(f) accumulates in mitotic tissues and *su(f)* temperature sensitive mutants show a block or delay in the metaphase-anaphase transition, it has been suggested that while the low levels of Su(f) are sufficient for 3'end processing of the majority of the mRNAs, the 3'end processing of mRNAs involved in cell cycle progression may require higher levels of Su(f) (Audibert and Simonelig 1999). The presence of weaker polyadenylation sites or competition between alternative polyadenylation sites in these genes, are factors that could explain the requirement for Su(f) accumulation in proliferative tissues (Audibert and Simonelig 1999).

Whatever the precise mechanism through which this autoregulation occurs, it is highly conserved between *D. melanogaster* and *D. virilis* (Audibert and Simonelig 1998). The Su(f) protein sequences are 96.3% identical and 97.7% similar throughout the entire length (Audibert and Simonelig 1998). Although analysis of the introns sequences shows that the introns in both species have diverged completely, an exception is made for intron 4 with a 137 nt region presenting a 75.8% of identity (Audibert and Simonelig 1998). In fact, in *D. virilis* two *su(f)* mRNAs are synthesized: a 2.6 and a 1.3 kb mRNA (Audibert and Simonelig 1998). While the longer transcript results from the use of a polyadenylation site present in the 3'most exon, the 1.3 kb mRNA results from the use of a polyadenylation site present in the intron 4. Therefore, there is a high similarity between the *su(f)* transcription units in both *Drosophila* species which suggest that the same autoregulatory mechanism may occur in *D. virilis* and reinforces its importance (Audibert and Simonelig 1998).

Interestingly, a recent study reported that the human *su(f)* homologue, the CstF-77 transcription unit, also produces three mRNAs (Pan *et al.,* 2006). Two of these transcripts differ only in the 3'UTR as a result of two polyadenylation sites present in the 3'most exon and therefore code for the full length CstF-77. The smaller mRNA is synthesized as a result of the use of a polyadenylation site present in intron 3 and it has a stop codon right before the polyadenylation site. As a result, the smaller transcript codes for a 103 amino acid protein. It is likely that this protein is not functional as it lacks the C-terminal domain necessary for interaction with CstF-64, CstF-50 and itself (Takagaki and Manley 2000; Pan *et al.,* 2006). Analysis of the genomic sequence surrounding the polyadenylation site

in intron 3 shows that it has no similarity with the intronic sequences of *su(f)* (Pan *et al.,* 2006). However, it is conserved in several other vertebrate organisms such as mouse. Moreover, analysis of the levels of these transcripts in mouse B cells and in plasma cells lines show that the levels of the transcript resulting from the use of the intronic polyadenylation site is 5-fold higher in mouse B cells compared to the mitotically active plasma cells (Pan *et al.,* 2006). This result is in agreement with the observations of accumulation of Su(f) in proliferative cells in *Drosophila melanogaster* (Audibert *et al.,* 1998; Audibert and Simonelig 1999).

Therefore, although there is no sequence similarity between *su(f)* and hCstF-77 there seems to be a strong conservation of the autoregulatory mechanism that operates in their expression.

5.3*. enhancer of rudimentary (e(r))*

A bioinformatic analysis of the human and mouse genomes has revealed that alternative polyadenylation of the 3'most exon is the most frequent type of alternative polyadenylation (Tian *et al.,* 2005). A considerable number of genes in these genomes, 32% and 24% of the human and mouse genes, respectively, undergo alternative polyadenylation of the 3'most exon.

The presence of more than one polyadenylation site in the 3'most exon results in the synthesis of several mRNAs that contain the same open reading frame but differ in the length of the 3' UTR. The presence in the different 3'UTRs of *cis*-acting elements that can differently modulate mRNA localization, stability or translation, in a developmental and/or tissue/cell-specific manner, have a direct role in gene expression regulation (reviewed in Edwalds-Gilbert *et al.,* 1997).

The *Drosophila enhancer of rudimentary* is an example of this type of genes. *e(r) codes for* a potential translational repressor (Pogge von Strandmann *et al.,* 2001). It has two polyadenylation signals present in the 3'most exon: a proximal non-canonical poly(A) signal, TATAAA, and the distal canonical AATAAA (Wojcik *et al.,* 1994). Both polyadenylation signals are used. However, the use of each polyadenylation signal is sex-specific (Wojcik *et al.,* 1994; Gawande *et al.,* 2006). While the smaller mRNA resulting from the use of the proximal poly(A) signal is expressed in both male and female adults, the longer mRNA is specifically expressed in females. Analysis of the expression of the female-specific *e(r)* (*e(r)*-fs) mRNA in both wild-type and in *tudor* mutants that lack a germline, shows that the *e(r)*-fs mRNA is primarily expressed in the female germline in the later stages of oogenesis (Gawande *et al.,* 2006).

The expression of the *e(r)*-fs mRNA is regulated by three GU-rich elements present downstream the proximal poly(A) signal, similar to DSEs (Gawande *et al.,* 2006). Deletion of these elements leads to expression of *e(r)*-fs mRNA in both males and females. RNA-binding assays showed that Sex-lethal (Sxl) binds with a high affinity to the GU-rich elements downstream of the proximal poly(A) signal (Gawande *et al.,* 2006). Moreover, it was shown to compete with CstF-64 for the binding to these elements (Gawande *et al.,* 2006). Sxl is specifically expressed in females and therefore, its expression correlates with the synthesis of *e(r)*-fs mRNA (reviewed in Penalva and Sanchez 2003). In fact, analysis of mutants in which Sxl function is impaired in the female germline shows that this protein is necessary for the expression of the *e(r)*-fs mRNA (Gawande *et al.,* 2006).

Therefore, a model has been proposed for the expression of *e(r)*-fs mRNA (Fig. 8) (Gawande *et al.,* 2006). In the female germline, Sxl binds the GU-rich elements downstream the proximal poly(A) signal in competition with CstF. As a result the binding of CstF to these elements is blocked, which prevents the use of the proximal polyadenylation site. Thus, the use of the distal poly(A) site and the synthesis of the *e(r)*-fs mRNA is allowed. Since males do not express Sxl, the binding of CstF to the GU-rich elements is not restricted and therefore the smaller non-sex specific *e(r)* mRNA (*e(r)-nss)* is the only transcript synthesized.

However, it is still to be explained why the *e(r)*-fs mRNA is barely detected in the female soma, where Sxl is also expressed (Gawande *et al.,* 2006). It was suggested that variation in the levels of Sxl during oogenesis or even different female germline-specific isoforms may play a role in the expression of the *e(r)*-fs mRNA (Gawande *et al.,* 2006). It is also possible that an additional germline-specific factor may be required to determine germline specificity (Gawande *et al.,* 2006).

Analysis of the expression of the *e(r)*-fs mRNA in isolated ovaries, shows that this transcript is less efficiently translated, suggesting that the 3'UTR sequence present in this transcript is necessary to repress translation of the *e(r)*-fs mRNA in the female germline (Gawande *et al.,* 2006). Therefore, alternative polyadenylation of the *e(r)* presents a mechanism through which regulation of translation can be achieved in a tissue-specific manner.

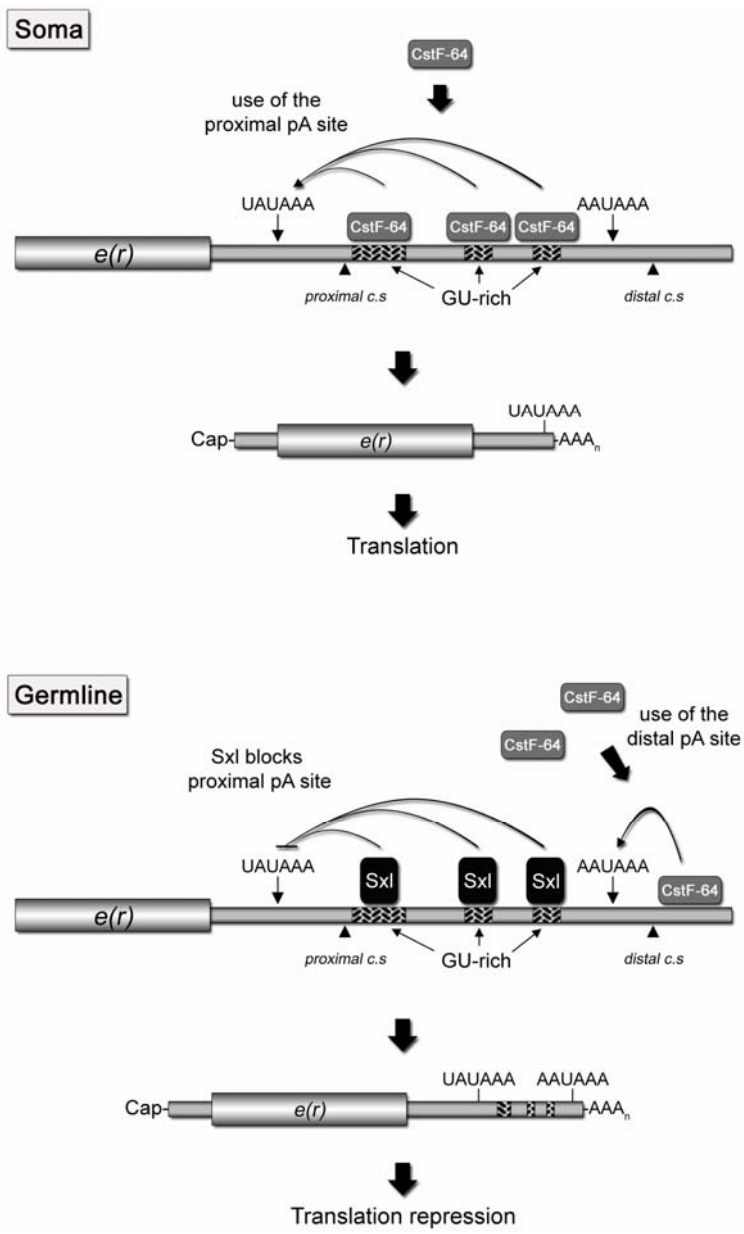

Figure 8. Model for alternative polyadenylation of *enhancer of rudimentary* in the *Drosophila melanogaster* female germline. In somatic tissues, CstF-64 binds preferentially to the GU-rich elements (hatched black boxes) downstream the proximal poly(A) signal promoting the assembly of the 3'end processing complex and use of the proximal poly(A) site leading to the synthesis of the *e(r)*-nss mRNA. In the germline, the levels of E(r) are regulated by alternative polyadenylation: Sex-lethal (Sxl) competes with CstF-64 for the binding to the GU-rich elements, preventing the recognition and use of the proximal poly(A) site. As a result, CstF-64 promotes the activation of the distal poly(A) site leading to the synthesis of the *e(r)*-fs mRNA. The *e(r)*-nss mRNA is translated while translation of the *e(r)*-fs mRNA is repressed, resulting in an overall decreased in the levels of E(r). The *c.s* represents the cleavage site (black arrowheads) and the (AAA_n) the poly(A) tail.

6. The *polo* gene

The *polo* gene codes for an important cell cycle kinase and is the founding member of the conserved Polo-like kinases (Plks) family which are important regulators with multiple key roles throughout the cell cycle (reviewed in Glover 2005; van de Weerdt and Medema 2006). Polo was originally discovered by identification of the mutant alleles *polo*^{1} in a genetic screen designed to select sterile females, and *polo*^{2} identified from a genetic screen of P-element-induced lethal mutants (Sunkel and Glover 1988).

Plks have been intensively studied during the two decades that followed the identification of *polo*. The Polo kinase plays multiple functions and it is involved in centrosome maturation, in metaphase-anaphase transition, cytokinesis and recently in the regulation of polarized transport in germline cells (reviewed in Glover 2005; Mirouse *et al.,* 2006). Although a great amount of data has been reported concerning the function of Polo, very little is known regarding the regulation of *polo* expression. *polo* transcription unit produces two mRNAs with 2.2 and 2.5 kb respectively, which result from the usage of two polyadenylation sites present in the 3'most exon (Llamazares *et al.,* 1991). Therefore, the two transcripts have the same open reading frame and code for the same protein. Both mRNAs are expressed throughout development with Polo accumulating in mitotic tissues (Llamazares *et al.,* 1991). So far the role of each transcript in the Polo function during *Drosophila* development remains undetermined. Early studies have shown that regulation of the mouse Plk during terminal erythrodifferentiation most likely involves polyadenylation and deadenylation of the poly(A) tail (Lake and Jelinek 1993). These results together with the observation that *polo* undergoes alternative polyadenylation suggest that Plks might be regulated through regulation of the mRNA 3'end formation.

7. Aims of the thesis

In this thesis, we aimed to understand the role of alternative polyadenylation of the *polo* 3'most exon in the regulation of Polo function during development. For this purpose two approaches were followed:

I. To analyse the physiological role and requirement of each *polo* mRNA in the Polo function during development.
II. Identification and characterization of *cis*-acting elements and the corresponding *trans*-acting factors involved in *polo* poly(A) site selection.

PART II

EXPERIMENTAL WORK

Chapter I

The longer 2.5 kb polo mRNA is required for abdominal histoblasts proliferation during metamorphosis

1. Introduction

It is widely accepted that modulation of gene expression at the mRNA level, can be achieved through alternative splicing, which allows expression of different protein isoforms (Black 2003; Ben-Dov *et al.,* 2007). Recent reports have shown that at least half of the human genes undergo alternative polyadenylation, with polyadenylation sites found in introns, internal exons and in the 3'most exons (Tian *et al.,* 2005; Yan and Marr 2005). This clearly suggests that the mechanisms and factors involved in regulation of alternative polyadenylation are essential in the modulation of gene expression. Polyadenylation sites present in internal exons are usually associated with alternative splicing and consequently the synthesis of different protein isoforms (reviewed in Edwalds-Gilbert *et al.,* 1997). A well studied example of this type of alternative polyadenylation is the synthesis of the membrane-associated and secreted forms of the IgM heavy chain in B cells and plasma cells (reviewed in Edwalds-Gilbert *et al.,* 1997; Edmonds 2002). The presence of two poly(A) sites with different strengths located in different exons is essential for the immune response of B cells in the presence of antigens (Galli *et al.,* 1987; Peterson and Perry 1989; Takagaki *et al.,* 1996). In resting B cells the stronger distal poly(A) site, positioned downstream the coding sequences for the transmembrane and cytoplasmic domains of the IgM heavy chain is preferentially used and as result a membrane-associated IgM isoform is synthesized. However, when activated by the presence of antigens, B cells undergo a change in gene expression and differentiate into plasma cells. The levels of CstF-64 expressed in limiting amounts in resting B cells are increased and as a consequence the recognition of the proximal poly(A) site is favoured (Takagaki *et al.,* 1996). As a result, in plasma cells, the proximal poly(A) site positioned upstream the exons coding for the transmembrane and cytoplasmic domains is now recognized and the synthesis of an mRNA lacking both domains is favoured resulting in the synthesis of the secretory isoform of the IgM heavy chain.

The presence of alternative polyadenylation sites in the 3'most exon direct the synthesis of mRNAs that share the same coding sequence but have different 3'UTRs. This is usually coupled with the presence of regulatory elements in the 3'UTR able to direct regulation of mRNA stability, translation or mRNA localization in a tissue or/and developmental-specific manner (reviewed in Edwalds-Gilbert *et al.,* 1997). Despite several evidences showing that alternative polyadenylation at the 3'most exon may play a role in gene expression (reviewed in Edwalds-Gilbert *et al.,* 1997), the role of this mechanism during development has been sparsely addressed. The recent study of the *Drosophila* gene *enhancer of rudimentary* (*e(r)*), has shown how this process may regulate gene expression during development (Gawande *et al.,* 2006). Two *e(r)* mRNAs are synthesized

due to two poly(A) signals present in the 3'most exon (Wojcik *et al.,* 1994; Gawande *et al.,* 2006). However, the longer *e(r)* mRNA is specifically expressed in females, with the bulk of expression occurring in the germline (Wojcik *et al.,* 1994; Gawande *et al.,* 2006). It was shown that the expression in the female germline of the sex-determining gene, *sex-lethal* (*sxl*), is essential for the synthesis of the longer *e(r)* transcript (Gawande *et al.,* 2006). This study suggests that in female somatic cells, the *Cleavage and stimulation Factor* (CstF) complex binds to GU-rich elements downstream the proximal poly(A) site and promotes 3'end processing and, consequently, the synthesis of the smaller *e(r)* transcript. However, in the female germline, Sxl competes with CstF for the GU-rich binding sites (Gawande *et al.,* 2006). As a result, CstF is now free to recognize the distal poly(A) site. This competition results in the downregulation of the usage of the proximal poly(A) site in favour of the usage of the distal one and consequently in the synthesis of an mRNA with a longer 3'UTR. As a result of the specific synthesis of the longer *e(r)* mRNA in the female germline, a downregulation of E(r) levels is observed in this tissue (Gawande *et al.,* 2006).

The *Drosophila polo*, shows some similarities with *enhancer of rudimentary*. Polo is a cell cycle kinase, required, among other functions, for the entry and exit of mitosis (reviewed in Glover 2005). Despite the knowledge about the function of this protein during cell cycle progression little is known about the mechanisms that regulate *polo* expression. *polo* contains two poly(A) signals that are used to produce two mRNAs with 2.2 and a 2.5 kb, respectively (Fig. 9A). The cleavage sites of the two pre-mRNAs were previously mapped by sequencing two cDNAs isolated from a poly(A)$^+$ 0-24h embryonic mRNA library (Llamazares *et al.,* 1991). Subsequently, these two transcripts were shown to be present mainly in cells that undergo proliferation (Llamazares *et al.,* 1991). The proximal poly(A) signal, that here we describe as pA1, is the fly variant AUUAAA, and the distal poly(A) signal, pA2, the variant AAUAUA (Fig. 9A). These two poly(A) signal variants are present in 10,3% and 5.1% of the *Drosophila* genes, respectively, representing the second and third most common variants of the canonical poly(A) signal in *Drosophila melanogaster* (Retelska *et al.,* 2006). Although it has been shown that both signals are used throughout development (Llamazares *et al.,* 1991), the role of these mRNAs in the function of Polo during development has not been addressed.

2. Results

2.1. Generating transgenic flies expressing only one *polo* mRNA

To address the role of alternative polyadenylation in the *polo* gene, we studied the function of each *polo* mRNA during development. To that end, we generated transgenic flies expressing only one of the two *polo* mRNAs. Thus, two transgenes were constructed using as a template the previously described transgene *gfp-polo* (Moutinho-Santos *et al.,* 1999). This transgene was constructed using a 7 kb genomic fragment that contains the *polo* gene and its promoter and was shown to be sufficient for complementation of $polo^1$ *and* $polo^2$ mutant alleles (Llamazares *et al.,* 1991). By PCR mutagenesis, a GFP coding sequence was fused with the *polo* initiation codon. This fragment was then subcloned into a pW8 plasmid and the resulting construct used for P-element-mediated germ line transformation (Fig. 9B). The resulting flies were viable and the transgene was sufficient to fully rescue the $polo^1$*/Df (3L) rdgC-co2* and $polo^9/polo^9$ allelic combinations (Moutinho-Santos *et al.,* 1999; Pearson *et al.,* 2006). Furthermore, the GFP-Polo protein was also shown to retain the kinase activity of Polo (Moutinho-Santos *et al.,* 1999).

Therefore, for this reason, this transgene was used as a template to construct the *polo* transgenes used in this study. By PCR mutagenesis, point mutations were introduced in the proximal poly(A) signal (ATTAAA to **G**TTAA**C**), thus generating the transgene *gfp-polo∆pA1*, as to allow the expression of only the longer *polo* transcript (Fig. 9A, B). Also by PCR mutagenesis, a second transgene, the *gfp-polo∆pA2*, was generated by deletion of the distal poly(A) signal and the entire downstream region. Therefore, flies carrying this transgene will only express the shorter *polo* mRNA (Fig. 9A, B).

In order to generate flies that express only one *polo* mRNA, we used the $polo^9$ allele. This allele was originally isolated in a search for strong hypomorphic alleles of *polo* while screening a collection of mutants generated by P-element mediated mutagenesis (Donaldson *et al.,* 2001). Homozygous individuals for $polo^9$ die at late third instar larval stage with larvae brains presenting a high mitotic index as a result of an arrest in prometaphase/metaphase stage of the cell cycle. Western blot analysis of third instar $polo^9$ brains shows extremely reduced levels of Polo. These low levels are most likely due to the insertion of the P-element downstream the second transcription initiation start site, thus affecting the synthesis of both *polo* mRNAs (Donaldson *et al.,* 2001).

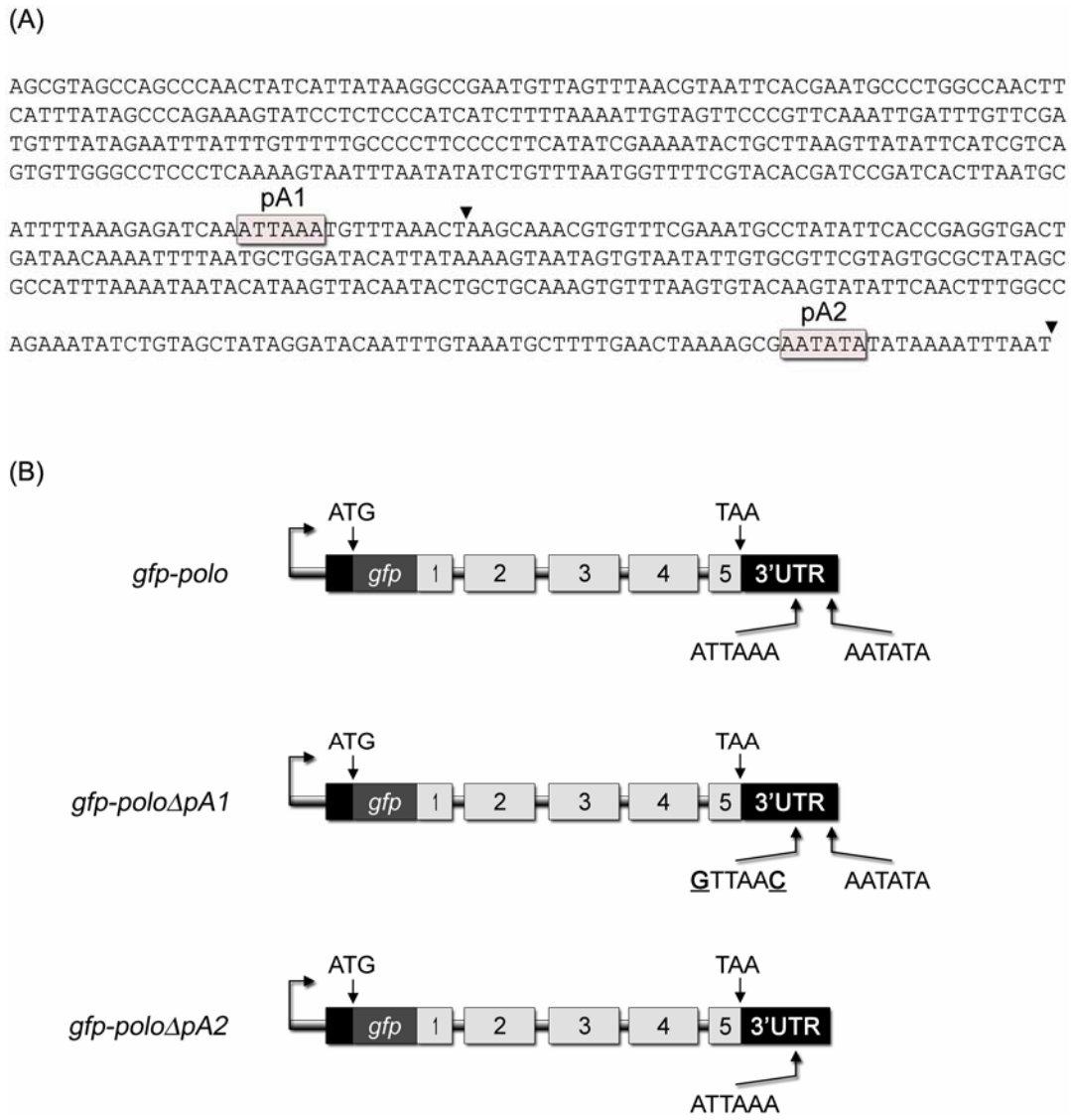

Figure 9. Generating transgenic flies expressing only one of each *polo* mRNAs. (A) *polo* 3'UTR sequence. The two grey boxes indicate the proximal (pA1) and distal (pA2) poly(A) signals. The arrowheads indicate both cleavage sites. **(B)** Schematic representation of the *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenes constructed to generate transgenic flies. The arrow at the beginning of the transgenes represents the endogenous *polo* promoter. The labelled 1-5 light grey boxes represent the five *polo* exons with the black boxes representing the 5' and 3'UTR of exons 1 and 5 respectively. The grey bar between the exons represents the introns. The *gfp* box represents the *gfp* coding sequence placed in frame with the *polo* initiation codon. The ATTAAA and AATATA are the proximal and distal poly(A) signals, respectively. In the *gfp-poloΔpA1* transgene point mutations were introduced in the proximal poly(A) signal changing the ATTAAA into **G**TTAA**C** (the point mutations are indicated in bold and underlined).

The strong hypomorphic nature of the *polo*^{9} allele makes it suitable to study the function of each *polo* transcript beyond the third instar stage. Moreover, the *gfp-polo* transgene is sufficient to rescue the *polo*^{9}/*polo*^{9} allelic combination (Pearson *et al.,* 2006), which allows the use of this transgene as a control in this study. Hence, flies carrying the

gfp-polo, *gfp-poloΔpA1* or *gfp-poloΔpA2* transgene in a *polo*^{9}/*TM6B* background were generated. This genetic background allows us to determine if each transgene is able to complement *polo*^{9} and if so, to analyse any phenotype on the adult flies that could elucidate on the function of each *polo* transcript.

The generated *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenes were used for P-element germline transformation. From the transformants obtained, lines with the transgene inserted on the second chromosome were selected by genetic mapping through matting with the *w*^{1118}; *Sco*/*SM6* strain. Finally, two lines homozygous for each transgene, viable and fertile, were selected to complement the *polo*^{9} allele: lines 9.1 and 12.1 for *gfp-poloΔpA1* and lines 2.3 and 5 for the *gfp-poloΔpA2* transgene.

2.2. Characterization of the *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenic flies

To determine if the point mutation of the proximal poly(A) signal and the deletion of the distal poly(A) signal were effective in inducing the synthesis of the longer (2.5 kb) and the shorter (2.2 kb) *polo* mRNAs respectively, a Northern blot was performed (Fig. 10). As *polo* is highly expressed in brains during the third instar larval stage (Llamazares *et al.,* 1991; Tavares *et al.,* 1996), brains from *gfp-polo, gfp-poloΔpA1, and gfp-poloΔpA2* third instar larvae were dissected and total RNA was extracted. Total RNA from *w*^{1118} third instar larvae was also prepared and used as a negative control.

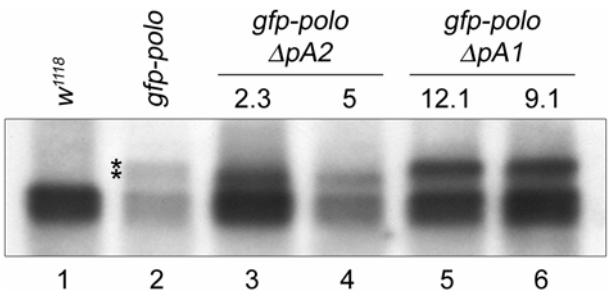

Figure 10. The *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenes express the longer and the shorter *gfp-polo* mRNAs, respectively. Third instar larvae from *w*^{1118} (lane 1), *gfp-polo* (lane 2), *gfp-poloΔpA2* (lines 2.3 (lane 3) and 5 (lane 4)) and *gfp-poloΔpA1* (lines 12.1 (lane 5) and 9.1 (lane 6)) were collected, brains were dissected and total RNA was extracted in order to perform a Northern blot. As probe, a radiolabelled *polo* cDNA fragment was used. The asterisks in lane 2 indicate the longer and shorter *gfp-polo* transcripts.

As illustrated in Fig. 10, the *gfp-polo* transgene expresses the two transcripts due to the usage of both poly(A) signals (lane 2; the two asterisks indicates the two *gfp-polo*

mRNAs). The lines of *gfp-polo∆pA1* (lanes 5 and 6) and *gfp-polo∆pA2* (lanes 3 and 4) express the longer and the shorter mRNA respectively, indicating that the introduced mutations are effective in generating transgenes capable of synthesizing only one *polo* mRNA.

Expression of GFP-Polo was also determined by Western blot analysis (Fig. 11). Third instar larvae brains (Fig. 11A), adult female ovaries (Fig. 11B) and adult male testis (Fig. 11C) from *gfp-polo* (lane 2), *gfp-polo∆pA1* (lanes 5 and 6) *and gfp-polo∆pA2* (lanes 3 and 4) flies were dissected and total protein extracts prepared. The GFP-Polo levels were detected using the MA294 monoclonal antibody (Llamazares *et al.,* 1991). α-Tubulin was used as a loading control for the quantification of the GFP-Polo levels expressed in each transgenic line. As show in Fig. 11, all transgenes express GFP-Polo. Despite some differences in expression levels between the different transgenic lines most likely due to insertion of the transgenes in different regions of the genome, all *gfp-polo∆pA2* transgenic flies express higher levels of Polo when compared with *gfp-polo∆pA1* flies (Fig. 11A, B and C, compare lanes 3 and 4 with 5 and 6). Transgenic flies carrying *gfp-polo∆pA2* express ~2 and 2.6-fold higher levels of GFP-Polo in brains and testis, respectively. In ovaries, *gfp-polo∆pA2* flies also appear to express higher levels of GFP-Polo. However, this is not the case for line 5 which has levels of expression similar to the *gfp-polo∆pA1* lines (Fig. 11B, compare lanes 3 and 4 with 5 and 6).

One of the main phenotypes that characterizes the $polo^9$ allele, is arrest at prometaphase/metaphase stage of the cell cycle (Donaldson *et al.,* 2001). To determine if the transcripts from the different transgenes could rescue this mitotic arrest, we performed complementation of the $polo^9$ allele. With this purpose, each transgene was crossed with flies carrying the $polo^9$ allele and flies harbouring each transgene in the second chromosome and $polo^9$ mutation in the third chromosome were generated. Brains squashes from *gfp-polo*, *gfp-polo∆pA1 and gfp-polo∆pA2* third instar larvae in a $polo^9$ background were prepared and the mitotic index quantified (Table I). Similar to what was reported, cells from homozygous $polo^9$ individuals showed a ~15-fold increase in the mitotic index when compared with hemizygous individuals ($polo^9/TM6C$) (Donaldson *et al.,* 2001).

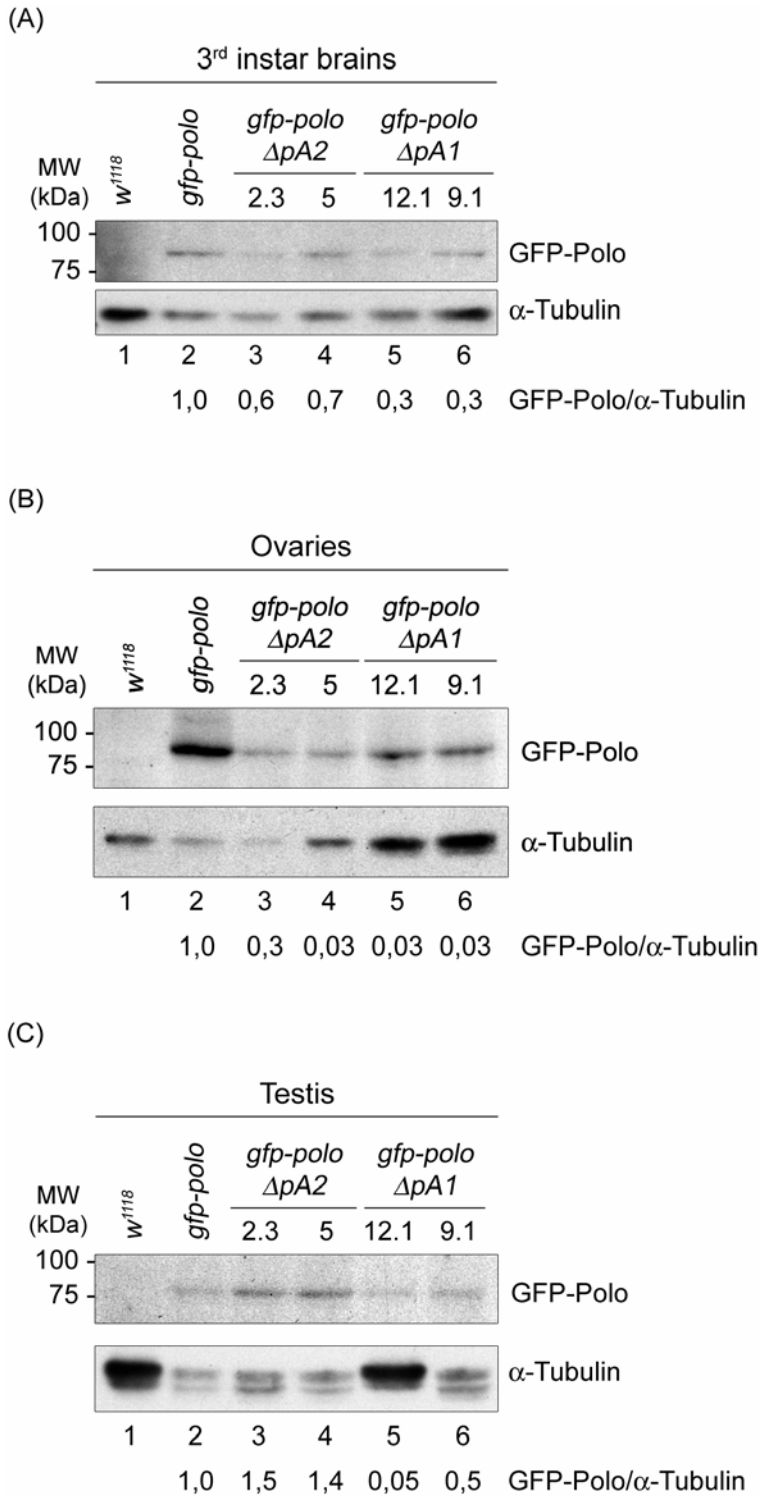

Figure 11. All transgenic flies express GFP-Polo. Western blot analysis against GFP-Polo was performed using total protein extracts prepared from **(A)** third instar larvae brains, **(B)** adult female ovaries and **(C)** adult male testis of w^{1118} (lane 1), *gfp-polo* (lane 2), *gfp-poloΔpA2* (lines 2.3 (lane 3) and 5 (lane 4)) and *gfp-poloΔpA1* (lines 12.1 (lane 5) and 9.1 (lane 6)) flies. α-Tubulin was used as loading control. The GFP-Polo and α-Tubulin were detected using Polo (MA294) and α-Tubulin (mAB DM1A) monoclonal antibodies. The ratios GFP-Polo/α-Tubulin were determined by densitometry scanning of the autoradiographs and normalized for *gfp-polo.*

As expected, complementation of *polo*^{9} by the *gfp-polo* transgene lowers the mitotic index to values similar to those presented by *polo*^{9} hemizygous, here used as a control strain. Importantly, both *gfp-polo∆pA1* and *gfp-polo∆pA2* transgenes are able to rescue the mitotic arrest associated with *polo*^{9} mutation. Therefore, despite the different levels of GFP-Polo expression in the different transgenes in third instar stage brains, both *gfp-polo∆pA1* and *gfp-polo∆pA2* are able to rescue the prometaphase/metaphase arrest of *polo*^{9} mutant brains.

Strain	Mitotic Index
polo^9*/TM6C*	1,5
polo^9*/polo*^{9}	22,2
gfp-polo;polo^9*/TM6B*	1,1
gfp-polo;polo^9*/polo*^{9}	1,3
gfp-polo∆pA1;polo^9*/TM6B* (Line 9.1)	1,3
gfp-polo∆pA1;polo^9*/polo*^{9} (Line 9.1)	1,5
gfp-polo∆pA1;polo^9*/TM6B* (Line 12.1)	1,4
gfp-polo∆pA1;polo^9*/polo*^{9} (Line 12.1)	1,6
gfp-polo∆pA2;polo^9*/TM6B* (Line 2.3)	1,2
gfp-polo∆pA2;polo^9*/polo*^{9} (Line 2.3)	1,2
gfp-polo∆pA2;polo^9*/TM6B* (Line 5)	1,1
gfp-polo∆pA2;polo^9*/polo*^{9} (Line 5)	1,1

Table I. Quantification of the mitotic index in cells from third instar larvae brains. *gfp-polo*, *gfp-polo∆pA1*, (lines 9.1 and 12.1) and *gfp-polo∆pA2* (lines 2.3 and 5) third instar larvae, either in a *polo*^{9}*/TM6B* or *polo*^{9}*/polo*^{9} genetic background, were dissected and the mitotic index quantified. As a control, the mitotic index of *polo*^{9}*/TM6C and polo*^{9}*/polo*^{9} third instar larvae was also determined.

Since both transgenes are able to rescue the *polo*^{9} mitotic phenotype we evaluated if the transgenes were also able to rescue the lethality of *polo*^{9} (Donaldson *et al.,* 2001). To this end, we evaluated the progeny from the fly lines carrying the different transgenes in a *polo*^{9} background. For each strain, individual crosses were performed using parents with a *polo*^{9}*/TM6B* background. From these crosses, flies homozygous for *polo*^{9} that reach adulthood were quantified over the total number of adult flies (Table II). The numbers obtained show that expression of both *gfp-polo* and *gfp-polo∆pA1* transgenes can rescue the lethality associated with *polo*^{9} allele. However, this is not observed for the *gfp-polo∆pA2* transgene, since *gfp-polo∆pA2; polo*^{9} flies do not reach adulthood. In fact, a close observation shows that *gfp-polo∆pA2; polo*^{9} die during the pupae stage. As a control, crosses were made to rescue the *polo*^{9}*/Df(3L) rdgC-co2* allelic combination to exclude any unexpected contribution of the *polo*^{9} chromosome in our quantifications. The same trend is observed. The *gfp-polo∆pA1* transgene is sufficient to the rescue the lethality of *polo*^{9}*/Df(3L) rdgC-co2* flies (49,7%), while only few *gfp-polo∆pA2; polo*^{9}*/Df(3L) rdgC-co2* flies survive (15,3%). Therefore, the *gfp-polo∆pA2* transgene is less efficient in

the rescue of *polo*^{9}. This data suggest that the longer *polo* mRNA is required for *Drosophila* development beyond the third instar larval stage.

Strain	Viability percentage of flies homozygous for *polo*^{9} carrying each of the transgenes.
gfp-polo;polo^9*/polo*^{9}	30,0%
gfp-poloΔ*pA1; polo*^{9}*/polo*^{9} (linha 9.1)	39,8%
gfp-poloΔ*pA1; polo*^{9}*/polo*^{9} (linha 12.1)	20,9%
gfp-poloΔ*pA2; polo*^{9}*/polo*^{9} (linha 2.3)	0,0%
gfp-poloΔ*pA2; polo*^{9}*/polo*^{9} (linha 5)	0,2%
gfp-poloΔ*pA1;polo*^{9}*/Df (3L) rdgC-co2* (Line 12.1)	49,7%
gfp-poloΔ*pA2;polo*^{9}*/Df (3L) rdgC-co2* (Line 2.3)	15,3%

Table II. Quantification of the ability of the different transgenes to rescue the lethality associated with mutations in *polo*. For each strain, crosses were performed to rescue the *polo*^{9}*/polo*^{9} and *polo*^{9}*/Df(3L) rdgC-co2* allelic combinations. From these crosses, flies homozygous for *polo*^{9} and *polo*^{9}*/Df(3L) rdgC-co2* were quantified over the total number of flies.

2.3 The longer *polo* mRNA is required for proper abdomen formation in *Drosophila*.

An analysis of the few *gfp-polo*Δ*pA2; polo*^{9} escapers shows a strong abdominal phenotype characterized by the absence or incorrect formation of tergites (Fig. 12, compare panel 1, 5 and 7). No phenotype was observed in other adult structures. We analysed whether improper abdomen formation was also present in *gfp-polo; polo*^{9} *and gfp-polo*Δ*pA1; polo*^{9} flies. Our analysis shows that while *gfp-polo* flies present no problems in the abdomen, flies carrying the *gfp-polo*Δ*pA1* transgene present a very mild phenotype (Fig.12, compare panel 1, 5 and 6). The same trend is observed in flies transheterozygous for *polo*^{9} and the *Df(3L) rdgC-co2.* While the *gfp-polo*Δ*pA2; polo*^{9}*/ Df(3L) rdgC-co2* flies have strong abdomen defects, *gfp-polo*Δ*pA1; polo*^{9}*/ Df(3L) rdgC-co2* flies do not show the same phenotype (data not shown).

These results suggest that the use of the distal poly(A) signal and therefore, the longer 2.5 kb *polo* mRNA, is required for the formation of the normal adult abdomen.

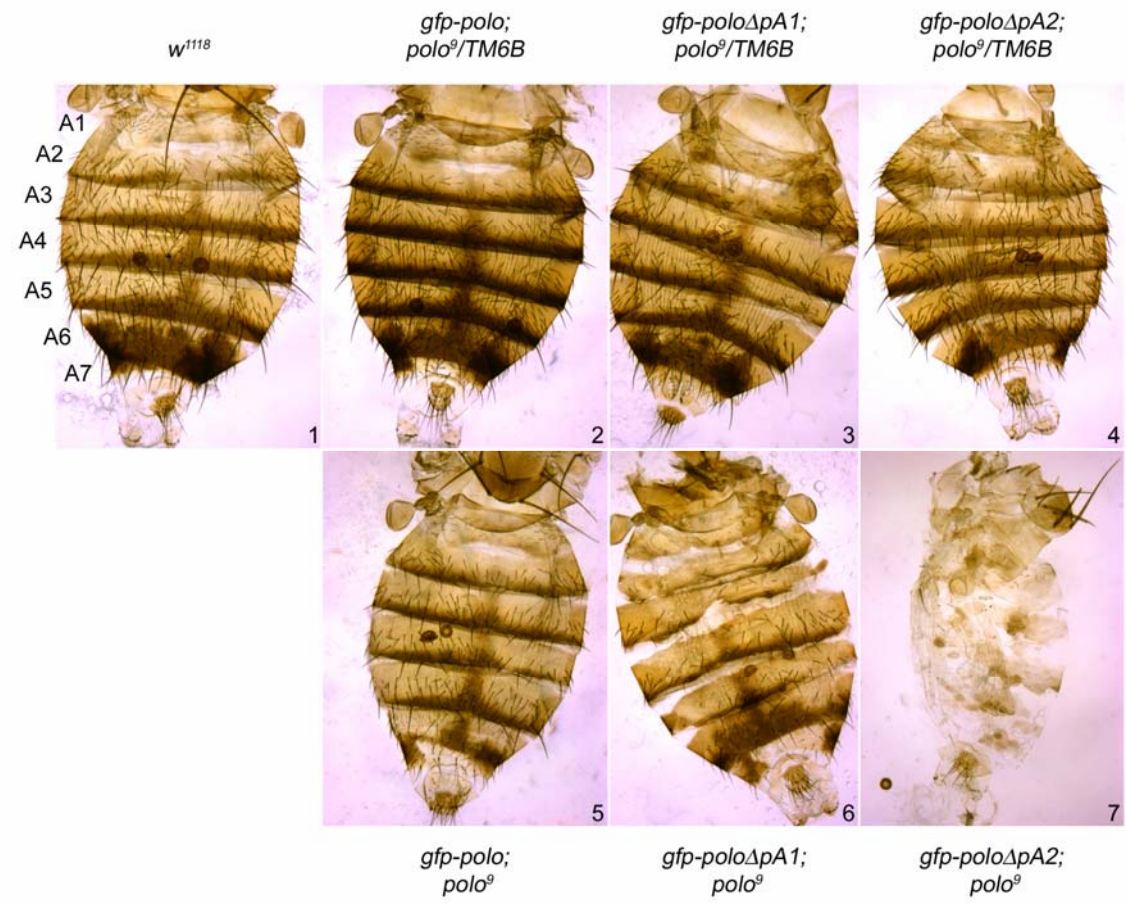

Figure 12. The longer *polo* mRNA is required for proper abdomen formation in *Drosophila*. Dorsal view of the adult abdomen of *gfp-polo*, *gfp-poloΔpA1* (line 12.1) and *gfp-poloΔpA2* (line 2.3) female flies. Abdomens were prepared from flies in a *polo*^{9}/*TM6B* (panel 2-4) or *polo*^{9} homozygous background (panel 5-7). An abdomen from a *w*^{1118} female was used as control (panel 1). The seven tergites of the adult abdomen are indicated as A1 to A7 (panel 1). Anterior is up.

To exclude the possibility that the phenotype could arise from gene disruption or interference with regulatory sequences as a result of the transgene insertion in the same place in the genome for both transgenic lines, we mapped the insertion of the *gfp-poloΔpA2* transgene in lines 2.3 and 5 (Fig. 13). As determined by *in situ* hybridisation in polytene chromosomes from salivary glands of third instar larvae, each line has one copy of the transgene. Two bands are observed for each line: one from the endogenous *polo* locus in the chromosome 3R, between region 77B2-B3 and one from the *gfp-poloΔpA2* transgene*.* The transgene in line 2.3 is mapped on chromosome 2R within the region 52A-C, while for line 5 the transgene is mapped on chromosome 2R between regions 41-42. Thus, the abdominal phenotype observed in *gfp-poloΔpA2; polo*^{9} is most likely the result of the absence of the longer 2.5 kb *polo* transcript rather than the result of the transgene insertion in the genome.

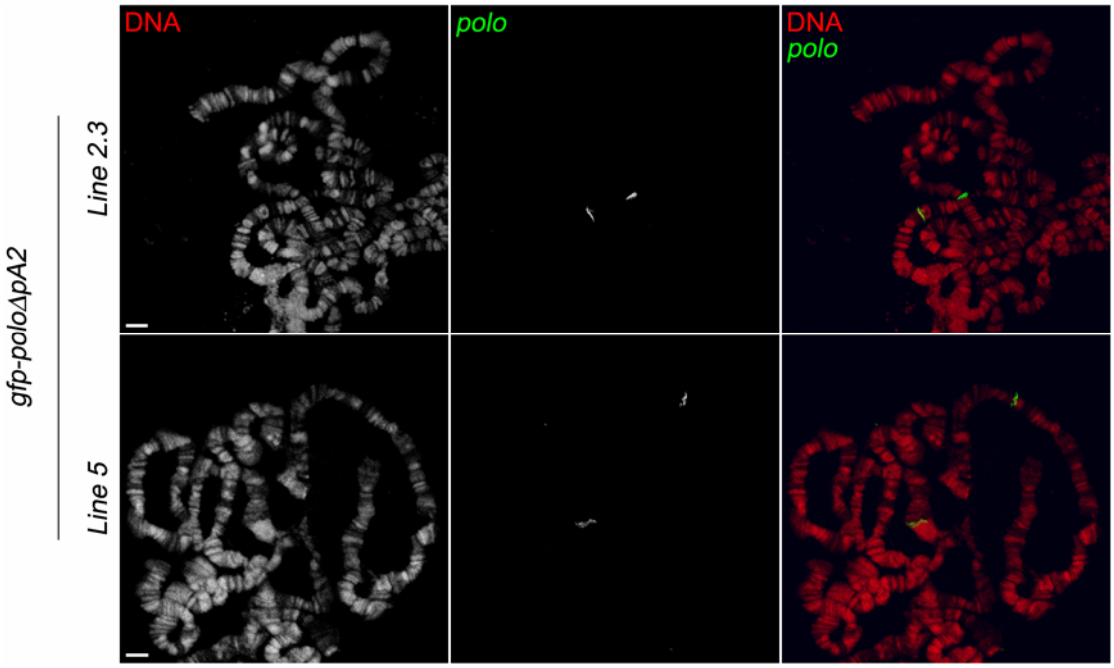

Figure 13. The *gfp-poloΔpA2* transgene is inserted in different regions in the genomes of *gfp-poloΔpA2* transgenic lines. The insertion of the *gfp-poloΔpA2* transgene was mapped by *in situ* hybridisation on polytene chromosomes from salivary glands of *gfp-poloΔpA2* third instar larvae. Both *gfp-poloΔpA2* lines (2.3 and 5) were mapped. To map the transgenes, a *polo* cDNA fragment was used as a probe. Two bands were mapped for each line: one from the *polo* locus in the chromosome 3R, 77B2-B3 and a second one from the *gfp-poloΔpA2* transgene. The transgene was mapped in chromosome 2R in the regions 52A-C and 41-42, for lines 2.3 and 5 respectively. The bar is 10 μm.

2.4. Both *polo* mRNAs are expressed during the last stages of embryogenesis.

3'UTRs often contain regulatory elements involved in mRNA stability, translation and localization and therefore are able to regulate gene expression (reviewed in Conne *et al.,* 2000). It is possible that the two *polo* transcripts have different regulatory elements with different roles in Polo expression. Formation of the abdominal segments starts during the last stages of embryogenesis with the formation of the germ band (stage 11), and are fully formed by the end of stage 16, with the amnioserosa closure (Hartenstein 1993).To determine if both *polo* transcripts can differentially direct Polo synthesis during abdominal segmentation, we analysed the GFP-Polo expression pattern in *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔpA2* embryos (Fig. 14).

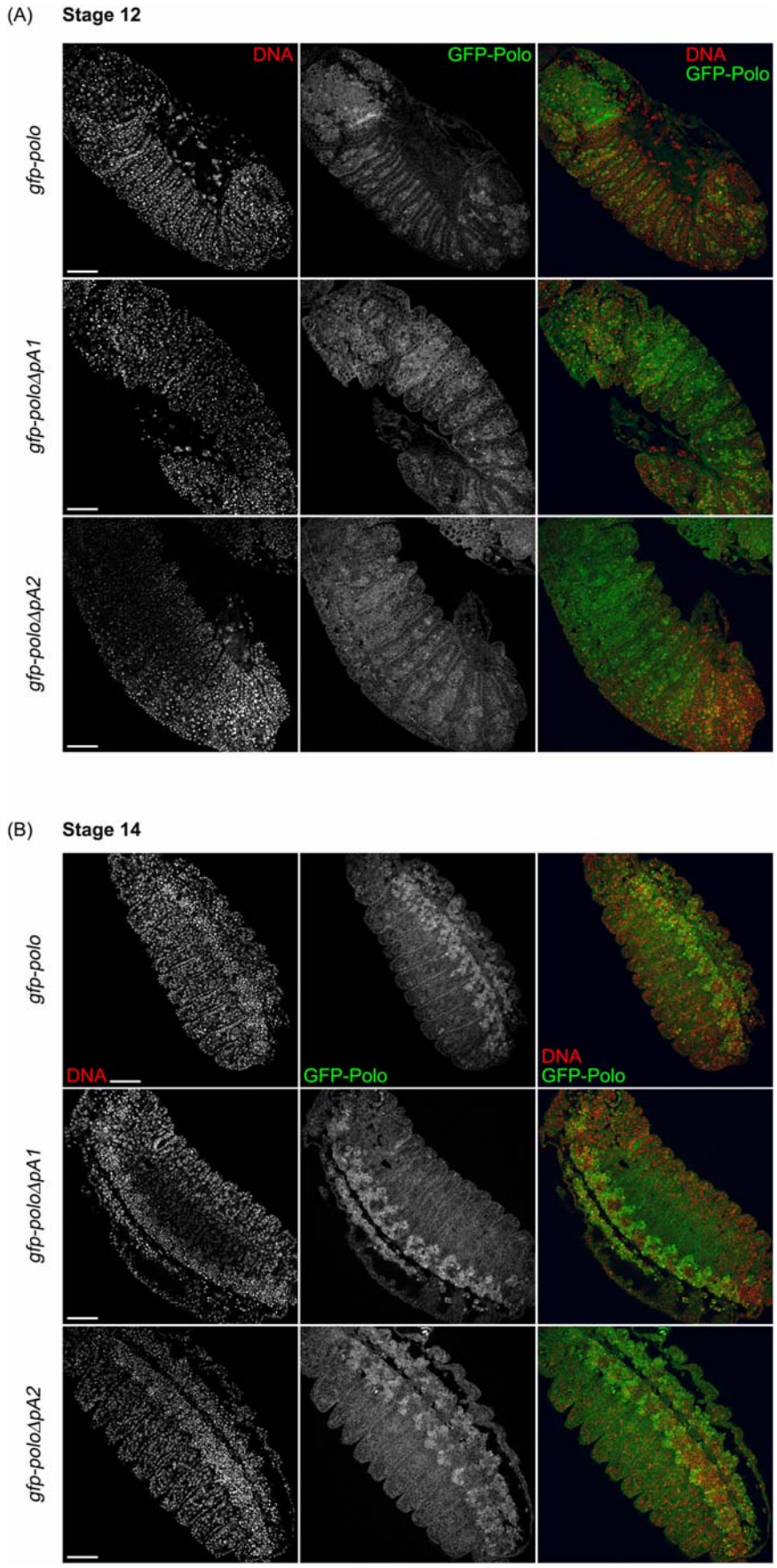

Figure 14. (see legend on next page).

Figure 14. *gfp-polo, gfp-poloΔpA1 and gfp-poloΔpA2* transgenes show the same pattern of expression during embryonic abdominal segmentation. GFP-Polo expression was determined in stage 12 **(A)** and stage 14 **(B)** embryos. *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔpA2* embryos were collected and GFP-Polo expression (green) analysed by confocal microscopy. DNA was stained with DAPI (red). The images show a ventral view of the embryos, with exception in (A) in which is shown a dorsal view of the embryo. In both panels, anterior is up. The bar is 50 μm.

It is possible to observe that the pattern of expression is very similar for all transgenes. From stage 12 to 14, GFP-Polo expression is observed throughout the thoracic and abdominal segments, central nervous system and a cluster of cells that judging by their numbers and location resembles the cluster of pole cells and follicle precursor cells. However, by stage 15, in the last steps of dorsal closure, GFP-Polo seems to be restricted to the central nervous system and the pole and follicle precursor cells (data not shown).

2.5. *gfp-poloΔpA2; polo*^{9}abdominal histoblasts are absent in pupae with 26-27 hours APF

The similarity of the GFP-Polo pattern of expression in late embryos in all transgenic flies suggests that the abdominal phenotype observed in *gfp-poloΔpA2; polo*^{9} flies reflects a defect in later stages of development. Indeed, the high levels of lethality during the pupa stage observed for *gfp-poloΔpA2; polo*^{9} pupae suggest a defect during this particular stage of development.

In *Drosophila* the adult abdominal epidermis is formed during metamorphosis through proliferation of the abdominal histoblasts nests (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980). The histoblasts are diploid cells organized in small clusters, called nests, with four nests present in each hemi-segment: an anterior dorsal nest (~15 cells), a posterior dorsal nest (~5 cells), a spiracular nest (~3 cells) and a ventral nest (~15 cells) (Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979). The histoblasts are formed during the last stages of embryogenesis and are arrested in the G2 phase of the cell cycle and will continue in this stage through first to third instar larva stages, during which they undergo a 60% increase in cell volume (Garcia-Bellido and Merriam 1971; Guerra *et al.,* 1973; Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979).

Upon entrance in metamorphosis, these cells undergo rapid proliferation and by 15 hours after pupa formation (APF) they will expand replacing all the larva polytene epidermis cells with occurring fusion between the different nests in each hemi-segment

and from adjacent hemi-segments (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). By 40 hours APF histoblasts from opposing hemi-tergites fuse and form a complete epidermis layer (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007).

Two genes were reported to be necessary for the formation of a normal adult abdomen: the *arrowhead* (*awh*) (Curtiss and Heilig 1995) and *escargot* (*esg*) (Hayashi *et al.,* 1993). *arrowhead* mutants die during metamorphosis with pharate adults showing no development of the abdominal epidermis (Curtiss and Heilig 1995). An analysis of these mutants showed an average of a 2-fold decrease in the number of abdominal histoblasts cells. Study of *escargot* mutants showed that this gene is required for the maintenance of diploidy in the abdominal histoblasts (Hayashi *et al.,* 1993). In *esg* mutants, histoblasts undergo DNA endoreplication resulting in polyploid cells which fail to proliferate, with adult flies presenting a abdominal phenotype very similar to the *gfp-polo∆pA2; polo*^{9} phenotype (Hayashi *et al.,* 1993).

As *gfp-polo∆pA2; polo*^{9} flies present a phenotype in agreement with an abnormal proliferation of histoblasts we investigated histoblast proliferation during the pupae stage. Pupae were collected 26-27 hours APF and dissection of the pupa epidermis was performed. At this stage histoblasts have already proliferate and form over 50% of the epidermis (Roseland and Schneiderman 1979). After dissection, the epidermis was stained with DAPI and analysed by fluorescence microscopy. As illustrated in Fig. 15 it is possible to observe that histoblasts proliferated in epidermis of both *gfp-polo; polo*^{9} and *gfp-polo∆pA1; polo*^{9} pupae (compare B with D). However, when we analyse *gfp-polo∆pA2; polo*^{9} epidermis, no histoblast proliferation is observed (Fig. 15, compare B with F). In fact, at this developmental stage the pupa epidermis is still formed by polytene larva epidermis cells. This clearly indicates that the absence or incorrect formation of the tergites in *gfp-polo∆pA2; polo*^{9} flies results from lack of histoblast proliferation. Moreover, a proliferation defect is observed for the *gfp-polo∆pA2* transgene only when in a *polo*^{9} homozygous background, again showing that this phenotype is a result of flies expressing only the smaller 2.2 kb mRNA (Fig. 15, compare E with F).

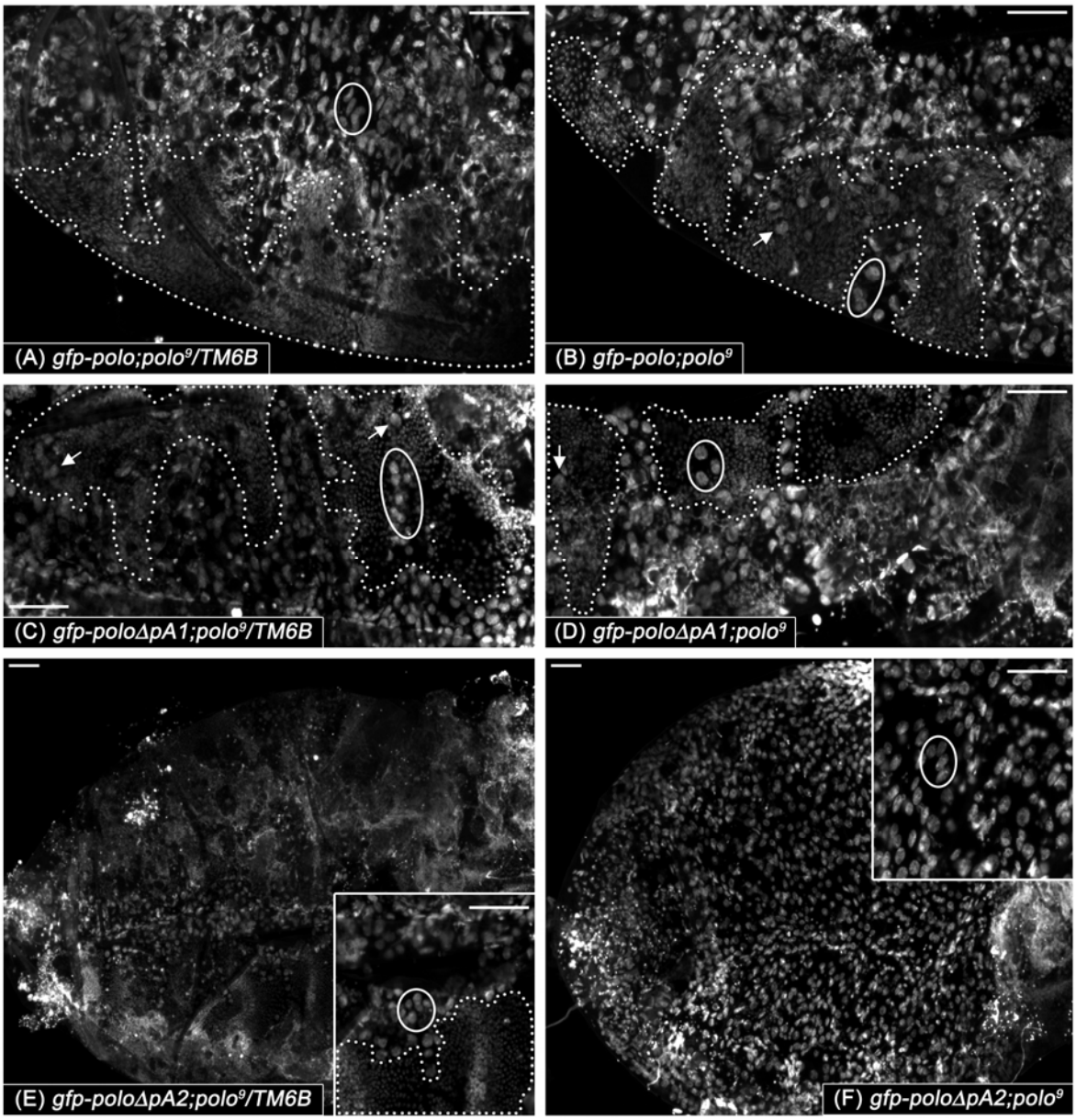

Figure 15. Deletion of the distal poly(A) signal prevents abdominal histoblast proliferation during metamorphosis. Histoblast proliferation was analysed in *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔA2* pupae in a *polo*^{9} homozygous background (B, D and F). Pupae were collected 26-27 hours APF and the pupae epidermis dissected. The epidermis was stained with DAPI and analysed by fluorescence microscopy. As a control *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔpA2* pupae in a *polo*^{9}*/TM6B* background were also collected and analysed 26-27 hours APF (A, C and E). In E and F, the lower right and upper right images, respectively, are high magnifications of selected areas. The white circles enclose some of the larva epidermis cells (LECs) and the white arrows indicate LECs surrounded by abdominal histoblast cells. The dotted areas encompass areas of histoblasts proliferation. The bar is 50 μm.

2.6. *gfp-poloΔpA2; polo*^{9} abdominal histoblasts fail to proliferate

Two possibilities could explain the absence of abdominal histoblast proliferation in *gfp-poloΔpA2; polo*^{9} pupae, 26-27 hours APF. One is that histoblasts fail to proliferate. The other possibility concerns the specification/formation of histoblasts. If histoblasts are not formed during the last stages of embryogenesis, there will be no cells to form the adult epidermis. Therefore, we investigated the presence of histoblasts at the onset of metamorphosis. Third instar larvae were dissected and an immunostaining was performed for Headcase (Hdc), an abdominal histoblast maker (Weaver and White 1995) (Fig. 16). It is possible to observe that histoblasts are present in *gfp-polo; polo*^{9}, *gfp-poloΔpA1; polo*^{9} and *gfp-poloΔpA2; polo*^{9} larvae and quantification of the number of abdominal histoblasts shows a similar number of cells in each nest for the different abdominal segments in all three transgenic lines (Fig. 16 and Table III).

Nest	Strain	Abdominal segments					
		A2	A3	A4	A5	A6	A7
ADH	*gfp-polo; polo*^{9}	10	15	15	12	12	10
	gfp-poloΔpA1; polo^{9} (Line 12.1)	13	12	16	12	16	14
	gfp-poloΔpA2; polo^{9} (Line 2.3)	16	16	15	12	12	12
PDH	*gfp-polo; polo*^{9}	5	6	5	3	3	3
	gfp-poloΔpA1; polo^{9} (Line 12.1)	4	4	4	3	3	2
	gfp-poloΔpA2; polo^{9} (Line 2.3)	4	4	4	4	3	2
VH	*gfp-polo; polo*^{9}	15	13	13	10	11	11
	gfp-poloΔpA1; polo^{9} (Line 12.1)	14	11	7	10	8	10
	gfp-poloΔpA2; polo^{9} (Line 2.3)	13	13	13	11	11	6

Table III. Quantification of the number of abdominal histoblast cells in the different transgenic flies. The number of abdominal histoblast cells was determined in the different nests for each abdominal segment of *gfp-polo*, *gfp-poloΔpA1* (line 12.1) and *gfp-poloΔpA2* (line 2.3) third instar larvae, in a *polo*^{9}*/polo*^{9} background. ADH-anterior dorsal nest, PDH-posterior dorsal nest, VH-ventral nest.

Moreover, all histoblasts stain strongly for Cyclin B in agreement with what is expected from G2 arrested cells (Fig.16). These results indicate that histoblasts are formed during embryogenesis and that are able to maintain the G2 arrest during the larvae development suggesting that the usage of the distal poly(A) site of *polo* is required for proliferation of these specific cells at the onset of metamorphosis.

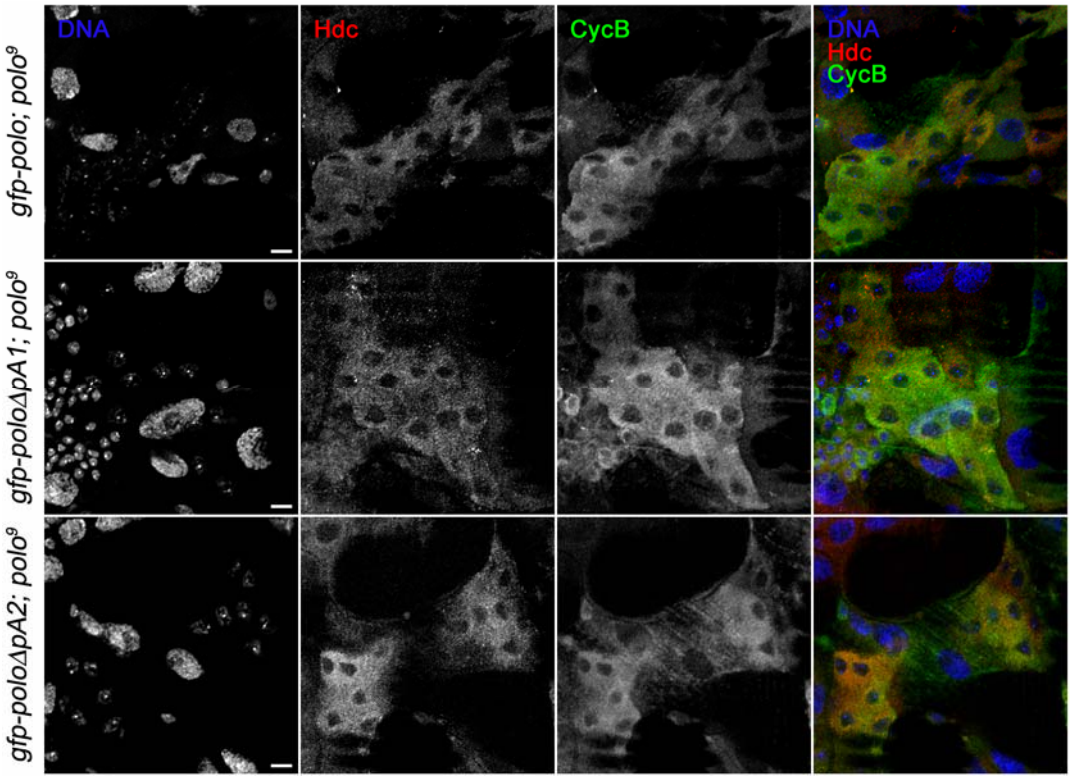

Figure 16. Abdominal histoblasts in *gfp-polo; polo*^{9}, *gfp-poloΔpA1; polo*^{9} and *gfp-poloΔpA2; polo*^{9} third instar larvae accumulate Cyclin B. *gfp-polo*, *gfp-poloΔpA1* (line 12.1) and *gfp-poloΔpA2* (line 2.3) third instar larvae in a *polo*^{9}/*polo*^{9} background were dissected and immunolabelled with Headcase (Hdc-red) and Cyclin B (Cyc B-green) antibodies. The picture shows an anterior dorsal nest. The DNA (blue) was stained with DAPI. The bar is 10 μm.

2.7. *gfp-poloΔpA2; polo*^{9} abdominal histoblasts express lower levels of GFP-Polo

To evaluate the requirement for the distal poly(A) site of *polo*, we analysed the expression of GFP-Polo in abdominal histoblasts. Pupae from different transgenes hemizygous for *polo*^{9} were collected 26-27 hours APF and dissected (Fig. 17). It is possible to observe that all transgenes are expressed in proliferating histoblasts. However, a quantitative analysis of the levels of GFP-Polo by measurement of the GFP fluorescence shows that while *gfp-polo* and the *gfp-poloΔpA1* transgenes express similar levels of GFP-Polo, a 30% decrease in fluorescence is observed for the *gfp-poloΔpA2* transgene (Table IV).

Altogether, this data shows that flies expressing the longer mRNA (*gfp-polo* and *gfp-poloΔpA1* flies), have abdominal histoblasts with similar levels of GFP-Polo. Only when the synthesis of mRNA is restricted to the smaller transcript (*gfp-poloΔpA2* flies), do histoblasts present lower levels of GFP-Polo.

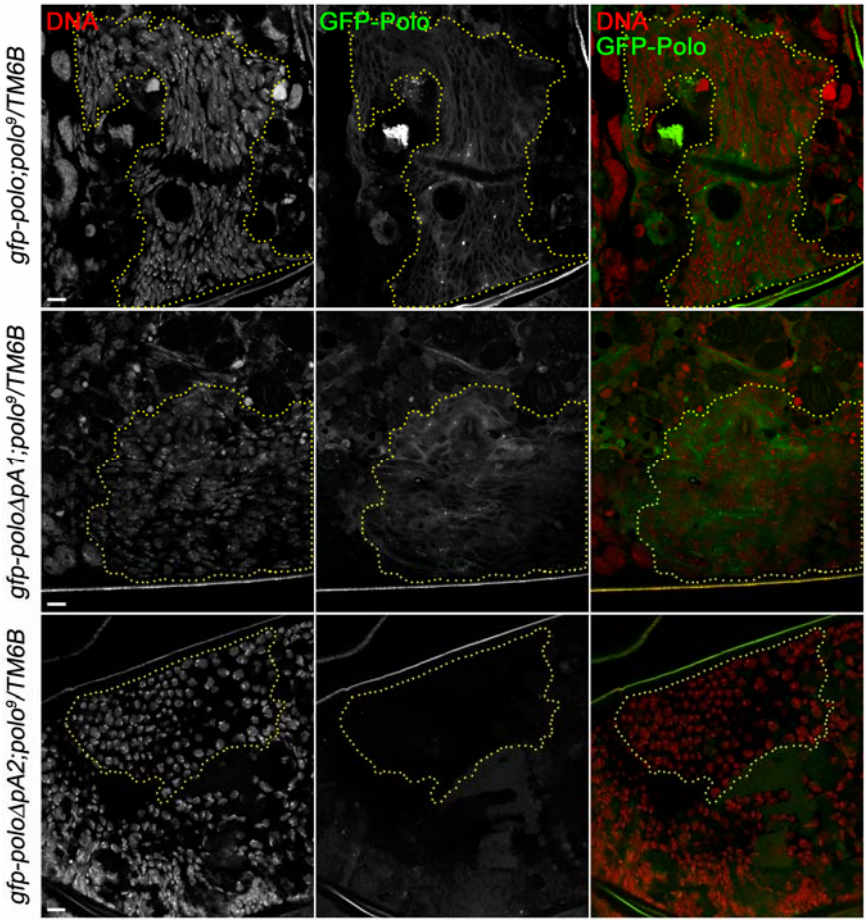

Figure 17. The *gfp-poloΔpA2* transgene express lower levels of GFP-Polo in abdominal histoblast cells. Pupae from *gfp-polo*, *gfp-poloΔpA1* (line 12.1) and *gfp-poloΔpA2* (line 2.3) in a *polo^9/TM6B* background were collected 26-27 hours APF. Dissection of the pupae epidermis was performed and GFP-Polo expression (green) quantified. The DNA (red) was stained with DAPI. The yellow dotted area encompasses some of the histoblast cells present in the epidermis. The bar is 10 μm.

Strain	GFP levels [(mean x area)/nº cells] +/- st error	
gfp-polo; polo^9/TM6B	44,2 +/- 4,0	(100%)
gfp-polo Δ pA1; polo^9/TM6B (Line 12.1)	46,1 +/- 4,5	(104%)
gfp-polo Δ pA2; polo^9/TM6B (line 2.3)	31,5 +/- 3,5	(71%)

Table IV. Quantification of the levels of GFP-Polo expressed in abdominal histoblasts. The levels of GFP-Polo were quantified from *gfp-polo*, *gfp-poloΔpA1* (line 12.1) and *gfp-poloΔpA2* (line 2.3) pupae in a *polo^9/TM6B* background, 26-27 hours APF. The levels of GFP-Polo were determined via quantification of the GFP fluorescence levels.

These results are opposite to those observed in third instar larvae brains and in adult male testis for the different transgenes. In these tissues *gfp-poloΔpA2* transgenic

flies express higher levels of GFP-Polo compared to *gfp-polo⊿pA1* transgenics (Fig.11A, C, compare lanes 3, 4 with lanes 5, 6). These observations seem to indicate that expression of GFP-Polo from each transgene is regulated in a tissue or/and developmental specific manner. Therefore, these evidences suggest that the distal poly(A) site of *polo* may be required for the appropriate levels of Polo in histoblasts at the entrance of the pupa stage and that it may be necessary for rapid proliferation upon the onset of metamorphosis.

3. Discussion:

The results obtained in this study, with transgenic flies expressing only one of the two *polo* mRNAs, show that the longer 2.5 kb mRNA, produced by utilization of pA2, is essential for *Drosophila* development. Individuals expressing only the small transcript (*gfp-polo⊿pA2; polo*^{9}) die during the pupa stage while individuals expressing only the longer mRNA are viable (*gfp-polo⊿pA1; polo*^{9}). Moreover, the few *gfp-polo⊿pA2; polo*^{9} flies able to reach adulthood show strong abdominal defects, with absent or incorrect formation of the tergites suggesting a problem during the development of the adult abdominal epidermis. In contrast, flies using the distal poly(A) signal (*gfp-polo⊿pA1; polo*^{9}) only present mild defects. Therefore, the results indicate that the longer 2.5 kb *polo* mRNA is required at a specific stage of development in a tissue specific way.

As mentioned earlier, the adult epidermis arises from a group of imaginal cells called abdominal histoblast cells (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). These diploid cells are organized in clusters of small number of cells called nests. The adult abdomen has 8 abdominal segments (Roseland and Schneiderman 1979). The last segment (the most posterior one) which comprises the genitalia or analia, arises from an imaginal disc while the remaining seven abdominal segments arise from the abdominal histoblasts (Roseland and Schneiderman 1979). For each abdominal hemi-segment there are four abdominal histoblast nests: an anterior dorsal nest with ~15 cells, a posterior dorsal nest with ~5 cells, a ventral nest of ~15 cells and a spiracular nest with ~3 cells (Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979). These cells arise in the last stages of embryogenesis and are arrested in the G2 phase of the cell cycle and will continue in this stage throughout the larval development until the onset of metamorphosis (Garcia-Bellido and Merriam 1971; Guerra *et al.,* 1973; Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979). Once metamorphosis starts, the histoblasts enter mitosis proliferating rapidly but

still remaining confined to their original location (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). By 15 hours APF, the histoblasts start to expand and replace the larval epidermal cells (LECs) with the consequently fusion of the different nests in each hemi-segment and fusion between neighbouring hemi-segments (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). This proliferative state continues until 40h APF when fusion of opposing hemi-segments occurs, completing the formation of the adult epidermal layer. The histoblasts will then undergo differentiation, secreting adult cuticle and forming abdomen adult structures such as hairs, micro and macrochaetae (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980).

Analysis of the developing adult epidermis 26-27 hours APF shows that in the absence of pA2 and consequently, of the longer 2.5 kb *polo* mRNA (*gfp-polo⊿pA2, polo*^{9} pupae), the histoblasts fail to proliferate during metamorphosis while in the control strain (*gfp-polo; polo*^{9}) and in pupae containing pA2 and therefore expressing the longer mRNA (*gfp-polo⊿pA1; polo*^{9}) proliferation occurs normally. Further supporting this observation is the analysis of *gfp-polo⊿pA2, polo*^{9} third instar larvae that shows that the absence of proliferation 26-27 hours APF is not due to lack of specification of histoblasts during embryogenesis. In fact, the use of Headcase, a specific marker for abdominal histoblasts, shows that these cells are present prior to the onset of metamorphosis. Quantification of the number of histoblasts in each nest shows that the control strain (*gfp-polo; polo*^{9}) as well as strains expressing only one *polo* mRNA (*gfp-polo⊿pA1; polo*^{9} and *gfp-polo⊿pA2; polo*^{9}) have similar number of cells in each nest in all abdominal segments.

Furthermore, all cells accumulated Cyclin B as expected from G2 arrested cells. This indicates that the phenotype observed in flies that do not possess pA2, and thus expressing the smaller *polo* mRNA (*gfp-polo⊿pA2, polo*^{9}) results from failure of histoblasts to proliferate once metamorphosis starts, suggesting that mRNA 3' end formation at the distal poly(A) signal is essential for abdominal histoblast proliferation and consequently the development of the adult epidermis.

Quantification of the GFP-Polo levels in the different pupae strains 26-27 hours APF shows that while pupae expressing the longer mRNA have similar levels of GFP-Polo (*gfp-polo; polo*^{9}/*TM6B* and *gfp-polo⊿pA1*;*polo*^{9}/*TM6B*), pupae expressing only the small mRNA (*gfp-polo⊿pA2; polo*^{9}*TM6B*) show a significant decrease in the levels of GFP.

In third instar larvae brains, in adult male testis and in adult female ovaries, the expression of the short 2.2 kb mRNA (the *gfp-polo⊿pA2* transgene) always results in higher levels of GFP-Polo when compared to the long 2.5 kb mRNA (the *gfp-polo⊿pA1* transgene). Only in abdominal histoblasts, higher levels of GFP-Polo are obtained with the

expression of the long 2.5 kb mRNA. Therefore, our results suggest that the 2.5 kb *polo* mRNA is essential for the proliferation of abdominal histoblasts upon the onset of metamorphosis and consequently for the proper development of the adult abdominal epidermis. Moreover, it is conceivable that the role of the long 2.5 kb transcript in these cells during this developmental stage might be to provide the necessary levels of Polo protein to ensure the proliferation of abdominal histoblasts.

In the earlier studies of *polo* it was reported that hemizygous $polo^1/polo^2$ individuals show an abnormal abdominal development, affecting both tergites and sternites (Sunkel and Glover 1988). $polo^1$ is an hypomorphic allele induced by EMS in which a point mutation in the kinase domain severely reduces the kinase activity without interfering with Polo protein levels (Tavares *et al.,* 1996). $polo^1/polo^1$ individuals, despite presenting a slower development, show normal abdomen morphogenesis in adults. $polo^2$ is a strong hypomorph allele resulting from a P-element insertion which produces reduced levels of protein (Sunkel and Glover 1988; Llamazares *et al.,* 1991; Herrmann *et al.,* 1998). Therefore, $polo^1/polo^2$ flies express lower levels of Polo protein as compared to wild-type (Herrmann *et al.,* 1998). Since flies bearing a mutation in the Polo kinase domain display a normal abdomen development, one can argue that the abdominal phenotype observed in $polo^1/polo^2$ flies is most likely due to lower levels of Polo in abdominal histoblasts during the pupa stage. This would support our findings that proper levels of Polo protein are necessary for abdominal development.

Finally, it was recently reported that Polo deficient cells present defects in the cortical recruitment of myosin and myosin regulatory light chain suggesting that Polo plays a specific role in regulating actinomyosin rearrangements during mitosis (Pearson *et al.,* 2006). Extrusion of the LECs during histoblast proliferation is initiated by apical constriction for which myosin II contraction was shown to be necessary (Ninov *et al.,* 2007). Failure in extrude LECs results in the abnormal abdomen formation, with LECs still present in pharate adults (Ninov *et al.,* 2007). However, since in this study no specific GFP-Polo was seen in LECs, it seems unlikely that this explanation can justify the phenotype observed in flies expressing only the shorter *polo* mRNA (*gfp-polo⊿pA2;* $polo^9$).

Altogether, these results show that *polo* mRNA 3' end formation at the distal poly(A) site producing the 2.5 kb long transcript is specifically required in abdominal histoblast cells during the pupae stage/metamorphosis.

Chapter II

Identification and characterization of Upstream Sequence Elements in the vicinity of the polo proximal poly(A) site

1. Introduction

Alternative polyadenylation at the 3'most exon is a regulatory mechanism through which gene expression is regulated in a tissue and/or developmental specific manner, through the synthesis of different mRNAs with different stabilities and/or translational efficiencies (reviewed in Edwalds-Gilbert *et al.,* 1997). The recently reported *enhancer of rudimentary*, *e(r)*, is a good example that illustrates the regulatory action of alternative polyadenylation. The *Drosophila melanogaster e(r)* has two polyadenylation signals in the 3'most exon (Wojcik *et al.,* 1994; Gawande *et al.,* 2006). Recognition of these signals results in the synthesis of two mRNAs with the same coding region. However, the longer mRNA is specifically expressed in the female germline (Gawande *et al.,* 2006). This regulation is dependent on the expression of Sex-lethal (Sxl) (Gawande *et al.,* 2006). This protein, which has been shown to be a master-gene in the sex determination pathway in *Drosophila*, is expressed in the female germline (reviewed in Penalva and Sanchez 2003). Analysis of the RNA elements necessary for 3'end formation of the proximal *e(r)* poly(A) site, shows that Sxl competes with CstF-64 for the binding to the DSE near the proximal poly(A) site (Gawande *et al.,* 2006). As a result, in the female germline, the proximal site is inactivated resulting in the recognition of the distal poly(A) site. The specific requirement for the longer mRNA in the female germline seems to be coupled with downregulation of the levels of E(r) in this tissue, since the longer mRNA seems to be translated with a lower efficiency (Gawande *et al.,* 2006).

Another example of regulation via alternative polyadenylation at the 3'most exon is the *cyclooxygenase-2* (*cox-2*) gene. Two mRNAs have been identified for *cox-2* (Yokoyama and Tanabe 1989; Funk *et al.,* 1991; Hla and Neilson 1992). Northern blot analysis suggested that these mRNAs have different half-lives, with the longer *cox-2* mRNA being more unstable (Lukiw and Bazan 1997). Hall-Pogar *et al.* 2005 showed that the synthesis of both *cox-2* mRNAs seems to be tissue-specific and that the expression of the smaller mRNA is dependent of an Upstream Sequence Element (USE) near the proximal poly(A) site. This element, may act by allowing the assembly of the 3'end machinery or by serving as a target for tissue-specific *trans*-acting factors that promote recognition of the proximal poly(A) site. The authors also suggested that alternative polyadenylation might be a key step in regulating *cox-2* expression *in vivo*.

Work on CFI_m has shown this complex to be one of the main factors required for both canonical and non-canonical poly(A) site recognition (Brown and Gilmartin 2003; Venkataraman *et al.,* 2005). The levels of CFI_m seem to be important for regulation of polyadenylation, as low levels stimulate 3'end formation while high levels have an inhibitory action (Brown and Gilmartin 2003). Moreover, RNAi of the 25 kDa subunit of

CFI_m (CFI_m25), shows that this complex has a major role in alternative poly(A) site selection in the 3'most exon of the human *tissue inhibitor of metalloproteinase 2* (*timp-2*), *syndecan2*, *excision repair cross-complementing rodent repair deficiency, complementation group 6* (*ercc6*) and *dihydrofolate reductase* (*dhfr*) genes (Kubo *et al.,* 2006). Interestingly, three *cfl_m25* mRNAs of 4.6, 2.0 and 1.1 kb respectively, are synthesized as a result to three polyadenylation sites in the 3'most exon of *cfl_m25* (Kubo *et al.,* 2006). Northern blot analysis shows that expression of these transcripts is tissue-specific (Kubo *et al.,* 2006). While the 4.6 kb transcript is expressed in heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas, the 1.1 kb mRNA is highly expressed in the heart, liver and skeletal muscle. As for the 2.0 kb mRNA, it is expressed in low levels in the heart, liver, skeletal muscle, kidney and pancreas. Consequently, this adds a new layer of complexity and flexibility to the regulation of gene expression through alternative polyadenylation.

Our results obtained from the study of the different transgenic flies clearly show a developmental- and/or tissue-specific requirement for the longer 2.5 kb *polo* mRNA. This requirement does not seem to rely on different translational properties of each mRNA since translational assays suggest that both mRNAs have similar rates of translation (data not shown). It is possible that in abdominal histoblasts, alternative polyadenylation might ensure the synthesis of a more stable 2.5 kb mRNA by enhancing the usage of the distal poly(A) signal in a similar way as described for the *e(r)* and *cox-2*. Therefore, to understand the molecular mechanisms underlying the usage of the distal poly(A) site of *polo*, we searched the 3'UTR for putative regulatory elements necessary for mRNA 3'end formation.

2. Results

2.1. Identification of putative regulatory elements in the *polo* 3'UTR

Since important regulatory elements tend to be conserved throughout evolution, we searched for strong conserved elements present in the *polo* 3'UTR. To this end, the genomes of twelve *Drosophila* species together with the mosquito (*A. gambiae*), the honeybee (*A. mellifera*) and the beetle (*T. casteneum)* genomes were aligned using UCSC Genome Bioinformatics database (Karolchik *et al.,* 2003) (http://genome.ucsc.edu/). Using the *PhastCons* software (Siepel *et al.,* 2005) five conserved elements were identified (Fig. 18, 19 and Table V).

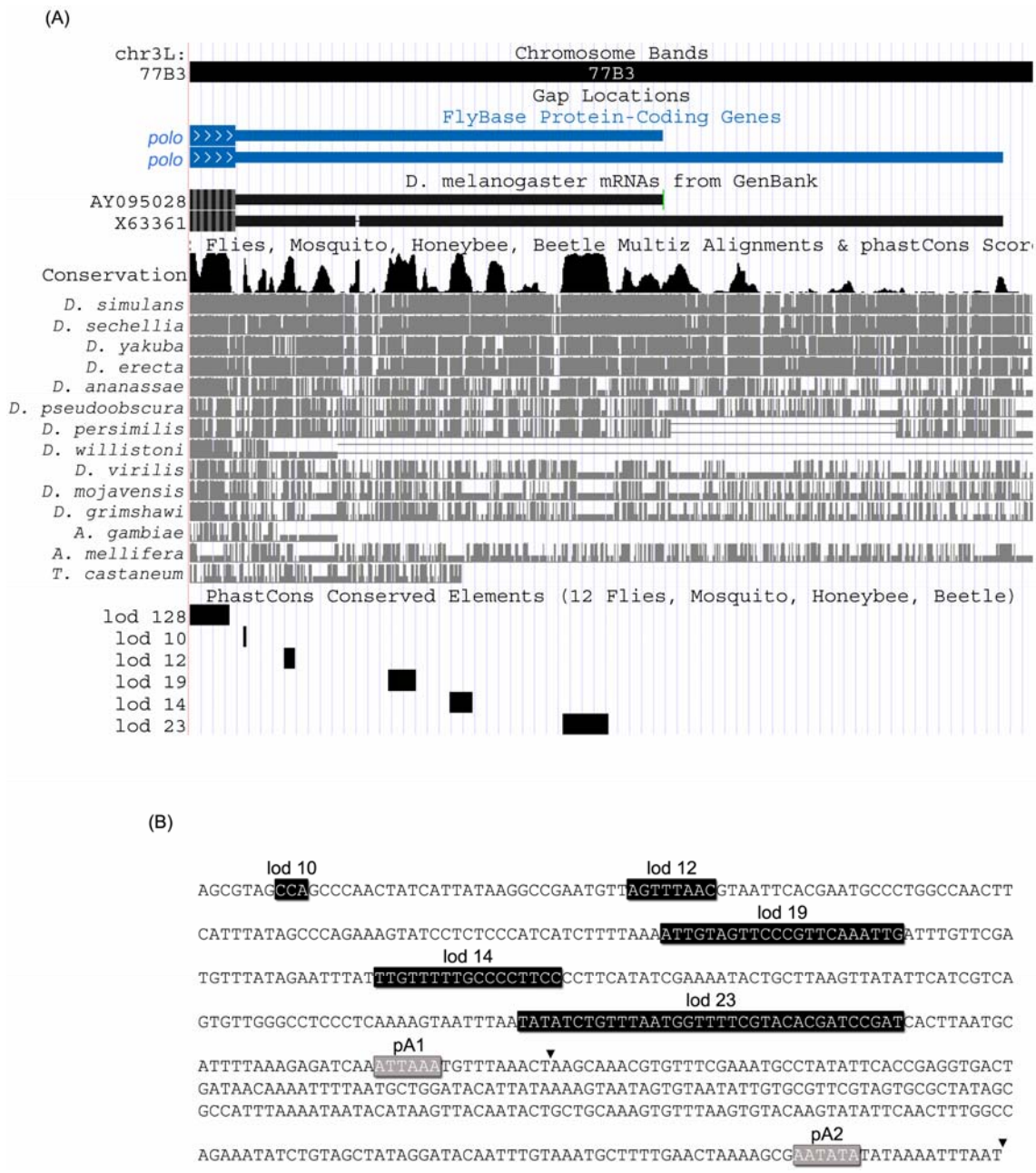

Figure 18. Conserved elements present in the *polo* 3'UTR. (A) The genomes of several *Drosophila* species, mosquito (*A.gambiae*), honeybee (*A.mellifera*) and the beetle (*T. castaneum*) were aligned using the UCSC Genome Browser software (http://genome.ucsc.edu/) and conserved elements searched using the *PhastCons* software. The chromosome bands indicate the position of the *polo* locus in the *D. melanogaster* genome. The blue and black bars underneath represent the two *polo* mRNAs deposited in the Flybase and GenBank databases (with the respective accession numbers) respectively. The grey bar attributed to each specie show the degree of homology between the *polo* gene of that specie and the *polo* from *D. melanogaster*. The black boxes on the bottom indicate the conserved elements identified accordingly to the *PhastCons* software. **(B)** Sequence and position of the *Phastcons* conserved elements in the *D. melanogaster polo* 3'UTR. The black boxes indicate the conserved elements (lod) and the grey boxes the proximal (pA1) and distal (pA2) poly(A) signals. The arrowheads indicate both cleavage sites.

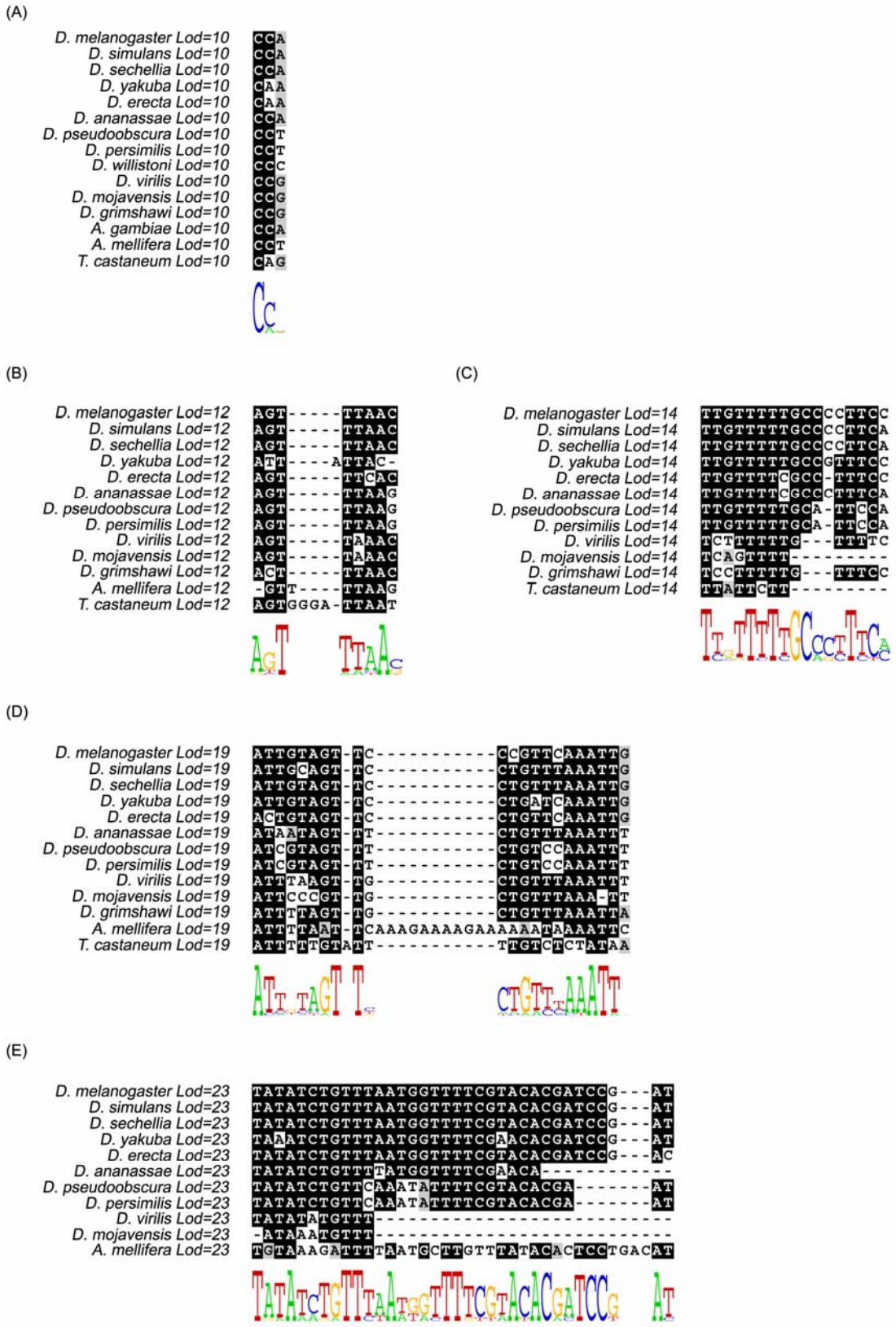

Figure 19. Conserved elements present in the *polo* 3'UTR. Conservation of the elements identified by the PhastCons software: **(A)** Lod 10, **(B)** Lod 12, **(C)** Lod 14, **(D)** Lod 19 and **(E)** Lod 23. The sequence alignment was obtained using the UCSC Genome Browser software (http://genome.ucsc.edu/). The sequence underneath each alignment represents the consensus sequence obtained with the Weblogo software (version 2.8.2) (http://weblogo.berkeley.edu/).

All five elements are located upstream the proximal poly(A) site. The two elements with the lowest conservation score, lod 10 and lod 12, are positioned 296 and 266 nt upstream the proximal poly(A) signal respectively, while the more conserved elements are positioned closer to the pA1, with the most conserved element, lod 23, just 27 nt from the poly(A) signal (Fig. 18B and Table V).

PhastCons conserved elements	Score	Lenght (nt)	Distance from pA1 (nt)
lod 10	144	3	296
lod 12	169	8	266
lod 14	190	17	132
lod 19	232	22	175
lod 23	258	35	27

Table V. *PhastCons* predicted conserved elements in the 3'UTR of *polo*.

Analysis of these elements shows that, with exception of lod 10, all elements are T-rich (Table VI). Moreover, the thymidylate content follows the conservation trend, *i.e.*, the higher the conservation, the higher the amount of thymine deoxynucleotides present in the elements (Table V and VI, exception made for lod 14).

PhastCons conserved elements	A	T	G	C
lod 10	33.3%	0%	0%	**66.7%**
lod 12	**37.5%**	**37.5%**	12.5%	12.5%
lod 14	0%	**52.9%**	11.8%	35.3%
lod 19	22.7%	**40.9%**	18.2%	18.2%
lod 23	22.9%	**42.9%**	17.1%	17.1%

Table VI. Nucleotide composition of the *PhastCons* conserved elements. The numbers in bold indicate the most represented nucleotide(s) in each element.

Thus, there seems to be a correlation between conservation, position towards the proximal poly(A) signal and the thymine deoxynucleotide content, with the more conserved elements positioned closer to the proximal poly(A) signal and more T-rich.

The three most conserved elements, lod 14, 19 and 23 are interesting, as not only their conservation suggests a functional role, but their nucleotide composition and position resembles an Upstream Sequence Element (USE). These poorly conserved but usually U-rich elements, have been characterized in virus and in cellular genes as elements required for efficient 3'end formation (Carswell and Alwine 1989; DeZazzo and Imperiale 1989; Russnak and Ganem 1990; Brown *et al.,* 1991; DeZazzo *et al.,* 1991; Russnak 1991; Sanfacon *et al.,* 1991; Valsamakis *et al.,* 1991; Cherrington and Ganem 1992; Rothnie *et al.,* 1994; Moreira *et al.,* 1995; Moreira *et al.,* 1998; Brackenridge and Proudfoot 2000; Natalizio *et al.,* 2002; Hall-Pogar *et al.,* 2005; Danckwardt *et al.,* 2007).

So far, USEs have only been described for a small number of cellular genes, but *in silico* studies of the human genome suggest the idea that these elements are more general (Legendre and Gautheret 2003). The study of 3'end signals shows that regions upstream the poly(A) signals are usually U-rich.

As described for the *IgM heavy chain* gene (Galli *et al.,* 1987; Peterson and Perry 1989; Peterson 1994; Takagaki *et al.,* 1996) and for *the enhancer of rudimentary* gene (Gawande *et al.,* 2006), RNA elements required for 3'end formation can have regulatory roles in gene expression. It is possible that the lod 14, 19 and 23 elements may function as USEs regulating the use of the proximal poly(A) site. The histoblasts dependence on the usage of the distal poly(A) site could result, in some extension, from down-regulation of the proximal poly(A) site. For these reasons, a detailed analysis of the role of these elements on the 3'end formation of the proximal poly(A) site was performed.

2.2. Mutation of lod 14/ USE 1 inhibits *in vitro* polyadenylation of the proximal poly(A) site

Of the three most conserved elements, the lod 14 element is particularly interesting. Although not the most conserved element, it is however, the most T-rich element (Table VI). In fact, when taking into account the vicinity of lod 14, an even richer region is observed (Fig. 20A). A strong thymine deoxynucleotide tract which includes lod 14 is conserved between the 25-30 million of year distant *D. melanogaster* and *D. persimilis* species. Moreover, the more conserved positions are the ones in which a thymine deoxynucleotide is present (Fig. 20A). Comparison of this region with known described USEs reveals some similarities with the SV40 late (Schek *et al.,* 1992), *cyclooxygenase-2* (*cox-2*) (Hall-Pogar *et al.,* 2005), *collagen* (Natalizio *et al.,* 2002) and *prothrombin* USEs (Danckwardt *et al.,* 2007). Therefore, we focused our attention in this extended sequence of lod 14 that in the following experiments will be referred as USE 1.

Figure 20. lod 14/USE 1 is a highly conserved element (see figure on next page). (A) Conservation of USE 1 in the *polo* 3'UTR, from *D. melanogaster* to *D. persimilis*. The top bar indicates the sequence of lod 14 in the USE 1. **(B)** Alignment of the USEs present in SV40 late, and in the *cyclooxygenase-2* (*cox-2*) and *collagen* genes. **(C)** Alignment of the 6-14 nt region of the *polo* USE 1, the SV40 late USE 2, the *cox-2* USE 3 and the *collagen* USE 2. **(D)** Alignment of the 2-14 nt region of the *polo* USE 1 and the *prothrombin* USE. The sequence underneath the alignments represents the consensus sequence obtained using the Weblogo software (version 2.8.2) (http://weblogo.berkeley.edu/). The sequences were aligned using the ClustalW software (http://www.ch.embnet.org/software/ClustalW .html).

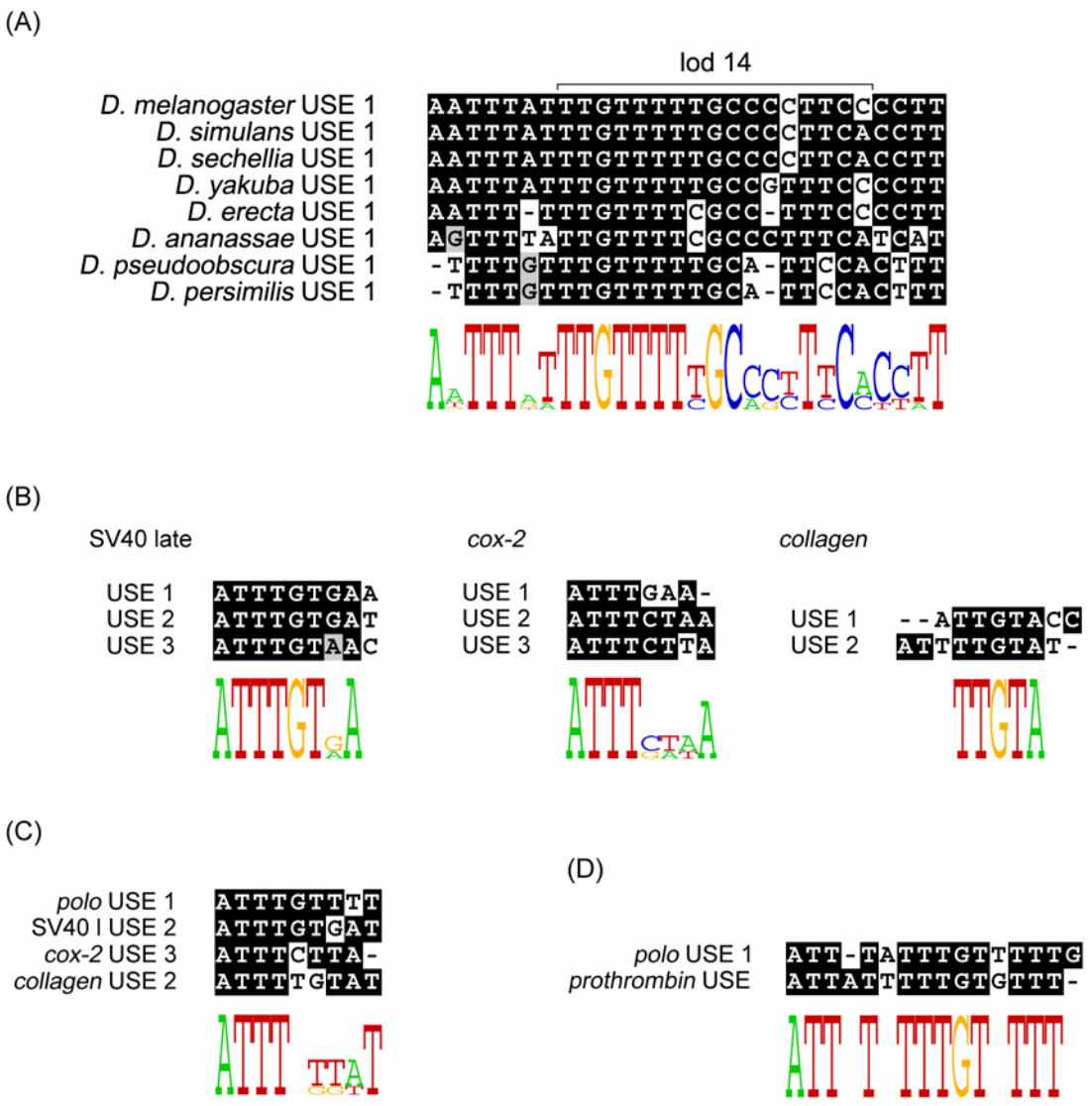

Figure 20. (see legend on previous page).

To study the function of the USE 1 in mRNA 3'end formation at the proximal poly(A) site, we cloned a 171 bp fragment from *polo* cDNA containing the USE 1 and the proximal poly(A) site. By PCR mutagenesis the USE 1 was mutated, changing the pyrimidines for purines, and purines for pyrimidines (Fig. 21A, B). Both constructs were then used to perform *in vitro* transcription in the presence of α-^{32}P-[UTP], to synthesize a radiolabelled pre-mRNA. After purification, the labelled transcripts were incubated with nuclear protein extracts to perform *in vitro* cleavage and polyadenylation assays. Although these reactions occur simultaneously *in vivo*, *in vitro* conditions allow us to separate both reactions (Moore and Sharp 1985). Using a full length pre-mRNA (with the poly(A) signal, the cleavage site and the downstream sequence) incubated with protein extracts in the presence of EDTA, we can assay for the cleavage reaction since EDTA sequesters the Mg^{2+} ions required for the function of the Poly(A) Polymerase (PAP). This way we ensure

that only cleavage of the pre-mRNA takes place. To assay for the polyadenylation reaction, pre-cleaved pre-mRNAs are used. These pre-mRNAs are synthesized from the plasmid DNA template, linearized with a restriction enzyme in a site close to the genuine cleavage site. This way the newly synthesized pre-mRNA will mimic a pre-cleaved pre-mRNA. The pre-mRNAs are then incubated with nuclear protein extracts in the presence of Mg^{2+} ions.

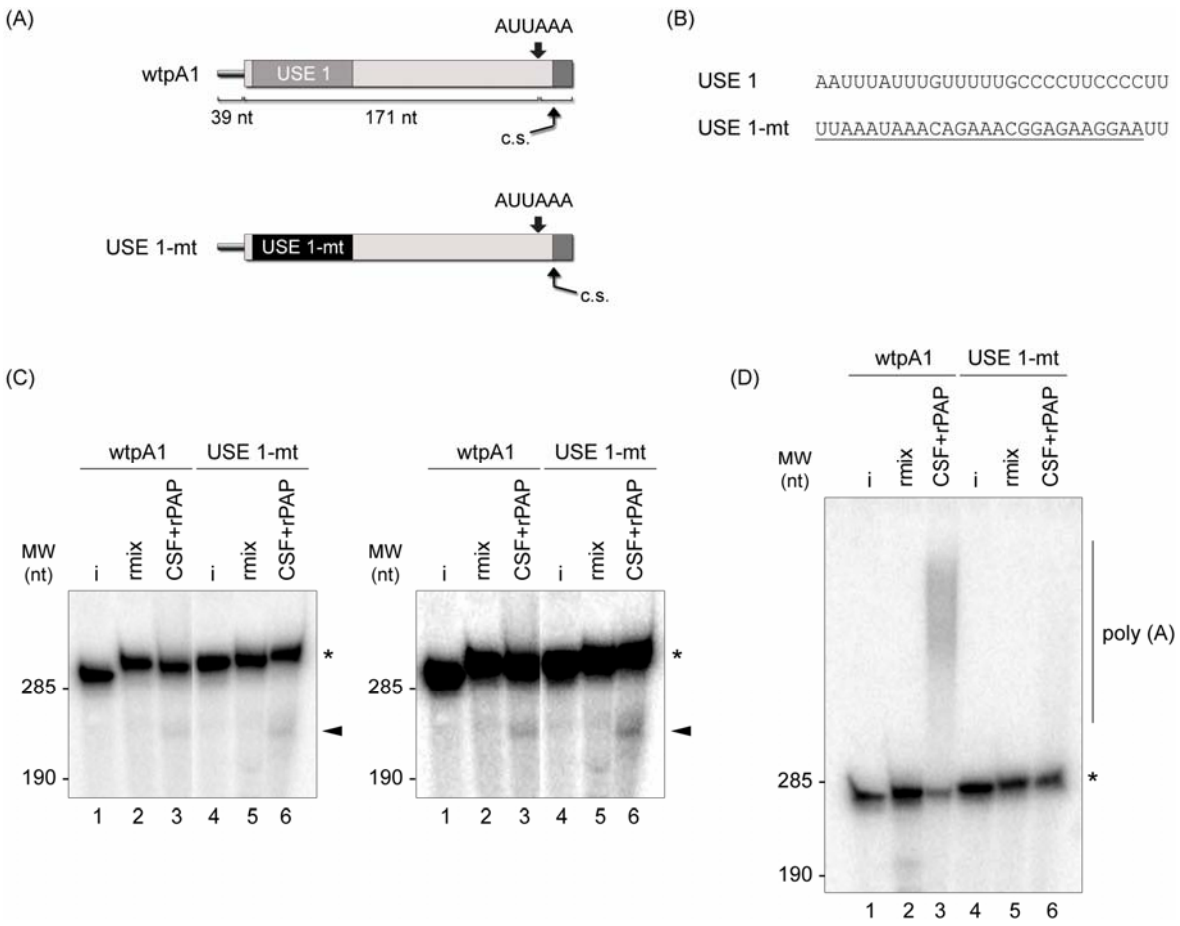

Figure 21. USE 1 is necessary for polyadenylation at the proximal poly(A) site. (A) Schematic representation of the wtpA1 and USE 1-mt pre-mRNAs used in the *in vitro* cleavage and polyadenylation assays. The light grey and black boxes represent the wild-type and mutated USE 1, respectively. The thick and thin black arrows indicate the proximal poly(A) signal and cleavage site (c.s), respectively. The dark grey box downstream the cleavage site represents a plasmid polylinker fragment used as downstream sequence in full length pre-mRNAs for the cleavage assays; in the pre-cleaved pre-mRNAs used in the polyadenylation assays this fragment was excluded by linearization of the plasmids close to the cleavage site. The thin bar at the beginning of the pre-mRNAs represents a 39 nt polylinker sequence that is included in the pre-mRNA **(B)** Point mutations in USE 1. The USE 1 and USE 1-mt sequences show the wild-type and mutated sequences present in the wtpA1 and USE 1-mt pre-mRNAs. **(C)** Cleavage assays with full-length wtpA1 (lanes 1-3) and USE 1-mt (lanes 4-6) pre-mRNAs using a Superose 6 gel filtration fraction prepared from HeLa cell nuclear extracts and recombinant Poly(A) Polymerase (CSF+PAP). The two gels represent different exposure times. Lanes 1 and 4 show the input full-length pre-mRNAs

and lanes 2 and 5, the incubation of the input full-length RNAs with all the reagents with exception of the CSF+PAP protein fraction. Lanes 3 and 6 show the incubation of the input full-length pre-mRNAs with all the reagents and the CSF+PAP protein fraction. The asterisk and the arrowhead indicate the full-length pre-mRNAs and the upstream cleavage product, respectively. **(D)** Polyadenylation assays with pre-cleaved wtpA1 (lanes 1-3) and USE 1-mt RNAs (lanes 4-6) using the CSF+PAP protein fraction. Lanes 1 and 4 show the input pre-cleaved pre-mRNAs and lanes 2 and 5, the incubation of the input pre-cleaved pre-mRNAs with all the reagents with exception of the CSF+PAP protein fraction. Lanes 3 and 6 show the incubation of the input pre-cleaved RNAs with all the reagents plus the CSF+PAP protein fraction. The asterisk represents the input pre-cleaved pre-mRNAs and the vertical bar the polyadenylated RNA.

Since mammals and *Drosophila* mRNA processing protein factors share a very high degree of homology, both structurally and functionally (Benoit *et al.,* 1999; Mount and Salz 2000; Murata *et al.,* 2001; Benoit *et al.,* 2002), we performed these experiments using a Superose 6 gel filtration fraction prepared from HeLa cell nuclear extracts (CSF) and recombinant PAP, both kindly provided by Dr. Yoshio Takagaki and Prof. James Manley (Columbia University, New York) (Takagaki *et al.,* 1988),

As seen from the cleavage assay (Fig. 21C), mutation of USE 1 has no effect on the cleavage reaction since no significant changes are observed between USE 1-mt and wtpA1 (Fig. 21C; compare lanes 3 and 6).

When polyadenylation reaction is performed, poly(A) addition to the USE 1-mt transcript is inhibited (Fig. 21D, compare lanes 3 and 6). Therefore, these results indicate that the conserved USE 1 has a specific role in the polyadenylation at the proximal poly(A) site.

A short GUGUUU element positioned 9 nt downstream of the cleavage site closely resembles a putative *Drosophila* Downstream Sequence Element (DSE) (Graber *et al.,* 1999) (Fig. 18B). Present in most genes, this element is involved in the cleavage of the pre-mRNA, through binding of CstF (reviewed in Zhao *et al.,* 1999). However, in our transcripts, this element is not present. Instead, the sequence downstream the +11 position (+ meaning position downstream the cleavage site) was replaced by a 32 nt plasmid polylinker sequence in which no putative DSE is present. Therefore, it seems that the putative GUGUUU DSE is not necessary for the cleavage reaction at the *polo* proximal poly(A) site.

2.3. USE 1 is required for the assembly of a protein complex

Since USE 1 is required for polyadenylation at the proximal poly(A) site, we investigated the RNA-binding proteins that could bind to this element. Pre-cleaved pre-

mRNAs were synthesized *in vitro,* double labelled with α-^{32}P-[ATP] and α-^{32}P-[UTP] and used to perform UV. cross-linking assays (Fig. 22). In these experiments, pre-mRNAs are incubated with HeLa cell nuclear extracts in conditions that allow the formation of RNA-protein complexes. Subsequently, the reaction mixture is irradiated with UV. light. Under these conditions proteins bound to RNA are covalently linked to the radiolabelled pre-mRNA. Following addition of RNase A, the pre-mRNA is degraded, except for the regions that are protected by the proteins covalently bound. After SDS-PAGE electrophoresis, the proteins bound to the pre-mRNA are visible in an autoradiography film as a clear distinct band.

Our analysis showed that several proteins can assemble onto the wtpA1 pre-mRNA (Fig. 22B, lane 1). Moreover, mutation of the USE 1 decreases the cross-link of three proteins bands with molecular weights of approximately 35, 40 and 55 kDa (Fig. 22B, lane 2). While a mild reduction of the cross-link is observed for the 55 kDa protein, a strong decrease is seen for the 35 and 40 kDa protein bands respectively. A strong 60 kDa protein also binds to the wtpA1 pre-mRNA. However, contrary to what is seen for the other proteins, it displays an increased binding to the USE 1-mt pre-mRNA (Fig. 22B, lane 2).

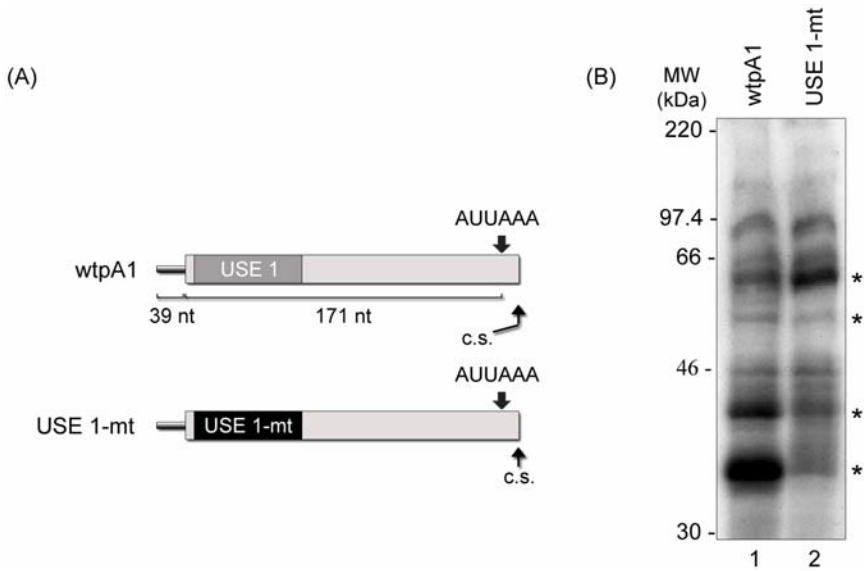

Figure 22. A protein complex assembles onto the wtpA1 pre-cleaved pre-mRNA. (A) Schematic representation of the wtpA1 and USE 1-mt pre-cleaved pre-mRNAs used in UV. cross-linking assays. The light grey and black boxes represent the wild-type and mutated USE 1, respectively. The thick and thin black arrows indicate the proximal poly(A) signal and cleavage site (c.s), respectively. The thin bar at the beginning of the pre-mRNAs represents 39 nt of polylinker sequence that is included in the pre-mRNA. **(B)** UV. cross-linking assays of wtpA1 (lane 1) and USE 1-mt (lane 2) pre-cleaved pre-mRNAs using HeLa cell nuclear extracts. The asterisks indicate the protein bands whose binding appears to be changed upon mutation of USE 1.

To identify these proteins we performed immunoprecipitation assays after UV. cross-linking, using antibodies against possible candidates based on the molecular weight of the protein bands and potential consensus binding sites present in USE 1 (Fig. 23). The molecular weight of the 60 kDa protein band suggested that this protein could be the 64 kDa subunit of the Cleavage and stimulation Factor (CstF) complex. Thus, using a monoclonal antibody against CstF-64 we tried to immunoprecipitate the 60 kDa protein band.

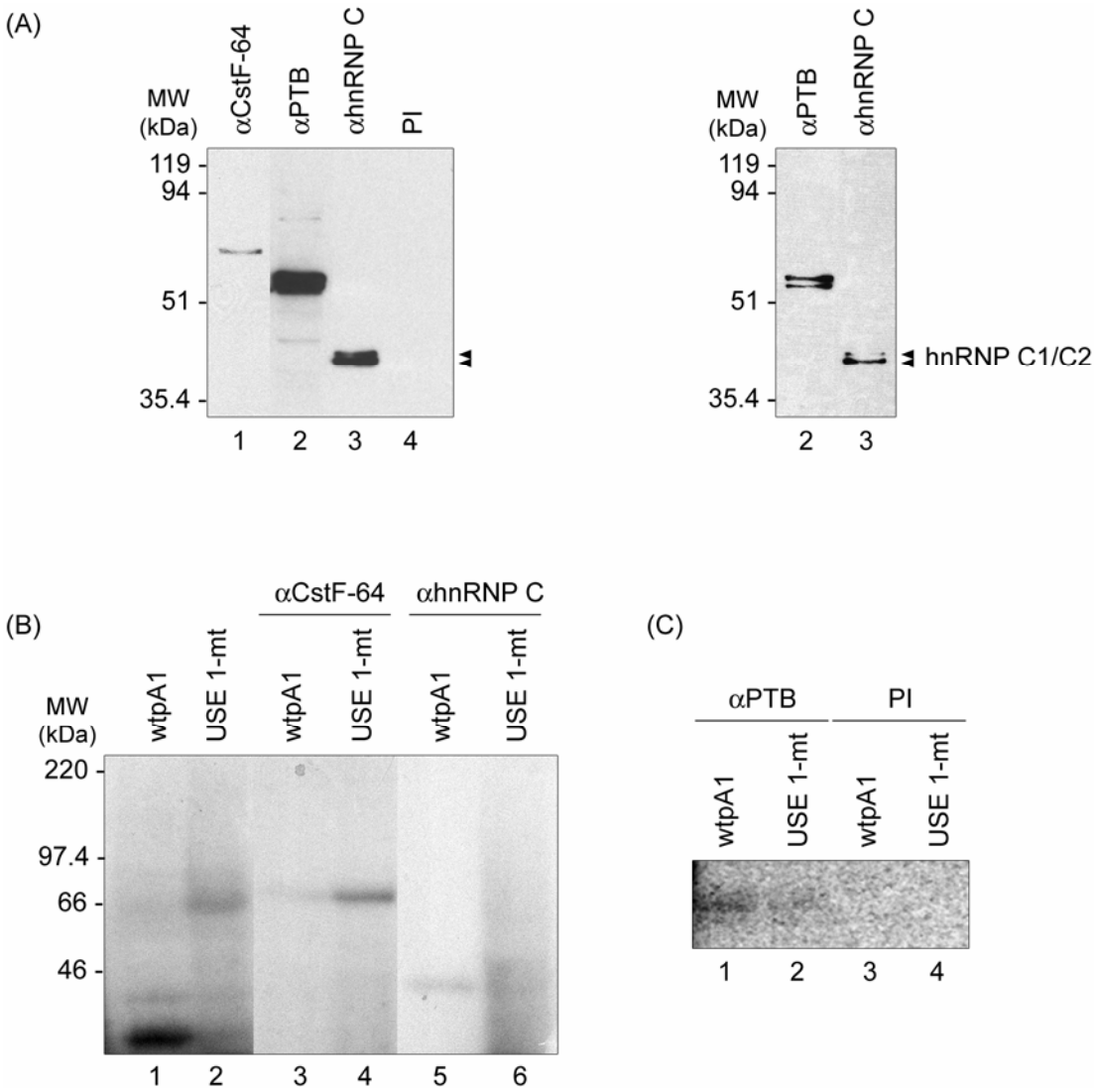

Figure 23. CstF-64, hnRNP C and PTB bind in a USE 1-dependent manner to the wtpA1 pre-mRNA. (A) Western blot against CstF-64, hnRNP C and PTB using HeLa cell nuclear extracts. For detection of CstF-64 and hnRNP C, specific monoclonal antibodies were used (αCstF-64 and αhnRNP C) while for detection of PTB, a polyclonal antibody (αPTB) and a pre-immune (PI) serum were used. The right panel shows a lower exposure time of the left panel. **(B, C)** Immunoprecipitations performed after UV. cross-linking with wtpA1 and USE 1-mt pre-cleaved pre-mRNAs. Immunoprecipitations were performed using the CstF-64 (lanes 3 and 4) and hnRNP C (lanes 5 and 6) monoclonal antibodies and the PTB polyclonal antibody (**C**, lanes 1 and 2). As a

control for the immunoprecipitation with the PTB polyclonal antibody, the pre-immune serum was used (**C**, lanes 3 and 4).

As seen in Fig. 23B, lanes 1-4, the 60 kDa protein band is precipitated using the CstF-64 antibody. Thus, CstF-64 binds upstream the proximal poly(A) site of the wtpA1 pre-mRNA.

Two well characterized protein binding sites are present in USE 1: the UUUUU (Wilusz and Shenk 1990; Gorlach *et al.,* 1994) and the AUUUA elements (Hamilton *et al.,* 1993) that constitute the binding sites for hnRNP C, and sequences similar to the two consensus binding sites determined for PTB: (U/G)C(A/Y)GCCUG(Y/G)UGCYYYYCYYY YG(Y/G)CCC (Singh *et al.,* 1995) and UCUU(C) (Perez *et al.,* 1997). Moreover, the molecular weight of hnRNP C and PTB is in agreement with the 40 and 55 kDa protein bands identified in Fig. 22B. Therefore, hnRNP C and PTB antibodies were used to immunoprecipitate the 40 and 55 kDa protein bands, respectively. Both antibodies, which are able to recognize specifically hnRNP C and PTB in nuclear extracts (Fig. 23A, lane 2-4), precipitated both 40 and 55 kDa protein bands identifying these proteins as hnRNP C and PTB, respectively (Fig. 23B, lanes 1, 2, 5, 6 and Fig. 23C). The binding of PTB to the wtpA1 pre-mRNA is also observed when UV. cross-linking assays are performed using recombinant human PTB1 (Fig. 24). PTB1 is able to bind to both wtpA1 and USE 1-mt pre-mRNAs in a similar manner as seen when HeLa nuclear extracts are used (Fig. 24, compare lanes 1-4).

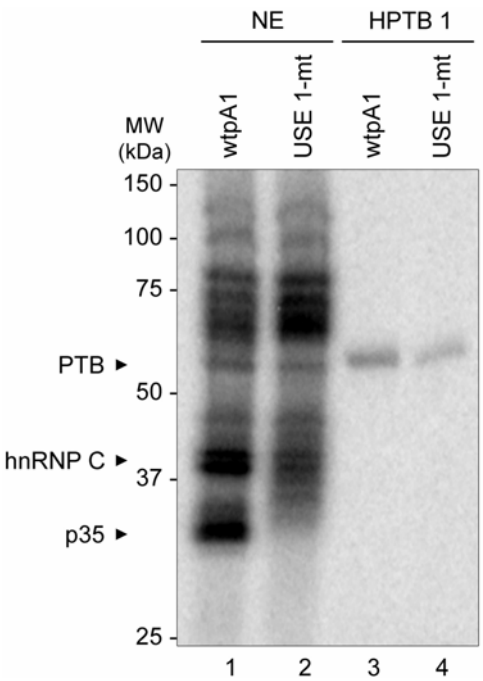

Figure 24. Recombinant PTB 1 binds to wtpA1 pre-cleaved pre-mRNA. UV. cross-linking assays were performed using wtpA1 (lanes 1 and 3) and USE 1-mt (lanes 2 and 4) pre-cleaved pre-mRNAs with HeLa nuclear extracts (NE) (lanes 1 and 2) and 50 ng of recombinant human PTB 1 (lanes 3 and 4).

The SV40 late USE has three highly similar elements: AUUUGUGAA, AUUUGUGAU and AUUUGUAAC, necessary for efficient 3' end formation of the SV40 late pre-mRNA (Carswell and Alwine 1989; Schek *et al.,* 1992) (Fig. 20B). The 31 kDa U1A protein has been shown to bind these elements and stimulate *in vitro* polyadenylation of the SV40 late RNA through interactions with CPSF-160 (Lutz and Alwine 1994; Lutz *et al.,* 1996). The highly similar AUUUGUUUU element present in USE 1 (Fig. 20C) raised the possibility of U1A being the 35 kDa protein band. Thus, we attempted to precipitate this protein using an U1A polyclonal antibody (Fig. 25).

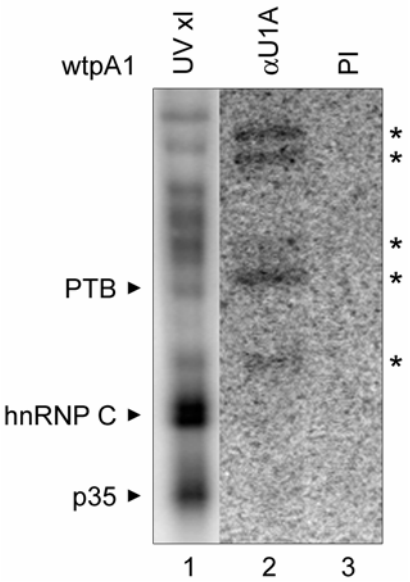

Figure 25. U1A is part of the protein complex assembled onto the wtpA1 pre-mRNA. A UV. cross-linking (UV xl) reaction with the wtpA1 pre-cleaved pre-mRNA using HeLa cell nuclear extracts was performed (lane 1) followed by immunoprecipitation of U1A using the polyclonal U1A antibody (lane 2); as a non-specific control an immunoprecipitation was performed using a pre-immune serum (PI) (lane 3). The asterisks indicate the immunoprecipitated protein bands. All lanes are from the same gel; since the immunoprecipitation signals from lane 2 and 3 were very weak, the signals were equally improved during imaging analysis.

As a result we are able to precipitate five other proteins with molecular weights of approximately 45, 55, 65, 100 and 120 kDa (Fig. 25, lane 2). However, the 35 kDa protein band is not present in our immunoprecipitations suggesting that the U1A protein is not the 35 kDa protein band. These results show that although U1A is not the 35 kDa protein band that is interacting with the wtpA1 pre-mRNA, it is nevertheless present in the protein complex assembled onto the transcript, interacting with other RNA-binding proteins.

The α-Complex Proteins (αCP) are members of the poly(C)-binding proteins (PCBPs) (reviewed in Makeyev and Liebhaber 2002). These RNA-binding proteins have a high affinity for poly(C) sequences and have been shown to play multiple functions in gene expression (reviewed in Makeyev and Liebhaber 2002). Observation of the USE 1 shows a strong C-tract interspersed by two uracil nucleotides (CCCCUUCCCC) (Fig. 20A). Moreover, this tract is contained within the consensus sequence described for the 35 kDa αCP2KL protein: $(U/A)_{2}$-$C_{3-5}(U/A)_{2-6}C_{3-5}(U/A)_{2-6}$-$C_{3-5}(U/A)_{2}$ (Thisted *et al.,* 2001). To determine if a PCBP is present in the protein complex assembled upstream the proximal poly(A) signal, we performed UV. cross-linking competition assays in which increasing

amounts of unlabelled poly(rC) oligoribonucleotide was added to the reaction. It is possible to see that addition of poly(rC) oligoribonucleotide decreases the binding of hnRNP C (Fig. 26A, compare lanes 1, 2-4 with 5-7). Moreover, increasing amounts of the oligoribonucleotide also prevented the binding of the 35 kDa protein.

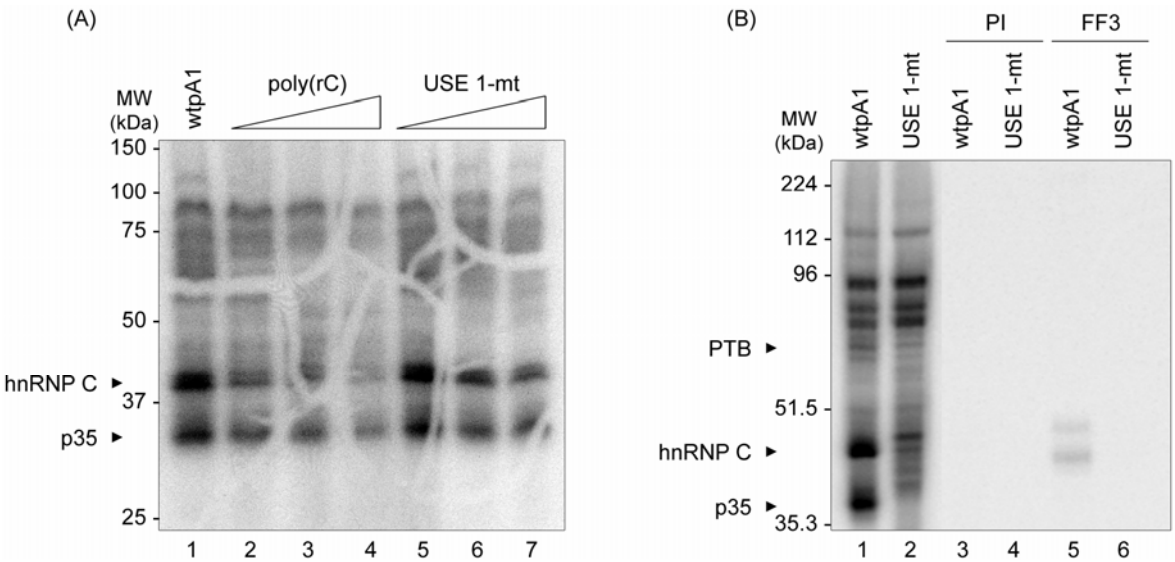

Figure 26. αCP2/2KL is part of the protein complex assembled onto the wtpA1 pre-mRNA. (A) UV. cross-linking competition assay using wtpA1 pre-cleaved pre-mRNA and HeLa cell nuclear extracts was performed with 0, (lane 1), 1 (lanes 2 and 5), 50 (lanes 3 and 6) and 150 (lanes 4 and 7) pmoles of a poly(rC) oligoribonucleotide (lanes 2-4) or with an oligonucleotide containing the USE 1-mt sequence as a non-specific control (lanes 5-7). **(B)** UV. cross-linking reactions with wtpA1 (lane 1) and USE 1-mt (lane 2) pre-cleaved pre-mRNAs were followed by immunoprecipitations with the FF3 polyclonal antibody (lanes 5 and 6) or with a pre-immune serum (PI) as a unspecific control (lanes 3 and 4).

Thus, we attempted to immunoprecipitate the 35 kDa protein band using the FF3 polyclonal antibody that specifically recognizes the αCP2 (38,5 kDa) and αCP2KL (35 kDa) isoforms (Chkheidze *et al.,* 1999). Two bands with molecular weights of approximately 40 kDa are precipitated in a USE 1-dependent manner (Fig. 26B, compare lanes 1 and 2 with 5 and 6). These results suggest that αCP2/2KL is not the 35 kDa protein band, but is however present in the protein complex that assembles onto the wtpA1 pre-mRNA, most likely via interactions with hnRNP C and p35.

In conclusion, a protein complex containing a 35 kDa protein (henceforward designated as p35), hnRNP C, PTB, CstF-64, U1A and αCP2/2KL assembles onto the wtpA1 pre-mRNA. Moreover, p35, hnRNP C, αCP2/2KL and PTB bind to the wtpA1 pre-mRNA in an USE 1-dependent manner.

2.4. USE 2 and USE 3 are necessary for efficient *in vitro* polyadenylation at the proximal poly(A) site

We identified a 16 nt conserved element that displays a high similarity with USE 1 (Fig. 27A, B, C). This GU-rich element (37.5% G, 31.3% U) is positioned 35 nt downstream the USE 1 and shares 50% identity with the USE (Fig. 27 A, C). Therefore, this element, designated as USE 2, was studied in order to determine its role on mRNA 3'end formation at the proximal poly(A) site. Moreover, of the several conserved elements predicted by the *PhastCons* software, lod 23 is the element with the highest conservation score (Fig. 19E, 27A and Table V). Its conservation, base composition and position toward the proximal poly(A) signal suggest that this element may also function as an USE. The alignment of the USE 1 and lod 23 emphasizes this idea (Fig. 27C). There is a high homology between USE 1 and the TCTGTTTAATGGTTTTCGT element present in lod 23 (58% identity). This thymine deoxynucleotide rich element accounts for most of the T content of lod 23 (73%). Thus, this element was designated USE 3 and studied together with USE 2.

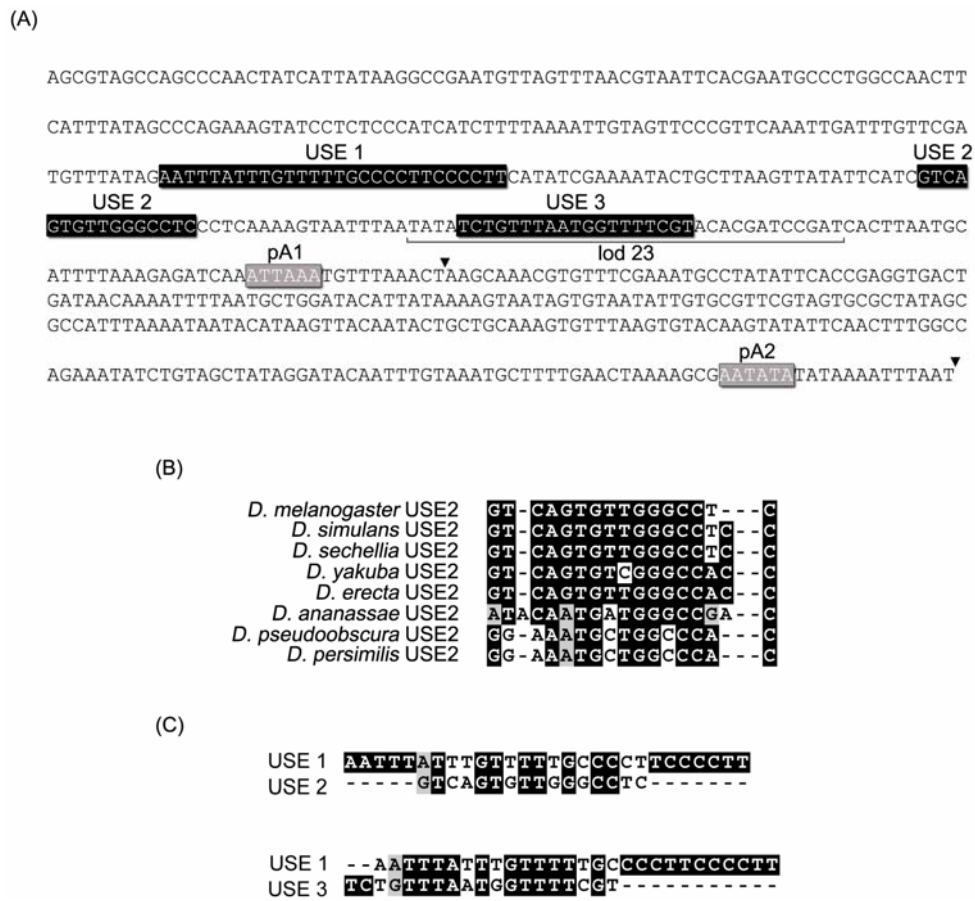

Figure 27. USE 2 and USE 3 share similarity with USE 1. (A) *polo* 3'UTR sequence. The black boxes indicate the USE 1, USE 2 and USE 3. The grey boxes indicate the proximal (pA1) and

distal (pA2) poly(A) signals. The sequence in brackets represents the *PhastCons* predicted conserved element lod 23.The arrowheads indicate the cleavage sites. **(B)** Conservation of USE 2 in the *polo* 3'UTR from *D. melanogaster* to *D. persimilis*. **(C)** Alignment of USE 1 with USE 2 and USE 3. All alignments were performed using the ClustalW software (http://www.ch.embnet.org/software/ClustalW.html).

To ascertain if both USE 2 and USE 3 were in fact upstream sequence elements involved in the mRNA 3'end processing at the proximal poly(A) site, polyadenylation competition assays were performed, using the wtpA1 pre-mRNA (Fig. 28B). From these experiments, it is possible to see that increase amounts of unlabelled USE 2 or USE 3 oligonucleotides results in a decrease of polyadenylation suggesting that both USE 2 and USE 3 are necessary for efficient mRNA 3'end processing at the proximal poly(A) site (Fig. 28B, compare lane 2 with lanes 3-5 and lanes 6 and 7).

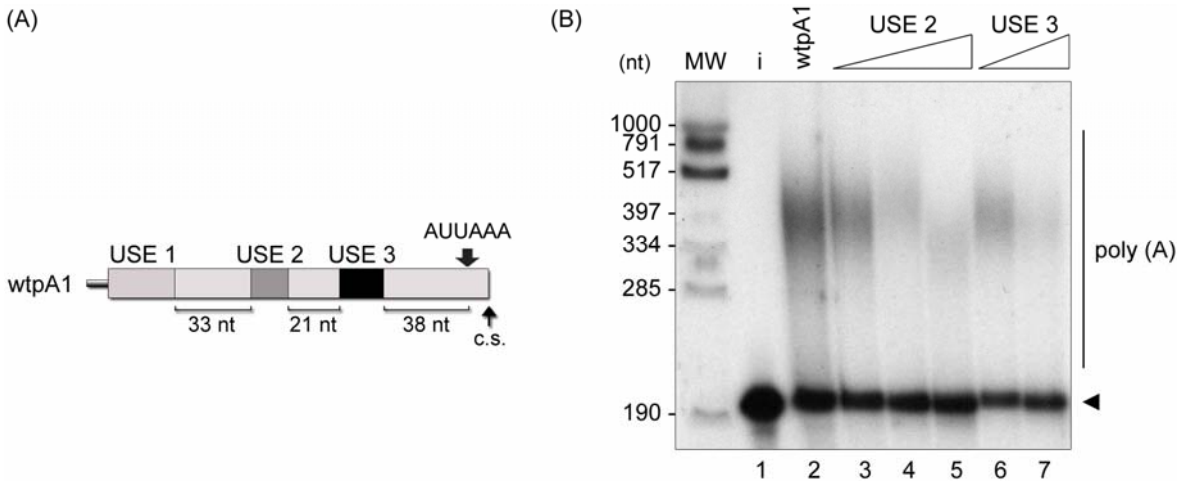

Figure 28. USE 2 and USE 3 are required for efficient polyadenylation at the proximal poly(A) site. (A) Schematic representation of the wtpA1 pre-cleaved pre-mRNA. The light grey, dark grey and black boxes represent the USE 1, USE 2 and USE 3, respectively. The thick and thin black arrows indicate the poly(A) signal (AUUAAA) and the cleavage site (c.s), respectively. The thin bar at the beginning of the pre-mRNA represents the 39 nt polylinker sequence included in the pre-mRNA. The spacers and the numbers underneath the wtpA1 pre-mRNA indicate the distance between the each USE and between USE 3 and the proximal poly(A) signal. **(B)** Polyadenylation competition assay using wtpA1 pre-cleaved pre-mRNA with CSF+rPAP were performed with 0, (lane 2), 1 (lanes 3 and 6), 5 (lanes 4 and 7) and 50 (lane 5) pmoles of unlabelled USE 2 (lanes 3-5) or USE 3 (lanes 6 and 7) oligonucleotides. i (lane 1) indicates the input wtpA1 pre-cleaved pre-mRNA. The arrowhead indicates the input pre-cleaved pre-mRNA and the vertical bar the polyadenylated pre-mRNA.

To investigate which proteins bind to these elements, we generated three constructs where a 29, a 16 and a 19 nt oligonucleotide with the USE 1, USE 2 and USE 3 sequences, respectively, was subcloned. These constructs were used as templates to perform *in vitro* transcription in order to synthesize the USE 1, USE 2 and USE 3 RNAs

and to proceed with UV. cross-linking assays (Fig. 29A). As seen in Fig. 29A, hnRNP C and PTB bind to USE 1 and more weakly to USE 3 (lanes 1, 2 and 4) while p35 binds specifically to USE 1 (lane 1 and 2). As for USE 2, we do not observe any specific binding to this element (compare lanes 1, 3 and 5). As an unspecific control, RNA transcribed from a linearized pGEM7 polylinker was used (lane 5).

These results are supported by UV. cross-linking competition assays (Fig. 29B). Addition of 50 pmoles of unlabelled USE 1 oligonucleotide completely prevents the binding of p35, hnRNP C and PTB to the wtpA1 pre-mRNA (Fig. 29B, lanes 1-3). The same effect, although not as strong, is observed when 150 pmoles of unlabelled USE 3 is used (Fig. 29B, lanes 8-11). The decrease in intensity of the p35 protein band when 150 pmoles of competitor USE 3 are used is somewhat unexpected as UV. cross-linking assays show that only hnRNP C and PTB can bind to USE 3 (Fig. 29A, lanes 1 and 4).

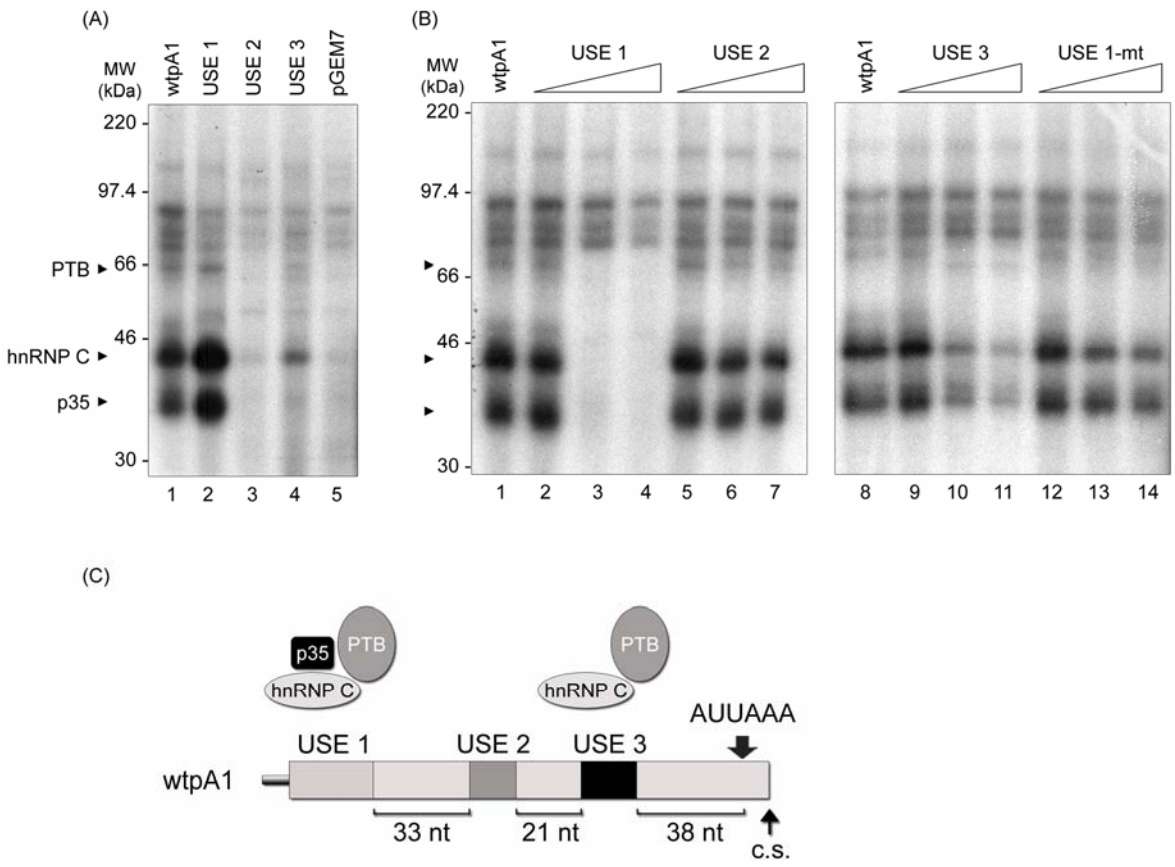

Figure 29. p35 binds specifically to USE 1. (A) UV. cross-linking reactions with HeLa cell nuclear extracts were performed using short RNAs containing the USE 1 (lane 2), USE 2 (lane 3) or USE 3 (lane 4) sequences. As positive and negative controls, UV. cross-linking reactions were performed with wtpA1 pre-cleaved pre-mRNA (lane 1) and with RNA transcribed from linearized pGEM7 polylinker (lane 5), respectively. **(B)** UV. cross-linking competition assay using wtpA1 pre-cleaved pre-mRNA and HeLa cell nuclear extracts was performed with 0, (lane 1 and 8), 1 (lanes 2, 5, 9 and 12), 50 (lanes 3, 6, 10 and 13) and 150 pmoles (lanes 4, 7, 11 and 14) of unlabelled USE 1 (lanes 2-4), USE 2 (lanes 5-7) and USE 3 (lanes 9-11) oligonucleotides. As a non-specific control,

an oligonucleotide containing the USE 1-mt sequence was used (lanes 12-14). **(C)** Schematic representation of the binding sites for p35, hnRNP C and PTB in the wtpA1 pre-cleaved pre-mRNA based on the results from the UV. cross-linking reactions in (A) and (B). The light grey, dark grey and black boxes, represent the USE 1, USE 2 and USE 3, respectively. The thick and thin black arrows indicate the poly(A) signal (AUUAAA) and the cleavage site (c.s), respectively. The thin bar at the beginning of the pre-mRNA represents a region of 39 nt of the polylinker sequence that is included in the pre-mRNA. The spacers and the numbers underneath the wtpA1 pre-mRNA indicate the distance between each USE and between USE 3 and the proximal poly(A) signal.

Increasing amounts of unlabelled USE 2 shows no effect in the wtpA1 cross-linking pattern (Fig. 29B, compare lane 1 with lanes 5-7), which is in agreement with the observation that no specific protein can bind to USE 2 (Fig. 29A, compare lanes 1, 3 and 5). As an unspecific control, competition assays were performed with oligonucleotides containing the USE 1 mutation (Fig. 29B, compare lane 8 with lanes 12-14).
Altogether, these results suggest that USE 1, USE 2 and USE 3 constitute a tripartite upstream sequence element.

2.5. Different protein complexes are assembled in the vicinity of each poly(A) site

The data obtained from the different experiments described above, suggest that the proteins bound to USE 1 and USE 3 play a specific role in mRNA 3'end formation at the proximal poly(A) site. Nevertheless, it is possible that these proteins may have a more general effect, by having also a role in the distal poly(A) site. To determine which proteins can act as potential specific regulators of the proximal site, the protein complexes assembled upstream both poly(A) sites were investigated. Several constructs were generated in order to synthesize pre-mRNAs with both poly(A) sites or only with the proximal or distal site (Fig. 30A). After *in vitro* transcription, UV. cross-linking assays using pre-cleaved pre-mRNAs were performed. As shown in Fig. 30B, some of the proteins, including hnRNP C and PTB, are able to bind upstream both poly(A) signals (lanes 1-3 and 5). However, p35 binds only upstream the proximal poly(A) site (compare lanes 3 and 5). Since the mutation of the USE 1 inhibits polyadenylation of the wtpA1 pre-cleaved pre-mRNA (see above), the highly specific binding of p35 to USE 1 suggests that it might play a regulatory role in mRNA 3'end processing of the proximal poly(A) site.

We also used a wtpA1-DSE pre-mRNA that has the putative GUGUUU DSE (Fig. 30A). It is possible to see that the UV. cross-linking protein binding pattern between the wtpA1 and wtpA1-DSE transcripts shows no differences (Fig. 30B, lanes 3 and 4), supporting our previous observation that the putative DSE is not essential for mRNA 3'end formation at the proximal site.

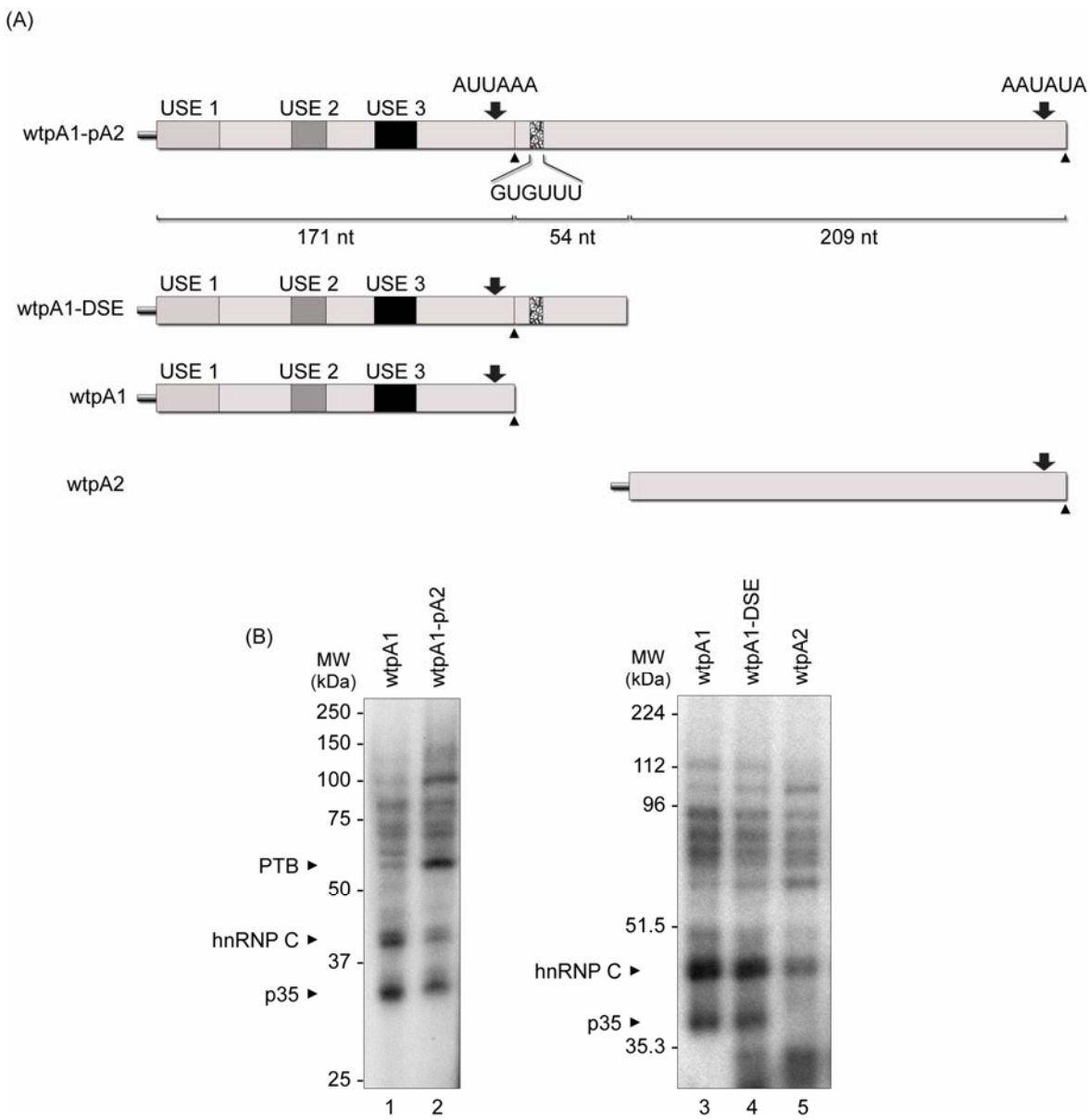

Figure 30. p35 binds specifically upstream the proximal poly(A) site. (A) Schematic representation of the wtpA1-pA2, wtpA1, wtpA1-DSE and wtpA2 pre-mRNAs used in the UV. cross-linking assays. The light grey, dark grey and black boxes, represent the USE 1, USE 2 and USE 3, respectively. The hatched box represents the putative DSE, downstream the proximal cleavage site. The thick arrows and the black arrowheads indicate the poly(A) signals and the cleavage sites, respectively. The thin bar at the beginning of the pre-mRNAs represents the 39 nt polylinker sequence included in the pre-mRNAs. **(B)** UV. cross-linking assays were performed using wtpA1 (lanes 1 and 3), wtpA1-pA2 (lane 2), wtpA1-DSE (lane 4) and wtpA2 (lane 5) pre-cleaved pre-mRNAs, with HeLa cell nuclear extracts.

2.6. Mapping the p35 binding site in the USE 1

Our results show a high specificity of p35 towards USE 1 with a possible regulatory function in polyadenylation at the proximal poly(A) site. Thus, to further characterize the potential regulatory elements present in the USE 1, we accurately mapped the p35 binding site. For this purpose we designed three constructs in which different segments of the USE 1 were mutated (Fig. 31A, B): the USE 1-mt (in which a new mutation of the USE 1, underlined in Fig. 31B, was introduced), the USE 1-5'mt and USE 1-3'mt. In the last two constructs, point mutations were introduced in the 5' or 3' region of the USE 1, respectively (Fig. 31A, B). Pre-cleaved pre-mRNAs were synthesized by *in vitro* transcription and UV. cross-linking assays were performed.

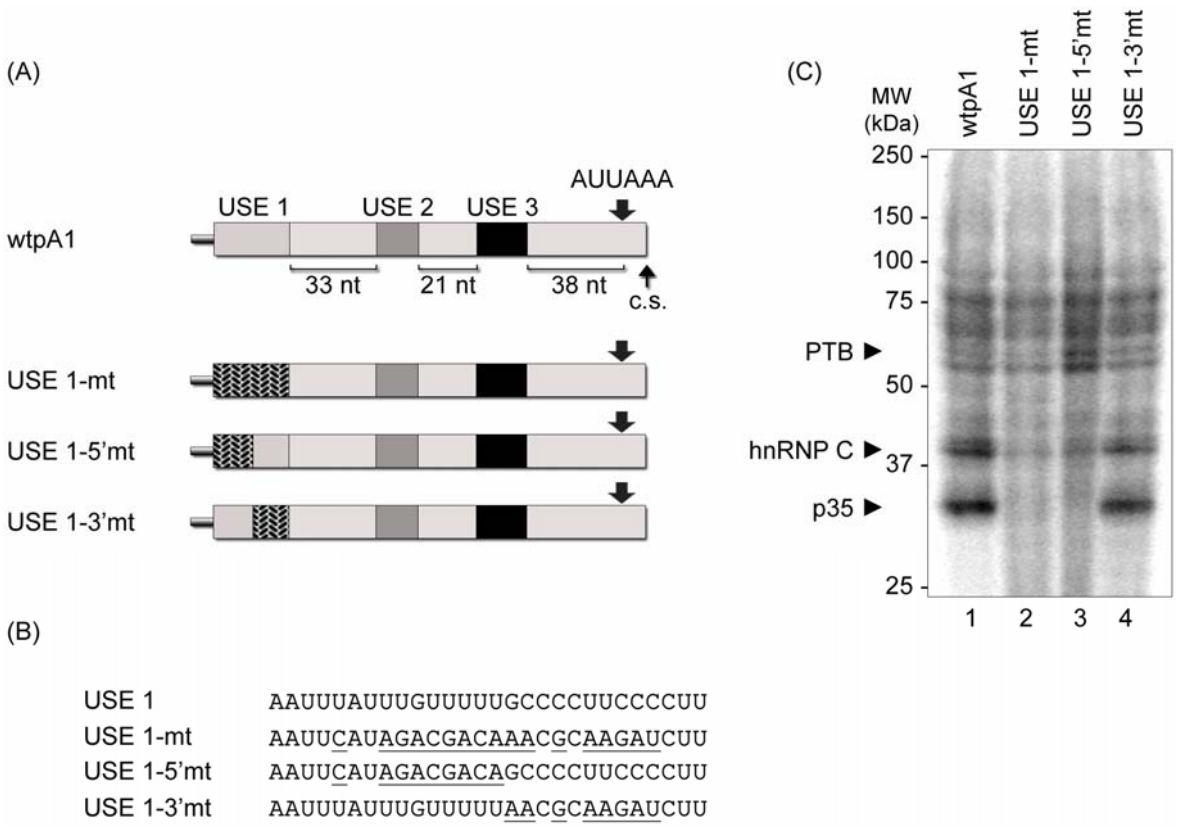

Figure 31. The UAUUUGUUUUU element in the USE 1 constitutes the binding site for p35 and hnRNP C. (A) Schematic representation of the USE 1, USE 1-mt, USE 1-5'mt and USE 1-3'mt pre-cleaved pre-mRNAs used in the UV. cross-linking assays. The light grey, dark grey and black boxes represent the USE 1, USE 2 and USE 3, respectively. The thick and thin black arrows indicate the poly(A) signals and the cleavage sites (c.s), respectively. The hatched boxes represent the regions of USE 1 that were mutated. The thin bar at the beginning of the pre-mRNAs represents the 39 nt polylinker sequence included in the pre-mRNAs. The spacers and the numbers underneath the wtpA1 pre-mRNA indicate the distance between each USE and between USE 3 and the proximal poly(A) signal. **(B)** USE 1 wild-type sequence, and point mutations introduced in the USE 1. The USE 1-mt, USE 1-5'mt and USE 1-3'mt sequences represent the

mutations present in the USE 1-mt, USE 1-5'mt and USE 1-3'mt pre-mRNAs. **(C)** UV. cross-linking assays were performed using wtpA1 (lane 1), USE 1-mt (lane 2), USE 1-5'mt (lane 3) and USE 1-3'mt (lane 4) pre-cleaved pre-mRNAs, with HeLa cell nuclear extracts.

As illustrated in Fig. 31C, the mutation in the USE 1-mt pre-mRNA, as expected, disrupts the protein complex assembled onto the pre-mRNA (compare lanes 1 and 2). The same result is obtained when the USE 1-5'mt pre-mRNA is used (compare lanes 1-3). Mutation of the 5' region of the USE 1 prevents the binding of p35 and hnRNP C. USE 1-3'mt pre-mRNA shows no differences compared with the wtpA1 pre-mRNA (compare lanes 1 and 4). Therefore, the binding of p35 and the function of USE 1 seems to be restricted to the UAUUUGUUUUU element contained within the USE 1.

In order to precisely map the p35 binding site in the USE 1 we divided the USE 1 into three fragments: USE 1-(a), USE 1-(b) and USE 1-(c) (Fig. 32A). Each fragment was cloned, *in vitro* transcribed and the transcripts used for UV. cross-linking (Fig. 32B).

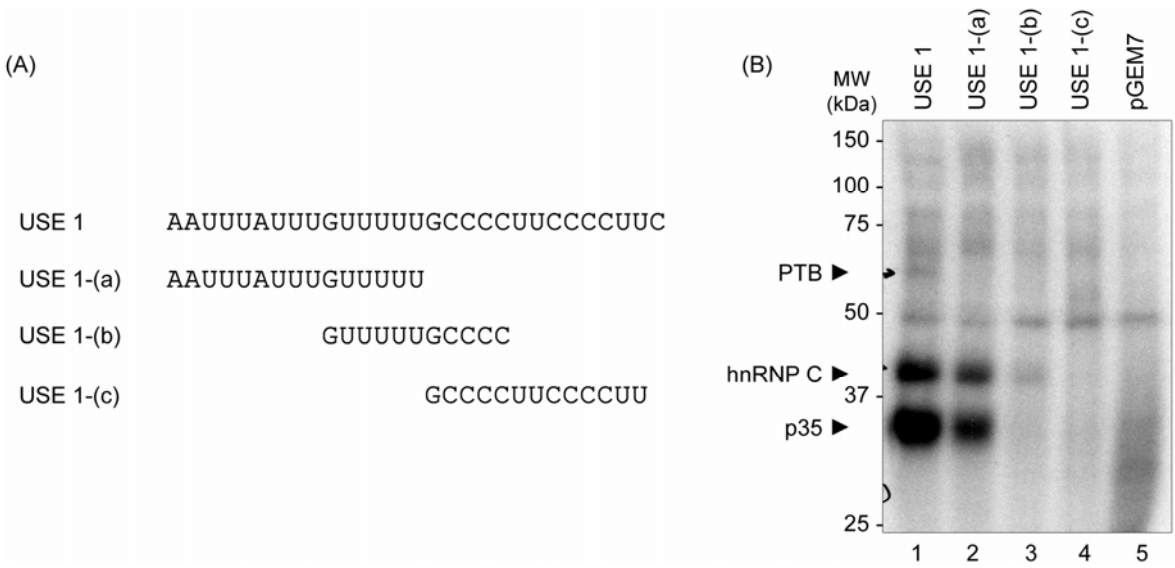

Figure 32. Mapping the p35 and hnRNP C binding sites in the USE 1. (A) To map the binding sites in USE 1, three short RNAs with USE 1 overlapping sequences were synthesized: USE 1-(a), USE 1-(b) and USE 1-(c). **(B)** UV. cross-linking assays were performed using the USE 1-(a), (lane 2), USE 1-(b) (lane 3), and USE 1-(c) (lane 4) RNAs, with HeLa cell nuclear extracts. As a positive and negative controls an USE 1 RNA, containing the wild-type USE 1 sequence (lane 1) and a RNA transcribed from a linearized pGEM7 polylinker (lane 5) were used, respectively.

As expected from our previous results, the USE 1-(a) RNA containing the UAUUUGUUUUU element is sufficient for the binding of p35 and hnRNP C (Fig. 32B, lanes 1-4; RNA transcribed from the pGEM7 polylinker-lane 5- was used as a non-specific control). When we compare USE 1-(a) and USE 1-(b) (lanes 2 and 3, respectively), the AAUUUAUUU sequence seems to be the minimal sequence required for the binding of

p35 while the GUUUUU element seems sufficient for the binding of hnRNP C. This observation is in agreement with the reported UUUUU binding site described for hnRNP C (Wilusz and Shenk 1990; Gorlach *et al.,* 1994). However, as the binding of hnRNP C is stronger for the USE 1-(a) RNA than for the USE 1-(b) RNA (Fig. 32B, lanes 2 and 3), these results cannot exclude the possibility that the hnRNP C optimal binding may require flanking sequences or that the binding of both p35 and hnRNP C might be cooperative.

2.7. Purification of p35

In order to identify the nature of p35, we decided to purify this protein based on its affinity towards the USE 1. For this purpose, an affinity chromatography was performed using a biotinilated oligonucleotide with three USE 1 copies in tandem (designated (USE 1)$_3$-oligo), bound to avidin coated agarose beads. The beads bound to the (USE 1)$_3$-oligo were then incubated with HeLa cell nuclear extracts. As control, incubation of nuclear extracts with avidin coated agarose beads without the (USE 1)$_3$-oligo was performed. After incubation, the beads were washed and the proteins eluted (Fig. 33A). As seen in Fig. 33A, several proteins bind specifically to the (USE 1)$_3$-oligo. Western blot analysis of the eluted fraction shows that PTB is specifically pulled down by the (USE 1)$_3$-oligo (Fig. 33B), which is in agreement with the previous results showing that PTB cross-links to the USE 1 (Fig. 29A).

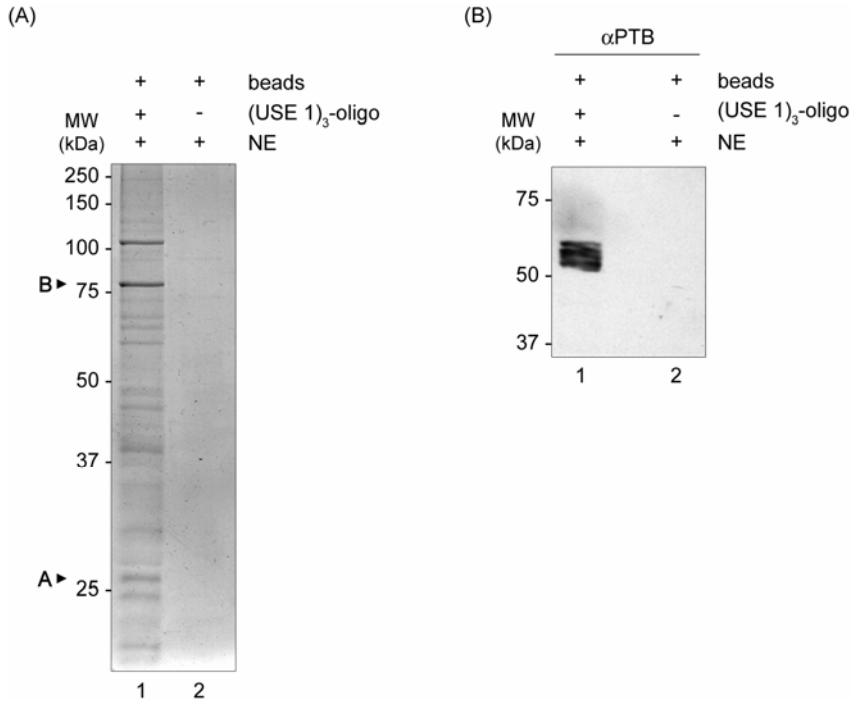

Figure 33. Purification of p35 by affinity chromatography. (A) HeLa cell nuclear extracts (NE) were incubated with avidin coated agarose beads bound to a biotinilated oligonucleotide containing

three USE 1 copies in tandem ((USE 1)$_3$-oligo) (lane 1) or with avidin coated agarose beads only (lane 2). After incubation, the beads were washed and the proteins bound to the oligonucleotide eluted. The protein bands A and B were isolated, purified and analysed by mass spectrometry analysis. **(B)** Detection of PTB by Western blot in the eluted fraction showed in (A), using a PTB polyclonal antibody.

In the 25-37 kDa range, a strong protein band designated protein A was selected for further purification (Fig. 33A, lane 1). This protein band was isolated and analysed by mass spectrometry analysis in collaboration with Dr. Alexandre Akoulitchev (Sir William Dunn School of Pathology, University of Oxford). In addition to this protein, a second protein, band B (Fig. 33A, lane 1), was also selected and analysed. From the mass spectrometry analysis, 4 and 7 peptides were obtained and sequenced for protein bands A and B respectively (Table VII).

Peptide	Band A	Band B
1	sfdllvk	vfgneik
2	skieteik	lelqgpr
3	tdtvlilcr	qgteidgr
4	iyedgdddmkr	esfdgsvr
5		tgisdvfak
6		sislyytgek
7		glsedtteetlk

Table VII. Peptides obtained from the mass spectrometry analysis of bands A and B.

The alignment of the peptide sequences with the NCBI database (http://www.ncbi.nlm.nih.gov/) identified the protein bands A and B as the 25 kDa Calcyclin Binding Protein (CacyBP) and the 76 kDa Nucleolin, respectively.

We tested if CacyBP was part of the protein complex assembled onto the wtpA1 pre-mRNA. Using a mouse recombinant CacyBP, kindly provided by Prof. Jacek Kuznicki (International Institute of Molecular and Cell Biology, Warsaw, Poland), UV. cross-linking assays using the wtpA1 and USE 1-mt pre-cleaved pre-mRNAs were performed. As seen in Fig. 34, CacyBP binds the wtpA1 pre-mRNA in a USE 1-dependent manner (compare lanes 5 and 10). However, CacyBP is not p35 since the p35 protein band displays a slower gel migration than CacyBP (compare lanes 1 and 5). Therefore, these results suggest that although CacyBP is not p35 it does however bind the wtpA1 pre-mRNA through the binding to the USE 1.

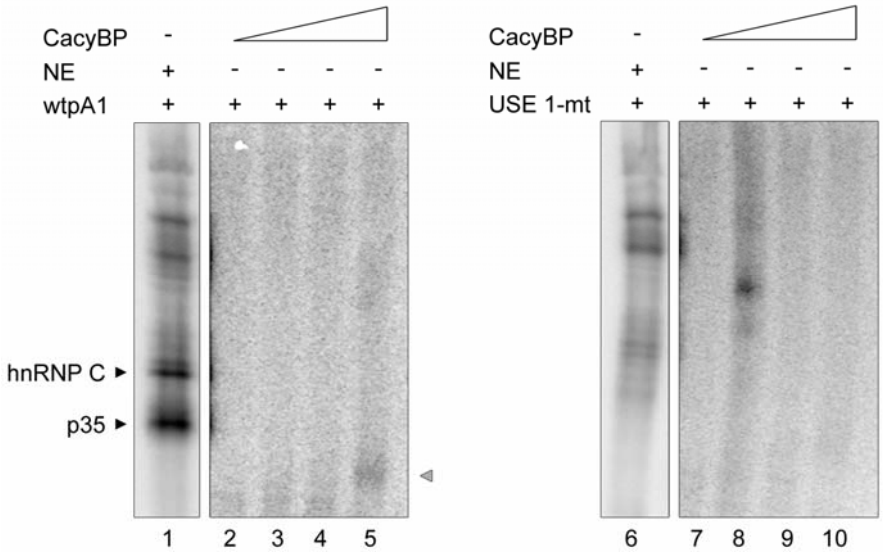

Figure 34. CacyBP binds to the wtpA1 pre-mRNA in an USE 1-dependent manner. UV. cross-linking assays were performed with wtpA1 (lanes 1-5) and USE 1-mt (lanes 6-10) pre-cleaved pre-mRNAs using HeLa cell nuclear extracts (NE) (lanes 1 and 6) or with increasing amounts of mouse recombinant CacyBP: 25 ng (lanes 2 and 7), 50 ng (lanes 3 and 8), 100 ng (lanes 4 and 9) and 1000 ng (lanes 5 and 10).

2.8. Analysis of the tripartite Upstream Sequence Element

To determine the contribution of each USE in mRNA 3'end formation at the proximal poly(A) site, several constructs were generated in which the USE 1, USE 2 or USE 3 were deleted (Fig. 35A). In addition, point mutations were also introduced in the USE 2 and USE 3 (Fig. 35A, B). After *in vitro* transcription, the radiolabelled RNAs were used in UV. cross-linking reactions (Fig. 35C). From these experiments, two main observations are possible. First, the effect of either mutation or deletion of each USE is the same, since the same disruption in the cross-linked protein pattern is observed (Fig. 35C, compare lanes 2 and 6, lanes 3 and 7 and lanes 4 and 8). This indicates that the effect observed is not due to the introduction of new sequences in the mutant, neither is due to spacing effects caused by deletion of each element. Second, the USE 1 is the major binding site for p35 and hnRNP C. The binding of both proteins to the pre-mRNA is inhibited only when USE 1 is either mutated or deleted (Fig. 35C, compare lanes 1 and 2 and 5 and 6). Mutation/deletion of any other USE has no effect on the binding of these two proteins (Fig. 35C, compare lane 1 with lanes 3 and 4 and lane 5 with lanes 7 and 8). Our previous results of UV. cross-linking assays using the USE 3 RNA showed that hnRNP C could bind, although weakly, to USE 3 (Fig. 29A). Since USE 1 is the major binding site

for hnRNP C, it is possible that the presence of this element in the USE 3-mt and ΔUSE 3 RNAs may mask any small variation resulting from the mutation/deletion of USE 3.

In the previous polyadenylation assays (Fig. 21D and 28B), a Superose 6 gel filtration fraction was used. Since this fraction was no longer available, HeLa cell nuclear extracts were prepared and used in the following experiments. However, these new extracts proved to be very inefficient in our polyadenylation assays (Fig. 36).

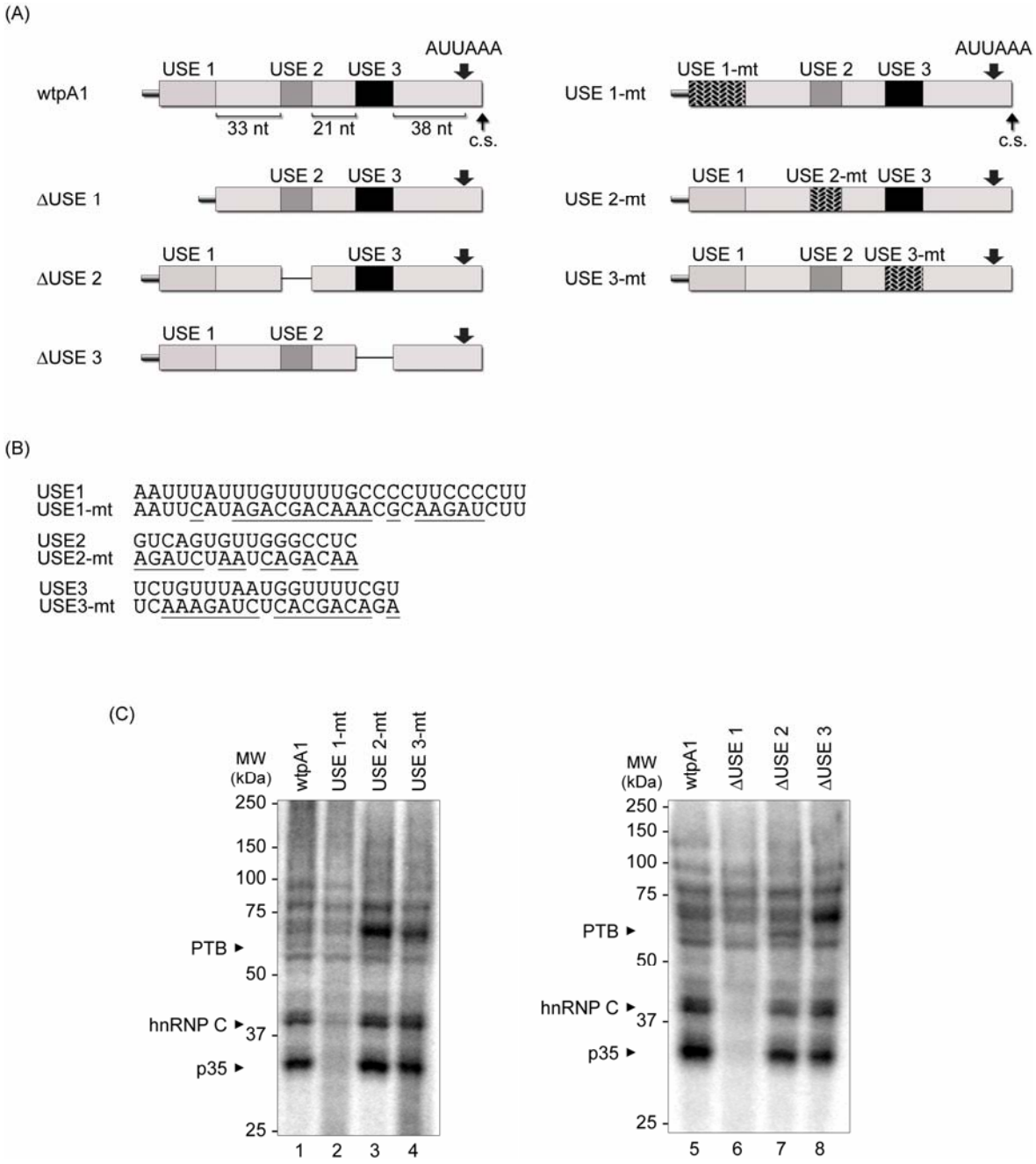

Figure 35. USE 1 is the major binding site for p35 and hnRNP C in wtpA1 pre-mRNA. (A) Schematic representation of the wtpA1 pre-mRNA and the pre-mRNAs in which each USE has been either deleted (ΔUSE 1, ΔUSE 2, and ΔUSE 3) or point mutated (USE 1-mt, USE 2-mt and

USE 3-mt). All pre-mRNAs are pre-cleaved. The light grey, dark grey and black boxes represent the USE 1, USE 2 and USE 3, respectively. The thick and thin black arrows indicate the poly(A) signals and the cleavage sites (c.s.), respectively. The hatched boxes represent the point mutated USEs. The thin bar at the beginning of the pre-mRNAs represents the 39 nt polylinker sequence included in the pre-mRNAs. The spacers and the numbers underneath the wtpA1 pre-mRNA indicate the distance between each USE and between USE 3 and the proximal poly(A) signal. **(B)** The point mutations present in the pre-cleaved pre-mRNAs USE 1-mt, USE 2-mt and USE 3-mt sequences are underlined. The USE 1, USE 2 and USE 3 sequences indicate the wild-type sequences of the corresponding USE. **(C)** UV. cross-linking assays with wtpA1 (lane 1 and 5), USE 1-mt (lane 2), USE 2-mt (lane 3), USE 3-mt (lane 4), ΔUSE 1 (lane 6), ΔUSE 2 (lane 7) and ΔUSE 3 (lane 8) pre-cleaved pre-mRNAs using HeLa cell nuclear extracts.

Despite the fact that these extracts are able to polyadenylate the Ad-L3 pre-cleaved pre-mRNA (used as a positive control) (Fig. 36, lane 3) we were unable to obtain sufficient polyadenylation of the wtpA1 pre-mRNA, which has a weaker poly(A) signal and therefore only used *in vitro* with extremely active nuclear extracts (Fig. 36, lanes 1 and 2). The low levels of polyadenylation were not sufficient to allow an analysis of the several mutated pre-mRNAs.

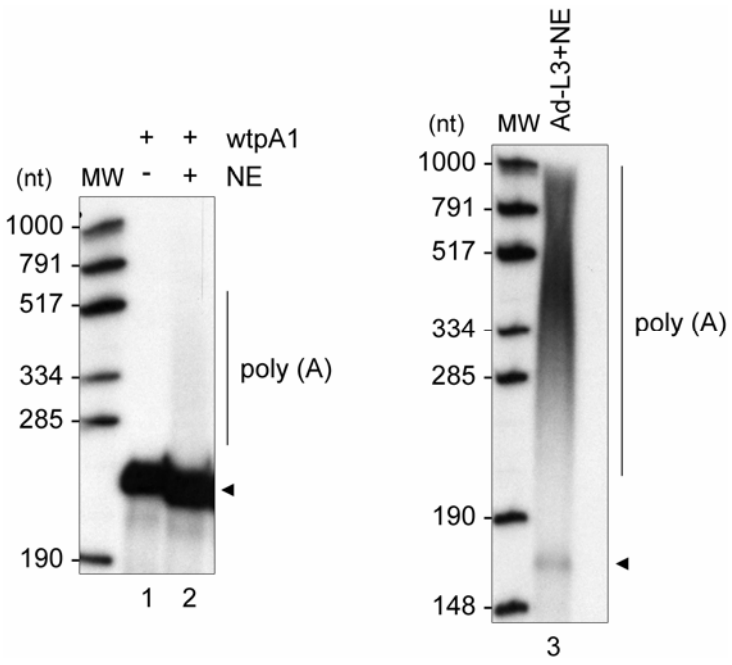

Figure 36. The wtpA1 pre-mRNA shows low levels of polyadenylation in HeLa cell nuclear extracts. Polyadenylation assays were performed with pre-cleaved wtpA1 pre-mRNA in the absence (lane 1) or presence (lane 2) of HeLa cell nuclear extracts (NE). As a positive control, a polyadenylation reaction was performed with the Ad-L3 pre-cleaved pre-mRNA (lane 3). The black arrowhead indicates the input pre-cleaved RNA and the vertical bars, the polyadenylated RNA.

To circumvent the low polyadenylation activity of our protein extracts, we tried to use commercial prepared HeLa cells nuclear extracts from Calbiotech. However, we were

unable to draw any conclusion as the extracts displayed unspecific polyadenylation activity (Fig. 37). This was tested by performing polyadenylation reactions using the Ad-L3 and Ad-L3MA pre-mRNAs (Takagaki *et al.,* 1988) as positive and negative controls, respectively (Fig. 37, lanes 1, 6 and 23-26). The Ad-L3 has a strong AAUAAA poly(A) signal that is mutated to AAUACA in the Ad-L3MA pre-mRNA.

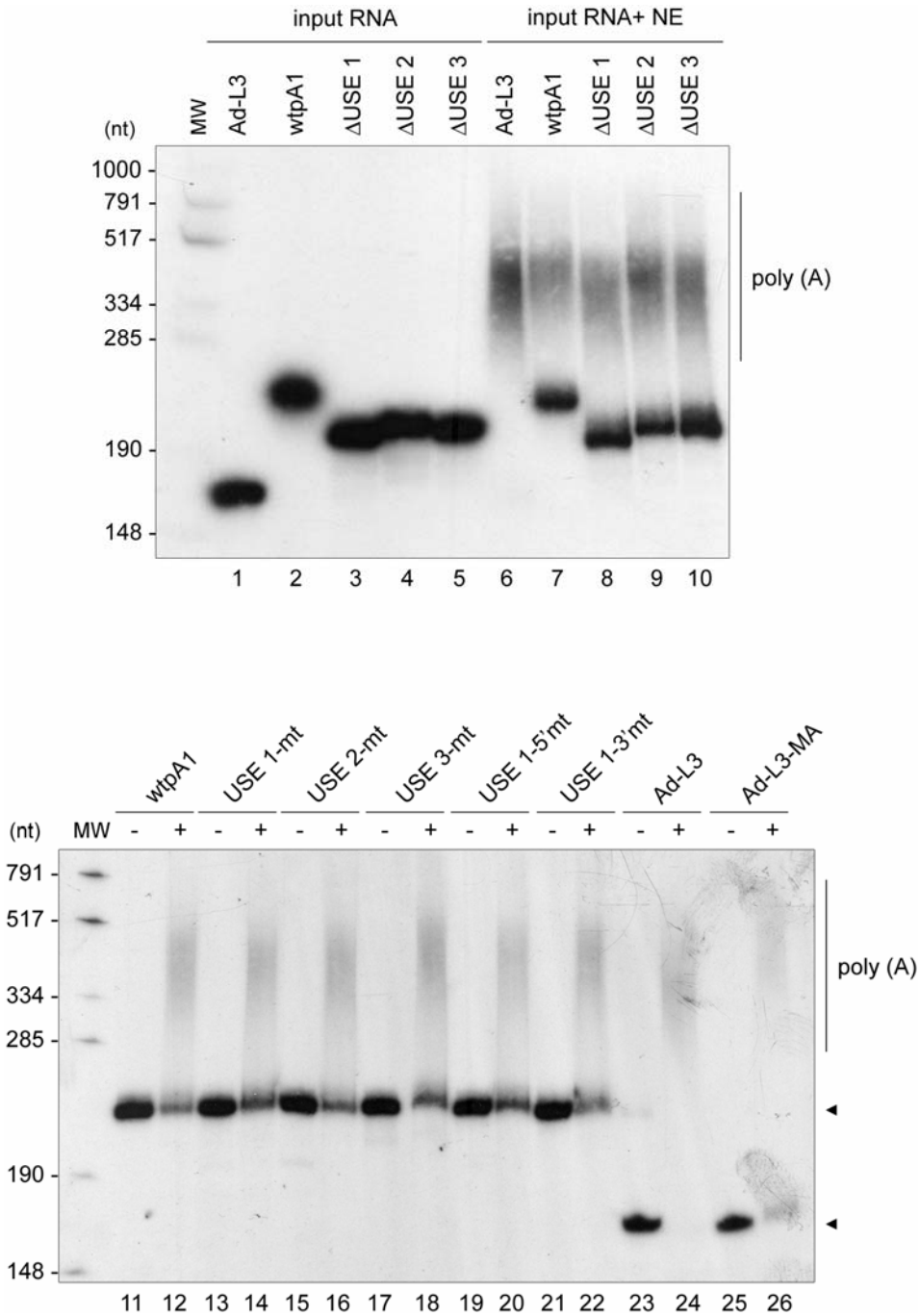

Figure 37. Unspecific polyadenylation activity. The commercial HeLa cell nuclear extracts (NE) exhibited unspecific polyadenylation activity. Polyadenylation assays were performed using the following pre-cleaved RNA substrates: wtpA1 (lanes 2, 7, 11 and 12), ΔUSE 1 (lanes 3 and 8) ,

ΔUSE 2 (lanes 4 and 9), ΔUSE 3 (lanes 5 and 10), USE 1-mt (lanes 13 and 14), USE 2-mt (lanes 15 and 16), USE 3-mt (lanes 17 and 18), USE 1-5'mt (lanes 19 and 20) and USE 1-3'mt (lanes 21 and 22). As a positive and negative controls, the Ad-L3 and Ad-L3-MA (in which the Ad-L3 poly(A) signal AAUAAA was mutated to AAUACA) pre-mRNAs were used (lanes 1, 6 and 23-26). The – and + indicate polyadenylation reactions performed in the absence or presence of HeLa cell nuclear extracts. The black arrowhead indicates the input pre-cleaved RNA and the vertical bars, the polyadenylated RNA.

As seen in Fig. 37, the input band corresponding to Ad-L3MA pre-cleaved pre-mRNA was almost completely polyadenylated (lanes 25 and 26, arrowhead) as well as all other pre-mRNAs used in this experiment.

In order to achieve specific polyadenylation and reproduce our initial results, we tried to optimize the polyadenylation reaction. Using the wtpA1 and USE 1-mt pre-cleaved pre-mRNAs, polyadenylation assays were performed with lower amounts of protein extracts (Fig. 38). Additionally, instead of the standard incubation of 90 min, incubation periods were also shortened to 15 and 30 min.

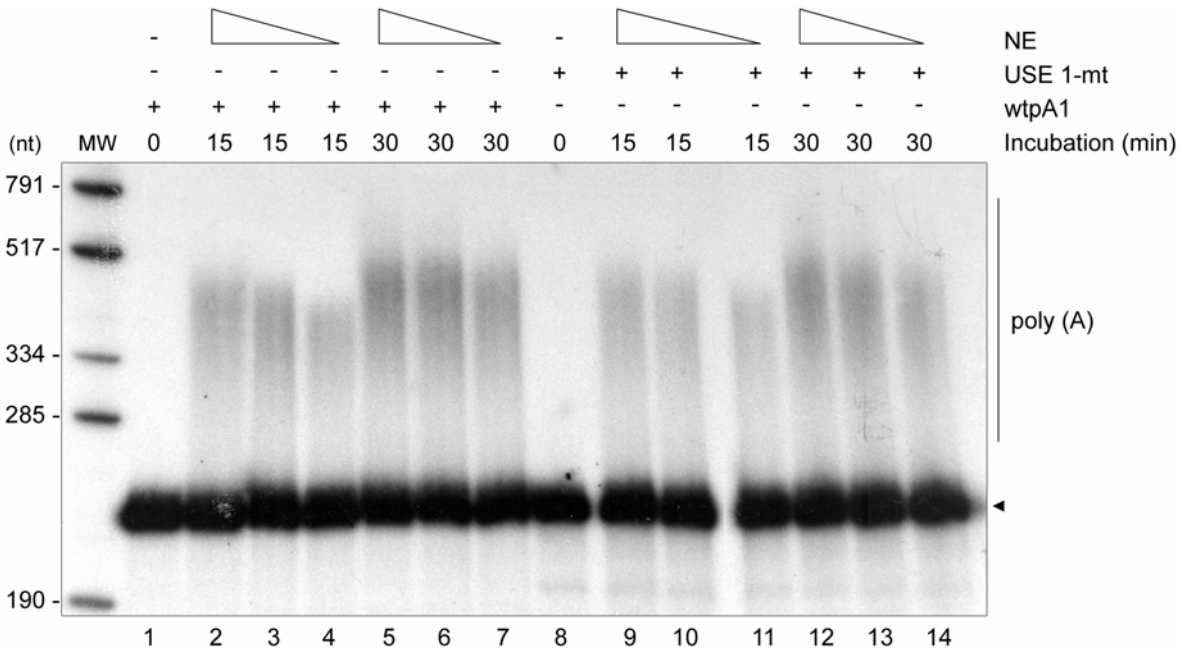

Figure 38. Optimization of the polyadenylation reaction. Polyadenylation reactions were performed with the wtpA1 (lanes 1-7) and USE 1-mt (lanes 8-14) pre-cleaved pre-mRNAs using commercial HeLa cell nuclear extracts (NE). Reactions were incubated for 15 (lanes 2-4 and 9-11) and 30 min (lanes 5-7 and 12-14), with decreasing amounts of nuclear extracts: 5 μl (lanes 2, 5, 9 and 12), 4 μl (lanes 3, 6, 10 and 13) and 3 μl (lanes 4, 7, 11, and 14). Lanes 1 and 8 indicate the wtpA1 and USE 1-mt polyadenylation reactions in the absence of nuclear extracts. The black arrowhead indicates the input pre-cleaved RNA and the vertical bar, the polyadenylated RNA.

Even under these conditions, all pre-mRNAs were still polyadenylated (Fig. 38, compare lanes 2-4 with 9-11 and 5-7 with 12-14).

In polyadenylation reactions, *E. coli* tRNA is added to prevent unspecificity. To try to reduce the unspecific polyadenylation, assays were performed with increasing amounts of tRNA. However, even a 4-fold increase in the tRNA concentration did not prevent the unspecific activity (data not shown). As PAP specificity is dependent on Mg^{2+} ions (Christofori and Keller 1988; Takagaki *et al.,* 1988), we tried to manipulate the Mg^{2+} concentration. Reactions were performed in a 3 mM and 5 mM M_gCl_2 concentration, instead of the standard 1mM. As a result, a decreased in polyadenylation efficiency was observed but not a decrease in the unspecific activity (data not shown).

Therefore, because our *in vitro* system did not allow a more detailed study of the role of each USE in mRNA 3'end formation at the proximal poly(A) site, we decided to pursue an *in vivo* strategy.

2.9. *In vivo* analysis of the Upstream Sequence Elements

In order to conduct a study of each USE, we proceeded to study these elements *in vivo*. To this end, mini-genes were constructed in which each USE was deleted. These mini-genes were used to transfect HeLa cells and the produced mRNAs analysed by S1 nuclease. In this assay, a 5'end radiolabelled DNA probe is used to hybridize and protect a specific region of the produced mRNAs. After hybridization under conditions that favour the formation of DNA:RNA hybrids, S1 Nuclease is added to the reactions. This enzyme is a endoribonuclease that degrades single-strand DNA and RNA thus protecting the DNA:RNA hybrids that are then resolved by electrophoresis in a denaturing polyacrylamide gel and visualized by autoradiography.

To generate the *polo* mini-genes, we used as a template the pUC-α*C2* plasmid. This plasmid, kindly provided by Prof. Nicholas J. Proudfoot (Sir William Dunn School of Pathology, University of Oxford) is used in polyadenylation competition assays, where the efficiency of a poly(A) signal (*C2 complement* in this plasmid*)* is measured in competition with the reference α*-globin* poly(A) signal (Moreira *et al.,* 1995) (Fig. 39A). When HeLa cells are transiently transfected with pUC-α*C2* and the produced mRNAs analysed by S1 Nuclease, two mRNAs are detected due to the usage of both poly(A) signals (Fig. 39B).

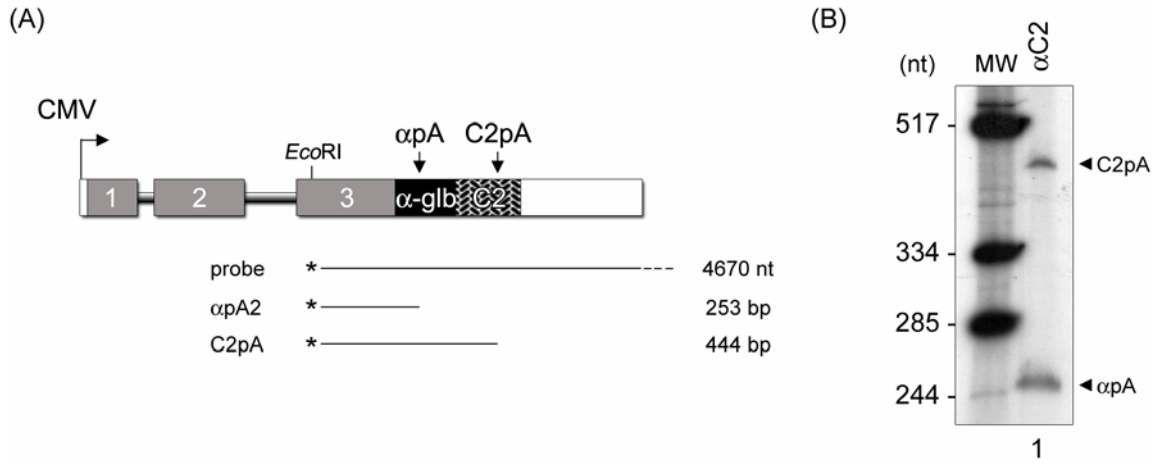

Figure 39. *In vivo* analysis of the pUC-*αC2* mini-gene. (A) Schematic representation of the pUC-*αC2* mini-gene. The mini-gene is under the CMV promoter, represented in the figure by an arrow. The grey numbered boxes represent the three *β-globin* exons and the grey thin bar between them, the corresponding introns. The black and the hatched boxes represent a fragment of the *α-globin* and *C2 complement* 3'UTR, respectively. The αpA and the C2pA are the *α-globin* and the *C2 complement* poly(A) signals respectively, with their position in the 3'UTRs indicated by thin black arrows. The white thick box represents pUC plasmid sequence. The lines underneath represent the band sizes correspondent to the probe and the protected RNA regions defined by the usage of the *α-globin* and *C2 complement* poly(A) signals. **(B)** S1 Nuclease analysis. Cytoplasmic RNA isolated from HeLa cells transfected with the pUC-*αC2* (lane 1). The bands indicated by arrowheads correspond to the usage of the *α-globin* and *C2 complement* poly(A) signals.

To generate the *polo* wild-type mini-gene, a genomic fragment containing the two last *polo* exons, the downstream intergenic region and a fragment of the downstream *snap* gene, was subcloned in frame with the *β-globin* exon 3 (Fig. 40A). Each USE was deleted to generate the following mini-genes: pUC-*poloΔUSE 1*, pUC-*poloΔUSE 2* and pUC-*poloΔUSE 3*. A negative control, pUC-*poloΔpA1*, was also generated by mutating the proximal poly(A) signal ATTAAA to GTTAAC (Fig. 40A). HeLa cells were transiently transfected with the different mini-genes. As a co-transfection control, each mini-gene was co-transfected with the pBR328βSV40 plasmid (Rβ) (Grosveld *et al.,* 1982). 48 hours after transfection, cells were collected and cytoplasmic RNA was extracted. The usage of each poly(A) signal was then determined by S1 Nuclease assay. However, as seen in Fig. 40B we were not able to detect any band of the expected size (lanes 1-5). Only two bands are visualized which are likely the result of unspecific hybridization, as they are also detected in cells transfected only with the Rβ plasmid (lane 6).

Figure 40 (see picture on next page). *In vivo* analysis of the upstream sequence elements in HeLa cells. (A) Schematic representation of the pUC-*polo wt* mini-gene. The mini-gene is under a CMV promoter represented in the figure by an arrow at the beginning of the mini-gene. The dark grey labelled 1-3 boxes represent the three β-*globin* exons in frame with the coding sequences of

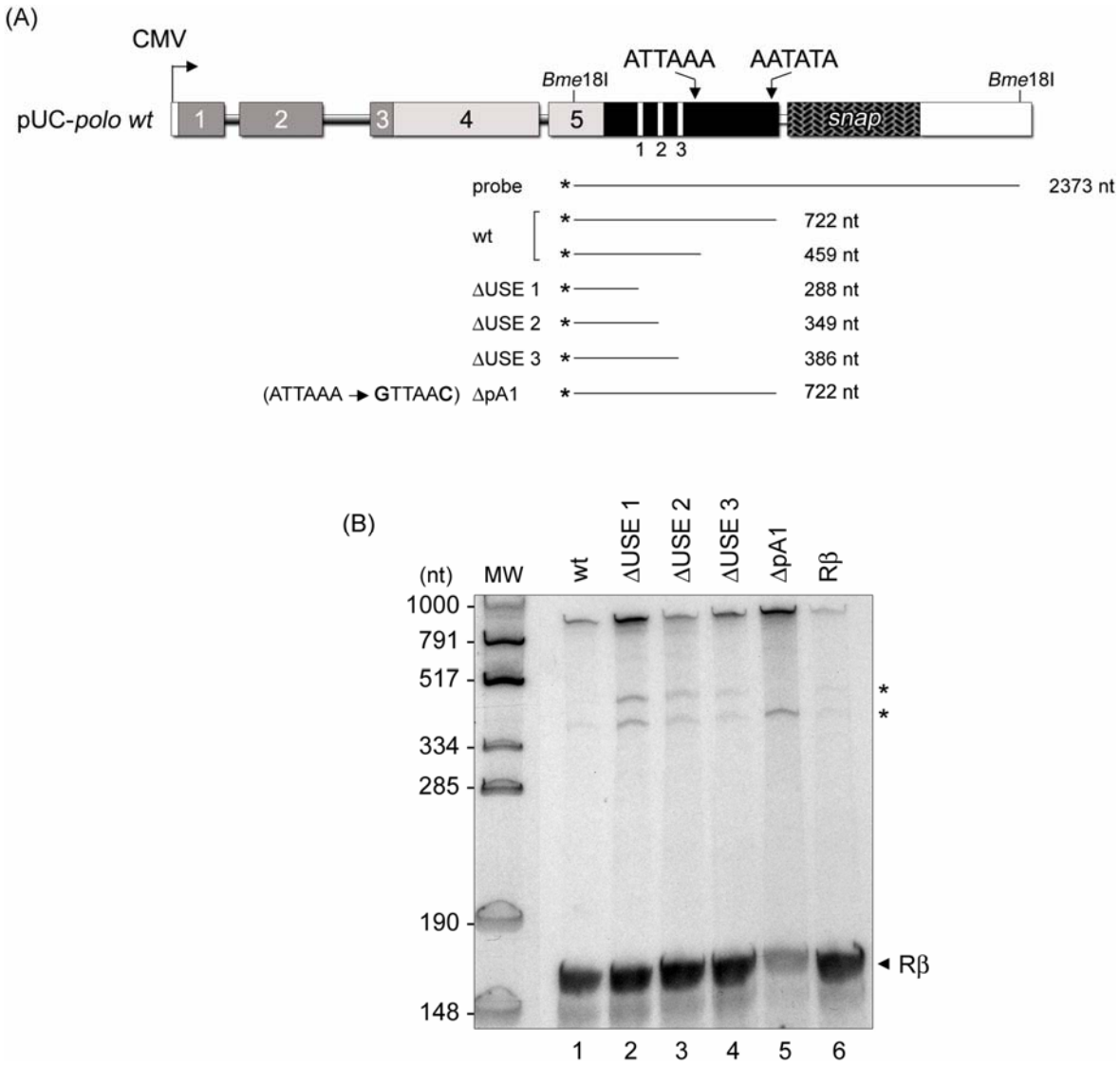

Figure 40 (continue from previous page). the *polo* exons 4 and 5, represented by the light grey labelled 4 and 5 boxes, respectively. The black box represents the 3'UTR in exon 5 and the three thin open boxes labelled underneath with 1, 2 and 3 represent USE 1, USE 2 and USE 3 respectively. The ATTAAA and the AATATA are the proximal and distal poly(A) signals. The downstream hatched box represents a fragment of the downstream *snap* gene. The white thick box represents pUC plasmid sequences. The thin horizontal grey bar between the exons represents the β-*globin* and *polo* introns, whereas the white bar between the 3'UTR and *snap*, represents the intergenic region. Three mini-genes were generated by deletion of each USE: pUC-*poloΔUSE 1*, pUC-*poloΔUSE 2* and pUC-*poloΔUSE 3*. As a negative control, pUC-*poloΔpA1* was constructed in which the proximal poly(A) signal ATTAAA was mutated to GTTAAC. To synthesize the probe used in the S1 Nuclease analysis, the minigene was cut with *Bme*18I and a 2373 nt fragment was radioactively labelled and used. The drawings underneath represent the probe and the expected fragments sizes from the S1 nuclease assay **(B)** S1 Nuclease analysis. Cytoplasmic RNA from HeLa cells transfected with pUC-*polo wt* (lane 1), pUC-*poloΔUSE 1* (lane 2), pUC-*poloΔUSE 2* (lane 3), pUC-*poloΔUSE 3* (lane 4) and pUC-*poloΔpA1* (lane 5) mini-genes was analysed. Each mini-gene was co-transfected with the Rβ plasmid. As a control, HeLa cells transfected only Rβ were also analysed (lane 6). The black arrowhead indicates the expressed Rβ mRNA. The asterisks indicate bands that are most likely the result of unspecific hybridization.

Earlier reports on the biochemistry of mRNA 3'end formation documented the effect of point mutations in the canonical poly(A) signal (Wickens and Stephenson 1984; Wilusz *et al.,* 1989; Sheets *et al.,* 1990). In those studies it was shown that non-canonical poly(A) signals are less efficiently processed, with lower rates of cleavage and polyadenylation compared to the AAUAAA hexamer. Thus, it is possible that the two non-canonical poly(A) signals of *polo* may not be efficiently processed and therefore not able to be detected in our assays. Additionally, we must consider the possibility that the mRNAs produced by the *polo wt* mini-gene may not be stable in mammalian cells. A lower mRNA stability could result in low levels of mRNA, which could account for our lack of success in detecting these transcripts. Therefore, we decided to pursue this approach using instead a *Drosophila* cell line.

Using the *Drosophila* expression vector pAc5.1-V5/His A, a wild-type mini-gene was constructed under the *actin* promoter (Fig. 41A). The same point mutations used in our *in vitro* assays were introduced in each USE (see Fig. 35B for the point mutations introduced in each USE). As a negative control a mini-gene was generated in which the proximal poly(A) signal was mutated (ATTAAA to GTTAAC). Schneider cells were then transiently transfected with the different mini-genes. The pAc5.1-V5/HisA plasmid which contains the SV40 late poly(A) signal was used as a co-transfection control in these experiments. 48 h after transfection, total RNA was extracted and analysed by S1 Nuclease assay (Fig. 41B). In these assays, we could not detect a clear band for the usage of the proximal poly(A) site (Fig. 41B, lanes 1-6). Instead, several weak bands are observed in the range of the expected size. However, for the distal poly(A) site a strong band that matches the expected size is observed (Fig. 41B, lanes 1-6). Although mutations of the USEs have no clear effect over the usage of the distal poly(A) site, mutation of pA1 seems to inhibit the usage of the distal site (Fig. 41B, compare lanes 1-4 with lane 5). Quantification of the usage of the distal poly(A) site shows a 10-fold decrease for *polo∆pA1* (Table VIII).

	[Distal pA/SV40L pA]	
Plasmid	**Assay 1**	**Assay 2**
wt	1,0	1,0
USE 1-mt	1,2	0,8
USE 2-mt	2,0	0,9
USE 3-mt	1,3	0,8
∆pA1	0,1	0,1
pAc5.1-V5/HisA	0,1	0,1

Table VIII. Quantification of the usage of the distal poly(A) signal. The [distal pA/SV40L pA] ratios were determined by densitometry scanning of the autoradiographs. pA: poly(A) signal.

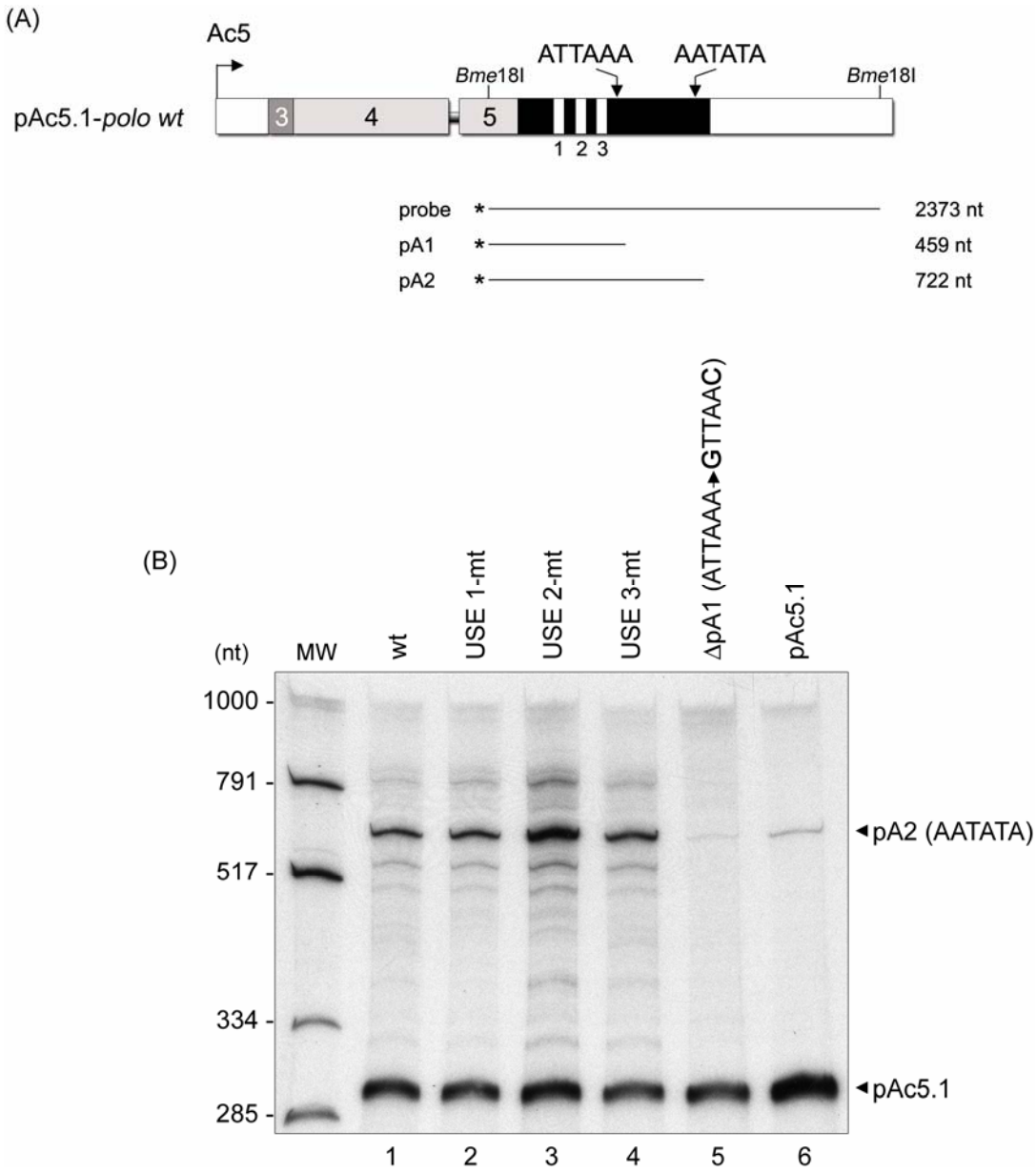

Figure 41. *In vivo* analysis of the upstream sequence elements in *Drosophila* Schneider cells. (A) Schematic representation of the pAc5.1-*polo wt* mini-gene. The mini-gene is under the *actin* promoter (Ac5) represented by an arrow at the beginning of the mini-gene. The dark grey box represents a fragment of the third *β-globin* exon in frame with the coding sequences of *polo* exons 4 and 5, represented by the light grey labelled 4 and 5 boxes, respectively. The black box represents the 3'UTR in exon 5 and the three white thin boxes labelled underneath with 1, 2 and 3 represent the USE 1, USE 2 and USE 3, respectively. The white thick boxes represent pAc5.1-V5/HisA plasmid sequences. The thin horizontal grey bar between the *polo* exons represents the *polo* intron 4. Three mini-genes were generated by mutation of each USE: the pAc5.1-*polo-USE 1-mt*, pAc5.1-*polo-USE 2-mt* and pAc5.1-*polo-USE 3-mt*. As a negative control, pAc5.1-*poloΔpA1* was constructed in which the proximal poly(A) signal ATTAAA was mutated to GTTAAC. The probe used in the S1 Nuclease analysis was the same probe described in Fig. 40. The drawings underneath represent the probe and the expected protected fragments sizes from the S1 nuclease assay for the usage of the proximal (pA1) and distal (pA2) poly(A) signals. **(B)** S1 Nuclease analysis. Total RNA isolated from Schneider cells transfected with pAc5.1-*polo wt* (lane 1), pAc5.1-

polo-USE 1-mt (lane 2), pAc5.1-*polo-USE 2-mt* (lane 3), pAc5.1-*polo-USE 3-mt* (lane 4) and pAc5.1-*poloΔpA1* (lane 5) mini-genes was analysed by S1 Nuclease. Each mini-gene was co-transfected with the pAc5.1-V5/HisA plasmid. As a control, Schneider cells transfected with pAc5.1-V5/HisA were also analysed (lane 6). The black arrowheads indicate the usage of the SV40L poly(A) signal in the pAc5.1-V5/HisA plasmid and pA2 in the mini-genes.

In conclusion, an evaluation of the precise function of each USE in cell lines was not possible, as after several efforts we could not clearly detect the band correspondent to the pA1 usage. Nevertheless, two interesting observations are possible from these experiments: First, in Schneider cells, the distal poly(A) site seems to be used more efficiently than the proximal site, which is the opposite result observed by Northern blot analysis. Second, the use of the distal poly(A) signal seems to depend on a functional proximal poly(A) signal.

2.10. *gfp-poloΔUSE 1* and *gfp-poloΔpA1* transgenic flies present a similar phenotype

Our initial findings suggested that all three USEs may have a role in mRNA 3'end formation at the proximal poly(A) site. USE 1 seems to be the most important of these USEs not only because mutation of this element inhibits polyadenylation, but because it contains the p35 binding site. Characterization of the protein complexes assembled upstream both poly(A) signals showed that p35 is highly specific for USE 1. This specificity suggests that this element may have a regulatory function in alternative polyadenylation. Therefore, we decided to generate transgenic flies in which USE 1 is deleted.

As in the case of the *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenes, PCR mutagenesis using the *gfp-polo* transgene as a template was performed to generate the *gfp-poloΔUSE 1* transgene (Fig. 42). The *gfp-poloΔUSE 1* was then used for P-element germline transformation. From the transformants obtained, lines with the transgene inserted on the second chromosome were selected by genetic mapping as described previously. As a result, one line homozygous for the transgene, viable and fertile was selected: *gfp-poloΔUSE 1* (line 11.5).

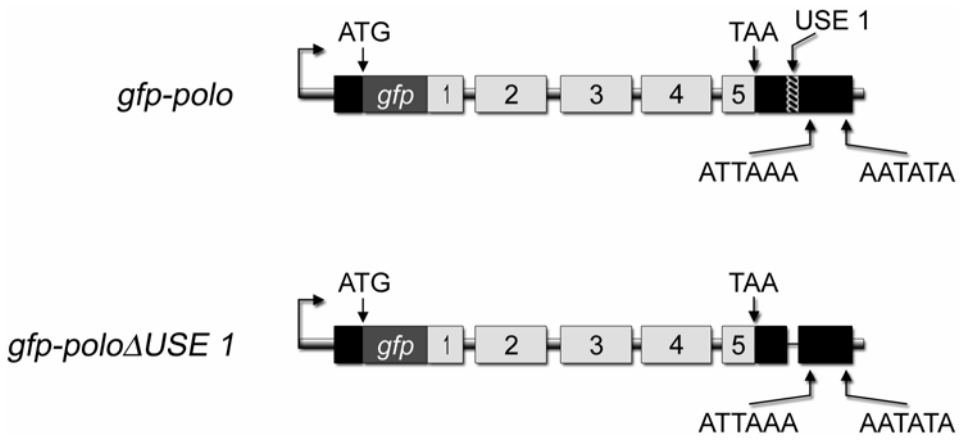

Figure 42. Generating *gfp-poloΔUSE 1* transgenic flies. Schematic representation of the transgenes constructed to generate *gfp-polo* and *gfp-poloΔUSE 1* transgenic flies. The arrow at the beginning of the transgenes represents the endogenous *polo* promoter. The labelled 1-5 light grey boxes represent the five *polo* exons with the black boxes representing the 5' and 3'UTR of exons 1 and 5 respectively. The grey bars between the exons represent the introns. The *gfp* box represents the *gfp* coding sequence placed in frame with the *polo* initiation codon. The hatched box in the 3'UTR represents USE 1. The ATTAAA and AATATA represent the proximal and distal poly(A) signals, respectively.

To generate flies expressing only *polo* mRNAs in which the USE 1 has been deleted, we used the same genetic approach as in the study of the *gfp-poloΔpA1* and *gfp-poloΔpA2* transgenes (Chapter I-Results). Therefore, *gfp-poloΔUSE 1; polo^9/TM6B* flies were generated and the ability to complement the *polo^9* mutation studied. *gfp-poloΔUSE 1, polo^9* flies were generated and brain squashes from third instar larvae in a *polo^9* background prepared and the mitotic index quantified (Table IX). As observed in the previous chapter, complementation of *polo^9* by the *gfp-polo* transgene lowered the mitotic index to values similar to *polo^9* hemizygous individuals. The same reduction is achieved when complementation of *polo^9* is performed using the *gfp-poloΔUSE 1* transgene. Therefore, despite deletion of the USE 1, *gfp-poloΔUSE 1* mRNAs are still able to rescue the prometaphase/ metaphase arrest observed in p*olo^9* mutant brains. Furthermore, the *gfp-poloΔUSE 1* transgene is also able to rescue the third instar larval lethality of *polo^9*. Analysis of the *gfp-poloΔUSE 1; polo^9/TM6B* stock population shows that *gfp-poloΔUSE 1; polo^9* individuals are viable and constitute 23,4% of the total population. Thus, the *gfp-poloΔUSE 1* transgene, as seen previously for the *gfp-polo*, *gfp-poloΔpA1* and *gfp-poloΔpA2,* is sufficient to rescue the described phenotype assigned to the *polo^9* allele.

Strain	Mitotic Index
polo ^{9} */TM6C*	1,5
polo ^{9} */polo* ^{9}	22,2
gfp-polo;polo ^{9} */TM6B*	1,1
gfp-polo;polo ^{9} */polo* ^{9}	1,3
gfp-poloΔUSE 1;polo ^{9} */TM6B* (Line 11.5)	1,4
gfp-poloΔUSE 1;polo ^{9} */polo* ^{9} (Line 11.5)	1,7

Table IX. Quantification of the mitotic index. *gfp-polo* and *gfp-poloΔUSE1* third instar larvae, either in a *polo*^{9}*/TM6B* or *polo*^{9}*/polo*^{9} genetic background, were dissected and the mitotic index quantified. As control, the mitotic index of *polo*^{9}*/TM6C* and *polo*^{9}*/polo*^{9} third instar larvae was also determined.

An analysis of *gfp-poloΔUSE 1; polo*^{9} adults shows an abdominal phenotype identical to the phenotype presented by *gfp-poloΔpA1; polo*^{9} individuals (Fig. 43, compare panels 6 and 7). Thus, deletion of the USE 1 generates the same phenotype as deletion of pA1. Moreover, the phenotype of the *gfp-poloΔUSE 1; polo*^{9} flies strongly supports our *in vitro* results obtained for the USE 1. Deletion of the USE 1 results in inhibition of the polyadenylation reaction at the proximal poly(A) site favouring the synthesis of the 2.5 kb *polo* mRNA.

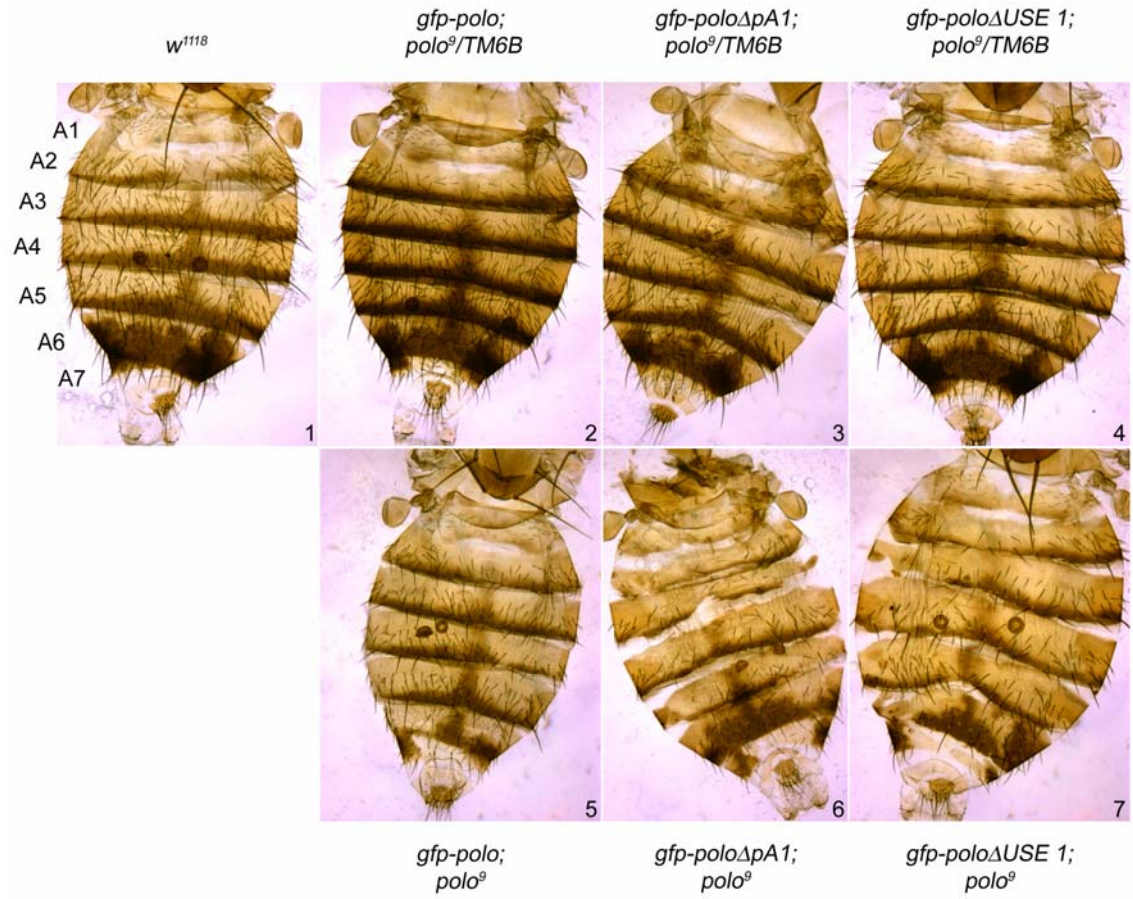

Figure 43. *gfp-poloΔUSE1; polo*^{9} flies have an abdominal phenotype identical to *gfp-poloΔpA1; polo*^{9}. Dorsal view of abdomen preparations from *gfp-polo*, *gfp-poloΔpA1* (line 12.1)

and *gfp-poloΔUSE 1* (line 11.5) adult female flies. Abdomens were prepared from flies in a *polo*^{9}/*TM6B* (panel 2-4) or *polo*^{9} background (panel 5-7). An abdomen from a *w*^{1118} female was also prepared and used as control (panel 1). The seven tergites of the adult abdomen are indicated as A1-A7 (panel 1). Anterior is up.

3. Discussion

To understand the role that poly(A) site usage may play in the overall requirement and expression of the longer *polo* mRNA in abdominal histoblasts, we searched the 3'UTR of *polo* for putative mRNA 3'end regulatory elements. Sequence conservation in non-coding regions between different species has been shown to predict possible regulatory sequences (Audibert and Simonelig 1998; Pan *et al.,* 2006; Danckwardt *et al.,* 2007). An alignment of the genomes of twelve *Drosophila* species plus the genomes of the mosquito, the honeybee and the beetle, identified several conserved elements present in the *polo* 3'UTR: lod 10, lod 12, lod 14, lod 19 and lod 23. Interestingly, all elements are positioned upstream the proximal poly(A) signal. No conserved elements were found between the two poly(A) sites. Analysis of these elements shows a trend in position, nucleotide composition and conservation. With exception of lod 10, all elements are T-rich, and in general, elements positioned closer to the proximal poly(A) signal have higher conservation scores and higher thymine deoxynucleotide contents. The finding that these elements have been conserved in respect to their genomic relative position and content is suggestive of a possible regulatory function.

Lod 14 is the third most conserved element predicted by the *PhastCons* program and the second closest to the poly(A) signal (132 nt upstream of pA1). It has a strong thymine deoxynucleotide tract that is highly conserved between *D. melanogaster* and *T. castaneum.* As these two species are more than 300 million years divergent, such high degree of conservation strongly suggests a functional role for this element. When we analyse the sequences in the immediate vicinity of lod 14, we observe an even more extended T-tract, with a strong conservation up to *D. pseudobscura*. This extended element was designated as USE 1 since it shows a strong similarity with the SV40 late (Schek *et al.,* 1992), *collagen* (Natalizio *et al.,* 2002), *cox-2* (Hall-Pogar *et al.,* 2005), *prothrombin* (Danckwardt *et al.,* 2007) and *C2 complement* (Moreira *et al.,* 1995) USEs. Our *in vitro* results show that this element is necessary for efficient polyadenylation at the proximal poly(A) site through the assembly of a protein complex upstream the poly(A) signal. Moreover, this observation is further reinforced by *in vivo* data, with analysis of *gfp-poloΔUSE 1* transgenic flies. Flies expressing *polo* transcripts in which USE 1 has been deleted show a phenotype similar to flies expressing only the longer *polo* mRNA. This

supports a role for USE 1 in the synthesis of the longer *polo* mRNA in abdominal histoblast, most likely through down-regulation of the proximal poly(A) site.

An USE 2 and USE 3 were also investigated. USE 2 is a GT-rich element (37.5% G, 31.3% T) that shares 50% identity with USE 1 and is conserved between *D. melanogaster* and *D. pseudoobscura* (62% conservation). USE 3 is contained in lod 23, which is a T-rich element (42,9% T) that has the highest conservation score of the five identified conserved elements. Interestingly, lod 23 has a pyrimidine tract that is highly similar to the USE 1 (58% identity). *In vitro* polyadenylation competition assays showed that increasing amounts of unlabelled competitors USE 2 and USE 3 oligonucleotides were able to significantly decrease the efficiency of polyadenylation of wtpA1 pre-mRNA, which indicates that both USE 2 and USE 3 can act as Upstream Sequence Elements.

Upon the identification of these USEs, to address their regulatory role we tried to characterize the proteins that would bind these elements and modulate their function. Through UV cross-linking assays we were able to identify a protein complex assembled upstream the proximal poly(A) site, containing hnRNP C, PTB, CstF-64, αCP2/2KL, U1A, CaCyBP and an unidentified 35 kDa protein (p35). Moreover, we could establish that hnRNP C and PTB bind to both USE 1 and USE 3, while p35 is highly specific for USE 1. Additionally, in UV. cross-linking competition assays, USE 1 oligonucleotide efficiently prevented the binding of p35, hnRNP C and PTB. Although with less efficiency, the same result was obtained with an USE 3 oligonucleotide. Since p35 only binds to USE 1, competition of all three proteins by USE 3 suggests an interaction between these proteins. Although we were not able to co-precipitate p35 with hnRNP C or PTB, this could reflect a weak interaction between p35 and any of these two proteins or, even that these interactions are mediated by an unidentified protein.

We then assayed if the usage of different poly(A) sites would reflect the existence of different protein complexes. Analysis of the proteins assembled onto wtpA1 and wtpA2 pre-mRNAs shows a very similar protein pattern: hnRNP C and PTB bind to both wtpA1 and wtpA2 pre-mRNAs. However, both protein bands exhibit different intensities: hnRNP C seems to bind more strongly to wtpA1, while PTB to wtpA2. Although UV. cross-linking assays are not quantitative, we can not exclude the possibility that these proteins might have different functions in mRNA 3'end formation at the different poly(A) sites. In fact, it has been shown that, depending on its levels, PTB can act as an activatory or inhibitory factor of polyadenylation (Castelo-Branco *et al.,* 2004).

In our assays p35 is specific for wtpA1, more specifically, to the USE 1. Therefore, it is possible that this protein is a specific regulatory factor of polyadenylation at the proximal poly(A) site through the binding to the USE 1. Mapping the protein binding sites in USE 1 shows that the UAUUUGUUUUU sequence constitutes the minimal element

necessary for the binding of p35 and hnRNP C. Mutation of this element in the USE 1-5'mt pre-mRNA is sufficient to reproduce the same results obtained with the USE 1-mt pre-mRNA. This suggests that the function of USE 1 may be restricted to the UAUUUGUUUUU element. Reinforcing this observation is the degree of conservation of this element. The UAUUUGUUUUU has a significant degree of conservation between *D. melanogaster* and *D. pseudoobscura* (91%, 10 out of 11 nt).

Since p35 is specific for wtpA1 and USE 1, we attempted to purify p35 by affinity chromatography. We were able to isolate a strong protein band (band A) in the 25-37 kDa range. Mass spectrometry analysis of this band identified it as Calcyclin Binding Protein (CacyBP). CacyBP is able to bind to the wtpA1 pre-mRNA in a USE 1-dependent manner, since no binding is observed when the USE 1-mt pre-mRNA is used in UV. cross-linking experiments. However, it does not correspond to p35 since CacyBP migrates faster in an SDS-PAGE electrophoresis. Thus, despite not being p35, CacyBP is present in our identified protein complex, most likely through interaction with *polo* USE 1. Binding of CacyBP to the wtpA1 pre-mRNA is only observed with 1 μg of recombinant protein. It is possible that a significant fraction of the recombinant CacyBP used in our experiments is not active and therefore the weak intensity of the protein band may reflect the low amount of active protein capable of binding to the pre-mRNA. It is also possible that CacyBP exhibits a very low RNA binding affinity, which is in agreement with the absence of any RNA binding domain in the protein. Alternatively, and since these assays were performed with only recombinant protein, an argument can be made that binding of CacyBP to RNA is enhanced by other RNA binding proteins which are lacking in this assay. However, we can not discard the possibility that binding of CacyBP to the wtpA1 pre-mRNA may not be physiologically relevant. Therefore, further experiments are necessary to determine the functional role of CacyBP in mRNA 3'end formation at the proximal poly(A) site.

In an attempt to characterize p35, based on its molecular weight and in the putative binding sites present in *polo* USE 1, several proteins were considered as candidates for the p35 protein band. Thus, several antibodies were tested. U1A was identified in one of these attempts. However, we could not detect binding of recombinant U1A to the wtpA1 pre-mRNA which suggests that p35 is not U1A (data not shown). Nevertheless, although U1A does not seem to bind directly to the wtpA1 pre-mRNA, co-immunopreciptation, using an U1A antibody, of five radiolabelled protein bands with molecular weights of approximately 45, 55, 65, 100 and 120 kDa, shows that U1A is present in the protein complex. This is in agreement with several reports that have attributed a role to U1A in mRNA 3'end formation (Castrillon *et al.,* 1993; Lutz and Alwine 1994; Lutz *et al.,* 1996; Lutz *et al.,* 1998; Phillips *et al.,* 2001; Liang and Lutz 2006; Ma *et al.,* 2006; Hall-Pogar *et al.,* 2007). It has been shown for the *cox-2* gene, that

polyadenylation at the proximal poly(A) site is dependent on a tripartite USE (Hall-Pogar *et al.,* 2005) involving a protein complex, containing U1A, $p54^{nrb}$, PTB, and PSF (Hall-Pogar *et al.,* 2007). A similar complex constituted by U1A, $p54^{nrb}$, p68 and PSF was also shown to be involved in mRNA 3'end formation of the SV40 late pre-mRNA (Liang and Lutz 2006). Considering the molecular weight of the protein bands co-immunoprecipitated with the U1A polyclonal antibody, it is tempting to speculate that the 55, 65 and 100 kDa protein bands could correspond to $p54^{nrb}$ or/and PTB, p68 and PSF, respectively. The similarity between the *cox-2* and SV40 late USEs, and between these USEs and *polo* USE 1, lead us to speculate if these protein complexes recognize similar USEs. Interestingly, the recently characterized *prothrombin* USE (Danckwardt *et al.,* 2007) also shows sequence similarities with the USEs of SV40 late and *collagen* and with *polo* USE 1. Again, purification of proteins that interact with *prothrombin* USE identified, among others, $p54^{nrb}$, PTB and PSF with *in vivo* assays establishing a role for these proteins in *prothrombin* mRNA 3'end formation (Danckwardt *et al.,* 2007).

Another protein present in the complex upstream the *polo* proximal poly(A) signal is αCP2/2KL. αCPs are members of the poly(C)-binding proteins (PCBPs) which are RNA-binding proteins with high affinity for poly(C) sequences that have been shown to play multiple functions in gene expression (reviewed in Makeyev and Liebhaber 2002). UV. cross-linking competition assays using increasing amounts of unlabelled poly(rC) oligonucleotide suggested the presence of αCP2/2KL in the protein complex. In these experiments, increasing amounts of poly(rC) oligonucleotide prevented the binding of hnRNP C and p35 to the wtpA1 pre-mRNA. It is unlikely that hnRNP C binds to the poly(rC) oligonucleotides, since this protein exhibits a high affinity for poly(U) polymers and a very low affinity for poly(C) polymers (Swanson and Dreyfuss 1988). Moreover, as determined by the mapping of hnRNP C and p35, these proteins bind to a U-rich element in *polo* USE 1. Thus, these results suggest an interaction between αCP2/2KL and p35 and hnRNP C. These interactions were further confirmed when immunoprecipitation was performed with the FF3 antibody. This antibody specifically recognizes the αCP2 (38.5 kDa) and αCP2KL (35 kDa) isoforms (Chkheidze *et al.,* 1999). Using the FF3 antibody, two ~40 kDa radiolabelled protein bands were precipitated in an USE 1-dependent manner. It is not possible to match these bands with the bands presented by the UV. cross-linking prior to the immunoprecipitation, since they display an altered mobility. Nonetheless, there is a resemblance between the precipitated bands and hnRNP C (C1/C2). Earlier immunoprecipitations identified the hnRNP C as the strong 40 kDa protein bound to the wtpA1 pre-mRNA. However, in general, a second slightly heavier protein band is seen in UV. cross-linking reactions, sometimes overshadowed by the

hnRNP C protein band. This second protein band seems to be hnRNP C2 and the previous identified hnRNP C, the C1 isoform. Therefore, these results seem to suggest an interaction between αCP2/2KL and hnRNP C (C1/C2), which is in agreement with the UV. cross-linking competition results. It is likely that αCP2/2KL binds to the CCCCUUCCCC tract in USE 1. Such close proximity to the p35 and hnRNP C binding sites is in agreement with an interaction between the three proteins. The fact that mutation of the C-tract in the USE 1-3'mt pre-mRNA did not have any effect in the cross-linked protein pattern, as well as the fact that USE 1-(c) RNA did not show any specific protein binding pattern, is most likely due to the labelling of the pre-mRNAs with α-P^{32-}[ATP] and α-P^{32-}[UTP]. As only As and Us were labelled, binding of αCP2/2KL to the CCCCUUCCCC tract would result in very weak labelling, too weak to be detected in our assays.

Although polyadenylation competition assays using USE 2 show a decrease in polyadenylation of wtpA1 pre-mRNA, UV. cross-linking assays with USE 2 and USE 2-mt RNAs seem to support that USE 2 does not contain binding sites for specific proteins. However, since USE 2 is only uniformly radiolabeled in the 5'end (in the first 9 of a 16 nt element), it is possible that a protein interaction might have eluded us. On the other hand, it is also possible that the existence of an RNA structure may account for the activatory role observed for USE 2. RNA annealing activities have shown to co-purify with several known hnRNPs (Portman and Dreyfuss 1994). HnRNP C, αCP2 (hnRNP E) and PTB (hnRNP I) are some of the hnRNPs present in fractions with RNA annealing activities. In the case of hnRNP C, it was shown that this protein is able to strongly promote RNA annealing (Portman and Dreyfuss 1994). Therefore, it is tempting to speculate that the assembly of these proteins in the wtpA1 pre-mRNA promotes the formation of a RNA structure containing USE 2. A secondary structure is in fact predicted by the MFOLD software (Mathews *et al.,* 1999; Zuker 2003) (Fig. 44). Disruption of this RNA structure by addition of an unlabelled USE 2 oligonucleotide competitor, would explain the decrease of polyadenylation in the wtpA1 pre-mRNA.

We further studied the function of the *polo* USEs *in vivo*, both by using tissue cultured cell lines, and transgenic flies. Several mini-genes were generated through individual deletion of each USE. The mini-genes were then used to transfect HeLa cells and the produced mRNAs were analysed by S1 Nuclease. However, we are unable to detect specific bands that matched the expected size for both mRNAs. It is possible that the absence of specific bands may result from the weak nature of both *polo* poly(A) signals. Both proximal and distal poly(A) signals are non-canonical signals that have been shown to be processed with low efficiency compared to the canonical poly(A) signal (Wickens and Stephenson 1984; Wilusz *et al.,* 1989; Sheets *et al.,* 1990). This possibility

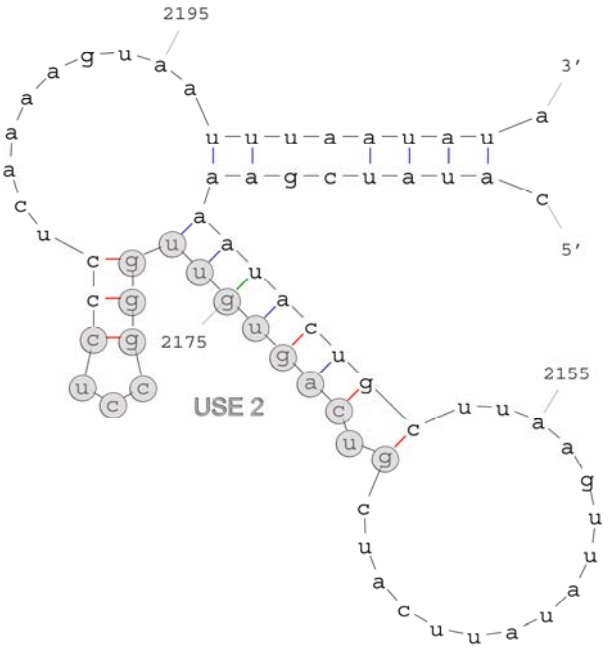

Figure 44. RNA structure predicted for the sequence between USE 1 and USE 3 using the MFOLD software (version 3.2). The RNA structure has a predicted ΔG of -5.9 Kcal/mol**.** The highlighted nucleotides indicate the USE 2. The numbers in the RNA structure correspond to the nucleotide position in the *polo* cDNA (accession number: X63361).

correlates well with the reduced levels of polyadenylation obtained for the wtpA1 pre-mRNA using the *in vitro* system with our prepared HeLa nuclear extracts. Thus, the inefficient recognition and processing of both *polo* poly(A) signals could result in low levels of mRNA. Additionally, it is also possible that the synthesized mRNAs may be unstable in HeLa cells which would also result in low levels of mRNA. Therefore, two different events, by themselves or in combination, may account for the absence of specific bands using the HeLa cell line.

Different results were obtained when the same strategy was approached in a *Drosophila* cell line. Several mini-genes were constructed by individual mutation of each USE and used to transfect Schneider *Drosophila* line 2 (SL2) cells. Using the *Drosophila* system we could detect a sharp strong band that matches the expected size for the usage of the distal poly(A) signal (pA2). We were unable to detect a specific band that results from the usage of proximal poly(A) signal (pA1). Instead, several similarly weak bands were detected in the molecular weight range where we would expect a band resulting from the processing of the proximal site. Despite the fact that both poly(A) signals are non-canonical and weak signals, studies have shown that, *in vitro,* the proximal poly(A) signal AUUAAA is the second most efficiently processed variant of the canonical poly(A) signal (Wilusz *et al.,* 1989; Sheets *et al.,* 1990). It is also the major variant of the AAUAAA hexamer both in humans and *Drosophila* (Graber *et al.,* 1999; Beaudoing *et al.,* 2000; Retelska *et al.,* 2006). Therefore, we expected pA1 to be a stronger poly(A) signal than pA2, and consequently to be processed more efficiently. In fact, this is what we observed

by Northern blot (Llamazares *et al.,* 1991). The lack of any specific band for the usage of pA1 and the strong signal observed for pA2 can be explained if we consider that the smaller mRNA is more unstable in Schneider cells. In this case, despite resulting from the processing of a weaker poly(A) site, the higher half-life of the longer mRNA could account for these results. The detection of several weak bands in the range where a band resulting from the usage of pA1 would be expected, could be the result of a lower stability of the smaller mRNA, with these bands resulting from the degradation of the smaller transcript. However, we cannot discard the possibility that these bands may be the result of inefficient processing of the proximal poly(A) site since in alternative polyadenylation, a strong poly(A) signal is not always a prerequisite to be the most processed site or resulting in the most abundant mRNA (reviewed in Edwalds-Gilbert *et al.,* 1997).

One intriguing result arises from these *in vivo* experiments. A negative control for the proximal poly(A) signal was generated by point mutation of the proximal signal AUUAAA into GUUAAC. Considering previous studies on poly(A) signal variants and their efficiency in mRNA 3'end processing (Wickens and Stephenson 1984; Wilusz *et al.,* 1989; Sheets *et al.,* 1990; Graber *et al.,* 1999; Beaudoing *et al.,* 2000), such mutation is expected to inhibit the use of the proximal poly(A) site. Since we are not able to detect mRNAs processed at the proximal poly(A) site, we could not observe the effect of this mutation. Nevertheless, it is interesting to see that mutation of the proximal poly(A) signal seems to inhibit processing at the distal poly(A) site. This suggests that recognition of the distal poly(A) signal is dependent of a functional proximal signal.

Transcriptional termination is dependent on a functional poly(A) site, and a well studied example of this is the *α2-globin* poly(A) signal and the development of α-thalassaemia (Higgs *et al.,* 1983; Whitelaw and Proudfoot 1986). Mutation of the *α2-globin* poly(A) signal reduces the *α2-globin* mRNA levels in erythroid cells and readthrough of transcripts extending beyond the normal poly(A) addition site is detected due to inefficient transcription termination by RNA polII (Higgs *et al.,* 1983). The intergenic region between the *polo* gene and the downstream *snap* gene is very small (184 nt) and, therefore, transcription of *polo* must efficiently terminate to prevent transcriptional interference with *snap*. It is formally possible that the presence of two poly(A) signals in the *polo* 3'UTR may be required for efficient transcription termination. The two poly(A) signals could be required for slowing down the elongating RNA polymerase II, allowing recognition and processing of both poly(A) signals. Mutation of the proximal poly(A) signal would prevent the slowing down of RNA polymerase II and recognition of the distal poly(A) signal, resulting in decreased levels of the longer mRNA. However, we do not favour this hypothesis due to two observations: one is that we fail to detect an increase in the readthrough transcript by RNA polymerase II in our S1 Nuclease assays. Second, if

mutation of the proximal poly(A) signal prevents transcription termination, both *gfp-polo∆pA1; polo*^{9} *and gfp-polo∆pA2; polo*^{9} flies should present the same abdominal phenotype. Moreover, if both poly(A) signals are required for transcription termination, we would expect the *gfp-polo∆pA2; polo*^{9} phenotype observed for the different lines to be different as a result from transcriptional interference with different downstream genes. Therefore, our analysis of the transgenic flies clearly indicates that the phenotype observed is due to incorrect pre-mRNA processing and not transcription termination.

Our results suggest that of all USEs present in the *polo* 3'UTR, USE 1 is likely to be the major regulatory element involved in 3'end formation at the proximal poly(A) site. Several evidences seem to support this observation. First, mutation of USE 1 inhibits polyadenylation at the proximal poly(A) site. Second, USE 1 is the major binding site for p35, hnRNP C and most likely αCP2/2KL. Third, p35 binds specifically to USE 1 and finally, of all proteins bound upstream both poly(A) signals, only p35 is specific for the proximal poly(A) site. Moreover, these observations are supported with the results obtained from the study of *gfp-polo∆USE 1; polo*^{9} flies. Expression of the *gfp-polo∆USE 1* transgene rescues the mitotic phenotype and the larval lethality of *polo*^{9} mutants, just as observed with *gfp-polo* and *gfp-polo∆pA1*. However, flies expressing *polo* mRNAs in which USE 1 has been deleted (*gfp-polo∆USE 1; polo*^{9}) show the same abdominal phenotype as flies expressing only the longer *polo* mRNA (*gfp-polo∆pA1; polo*^{9}). Therefore it seems that USE 1 is required for the synthesis of the small *polo* transcript. The similarity between the phenotypes of *gfp-polo∆pA1* and *gfp-polo∆USE 1* flies is consistent with a role for USE 1 in activation of polyadenylation at the proximal poly(A) site. Moreover, the specific binding of p35 to the UAUUUGUUUUU element in USE 1 plus the high degree of conservation of this element between *D. melanogaster* and *D. pseudoobscura* (10 out 11 nt; 91% conservation) further supports this hypothesis. Therefore, it is possible that USE 1 functions as a regulatory element involved in alternative polyadenylation in abdominal histoblasts. Downregulation of the proximal poly(A) site may be a simple mechanism through which, via USE 1, the usage the distal poly(A) site might be favoured.

PART III

GENERAL DISCUSSION

General Discussion:

The only available data regarding *polo* mRNA expression during development comes from Llamazares *et al.* (1991). The authors showed that *polo* has two transcripts, both expressed throughout development in a very similar way, being particularly abundant during embryogenesis and to a lesser extent in pupae and adult females. During embryo development the highest levels of *polo* mRNAs are observed during the first four hours (Llamazares *et al.,* 1991) which reflects the need for Polo in the initial mitotic divisions (Moutinho-Santos *et al.,* 1999; reviewed in Gilbert 2000). During the pupa stage, proliferation of a subset of imaginal cells takes place to form adult structures (reviewed in Gilbert 2000), while the levels of *polo* seen in adult females are probably due to accumulation of *polo* mRNAs in the oocyte which will be used in the early steps of embryogenesis (Sunkel and Glover 1988). This expression pattern reflects and is in agreement with the requirement of Polo during the proliferative stages of development (reviewed in Gilbert 2000; Glover 2005).

The expression of both *polo* mRNAs during development suggests that most likely, expression of Polo is not developmentally regulated (Llamazares *et al.,* 1991). Instead, it suggests that the synthesis of these mRNAs could be tissue-specific. Furthermore, it indicates that both transcripts may have different properties. From these early experiments it is possible to observe that the small 2.2 kb mRNA is almost always twice more abundant than the long 2.5 kb mRNA (Llamazares *et al.,* 1991). Therefore, these results suggest that the levels of the smaller *polo* mRNA could specify a more important requirement or role for this transcript. Nevertheless, we cannot discard the hypothesis of a possible redundancy between both transcripts. In this hypothesis, the different levels of both transcripts could easily be explained from close analysis of both poly(A) signals. Although both poly(A) signals are non-canonical, the proximal poly(A) signal, AUUAAA, is the most common variant of the AAUAAA signal both in humans (Beaudoing *et al.,* 2000) and in *D. melanogaster* (Graber *et al.,* 1999; Retelska *et al.,* 2006), whereas the distal poly(A) signal, AAUAUA, shows a higher divergency from the canonical signal, being present only in 5.1% of *Drosophila* genes (Retelska *et al.,* 2006). Moreover, the proximal poly(A) signal is stronger compared to the distal poly(A) signal (Wilusz *et al.,* 1989; Sheets *et al.,* 1990). This and the fact that the proximal signal is the first signal to be transcribed by the elongating RNA polymerase II and, therefore, the first to be exposed to the 3'end processing complex that travels coupled to the elongating polymerase (reviewed in Proudfoot *et al.,* 2002; Proudfoot 2004; Buratowski 2005), are arguments that could explain why the smaller *polo* mRNA is the most abundant transcript. If there is redundancy between both transcripts, the efficiencies of both poly(A) signals plus their position in the

3'UTR, could be responsible for the different expression levels observed for the *polo* transcripts. The results obtained from our study support a mix of both hypotheses, *i.e*, there appears to be redundancy between both transcripts in most proliferating tissues with the exception of a specific requirement for the long 2.5 kb transcript in the abdomen precursor cells, the abdominal histoblasts. This hypothesis is supported by the finding that flies expressing only the smaller *polo* mRNA (*gfp-polo⊿pA2; polo*^{9}) are less viable and that in these flies, abdominal histoblasts fail to undergo proliferation during metamorphosis. The fact that no other clear phenotype is observed in flies expressing the smaller mRNA (*gfp-polo⊿pA2; polo*^{9}) and the observation that, in all tissues analysed (third instar brains, testis and ovaries), flies expressing only the smaller mRNA (*gfp-polo⊿pA2*) have higher levels of GFP-Polo, is in agreement with the possibility that the usage of the proximal and distal poly(A) signals might be redundant in most proliferative tissues, with the proximal poly(A) signal being favoured. Moreover, the viability and the apparent normal development of flies expressing the long 2.5 kb mRNA (*gfp-polo⊿pA1; polo*^{9}) suggests a redundancy of both *polo* transcripts. In addition, *gfp-polo⊿pA1; polo*^{9}*/Df (3L) rdgC-co2* flies expressing only the longer *polo* mRNA are viable and fertile, which further supports the redundancy of both *polo* mRNAs in the majority of the proliferating tissues.

One would expect that during the course of evolution, the system had evolved to select only the proximal poly(A) signal. On the other hand, it is possible that the very nature of the abdominal histoblasts cell cycle, with its long G2 arrest and their rapid proliferation during the pupa stage (Garcia-Bellido and Merriam 1971; Guerra *et al.,* 1973; Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979; Ninov *et al.,* 2007), may require a special feature provided by the longer mRNA. The fact that flies expressing only the smaller *polo* mRNA (*gfp-polo⊿pA2*) have lower levels of GFP-Polo in abdominal histoblasts during metamorphosis, is in agreement with this hypothesis. One possibility is that the synthesis of the longer mRNA could be coupled to higher translation efficiency. Analysis of both *polo* cDNAs, shows that besides having different 3'ends, both transcripts also have different 5'UTRs (Llamazares *et al.,* 1991). As a result of alternative transcription start sites, the longer mRNA also has a longer 5'UTR. It is possible that RNA-binding proteins expressed or active in abdominal histoblasts could bind to specific 5' and 3'UTR elements, resulting in a more efficient cross-talk between both ends and consequently in an enhanced translational activity. Alternatively, the short transcript could be more unstable than the longer *polo* mRNA. In other tissues or cell types a more unstable transcript may not represent a problem since it could be compensated by higher transcription rates or by the long 2.5 kb transcript. But in cells that have, among the imaginal cells, the highest proliferation rate (Madhavan and Schneiderman 1977;

Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007), the instability of the smaller mRNA may present a problem to the proliferating histoblasts. In such a scenario, the longer *polo* mRNA could become the limiting factor in the synthesis of Polo and of rapid cell divisions. An analysis of the 3'UTR of *polo* reveals several elements similar to the AU-rich elements (AREs) involved in mRNA stability (reviewed in Barreau *et al.,* 2005). Upstream the proximal poly(A) signal the UCAUUUAUA*,* UAAUUUAAU and AUUUA elements resemble the class I (AUUUA) and class II (UUAUUU(A/U)(A/U)) AREs respectively. Upstream the proximal poly(A) site four UUUU/U sequences are also present. This sequence, that match the described binding site of hnRNP C, has been shown to be involved in mRNA stability (Wilusz and Shenk 1990; Hamilton *et al.,* 1993; Gorlach *et al.,* 1994). Moreover, hnRNP C has also been described as an AUBP (ARE binding protein) (Hamilton *et al.,* 1993). Furthermore, downstream the proximal poly(A) site, two AUUUA elements and two UUUU elements are also present. The binding of hnRNP C, both upstream and downstream the proximal poly(A) site, supports the hypothesis of different mRNA stabilities for both *polo* mRNAs. As previously mentioned, UV. cross-linking assays are not quantitative. However, the different protein band intensities observed upon the binding of hnRNP C to the wtpA1 and wtpA2 pre-mRNAs raises the possibility that the levels of hnRNP C, as well as the affinity for its binding sites, may play a role in the stability of these mRNAs. This hypothesis is further supported by the presence of αCP2/2KL and its interaction with hnRNP C. αCP2/2KL has been extensively studied as a protein factor involved in mRNA stability (reviewed in Makeyev and Liebhaber 2002). Although our results suggest a role for these proteins in 3'end formation at the proximal poly(A) site, their precise function in this process cannot be determined from our experiments. Thus, we cannot exclude the possibility that hnRNP C and αCP2/2KL may play a role in mRNA stability as well as in 3'end formation. Hence, it is possible that both mRNAs may have different mRNA stabilities that could be required for the regulation of the levels of Polo in the abdomen precursor cells.

Whatever the feature(s) responsible for the requirement of the longer *polo* mRNA in abdominal histoblasts, the synthesis of this transcript is likely to be coupled with the regulation of alternative polyadenylation. It is possible that in abdominal histoblasts a regulatory mechanism could have developed to ensure the synthesis of the longer mRNA. Enhancement of the recognition of the distal poly(A) site or simply, the downregulation of the proximal site, are mechanisms by which tissue-specificity can be achieved. The presence of USEs positioned upstream the proximal poly(A) site may be required for the regulation of poly(A) site usage in abdominal histoblasts. Our results show that these elements are required, *in vitro*, for efficient 3'end formation at the proximal poly(A) site. As

several reports have shown, poly(A) site usage is usually coupled with modulation of levels of protein factors involved in 3'end formation or/and competition for the same regulatory elements (Takagaki *et al.,* 1996; Takagaki and Manley 1998; Brown and Gilmartin 2003; Castelo-Branco *et al.,* 2004; Venkataraman *et al.,* 2005; Gawande *et al.,* 2006; Kubo *et al.,* 2006). As shown for the *enhancer of rudimentary*, usage of the distal poly(A) site in the female germline is achieve through downregulation of the proximal site by means of competition between Sex-lethal and CstF for the binding to the GU-rich elements present downstream of the proximal poly(A) site (Gawande *et al.,* 2006). A similar situation may occur for *polo* in abdominal histoblasts. It is possible that expression of specific protein factor(s) in abdominal histoblasts may compete for the binding to the *polo* USEs, in particular to the USE 1. Such a competition mechanism would prevent the assembly of the protein complex required for 3'end formation at the proximal poly(A) site resulting in the downregulation of the proximal site favouring the synthesis of the longer *polo* mRNA. Supporting this hypothesis are the results obtained with *gfp-poloΔUSE 1* flies. Flies expressing *polo* mRNA in which USE 1 has been deleted (*gfp-poloΔUSE 1; polo*^{9}), have a phenotype that is very similar to the phenotype of flies expressing only the longer *polo* mRNA (*gfp-poloΔpA1; polo*^{9}). The absence of an available USE 1 appears to favour the synthesis of the longer *polo* mRNA. Therefore, these results suggest that in abdominal histoblasts competition for the binding to USE 1 could represent a possible mechanism through which the recognition and usage of the distal poly(A) site could be favoured.

Additionally, it is also possible that modulation of the levels of the proteins involved in 3'end formation at the proximal poly(A) site might regulate the synthesis of the longer *polo* mRNA. It has been shown that, depending on its levels, PTB can have different functions in 3'end processing (Castelo-Branco *et al.,* 2004). While low levels of PTB enhance 3'end processing, higher levels downregulate it (Castelo-Branco *et al.,* 2004). Despite the fact that in Castelo-Branco *et al.* (2004), the function of PTB in 3'end processing is the result of a competition with CstF-64, it is possible that similar processes may occur in abdominal histoblasts during the synthesis of the longer *polo* mRNA. UV. cross-linking assays using wtpA1 and wtpA2 pre-mRNAs show that, with exception of p35 that is highly specific for USE 1, the majority of the proteins are able to bind upstream both poly(A) signals. It is interesting that hnRNP C and PTB present different protein band intensities whether they are bound to wtpA1 or wtpA2 pre-mRNAs. We can speculate that depending on the nature of the protein-protein interactions and the affinity of the different proteins to the corresponding RNA binding sites, the synthesis of the longer *polo* mRNA in histoblasts could be achieved by regulating the levels of the protein factors involved in 3'end processing of both poly(A) sites. Again, it is possible that the USEs may play a role

in the synthesis of the longer *polo* mRNA in abdominal histoblast by favouring the usage of the distal poly(A) site through downregulation of the proximal site.

In addition, other mechanisms may act simultaneously to ensure the synthesis of the longer mRNA. It is possible that upregulation of the usage of the distal poly(A) site may occur as well. The presence of U1A upstream the proximal poly(A) site supports this view. U1A has been shown to bind the USE present in the SV40 late pre-mRNA and to enhance 3'end formation (Lutz and Alwine 1994). Moreover, it was shown that such effect occurs through interaction with CPSF-160 resulting in an enhanced binding of CPSF to the pre-mRNA (Lutz *et al.,* 1996). However, high levels of U1A inhibit 3'end formation, most likely through interaction with PAP as observed in the auto-regulation of the U1A mRNA (Gunderson *et al.,* 1994; Lutz *et al.,* 1996). Although there is still some controversy, it has been reported that the levels of U1A may play an important role in the expression of the secreted-form of IgM (Phillips *et al.,* 2001; Ma *et al.,* 2006). In these studies a model has been proposed in which a decrease in the levels of non-snRNP U1A during B cell differentiation into plasma cells results in an increased synthesis of the μ_s mRNA. Two elements, UCUGUAGCU and AUUUGUAAA, positioned 39 and 19 nt upstream the *polo* distal poly(A) signal, show a high similarity with the SV40 late USE (AUUUGUGAA, AUUUGUAAC, AUUUGUGAU). We have no data regarding the binding of U1A to these elements. However, in light of the examples presented, it would be interesting to determine the function of these elements as putative binding sites for U1A and the relationship between these elements and poly(A) signal selection.

Recent reports have suggested that Cleavage Factor I_m (CFI_m) is a key complex in alternative polyadenylation (Brown and Gilmartin 2003; Venkataraman *et al.,* 2005; Kubo *et al.,* 2006). This complex was shown to be one of the main protein factors required for canonical and non-canonical poly(A) site recognition (Brown and Gilmartin 2003; Venkataraman *et al.,* 2005). CFI_m binds specifically to UGUAN (N = A > U ≥ C/G) elements upstream the poly(A) site and directs the assembly of 3'end processing complexes through the recruitment of CPSF and PAP via the CPSF subunit hFip1 (Brown and Gilmartin 2003; Venkataraman *et al.,* 2005). Moreover, levels of CFI_m seem to be important, since low levels are required for stimulation of 3'end formation while high levels have an inhibitory role (Brown and Gilmartin 2003). Furthermore, RNAi of the 25 kDa subunit of CFI_m (CFI_m25), shows that CFI_m plays a major role in alternative polyadenylation at the 3'most exon of several genes, with CFI_m being required for the usage of distal poly(A) sites (Kubo *et al.,* 2006). In *polo*, three UGUAN elements (UGUAC, UGUAG and UGUAA) are present upstream the distal poly(A) signal, organized in a similar arrangement as in genes shown to have CFI_m binding sites (Brown and Gilmartin

2003; Venkataraman *et al.,* 2005). Whether these CFI_m binding sites play a role in alternative polyadenylation of *polo* is a subject of future work.

Our characterization of the *cis*-acting elements and the corresponding *trans*-acting factors involved in 3'end formation of the *polo* mRNA was performed using a HeLa cell system. However, the high degree of homology, both structurally and functionally, between mammals and *Drosophila* mRNA processing protein factors, provides o good level of confidence in our results (Benoit *et al.,* 1999; Mount and Salz 2000; Murata *et al.,* 2001; Benoit *et al.,* 2002). Supporting our observations is the result obtained with *gfp-polo∆USE 1* transgenic flies. Moreover, the *Drosophila* homologues of the proteins identified in this study share a high percentage of identity with the human proteins, with values ranging from 37-49% for the majority of the proteins, exception made for the *Drosophila* hnRNP C and U1A as both proteins have 16% and 70% identity, respectively.

So what is it about Polo that makes the abdominal histoblasts development so dependent on the longer *polo* mRNA? The answer to this question could lie in the specific nature of the cell cycle of histoblasts. The control of proliferation of imaginal cells is of the outmost importance in *Drosophila* development. During embryogenesis, after cell proliferation ceases by the end of the embryonic stage 5, the larva cells will cease dividing and begin cycles of endoreplication and in the process become polytene cells. The precursor imaginal cells, remain diploid and undifferentiated and in later stages of embryogenesis become committed to an imaginal state, reinitiating proliferation during the larva development or after the onset of metamorphosis. The timing of the reinitiation of the cell cycle and the extent of proliferation vary accordingly to the different populations of imaginal cells (Madhavan and Schneiderman 1977). Abdominal histoblasts arise in the last stages of embryogenesis and contrary to the development of the imaginal discs which arise from invaginations of the embryonic epithelium, the histoblasts are intimately associated with the larval polytene cells (Garcia-Bellido and Merriam 1971; Guerra *et al.,* 1973; Madhavan and Schneiderman 1977). The location of the abdominal histoblasts nests in the larva integmentum, surrounded by larval polytene cells must require a tight control of the cell cycle by the histoblasts. Failure to maintain the G2 arrest and premature entry into mitosis could lead to early proliferation of the histoblasts during the larval stages and thus to disrupt the larva epidermis by extrusion of the larvae epidermal cells. Directed expression of a dominant-negative form of the Ecdysone receptor in LECs during metamorphosis, prevented extrusion and apoptosis of these cells, which lead to the presence of LECs in the epidermis of the pharate adults, thus preventing abdominal morphogenesis (Ninov *et al.,* 2007). Moreover, inhibition of apoptosis in LECs during the pupa stage prevented histoblast proliferation with some histoblasts undergoing apoptosis (Ninov *et al.,* 2007)

There are several studies that show the need to correctly control the mechanisms required for the maintenance of the G2 arrest observed in histoblasts and to tightly control the entrance in mitosis during the pupa stage (Hayashi *et al.,* 1993; Fuse *et al.,* 1994; Curtiss and Heilig 1995). Mutants for the transcription factor *escargot* display problems in abdomen formation (Hayashi *et al.,* 1993; Fuse *et al.,* 1994). Interestingly, the phenotype of *esg* mutants is very similar to the phenotype of flies expressing only the short *polo* mRNA (*gfp-polo⊿pA2; polo*^{9}). Initial studies showed that in *esg* mutants, abdominal histoblasts fail to remain diploid, enter in an endocycle and become polytene cells (Hayashi *et al.,* 1993; Fuse *et al.,* 1994). As a result, these cells fail to undergo proliferation during the pupa stage and consequently adult individuals show defects in the adult epidermis. Further studies showed that *escargot* genetically interacts with *cyclin A* and *cdk1/cdc2* (Hayashi 1996). Moreover, *cdc2* mutants show a phenotype similar to *esg* mutants, with polytene abdominal histoblasts as a result of several rounds of DNA endoreplication (Hayashi 1996). Co-expression of Cyclin A and Cdc2 in salivary glands prevent endoreplication and growth of third instar larvae salivary glands and in a *cdc2* mutant background, expression of both proteins rescued the histoblast polytene phenotype (Hayashi and Yamaguchi 1999). Thus, a tight control over the cell cycle, in particular the correct control of the G2 arrest through inhibition of S phase and by preventing early mitosis, is essential for proper abdominal morphogenesis.

It has been shown that Polo is required during mitosis for centrosome maturation, microtubule nucleation, metaphase-anaphase transition and cytokinesis (reviewed in Glover 2005). Taking into consideration the multiple functions of Polo during mitosis, it is feasible that lower levels of Polo in abdominal histoblasts may impar the rapid proliferarion of these cells during metamorphosis. However, not much is known regarding Polo function outside the mitotic cycle. The protein phosphatase encoded by *cdc25*/*string*, is one of the main regulators of the entry into mitosis (reviewed in Bruce Alberts 2002). In the G2/M transition it dephosphorylates Cdk1 activating the Cyclin B/Cdk1 complex required for entry into mitosis (reviewed in Bruce Alberts 2002). One of the functions attributed to Polo occurs during the G2/M transition (Kumagai and Dunphy 1996; Qian *et al.,* 1998). It was reported that the *Xenopus* Polo homologue, Plx1, is co-purified with XCdc25 (Kumagai and Dunphy 1996). In this study, Plx1 was shown to phosphorylate XCdc25 *in vitro* and to activate its phosphatase activity. Moreover, injection of antibodies against Plx1 in *Xenopus* oocytes delayed phosporylation of XCdc25 and activation of Cyclin B/Cdk1 (Qian *et al.,* 1998). Another study, by Hayashi (1996), showed that ectopic expression of *string* was sufficient to induce premature entry into mitosis in histoblasts, suggesting that Cdc25/String is required for activation of the Cyclin B/Cdk1 complex. It is possible that just as suggested in the *Xenopus* system, in histoblast cells Polo is required for activation of

Cdc25 which would explain that despite being present at the onset of metamorphosis, histoblasts fail to proliferate. *polo*^{2} (Sunkel and Glover 1988) and *polo*^{9} (Donaldson *et al.,* 2001) mutants are characterized by a pro-metaphase/metaphase arrest which would not be expected if Polo would normally be required for activation of String. However, it is possible that, due to the particular nature of the cell cycle of histoblasts, a Polo-Cdc25 interaction might be tissue and developmental-specific. Since *polo*^{2} (Sunkel and Glover 1988) and *polo*^{9} (Donaldson *et al.,* 2001) alleles are lethal at third instar stage, such an interaction may have been eluded so far.

Whatever the precise nature of the role of Polo in abdominal morphogenesis, the control of histoblast proliferation is very tight. The *arrowhead* gene encodes a LIM-homeodomain transcription factor that seems to be necessary for the establishment or proliferation of abdominal histoblasts (Curtiss and Heilig 1995; Curtiss and Heilig 1997). Among other phenotypes, *awh* mutants show an abnormal abdomen formation as a result of abdominal histoblast nests containing only half the number of cells compared to wild-type (Curtiss and Heilig 1995). The inability of histoblasts to compensate the number of cells present at the beginning of metamorphosis indicates that histoblasts go through a well coordinated proliferation during the pupa stage. Such coordination is also apparent when histoblast proliferation is analysed. Histoblasts begin to proliferate 1-3 hours APF undergoing rapid proliferation until 12-15 hours APF (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980). During this time, cells proliferate co-ordinately, with the same spindle orientation, and with and average doubling time of 2-3 hours, dividing three times as fast as imaginal cells at any stage in postembryonic development (Madhavan and Schneiderman 1977; Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). After 15 hours APF, histoblast begin to expand from their original locations and begin replacing LECs (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). During this period and until the fusion of opposing hemi-segments, around 40 hours APF, the proliferation rate decreases with doubling times between 6 and 9 hours (Roseland and Schneiderman 1979; Madhavan and Madhavan 1980; Ninov *et al.,* 2007). Such rapid proliferation during the first 15 hours APF is a result of histoblast undergoing very short G1 phase (Ninov *et al.,* 2007). An increase in 60% cell volume during the larva stages allows histoblasts to rapid proliferate upon entrance in metamorphosis and to compensate the rapid decrease in cell volume during this time period (Madhavan and Schneiderman 1977; Ninov *et al.,* 2007). It is likely that in cells undergoing such rapid proliferation, Polo may play an important role during the several mitotic cycles. Insufficient levels of Polo may impair proliferation during this developmental period resulting in failure to correct develop the adult epidermis. This

hypothesis is supported by the observation that *polo*^{1}/*polo*^{2} individuals show an abnormal abdominal development, affecting both tergites and sternites (Sunkel and Glover 1988).

In conclusion, the work presented in this study shows that the long 2.5 kb *polo* mRNA is specifically required for proliferation of abdominal histoblasts during metamorphosis. Moreover, the expression of this mRNA in these cells is coupled to higher levels of Polo. Whether, this is due to the fact that the longer *polo* mRNA has a higher translation rate or a higher mRNA stability, further experiments are necessary to answer these questions. Finally, although our data does not show that expression of the longer *polo* mRNA in abdominal histoblasts results from alternative polyadenylation, it does however suggest that *cis*-acting elements present upstream the proximal poly(A) site may be involved in a regulatory mechanism that ensures the synthesis of the long 2.5 kb *polo* mRNA in abdominal histoblasts.

PART IV

MATERIALS AND METHODS

A. Materials

A1. Enzymes

The restriction enzymes used were obtained from New England Biolabs, Invitrogen and BIORON. Other enzymes were obtained from the following suppliers:

Alkaline Phosphatase, from calf intestinal	(Roche Applied Science)
DFS-Taq DNA polymerase (DNA-minus)	(BIORON)
DNase I (bovine pancreas) RNase free, FPLC pure	(GE Healthcare Life Sciences)
Klenow	(Roche Applied Science)
Lysozyme	(SIGMA)
Proteinase K	(SIGMA)
RNAguard™ Ribonuclease Inhibitor (Human Placenta)	(GE Healthcare Life Sciences)
RNase A	(SIGMA)
S1 Nuclease (*Aspergillus oryzae*)	(GE Healthcare Life Sciences)
SP6 RNA polymerase	(Roche Applied Science)
T4 DNA Ligase	(Invitrogen)
T4 Polynucleotide Kinase	(GE Healthcare Life Sciences)
T7 RNA polymerase	(Roche Applied Science)
Vent$_R$® DNA polymerase	(New England Biolabs)

A2. Kits

BioNick™ Labeling System	(Invitrogen)
ECL Plus™ Western Blot Detection Reagents	(GE Healthcare Life Sciences)
GENECLEAN® Turbo Kit	(Q-BIOgene)
Illustra ProbeQuant™ G-50 Micro Columns	(GE Healthcare Life Sciences)
PhoenIX™ Midiprep Kit	(Q-BIOgene)
RapidPURE™ Plasmid Mini Kit	(Q-BIOgene)
Ready-To-Go™ DNA Labeling Beads (-dCTP)	(GE Healthcare Life Sciences)

A3. Media

With exception of mediums that were purchased, all mediums used were prepared accordingly to Sambrook and Russell (2001).

LB Medium (Luria-Bertani Medium)

Per liter:

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Adjust the pH to 7.0 with 5 N NaOH

LB Agar

Per liter of LB:

Agar	15 g

2x TY Medium

Per liter:

Tryptone	16 g
Yeast extract	10 g
NaCl	5 g

Adjust pH to 7.0 with 5 N NaOH

Ampicillin plates

Autoclaved LB Agar was melted and ampicillin added to a final concentration of 50 μg/ml.

D-MEM (Dulbecco's Modified Eagle's Medium)

D-MEM, with 4500 mg/L D-glucose, Non Essential Amino Acid (NEAA), without L-glutamine and sodium pyruvate was obtained from Invitrogen. This medium was supplemented with 10% (v/v) FBS (Invitrogen), 2mM L-glutamine (Invitrogen), 1mM sodium piruvate (Invitrogen) and 50 U/ml penicillin and 50 μg/ml streptomycin (Invitrogen).

Schneider's Insect Medium

Schneider's Insect Medium (with L-glutamine and sodium bicarbonate) was obtained from Invitrogen and supplemented with 10% (v/v) FBS (Invitrogen)

Fly Media

Per 3 liters:

Mixture A

ddH_2O	2000 ml
Honey	103 ml
Agar	24 g

Mixture B

ddH_2O	1000 ml
Cornmeal	240 g
Yeast extract	54 g
Flour of soy	30 g
Malte	60 g

After boiling mixture A, the mixture B was added and let to boil for 35 min. The mixture was let to cool down at RT to a temperature bellow 60ºC at which point 17.4 ml of a propionic acid: 85% (v/v) phosphoric acid (15:1) solution was added.

A4. General solutions and buffers

The majority of the solutions and buffers used were prepared accordingly to Sambrook and Russell (2001).

Buffer D

20 mM Hepes-NaOH pH 7.9
20% (v/v) glycerol
50 mM $(NH_4)_2SO_4$
0.2 mM EDTA
0.5 mM PMSF
0.5 mM DTT

50x Denhardt's solution

1% (w/v) Ficoll 400
1% (w/v) PVP
1% (w/v) BSA (fraction V) (SIGMA)

DNA 6x gel-loading buffer (type II)

0.25% (w/v) bromophenol blue

0.25% (w/v) xylene cyanol FF

15% (w/v) Ficoll (Type 400; Pharmacia) in H_2O

G7654 6 x Gel Loading Solution (Type I) (SIGMA)

0.25% (w/v) bromophenol blue

0.25% (w/v) xylene cyanol FF

40% (w/v) sucrose

1x Phosphate-buffered Saline (PBS)

137 mM NaCl

2.7 mM KCl

10 mM Na_2HPO_4

2 mM KH_2PO_4

RNA gel-loading buffer

80% (v/v) formamide

10 mM EDTA pH 8.0

Gel G7654 1x Gel Loading Solution (Type I) (SIGMA)

RNase-free solutions

RNase-free solutions were prepared by incubating solutions with 0.1% DEPC (SIGMA) at 37ºC, ON. The solutions were then autoclaved to destroy the DEPC. Solutions that were not possible to treat with DEPC, were prepared using DEPC-treated water and RNase-free reagents.

2x SDS Gel-loading buffer

100 mM Tris-HCl pH 6.8

4% (w/v) SDS (electrophoresis grade)

0.2% (w/v) bromophenol blue

20% (v/v) glycerol

200 mM β-mercaptoethanol

20x SSC

3 M NaCl

0.3 M sodium citrate

Adjust pH to 7.0 with HCl

20x SSPE

3 M NaCl

0.2 M NaH_2PO_4

20 mM EDTA

Adjust pH to 7.4 with NaOH

STET

10 mM Tris-HCl pH 8.0

0.1 M NaCl

1 mM EDTA pH 8.0

5% (v/v) Triton X-100

50x TAE (Tris-Acetate-EDTA)

Per liter:

242 g Tris base

57.1 ml glacial acetic acid

100 ml 0.5 M EDTA pH 8.0

5x TBE (Tris-Borate-EDTA)

Per liter:

54 g Tris base

27.5 g boric acid

20 ml 0.5 M EDTA pH 8.0

A5. Radioisotopes

All radioisotopes were obtained from GE Healthcare Life Sciences:

. α-^{32}P-[dATP], 10 mCi/ml (3000 Ci/mmol)

. α-^{32}P-[dTTP], 10 mCi/ml (3000 Ci/mmol)

. α-^{32}P-[dGTP], 10 mCi/ml (3000 Ci/mmol)

. α-^{32}P-[dCTP], 10 mCi/ml (6000 Ci/mmol)
. α-^{32}P-[ATP], 10 mCi/ml (400 Ci/mmol)
. α-^{32}P-[UTP], 10 mCi/ml (400 Ci/mmol)

A6. Antibodies

The antibodies used were kind gifts from several laboratories:

. *rabbit polyclonal anti-PTB antibody* (and the rabbit pre-immune serum): Prof. Christopher W. J. Smith, Department of Biochemistry, University of Cambridge (Kaminski *et al.,* 1995).
. *mouse monoclonal anti-CstF-64 antibody*: Prof. James L. Manley, Department of Biological Sciences, Columbia University (Takagaki *et al.,* 1990).
. *mouse monoclonal U33 anti-Headcase antibody*: Prof. Robert A. H. White, Department of Physiology, Development and Neuroscience, University of Cambridge (Weaver and White 1995).
. *rabbit polyclonal anti-Cyclin B antibody*: Prof. Christian F. Lehner, Department of Genetics, BZMB, University of Bayreuth (Jacobs *et al.,* 1998).
. *mouse monoclonal MA294 Polo antibody*: Prof. Claudio E. Sunkel, Department of Molecular Genetics, Instituto de Biologia Molecular e Celular, University of Porto (Llamazares *et al.,* 1991).
. *mouse monoclonal 4F4 anti-hnRNP C antibody*: Prof. Gideon Dreyfuss, Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine (Choi and Dreyfuss 1984).
. *rabbit polyclonal anti-U1A antibody*: Prof. Iain W. Mattaj, EMBL Heidelberg (Kambach and Mattaj 1992).
. *rabbit polyclonal FF3 anti-αCP2/αCP2KL antibody*: Prof. Stephen A. Liebhaber, Department of Genetics and Medicine, University of Pennsylvania (Chkheidze *et al.,* 1999).

A7. Oligonucleotides

The oligonucleotides used were obtained from Amersham Pharmacia Biotech, MWG-Biotech AG, and GIBCO BRL and Invitrogen (see Appendix, Table AI).

B. Methods

B1. Plasmid constructions

The cloning strategies used to generate the various constructs are shown in Tables AII-AV (see Appendix). The pBs-*polo* and pxb7 plasmids were previously described in Llamazares *et al.* 1991 and the pW8-*gfp-polo* plasmid described in Moutinho-Santos *et al.* 1999.The wtpA1 pre-mRNA was synthesized by *in vitro* transcription using either the pSPT19-*polo* wt (Figs. 21-23 and 28), the pGEM7-*polo* wt (Figs. 24-26, 29, 30 (lane 3), 34 and 36) or the pGEM7-*polo* wt-DSE (Figs. 30 (lane 1), 31, 35, 37 and 38) constructs. The USE 1-mt pre-mRNA was synthesized by *in vitro* transcription using the pSPT19-*polo* mt (Figs. 21-23), pGEM7-*polo* mt (Figs. 24, 26 and 34) or the pGEM7-USE 1mt (Figs. 31, 35, 37 and 38) constructs. The USE 1, USE 2, USE 3, wtpA1-pA2, wtpA2, USE 1-5'mt, USE 1-3'mt, USE 1-(a), USE 1-(b), USE 1-(c), USE 2-mt, USE 3-mt, ΔUSE 1, ΔUSE 2, ΔUSE 3 pre-mRNAs were synthesized using the pGEM7-USE 1, pGEM7-USE 2, pGEM7-USE 3, pGEM7-*polo* wtpA1-pA2, pGEM7-wtpA2, pGEM7-USE 1-Rmt-DSE, pGEM7-USE 1-Lmt-DSE, pGEM7-left pyr, pGEM7-middle pyr, pGEM7-righr pyr, pGEM7-USE 2mt-DSE, pGEM7-USE 3mt-DSE, pGEM7-Δpyr, pGEM7-wt-DSE-ΔUSE 2 and pGEM7-wt-DSE-ΔUSE 3 constructs, respectively.

To generate the pGEM7-USE 1, pGEM7-USE 2, pGEM7-USE 3, pGEM7-left pyr, pGEM7-middle pyr, pGEM7-right pyr, the USE 1, USE 2, USE 3, left pyr, middle pyr and right pyr dsDNA oligos were used, respectively. To prepare these dsDNA oligos, the single strand sense pyr sense, polo U1, polo U2, left pyr sense, middle pyr sense and right pyr sense oligos and corresponding anti-sense oligos, the pyr anti-sense, compl polo U1, compl polo U2, left pyr anti-sense, middle pyr anti-sense and right pyr anti-sense, were purchased and hybridized (see Appendix, Table AI). Hybridization was performed with 1 nmol of each primer in 1x NEBuffer 3 (New England Biolabs) solution in a total volume of 10 μl. The solution was boiled for 5 min at 80ºC and then placed in water at 80ºC and let it to cool down at RT.

All cloning vectors were treated with Alkaline Phosphatase with exception to the construction of the pUC-*polo*3'UTR plasmid.

B2. Subcloning and DNA preparation

B2.1. Preparation of competent bacteria

One NM554 colony from a fresh LB plate was grown in 10 ml of 2x TY medium in an orbital incubator at 37ºC, ON. 1 ml of this culture was diluted to 10 ml with 2x TY and grown for 30 min. 5 ml of this culture was used to inoculate 500 ml of pre-warmed 2x TY and grown until A_{550} reached 0.3. The flask was cooled in ice for 10 min, after which the bacteria were pelleted at 600 g. Cells were resuspended in 250 ml ice cold 50 mM $CaCl_2$, left on ice for 20 min and pelleted at 4ºC. The pellet was gently resuspended in 20 ml ice cold 100 mM $CaCl_2$ containing 15% (v/v) glycerol. Aliquots were stored at -70ºC.
NM554 genotype: *AraD*139 *galK galU hsd R*2 (rK^-Km^+) *recA*13 *rpsL thi*-1 Δ(*ara-leu*)7696 Δ*lacX*74 F^-.

B2.2. Restriction Enzyme digestion and DNA modifications

Restriction and DNA modification reactions were carried out according to the manufacturer instructions. For restriction enzyme digestion, DNA was routinely digested with 1 U of enzyme per 1 μg of DNA with 1x restriction enzyme buffer in a minimal volume of 20 μl. The Klenow enzyme was used to create blunt ends by filling-in DNA 5' overhangs or to remove 3' overhangs. 4 U of Alkaline Phosphatase were used to remove 5'-phosphate groups from 1-5 μg of restriction enzyme digested cloning vector in order to decrease vector background and consequently increase cloning efficiency. The T4 Polynucleotide Kinase was used to add 5´-phosphates to oligonucleotides or to PCR products amplified with unphosphorylated primers to allow subsequent ligation. The digested or/and modified DNA fragments were purified using the GENECLEAN® Turbo Kit (according to the manufacturer instructions) following gel electrophoresis on a 0.7%-3% TAE-agarose gel, depending on the size of the DNA fragment.

B2.3. Ligation of inserts into cloning vectors

Ligations reactions were performed with a 1:6 vectort:insert ratio in a reaction containing 1x T4 DNA Ligase buffer and 1 U of T4 DNA ligase in a total volume of 20 μl. Ligation reactions were incubated at RT for 3 h for cohesive-cohesive ligations or 5 h for cohesive-blunt and blunt-blunt ligations.

B2.4. Polymerase Chain Reaction amplification

Reactions were set up containing 50 ng of DNA template, 0.2 pmol/μl of each primer, 0.8 mM dNTPs, 6 mM $MgSO_4$, 1x $Vent_R$® Polymerase buffer and 1 U of $Vent_R$® Polymerase in a total volume of 50 μl. The PCR programs used for construction of the several plasmids are indicated in Tables AIII and AIV (see Appendix). 5 μl of the PCR reaction were analysed by gel electrophoresis on a 0.7%-3% TAE-agarose gel, depending on the size of the amplified product. When the PCR products were to be digested by restriction enzymes, purification of the PCR products was performed using the GENECLEAN® Turbo Kit according to the manufacturer instructions.

For colony PCR, one colony was grown in 20 μl of LB with 50 μg/ml ampicillin for 1 hour in an orbital incubator at 37ºC. 1 μl of culture was used in a PCR reaction containing 0.2 pmol/μl of each primer, 0.8 mM dNTPs, 6 mM $MgCl_2$, 1x DSF-Taq DNA Polymerase buffer and 1.25 U of DSF-Taq DNA Polymerase (DNA-minus) in a total volume of 25 μl. The PCR programs varied accordingly with the set of primers used. After identification of positive clones, the remaining 19 μl of culture were used to inoculate 50-100 ml of LB with 50 μg/ml of ampicillin and grown in an orbital incubator at 37ºC, ON.

B2.5. Site-directed mutagenesis

The site-directed mutagenesis strategy used was adapted from Makarova *et al.* 2000, with minor modifications. 25 μl PCR reactions were prepared with 200 ng of DNA template, 0.2 μM of each primer, 0.8 mM dNTPs, 6 mM $MgSO_4$, 1x $Vent_R$® Polymerase buffer and 1 U of $Vent_R$® Polymerase. The primers used did not contain the 5'-phosphate group and with the exception for the pPCR-[*gfp-poloΔUSE 1*] construct, the PCR reactions were carried out with only the forward primer [5'-3' primer] (see Appendix, Table AV). The PCR programs used are shown in Table AV. After the PCR reaction, to digest the DNA template 5 U of *Dpn*I (Stratagene) was added and the reaction incubated for 4-5 h at 37ºC. Half volume of the reaction was then used to transform competent bacteria.

B2.6. Transformation of competent bacteria

The 20 μl ligation reaction was mixed with 200 μl of competent bacteria. Cells were incubated on ice for 15 min and then heat shocked at 42ºC for 90 seconds and put back on ice for 5 min. 800 μl of LB were added and cells incubated for 30 min at 37ºC in an orbital incubator. Cells were centrifuged at 4000 rpm for 2 min and ~900 μl of the volume

removed. The cells were resuspended in the remaining ~100 μl of LB and plated and grown in LB plates with 50 μg/ml of ampicillin at 37ºC, ON. The colonies were then screened by restriction enzyme analysis or by colony PCR analysis.

B2.7. Plasmid preparation

"Miniprep" DNA was prepared by centrifugation of 1 ml of bacteria culture for 3 min at 13200 rpm, RT. The pellet was resuspended in a 200 μl of STET buffer with 4 μl Lysozyme (50 mg/ml) solution and boiled at 95ºC for 1 minute. The samples were then centrifuged for 10 min at 13200 rpm and the pellet removed with toothpick. Precipitation of DNA plasmid was performed with addition of 200 μl of isopropanol to the supernatant; the solution was mixed and centrifuged for 10 min at 13200 rpm, RT. The supernatant was discarded and the pellet air dried. The precipitated plasmid DNA was resuspended in 20 μl of ddH_2O.

Large scale DNA plasmid preparations were prepared using the PhoenIX™ Midiprep Kit according to the manufacturer instructions.

B3. Cell transfections

HeLa cells were cultured in complete D-MEM medium in standard cell culture conditions and trypsinised when confluent. Cells were split usually in a 1:6 ratio (150 mm^2 plates) and transfected by the Calcium Phosphate precipitation technique when plates were 60% confluent. For transfection of HeLa cells, 23 μg of plasmid was used with 5 μg of the co-transfection plasmid, pBR328βSV40 (Grosveld *et al.,* 1982). The plasmids were mixed with 1 ml of 2x HBS and 1ml of 250 mM $CaCl_2$ was added dropwise while the tube containing the DNA plus HBS was agitated in the vortex, to allow a precipitate to form. The mixture was then added to the cells. 8-12 h after addition of the precipitate, the medium was removed and fresh medium was added. Cells were harvested 48 h after beginning of transfection.

Schneider's *Drosophila* Line 2 (SL2) cells were cultured in complete Schneider's Insect medium at 25ºC. For transfection of SL2 cells, 15 μg of plasmid was used with 5 μg of the co-transfection plasmid, pAc5.1-V5-HisA (Invitrogen). Cells were prepared for transfection by seeding $3x10^6$ SL2 cells in a 35 mm plate in 3 ml of medium. SL2 cells were grown for 24 h until cells reach the density of $2\text{-}4x10^6$ cells/ml and then transfected using the Calcium Phosphate precipitation technique as described above, using 300 μl of

2x HBS and 300 μl of 250 mM $CaCl_2$. 24 h after transfection, cells were collected and centrifuged for 2 min at 1000 g. The supernatant was removed and cells were resuspended in fresh medium. Cells were incubated for another 24 h and then harvested.

B4. RNA extraction

For cytosolic RNA extraction of HeLa cells, the medium was removed and cells washed with 1x PBS. Cells were scraped in 1x PBS, transfered to 50 ml tubes and centrifuged at 1400 rpm for 5 min at 4ºC. The supernatant was removed and the pellet drained. Cells were resuspended in 500 μl of RNA lysis buffer [140 mM NaCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCL pH 7.5, 0.5% (v/v) NP-40] using a 1 ml pipet tip. The lysate was transferred to eppendorf tubes, underlayered with 500 μl of RNA lysis buffer/24% sucrose and centrifuged at 14000 rpm for 10 min at 4ºC. The upper fraction containing the cytosolic RNA was isolated and two phenol/chloroform extractions were performed using phenol:chloroform 5:1 acid-equilibrated: pH 4.7 (SIGMA). The RNA was precipitated with 1/10 volume of 3 M sodium acetate pH 5.2 and 2.5 volumes of ethanol. The RNA was resuspended in 50 μl of DEPC-treated water.

For total RNA extraction, TRIzol® Reagent (Invitrogen) was used accordingly to manufacture instructions.

RNA was quantified spectrophotometrically by measuring the absorbance of appropriately diluted samples at 260 nm (A_{260}). One unit of A_{260} corresponds to 40 μg/ml of RNA (Sambrook and Russell 2001)

B5. RNA mapping

B5.1. DNA 3'end-labelled probes

The pUC-*polo*3'UTR, pUC-*polo*3'UTR-ΔUSE 1, pUC-*polo*3'UTR-ΔUSE 2, pUC-*polo*3'UTR-ΔUSE 3 and pUC-*polo*3'UTR-ΔpA1 plasmids were linearized using the restriction enzyme *Bme*18I. By gel electrophoresis on a 0.7 % TAE-agarose gel, a 2373 bp DNA fragment was isolated and purified. 1 μg of the DNA fragment was labelled with 17 μCi of α-^{32}P-[dCTP], 33 μCi of α-^{32}P-[dGTP], 1 mM dTTP and 4 U of Klenow enzyme in 1x Neb *Eco*RI buffer in a total volume of 30 μl.

The pBR328βSV40 and pAc5.1-V5/HisA were linearized with the restriction enzymes *Eco*RI and *Xba*I respectively. 1 μg of linearized plasmid was labelled with 20 μCi of α^{32}P-[dATP] and 1mM dTTP (for pBR328βSV40) or with 20 μCi of α-^{32}P-[dCTP], 1 mM of dATP, 1mM dGTP and 1 mM dTTP (for pAc5.1-V5/HisA) and 4 U of Klenow enzyme.

The reactions were incubated for 1 h at 37ºC and the probes purified using Illustra ProbeQuant™ G-50 Micro Columns according to manufacture instructions.

B5.2. S1 Nuclease protection assay

75-100 μg of cytosolic RNA or total RNA were precipitated with DNA radiolabelled probes at RT with 1/10 volume of 3 M sodium acetate and 2.5 volumes of ethanol. The mixture was centrifuged at 13000 rpm for 15 min at RT. The pellet was resuspended in 30 μl of Hybridisation Buffer [40 mM PIPES pH 7.4, 1 mM EDTA, 400 mM NaCl, 80% (v/v) formamide], denatured for 10 min at 80ºC and then incubated ON at 52ºC. This temperature favours the production of DNA:RNA hybrids. 300 μl of S1 Nuclease mix (180 U S1 Nuclease in 1.1x S1 Nuclease buffer), was added and the reaction incubated for 1 h at 30ºC. The reaction was stopped with addition of 100 μl of S1 Nuclease Stop buffer (4 M NH_4CH_3COO, 50 mM EDTA) and the products precipitated at RT with 20 μg of tRNA (type XX- *E. coli* strain w) (SIGMA) and 1 ml of ethanol. The mixtures were centrifuged at 13000 rpm for 15 min at RT, and the pellet resuspended in 10 μl of RNA gel-loading buffer. The DNA:RNA hybrids were denatured for 5 min at 95ºC and the digestion products separated in a denaturing 8.3 M urea-5% polyacrylamide gel at 20 W. The radiolabelled bands were visualized by autoradiography at after exposure at -70ºC with intensifying screens.

B6. *In vitro* transcription

The restriction enzymes used to linearize the DNA constructs for the synthesis of pre-mRNAs used in cleavage, polyadenylation or UV. cross-linking reactions, are presented in Table AVI (see Appendix). For UV. cross-linking assays, *in vitro* transcription was performed incubating 1 μg of linearized plasmid with 1x transcription buffer (Roche Applied Science), 10 mM DTT, 28.79 U of RNAguard™ Ribonuclease Inhibitor, 0.1 mM GTP, 0.5 mM CTP, 0.05 mM ATP, 0.05 mM UTP, 1 mM Cap Analogue (Ambion), 8 μCi of α-^{32}P-[ATP], 8 μCi of α-^{32}P-[UTP] and 20 U of T7 RNA Polymerase (Roche Applied Sciences) or SP6 RNA Polymerase (Roche Applied Sciences). The reaction was

incubated for 1 h at 37ºC. The plasmids where then digested with addition of 5 U of DNase I-RNase free and incubation at 37ºC for 15 min. The transcripts were purified using Illustra ProbeQuant™ G-50 Micro Columns according with the manufacture instructions.

For Cleavage and Polyadenylation assays, *in vitro* transcription was performed as described for UV cross-linking with modifications in the nucleotide concentrations: 0.5 mM ATP, 0.5 mM CTP, 0.1 mM GTP, 0.05 mM UTP, and 5 μCi of α-^{32}P-[UTP]. The radiolabelled transcripts were then purified by polyacrylamide electrophoresis. The RNAs were eluted from polyacrylamide with incubation of the RNA-containing polyacrylamide bands with 400 μl of Elution buffer [500 mM NH_4CH_3COO, 10 mM $Mg(CH_3COO)_2$, 1mM EDTA, 0.1% (w/v) SDS] at 4ºC ON with agitation. A Phenol:Chloroform extraction was performed with addition of an equal volume of Phenol:Chloroform 5:1 acid-equilibrated: pH 4.7 (SIGMA) to the eluate and the RNAs precipitated with 15 μg of tRNA (Type XX- *E. coli* strain w) (SIGMA) and 800 μl of ethanol. The RNAs were resuspended in 10 μl of DEPC-treated water.

B7. *In vitro* cleavage and polyadenylation analysis

For *in vitro* cleavage analysis a reaction mixture containing 2 mM EDTA, 1 mM ATP, 20 mM CP, 11.5 U RNAguard™ Ribonuclease Inhibitor, 2.5% (w/v) PVA and 40 μg/ml of tRNA (type XX – *E. coli* strain w) (SIGMA) was incubated with 100 cps of radiolabelled RNA and 5 μl of purified protein fractions in buffer D (3 μl of CSF plus 1.3 μl of rPAP) or 5 μl of HeLa cell NE in a final volume of 12.5 μl. The reactions were incubated at 30ºC for 90 min after which 112.5 μl of proteinase K solution [0.33 mg/ml proteinase K (SIGMA) in 1% (w/v) SDS] was added and the reactions incubated for further 15 min at 30ºC. Proteins were removed by extraction with 125 μl Phenol:Chloroform 5:1 acid-equilibrated: pH 4.7 (SIGMA) and the RNAs precipitated with 14 μl of a 2 M NH_4CH_3OO, 0.36 mg/ml tRNA (type XX-*E. coli* strain w) solution and 350 μl of ice-cold ethanol. The RNAs were resuspended in 10 μl of RNA gel-loading buffer and denatured for 3 min at 60ºC. The cleavage products were separated in a denaturing 8.3 M urea-5% polyacrylamide gel at 20 W. The radiolabelled bands were visualized by autoradiography at -70ºC with intensifying screens.

For *in vitro* polyadenylation analysis, the same reactions were performed as described for Cleavage analysis but using 1 mM $MgCl_2$ instead of 2 mM EDTA.

B8. UV cross-linking of proteins to RNA

For UV cross-linking analysis, the same reaction mixture used for cleavage analysis was prepared. The reactions were incubated for 10 min at 30ºC after which 1 μl of 2.5 μg/ml tRNA (type XX – *E. coli* strain w) (SIGMA) was added. The reactions were irradiated twice for 3 min with $96x10^4$ $\mu J/cm^2$ of UV. light using a Hoefer UVC 500 Ultraviolet Crosslinker (GE Healthcare Life Sciences). The radiolabelled RNAs were digested with 15 μg of RNase A for 30 min at 37ºC. 2x SDS gel-loading buffer was added to the reactions and proteins denatured at 95ºC for 5 min. The proteins were separated by electrophoresis in a 10% SDS-PAGE and the gel fixed for 30 min in a 10% (v/v) acetic acid, 10% (v/v) glycerol solution. The gel was then dried at 80ºC under vacuum for 2 h. The radiolabelled protein bands were visualized by autoradiography at -70ºC with intensifying screens.

B9. Immunoprecipitations

For immunoprecipitation of UV. cross-linked proteins with monoclonal antibodies, 400 μl of 10% (v/v) protein A Sepharose CL-4B beads (GE Healthcare Life Sciences) in IP-2 buffer [50 mM Tris pH 7.9, 50 mM NaCl, 0.1% (v/v) NP-40] was incubated with 40 μg of rabbit anti-mouse antibody (DakoCytomation) in a vertical wheel for 90 min at 4ºC. The beads were washed three times with ice-cold IP-2 buffer. One UV. cross-linking reaction and 30 μl of anti-CstF-64 or 2 μl of anti-hnRNP C were added to the beads and the mixture was rotated for 1 h at 4ºC. The beads were then washed three times with ice-cold IP-2 buffer and dried with the use of a Hamilton syringe. 20 μl 2x SDS gel-loading buffer was added to the beads and the proteins denatured at 95ºC for 5 min. For immunoprecipitation of PTB and αCP2/2KL, 5 μl of anti-PTB and 5 μl of FF3 rabbit serums were used, respectively. As a negative control, 5 μl of pre-immune serum was used. Each antibody, together with an UV. cross-linking reaction, was added to 100 μl of 50% (v/v) protein A-Sepharose beads in 1x PBS. The mixture was rotated for 1 h at 4ºC and then washed twice with ice-cold Binding Buffer I (20 mM Hepes pH 7.9, 150 mM NaCl, 0.05% (v/v) Triton X-100) and two times with Binding Buffer II (20 mM Hepes pH 7.9, 150 mM NaCl, 1% (v/v) Triton X-100). The beads were then dried with the use of a Hamilton seringe. 20 μl 2x SDS gel-loading buffer was added to the beads and the proteins denatured at 95ºC for 5 min. For immunoprecipitation of U1A, 17.5 μl of anti-U1A antibody and one UV. cross-linking reaction were added to 100 μl of 50% (v/v) protein A-

Sepharose beads in IPP150 buffer [10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% (v/v) NP40]. The mixture was rotated for 2 h at 4ºC. The beads were then washed three times with ice-cold IIP150 buffer and dried with the use of a Hamilton syringe. 20 μl 2x SDS gel-loading buffer was added to the beads and the proteins denatured at 95ºC for 5 min. After denaturing, the immunoprecipitated proteins were separated by gel electrophoresis in a 10% SDS-PAGE. The gel was fixed in a 10% (v/v) acetic acid, 10% (v/v) glycerol solution for 30 min at RT and dried at 80ºC under vacuum for 2 h. The radiolabelled proteins were visualized by autoradiography at -70ºC with intensifying screens.

B10. Preparation of HeLa cell nuclear extracts

HeLa nuclear extracts were prepared accordingly to Kleiman and Manley 2001. $3x10^7$ cells were grown and harvested in 5 ml of 1x PBS and centrifuged for 5 min at 4000 rpm, 4ºC. The cell pellet was resuspended in 4 ml of Buffer A [10 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, and 0.5 mM PMSF]. The cells were left on ice for 10 min and then lysed by douncing. The lysate was centrifuged for 5 min at 4000 rpm, 4ºC and the pellet resuspended in 200 μl of Buffer C [20 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 25% (v/v) glycerol, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, and 0.3 M NaCl]. Preparations were rocked for 30 min at 4°C and centrifuged for 15 min at 10000 rpm, 4ºC. Supernatants were quickly frozen at -80C.

B11. Affinity chromatography

100 μl of ImmunoPure© Immobilized Avidin (Pierce Biotechnology) were centrifuged for 4 s at 13200 rpm (the timer and the centrifuge start both at the same time). The supernatant was removed and the avidin-agarose beads where incubated (with gentle mixing) with 5 volumes of Binding buffer (20 mM Sodium Phosphate pH 7.5, 0.5 M NaCl) for 5 min, RT. The mixture was centrifuged for 4 s at 13200 rpm, RT, and the supernatant discarded. The biotinilated oligo (5.9 nmol/500 μl ddH_2O) was added to the pellet and incubated for 1 h at RT, with gentle mixing. The biotin-oligo:avidin-agarose beads complex was centrifuged for 4 s at 13200 rpm and washed with 5 volumes of Binding buffer. This wash was repeated twice. 250 μg of HeLa NE in 500 μl of Binding buffer was added to the biotin-oligo:avidin-agarose beads and incubated for 1 h at 4ºC, with gentle mixing. The mixture was then centrifuged for 4 s at 13200 rpm and the

supernatant removed. The protein:oligo-biotin:avidin-agorose beads complexes were washed with 500 μl of Binding buffer and centrifuged for 4 s at 13200 rpm. This wash was repeated twice. For mass spectrometry analysis, the proteins bound to the oligo-biotin:avidin-agarose beads were eluted with 5 volumes of 8 M Guanidine-HCl, pH 1.5. For protein analysis in a 10% SDS-PAGE, 20 μl of 2x SDS gel-loading buffer was added to the complexes and proteins denatured for 5 min at 95ºC.

B12. Protein precipitation with TCA

1/10 volume of TCA 100% was added to the protein samples and incubate on ice for 30 min. The mixture was centrifuged for 10 min at 14000 rpm, 4ºC. The supernatant was removed and the pellet washed with 500 μl of 50% (v/v) acetone in ddH_2O. The samples were centrifuged for 10 min at 14000 rpm, 4ºC and the pellet resuspended in 2x SDS gel-loading buffer.

B13. *Drosophila* stocks

w^{1118} was obtained from the Bloomington *Drosophila* Stock Center (Indiana University) and used as control strain. The $polo^{9}$ allele was kindly provided by David Glover (Department of Regulation of Mitosis and Meiosis, University of Cambridge) (Donaldson *et al.,* 2001). It was isolated in a genetic screen of a collection of mutants generated by P-element-mediated mutagenesis (Deak *et al.,* 1997). It is a late larval lethal allele with a P-element inserted in the *polo* 5'UTR, downstream of the second transcription start site. The mutagenised chromosome was balanced over the *TM6C* carrying the dominant larval/pupal marker *Tubby* (*Tb*), allowing selection of homozygous individuals based on the non-*Tubby* phenotype. The second chromosome was balanced by mating with w^{1118}; *If*/*CyO*; *MKRS*/*TM6B* (with the *TM6B* balancer also carrying the *Tubby* marker), obtaining the strain w^{1118}; *If*/*CyO*; $polo^{9}$/*TM6C*. *Df (3L) rdgC-co2,* th^{1} st^{1} in^{1} kni^{ri-1} p^{p} (breakpoints 77A1; 77D1) was obtained from the Bloomington *Drosophila* Stock Center and balanced over *TM6B* carrying the dominant larval/pupal marker *Tubby* (*Tb*) by mating with w^{1118}; *If*/*CyO*; *MKRS*/*TM6B*. Transgenic flies carrying the transgene pW8-*gfp-polo* were obtained from Claudio E. Sunkel (Department of Molecular Genetics, Instituto de Biologia Molecular e Celular, University of Porto) (Moutinho-Santos *et al.,* 1999).

Transgenic flies carrying the pW8-*gfp-polo∆pA1*, pW8-*gfp-polo∆pA2* or the pW8-*gfp-polo∆USE 1* transgene, were obtained by injection of w^{w1118} embryos with the respective transgenes according to Roberts 1986. Selection of transgenic lines with the transgene on the second chromosome was performed by matting the transgenic flies with a strain carrying dominant markers and a balancer chromosome, w^{1118}; *Sco*/*SM6*. These and the w^{1118}; *gfp-polo* strain were then matted with w^{1118}; *If*/*CyO*; *MKRS*/*TM6B and* w^{1118}; *If*/*CyO*; $polo^9$/*TM6C* to generate the lines w^{1118}; *gfp-polo;* $polo^9$/*TM6B*, w^{1118}; *gfp-polo∆pA1*; $polo^9$/*TM6B*, w^{1118}; *gfp-polo∆pA2*; $polo^9$/*TM6B* and w^{1118}; *gfp-polo∆USE 1*; $polo^9$/*TM6B*.

All stocks were grown at 25ºC on standard culture conditions and media.

B14. Quantification of the viability of the transgenic flies

For the w^{1118}*; gfp-polo*; $polo^9$/*TM6B*, w^{1118}*; gfp-polo∆pA1*; $polo^9$/*TM6B* and w^{1118}*; gfp-polo∆pA2*; $polo^9$/*TM6B*, crosses were performed to rescue $polo^9$/$polo^9$. To rescue the $polo^9$/*Df (3L) rdgC-co2* allelic combination, the w^{1118}*; gfp-polo∆pA1*; $polo^9$/*TM6B* and the w^{1118}*; gfp-polo∆pA2*; $polo^9$/*TM6B* were crossed with w^{1118}*; gfp-polo∆pA1*; *Df (3L) rdgC-co2/TM6B* and w^{1118}*; gfp-polo∆pA2*; *Df (3L) rdgC-co2/TM6B*, respectively. From these crosses, flies homozygous for $polo^9$ and $polo^9$/*Df (3L) rdgC-co2* were quantified over the total number of flies.

The crosses were performed at 25ºC on standard culture conditions and media.

B15. Mitotic index

Third instar larvae brains were dissected in 1x PBS and fixed in 45% (v/v) acetic acid followed by 60% (v/v) acetic acid for 1 min. The fixed brains were squashed and immersed in liquid nitrogen. DNA was counterstained with DAPI in VectaShield (Vector Laboratories). For each genotype, eight brains were prepared and 1000 cells scored for each brain. The mitotic index was then determined as the number of mitotic cells present over the total number of cells. All quantifications were performed on the upright microscope Axioskop (Carl Zeiss, Germany) with a b/w or color CCD camera Spot 2 (Diagnostic Instruments, USA).

B16. *In situ* hybridisation on polytene chromosomes from salivary glands

B16.1. Biotin-labelled DNA probe

A 1.2 kb *polo* DNA fragment, obtained from the digestion of pBs-*polo* (Llamazares *et al.,* 1991) with *Eco*RI, was used to generate a biotin-labelled DNA probe using the Bionick™ Labelling System (Invitrogen) accordingly to the manufacturer instructions with minor modifications: the labelling reaction was incubated for 3 h at 16ºC.

B16.2. Preparation of polytene chromosomes

Third instar larvae salivary glands were dissected in 0.7% (v/v) NaCl, and the fat tissue attached to it removed. The salivary glands were incubated in 45% (v/v) acetic acid for 30 s and then transferred to a 5 μl drop of 1:2:3 solution (1 volume lactic acid, 2 volumes of ddH_2O, 3 volumes of acetic acid) on a clean siliconised coverslip. A slide was placed over the coverslip. The polytene chromosomes were spread by tapping with the back of a pencil until the desirable spreading was achieved. The coverslip was then pressed hard to remove the excess of liquid and to flatten the chromosomes. Slides were transferred to liquid nitrogen and the coverslip immediately removed. The slides were then incubated in ethanol for 15 min at RT, and then air dried.

B16.3. *In situ* hybridisation

Unless stated otherwise, all incubations were performed at RT.

The slides with the polytene chromosomes were incubated for 30 min in 2x SSC at 65ºC, followed by a 5 min incubation in 70% (v/v) ethanol and 5 min in 100% ethanol. The slides were air dried and then incubated for 2 min in 70 mM NaOH. Incubations of 3 min in 70% (v/v) ethanol and then in 100% ethanol were performed and the slides air dried. 5 μl of biotin-labelled DNA probe was mixed with 5 μl of 2x hybridisation buffer [8x SSC, 2x Denhardt's, 20% (w/v) dextran sulphate, 0.4% (w/v) denatured salmon sperm DNA] and the mixture denatured for 10 min at 95ºC. The 10 μl solution was added to the region on the slides that contained the chromosomes and covered with a coverslip. The slides were incubated in a humid chamber at 58ºC ON. Next day, the following incubations were performed: 2 min in 2x SSC at 53ºC, 5 min in 4x SSC, 5 min in 4x SSC, 0.1% (v/v) Triton X-100 and 5 min in 4x SSC. 50 μl of 2% (v/v) Fluorescein Avidin D (Vector Laboratories) in 4x SSC was added to the chromosomal region of the slide, covered with a coverslip

and incubated for 30 min. A series of incubations were then performed: 5 min in 4x SSC, 5 min in 4x SSC, 0.1% (v/v) Triton X-100 and 5 min in 4x SSC. VectaShield (Vector Laboratories) with DAPI was added to the chromosomal region on the slide and covered with a coverslip and the edges sealed with nail varnish. Slides were stored at 4ºC to avoid degradation Fluorescein Avidin D.

B17. Fluorescent analysis of *Drosophila* embryos

Embryos with 0-24 h of development were collected from agar plates into an egg basket using ddH_2O and a small paintbrush and rapidly washed with ddH_2O to avoid anoxia. The embryos were dechorionated by placing the egg basket in a Petri dish partially filled with 30% (v/v) bleach (sodium hypochlorite-in solution 13% active chlorine) (LIMPOLAR) solution for 3 min and immediately washed with Embryo Wash solution (0.7% (w/v) NaCl, 0.05% (v/v) Triton X-100). Fixation was performed using the Slow Formaldehyde fixation method. The excess liquid in the egg basket was removed by gently blot the egg basket with a paper towel. With a small paintbrush, embryos were collected and added to a vial containing 1 ml of heptane. An equal volume of 3.7% (v/v) formaldehyde in PEM buffer (0.1 M PIPES pH 6.9, 1 mM $MgCl_2$, 1 mM EGTA) was added and the vial shook vigorously for 15 s. The vial was left stand for 20 min, RT. The formaldehyde phase was then removed and 1 ml of methanol was added. The vial was again shooked vigorously for 15 s and left to stand for 1 min after shaking. The upper heptane layer was then removed, along with the embryos that did not sink. Methanol was added until the vial was approximately two-thirds full. The methanol was quickly removed and the embryos were washed three times with 1 ml of PBT [1x PBS, 0.1% (v/v) Triton X-100]. The embryos were mounted in VectaShield with DAPI (Vector Laboratories). The images were acquired with an inverted Laser Scanning Confocal Microscope Leica SP2 AOBS SE (Leica Microsystems, Germany) and treated using the Adobe Photoshop 7.0 Software (Adobe Systems, Inc).

B18. Adult abdomen preparations

With a blade, flies were cut between the thorax and the abdomen, carefully removing all appendages. Abdomens were then incubated in a lactic acid:ddH_2O (3:1) solution and incubated ON at 60ºC. The next day, abdomens were mounted in a fresh lactic acid:ddH_2O (3:1) solution and incubated ON in a 60ºC oven. The images were

acquired with a Leica DM LB microscope (Leica Microsystems, Germany) with a 3CCD color camera vision module (DONPISHA) and treated using the Adobe Photoshop 7.0 Software (Adobe Systems, Inc).

B19. Dissection of pupa epidermis

Staging of pupae was determined accordingly to Bainbridge and Bownes (1981) and Sullivan *et al.* (2000). Dissection of the pupa epidermis was performed accordingly to Gompel and Carroll (2003). Pupae with 26-27 h APF were collected and lined-up, on their side, in doubled-side tape. The pupae were cut with a blade along the anterior-posterior axis. The dorsal cuticle was collected and placed in 1x PBS. Using fine forceps, the epidermis was detached from the cuticle and all internal tissues and fat removed. The epidermis was then fixed in 200 μl of 1.85% (v/v) formaldehyde in fixation buffer [0.1 M PIPES pH 6.9, 1 mM EGTA pH 6.9, 1% (v/v) Triton X-100, 1 mM $MgCl_2$] for 10 min at RT without shaking. After fixation, the epidermis was washed twice for 5 min at RT in PBT [1x PBS, 0.03% (v/v) Triton X-100]. The epidermis was then mounted in 6 μl of VectaShield with DAPI (Vector Laboratories). The images were acquired with the upright microscope AxioImager Z1 (Carl Zeiss, Germany), with a b/w CCD camera Axiocam MR ver.3.0 (Carl Zeiss, Germany). The images were treated using the Adobe Photoshop 7.0 Software (Adobe Systems, Inc).

B20. Immunostaining of third instar larvae abdominal histoblasts

This protocol was adapted from Weaver and White (1995) and Sullivan *et al.* (2000). Third instar larvae were collected and killed by heat-shock, by putting larvae in 1x PBS at 55ºC for 10 s. The larvae were then cut with a blade along the dorsal-ventral axis. Each half was placed in 1x PBS and all internal tissues were carefully removed. The cuticles were then fixed in 100 μl of 3.7% (v/v) formaldehyde in PBT [1x PBS, 0.1 % (v/v) Triton X-100] for 20 min at RT (without shaking) and then washed three times, 10 min each, in 100 μl of PBT, RT. After washing, blocking was performed for 1 h, RT, in 200 μl of PBT with 10% (v/v) FBS. The cuticles were then incubated ON at 4ºC with 200 μl of α-Hdc (1:5) and α-CycB (1:3000) in PBT with 10% (v/v) FBS. Next day the cuticles were washed three times, for 10 min each, in 200 μl of PBT and then incubated with 200 μl of α-rabbit Alexa 647 (Molecular Probes) (1:2000) and α-mouse Alexa 555 (Molecular

Probes) (1:2000) in PBT with 10% (v/v) FBS for 1 h at RT. The cuticles were then washed three times, 10 min each, with 200 μl of PBT, RT, and then mounted in 20 μl of VectaShield with DAPI (Vector Laboratories). The images were acquired with an inverted Laser Scanning Confocal Microscope Leica SP2 AOBS SE (Leica Microsystems, Germany) and deconvolved and projected onto a single plan using the Huygens Deconvolution Software (SVI, Netherlands). The images were treated using the Adobe Photoshop 7.0 Software (Adobe Systems, Inc). The number of histoblasts present in each nest for the different abdominal segments represents the average number of cells in two w^{1118}; *gfp-polo*; $polo^{9}$ and w^{1118}; *gfp-poloΔpA1*; $polo^{9}$ (line 12.1) larvae and three w^{1118}; *gfp-poloΔpA2*; $polo^{9}$ (line 2.3) larvae.

B21. Quantification of the GFP levels in 26-27h APF pupae epidermis

The images were acquired using a Laser Scanning Confocal Microscope Leica SP2 AOBS SE (Leica Microsystems, Germany) and the GFP fluorescence measured using the Leica Confocal Software, version 2.61 Build 1538 LCS Lite (Leica Microsystems, Germany). For each genotype, two pupae were analysed. For each sample, one plan was chosen and several GFP measurements were made throughout the epidermis. The GFP levels were quantified using the following formula: (mean x area)/ (number of cells), were the mean is the average of GFP fluorescence measured for each sample.

B22. Western blot analysis

For Western blot analysis, third instar larvae brains were dissected and homogenized in 1x PBS supplemented with “Complete” protease inhibitors (Roche Applied Science) at 4ºC. 40 μg of total protein extracts were resolved in a 10% SDS-PAGE, and transferred to a Hybond-C Extra nitrocellulose membrane (GE Healthcare Life Sciences) using a transfer buffer containing 20% (v/v) methanol (25 mM Tris base; 192 mM glycine; 20% (v/v) methanol) according to the manufacturer instructions. Membranes were blocked ON at 4ºC with 5% (w/v) dry milk in 1x PBS, 0.1% (v/v) NP-40. All primary and secondary antibodies were diluted in 1x PBS, 0.1% (v/v) NP-40 with 3% (w/v) dry milk. For analysis of Polo, MA294 antibody was diluted 1:80 and for detection of GFP-Polo, anti-GFP (Zymed) was diluted 1:500. α-Tubulin was detected using mAB DM1A (SIGMA) diluted 1:10000. All washes, unless otherwise stated, were performed for 10 min

at RT. After incubation with primary antibodies, the membranes were washed three times with 1x PBS, 0.1% (v/v) NP-40. Anti-mouse secondary antibodies conjugated to HRP (Vector Laboratories) were used according to the manufacturer instructions. After incubation with the secondary antibodies, the membranes were washed twice with 1x PBS, 0.1% (v/v) NP-40 and twice with 1x PBS. Blots were developed with ECL Plus™ Western Blot Detection Reagents according to manufacturer protocol.

Western Blot against PTB, hnRNP C and CstF-64 were performed as described above using 20 μg of HeLa cell nuclear extracts. For detection of PTB, the PTB antibody was diluted at 1:4000 whereas for detection of hnRNP C and CstF-64, the hnRNP C and CstF-64 antibodies were diluted at 1:5000 and 1:500 respectively. The pre-immune serum was diluted at 1:4000. The anti-mouse and anti-rabbit secondary antibodies conjugated with HRP (Vector Laboratories) were used accordingly to manufacturer instructions.

Western blot against PTB eluted from the affinity chromatography was performed as described above.

B23. Northern blot analysis

Northern blot analysis was performed accordingly to Sambrook and Russell (2001). A 1.2 kb *polo* DNA fragment obtained from the digestion of pBs-*polo* with *Eco*RI was used to generate a radiolabelled probe using the Ready-To-Go™ DNA Labeling Beads kit (-dCTP) (GE Healthcare Life Sciences) accordingly to the manufacturer instructions.

PART V

APPENDIX

Tables

Table AI

Primer	Sequence (5'-3')	Lenght (nt)
compl polo U1	GAG GCC CAA CAC TGA C	16
compl polo U2	ACG AAA ACC ATT AAA CAG A	19
compl polo U3	CTT TAA AAT GCA	12
left pyr anti-sense	AAA AAC AAA TAA ATT C	16
left pyr sense	GAA TTT ATT TGT TTT T	16
middle pyr anti-sense	GGG GCA AAA AC	11
middle pyr sense	GTT TTT GCC CC	11
NotpA1	ATA GTT TAG CGG CCG CCG CTT TTA GTT CAA AAG CA	35
polo U1	GTC AGT GTT GGG CCT C	16
polo U2	TCT GTT TAA TGG TTT TCG T	19
polopAmt	TTA AAT AAA CAG AAA CGG AGA AGG AAT TCA TAT CGA AAA TAC TGC	45
PshAI genomico	GCC CGA CGA CAC TCG TCT GGA GTC	24
pyr anti-sense	GAA GGG GAA GGG GCA AAA ACA AAT AAA TT	29
pyr sense	AAT TTA TTT GTT TTT GCC CCT TCC CCT TC	29
right pyr anti-sense	AAG GGG AAG GGG C	13
right pyr sense	GCC CCT TCC CCT T	13
T7 promoter	TAA TAC GAC TCA CTA TAG GG	20
USE 1-mt	GAG GAA TTC ATA GAC GAC AAA CGC AAG ATC TTC ATA TCG AAA ATA CTG CTT A	52
USE 1-Lmt	GAG GAA TTC ATA GAC GAC AGC CCC TTC CCC TTC ATA TC	38
USE 1-Rmt	CGA GGA ATT TAT TTG TTT TTA ACG CAA GAT CTT CAT ATC GAA AAT ACT GCT TA	53
USE 2-mt	AAT ACT GCT TAA GTT ATA TTC ATC AGA TCT AAT CAG ACA ACC TCA AAA GTA ATT TAA TAT ATC TG	65
USE 3-mt	CCT CAA AAG TAA TTT AAT ATA TCA AAG ATC TCA CGA CAG AAC ACG ATC CGA TCA CTT AAT G	61
USE 3-mt-Rev	TTG GAA TTC GTC GTG AGA TCT TTG ATA TAT TAA ATT ACT TTT GAG GGA GG	50
ΔpA1	GCA TTT TAA AGA GAT CAA GTT AAC TGT TTA AAC TAA GCA AAC GTG	45
Δpyr	ATA TCG AAA ATA CTG CTT AAG	21
Δpyr-F	GAT TTG TTC GAT GTT TAT AGC CCT TCA TAT CGA AAA TAC TGC TTA AG	47
Δpyr-R	CTA AAC AAG CTA CAA ATA TCG GGA AGT ATA GCT TTT ATG ACG AAT TC	47
ΔUSE 2	AAT ACT GCT TAA GTT ATA TTC ATC CCT CAA AAG TAA TTT AAT ATA TCT G	49
ΔUSE 3	CCT CAA AAG TAA TTT AAT ATA ACA CGA TCC GAT CAC TTA ATG	42
$(USE\ 1)_3$ -oligo	biotin-AAT TTA TTT GTT TTT GCC CCT TCC CCT TCA ATT TAT TTG TTT TTG CCC CTT CCC CTT CAA TTT ATT TGT TTT TGC CCC TTC CCC TTC	87

Table AI. Oligonucleotides used for the construction of all plasmids used in this thesis. The oligonucleotides were obtained from Amersham Pharmacia Biotech, MWG-Biotech AG, and GIBCO BRL and Invitrogen.

Table AII

	Insert			Cloning vector		
Construct	**Plasmid/PCR product**	**Restriction and DNA modifying enzymes**	**Isolated fragment size**	**Plasmid**	**Restriction and DNA modifying enzymes**	**Isolated fragment size**
pSPT19-*polo* wt	pBS-*polo*	(1) *Afl* III; (2) Klenow (5'-3'); (3) *Apo* I	183 bp	pSPT19	(1) *Eco* RI; (2) *Sma* I	3086 bp
pSPT19-*polo* mt	[polopAmt : T7]-247	(1) *Bam* HI	186 bp	pSPT19	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Bam* HI	3087 bp
pGEM7-*polo* wt	pSPT19-*polo* wt	(1) *Apo* I; (2) *Bam* HI	186 bp	pGEM-7Zf(+)	(1) *Eco* RI; (2) *Bam* HI	2962 bp
pGEM7-*polo* mt	[polopAmt : T7]-247	(1) *Bam* HI	186 bp	pGEM-7Zf(+)	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Bam* HI	2966 bp
pGEM7-*polo* wt-DSE	pBs-*polo*	(1) *Apo* I	221 bp	pGEM-7Zf(+)	(1) *Eco* RI	2997 bp
pGEM7-USE 1mt-DSE	[USE 1-mt : NotpA1]-434	(1) *Eco* RI; (2) *Acs* I	221 bp	pGEM-7Zf(+)	(1) *Eco* RI	2997 bp
pGEM7-USE 1 Lmt-DSE	[USE 1-Lmt : NotpA1]-434	(1) *Eco* RI; (2) *Acs* I	221 bp	pGEM-7Zf(+)	(1) *EcoRI*	2997 bp
pGEM7-USE 1 Rmt-DSE	[USE 1-Rmt : NotpA1]-434	(1) *Eco* RI; (2) *Acs* I	221 bp	pGEM-7Zf(+)	(1) *Eco* RI	2997 bp
pGEM7-[T7-USE 3mt-Rev]	[T7 : USE 3mt-Rev]-183	(1) *Eco* RI; (2) *Sfr* 274I	124 bp	pGEM-7Zf(+)	(1) *Eco* RI; (2) *Sfr* 274I	2991 bp
pGEM7-USE 3mt-DSE	[USE 3mt : NotpA1]-353	(1) *Bgl* II; (2) *Apo* I	118 bp	pGEM7-[T7-USE 3mt-Rev]	(1) *Bgl* II; (2) *Eco* RI	103 bp
pGEM7-Δpyr	[Δpyr : T7]-475	(1) *Apo* I; (2) Klenow (5'-3'); (3) T4 PNK	197 bp	pGEM-7Zf(+)	(1) *Eco* RI; (2) Klenow (5'-3')	2997 bp
pGEM7-*polo* wtpA1_pA2	pGEM7-ΔBHI-luc(3UTR-pA2Rev)	(1) *Bpu* 14I; (2) *Bam* HI	294 bp	pGEM7-*polo* wt-DSE	(1) *Bpu* 14I; (2) *Bam* HI	3147 bp
pGEM7-wtpA2	pBs-*polo*	(1) *Apo* I	205 bp	pGEM-7Zf(+)	(1) *Eco* RI	2997 bp
pGEM7-USE 1	USE 1 oligo	_	29 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp

Table AII. Constructs generated during the course of this thesis and the respective subcloning strategies employed. The Klenow enzyme was used to create blunt ends by filling-in DNA 5' overhangs (5'-3') or to remove 3'overhangs (3'-5').

Table AII (continue)

	Insert			Cloning vector		
Construct	**Plasmid/PCR product**	**Restriction and DNA modifying enzymes**	**Isolated fragment size**	**Plasmid**	**Restriction and DNA modifying enzymes**	**Isolated fragment size**
pGEM7-USE 2	USE 2 oligo	_	16 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp
pGEM7-USE 3	USE 3 oligo	_	19 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp
pGEM7-left pyr	left pyr oligo	(1) T4 PNK	16 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp
pGEM7-middle pyr	middle pyr oligo	(1) T4 PNK	11 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp
pGEM7-right pyr	right pyr oligo	(1) T4 PNK	13 bp	pGEM-7Zf(+)	(1) *Sma* I	2997 bp
pUC-*polo* 2pAwt	pxb7	(1) Acc*I*; (2) Klenow (5'-3'); (3) Eco*RI*	1875 bp	pUC-αC2	(1) Xba*I*; (2) Klenow (5'-3'); (3) Eco*RI*	4180 bp
pUC-*polo* 3'UTR	_	_	_	pUC-*polo* 2pAwt	(1) *Eco* RI; (2) Klenow (5'-3')	6061 bp
pUC-*polo* 3'UTR-ΔpA1	pPCR[*gfp-polo ΔpA1*]	(1) *Afl* II; (2) *Bpu* 14I	140 bp	pUC-*polo* 3'UTR	(1) *Afl* II; (2) *Bpu* 14I	5921 bp
pUC-*polo* 3'UTR-ΔUSE 1	pPCR[*gfp-polo ΔUSE 1*]	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Afl* II	863 bp	pUC-*polo* 2pAwt	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Afl* II	5173 bp
pUC-*polo* 3'UTR-ΔUSE 2	pGEM7-*polo* wt-DSE-ΔUSE2	(1) *Afl* II; (2) *Bpu* 14I	124 bp	pUC-*polo* 3'UTR	(1) *Afl* II; (2) *Bpu* 14I	5921 bp
pUC-*polo* 3'UTR-ΔUSE 3	pGEM7-*polo* wt-DSE-ΔUSE3	(1) *Afl* II; (2) *Bpu* 14I	121 bp	pUC-*polo* 3'UTR	(1) *Afl* II; (2) *Bpu* 14I	5921 bp
pAc5.1-*polo* 3'UTR	pUC-*polo* 3'UTR	(1) *Bam* HI; (2) *Xba* I; (3) Klenow (5'-3')	2802 bp	pAc5.1/V5-HisA	(1) *Hpa* I; (2) *Eco* RI; (3) Klenow (5'-3')	5047 bp
pAc5.1-*polo* 3'UTR-ΔpA1	pUC-*polo* 3'UTR-ΔpA1	(1) *Bam* HI; (2) *Xba* I; (3) Klenow (5'-3')	2802 bp	pAc5.1/V5-HisA	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Hpa* I	5047 bp
pAc5.1-*polo* 3'UTR-USE 1mt	pGEM7-*polo* 3UTR-USE 1mt	(1) *Eco* RV; (2) *Bpu* 14I	965 bp	pAc5.1-*polo* 3'UTR	(1) *Eco* RV; (2) *Bpu* 14I	6884 bp

Table AII (continue)

Construct	Insert			Cloning vector		
	Plasmid/PCR product	**Restriction and DNA modifying enzymes**	**Isolated fragment size**	**Plasmid**	**Restriction and DNA modifying enzymes**	**Isolated fragment size**
pAc5.1-*polo* 3'UTR-USE 2mt	pUC-*polo* 3'UTR-USE 2mt	(1) *Bam* HI; (2) *Xba* I; (3) Klenow (5'-3')	2802 bp	pAc5.1/V5-HisA	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Hpa* I	5047 bp
pAc5.1-*polo* 3'UTR-USE 3mt	pUC-*polo* 3'UTR-USE 3mt	(1) *Bam* HI; (2) *Xba* I; (3) Klenow (5'-3')	2802 bp	pAc5.1/V5-HisA	(1) *Eco* RI; (2) Klenow (5'-3'); (3) *Hpa* I	5047 bp
pPCR[*gfp-polo* Δ *pA1]*	pxb7ΔpA1	(1) *Bst* API; (2) *Eco* RI	1576 bp	pPCR[*gfp-polo*]	(1) *Bst* API; (2) *Eco* RI	9032 bp
pW8-*gfp-polo* Δ *pA1*	pPCR[*gfp-polo* Δ *pA1*]	(1) *Cci* NI; (2) *Xho* I	7717 bp	pW8	(1) *Cci* NI; (2) *Xho* I	8282 bp
pxb7ΔUSE 1	pPCR[*gfp-polo* Δ *USE 1*]	(1) *Psh* AI; (2) *Bst* API	1760 bp	pxb7	(1) *Psh* AI; (2) *Bst* API	8105 bp
pW8-*gfp-polo* Δ *USE 1*	pxb7ΔUSE 1	(1) *Psh* AI; (2) *Cci* NI	3538 bp	pW8-*gfp-polo*	(1) *Psh* AI; (2) *Cci* NI	12431 bp
pW8-*gfp-polo* Δ *pA2*	[pshA1 genomico : NotpA1]· 1494	(1) *Psh* A1; (2) *Not* I	1468 bp	pW8-*gfp-polo*	(1) *Psh* A1; (2) *Not* I	12431 bp

Table AIII

PCR product	DNA template	Primer set
[polopAmt : T7]-247	pSPT19-*polo* wt	(1) polopAmt; (2) T7
[USE 1-mt : NotpA1]-434	pxb7	(1) USE 1-mt; (2) NotpA1
[USE 1-Lmt : NotpA1]-434	pxb7	(1) USE 1-Lmt; (2) NotpA1
[USE 1-Rmt : NotpA1]-434	pxb7	(1) USE 1 -Rmt; (2) NotpA1
[T7 : USE 3-mt-Rev]-183	pGEM7-*polo* wt-DSE	(1) T7; (2) USE 3-mt-Rev
[USE 3-mt : NotpA1]-353	pxb7	(1) USE 3-mt; (2) NotpA1
[Δpyr : T7]-475	pBs-*polo*	(1) Δpyr; (2) T7
[pshA1 genomico : NotpA1]-1494	pPCR[*gfp-polo*]	(1) pshA1 genomico; (2) NotpA1

Table AIII. PCR products used to generate some of the constructs in Table AII with the corresponding DNA templates and primer sets used.

Table AIV

PCR product	PCR reaction					PCR product size
	Initial denaturing step (ºC)	Denaturing temp (ºC) (step 2)	Annealing temp (ºC) (Step 3)	Extension temp (ºC) (step4)	nº of cycles (step 2-4)	
[polopAmt : T7]-247	94 (5 min)	94 (1 min)	55 (1min)	72 (2 min)	35	247 bp
[USE 1-mt : NotpA1]-434	94 (5 min)	94 (1min)	56.2 (1min)	72 (30 s)	35	434 bp
[USE 1-Lmt : NotpA1]-434	94 (5 min)	94 (1min)	63.7 (1min)	72 (30 s)	35	434 bp
[USE 1-Rmt : NotpA1]-434	94 (5 min)	94 (1min)	61.2 (1min)	72 (30 s)	35	434 bp
[T7 : USE 3-mt-Rev]-183	94 (5 min)	94 (1min)	57 (1min)	72 (12 s)	35	183 bp
[USE 3-mt : NotpA1]-353	94 (5 min)	94 (1min)	70 (1min)	72 (25 s)	35	353 bp
[Δpyr : T7]-475	94 (5 min)	94 (1min)	55 (1min)	72 (1min)	35	475 bp
[pshA1 genomico : NotpA1]-1494	94 (5 min)	94 (1min)	72 (1min)	72 (2 min)	35	1494 bp

Table AIV. Settings used in the PCR reactions described in Table AIII.

Table AV

Construct	PCR reaction						
	DNA template	**Primer set**	**Initial denaturing step (ºC)**	**Denaturing temp (ºC) (step 2)**	**Annealing temp (ºC) (Step 3)**	**Extension temp (ºC) (step4)**	**nº of cycles (step 2-4)**
pGEM7-USE 2mt-DSE	pGEM7-*polo* wt-DSE	USE 2-mt	95 (3 min)	95 (15 s)	61.2 (1min)	72 (3 min. 30 s)	18
pGEM7-*polo* wt-DSE ΔUSE 2	pGEM7-*polo* wt-DSE	ΔUSE2	95 (3 min)	95 (15 s)	56.2 (1min)	72 (3 min. 30 s)	18
pGEM7-*polo* wt-DSE ΔUSE 3	pGEM7-*polo* wt-DSE	ΔUSE3	95 (3 min)	95 (15 s)	56.2 (1min)	72 (3 min. 30 s)	18
pPCR[*gfp-polo Δ USE 1*]	pPCR[*gfp-polo*]	(1) Δpyr-R; (2) Δpyr-F	95 (3 min)	95 (15 s)	54 (1min)	68 (11 min. 30 s)	18
pxb7ΔpA1	pxb7	ΔpA1	95 (3 min)	95 (15 s)	62 (1min)	72 (11 min)	18

Table AV. Constructs generated by site-directed mutagenesis and the respective PCR reactions settings.

Table AVI

Pre-mRNA	Construct used	Restriction enzymes used to linearize the construct for	
		Cleavage reactions	Polyadenylation and UV. cross-linking reactions
	pSPT19-*polo* wt	*Hind* III	*Bam* HI
wtpA1	pGEM7-*polo* wt	—	*Bam* HI
	pGEM7-*polo* wt DSE	—	*Afl* III
	pSPT19-*polo* mt	*Hind* III	*Bam* HI
USE 1-mt	pGEM7-*polo* mt	—	*Bam* HI
	pGEM7-USE 1mt	—	*Afl* III
USE 2-mt	pGEM7-USE 2mt	—	*Afl* III
USE 3-mt	pGEM7-USE 3mt	—	*Afl* III
ΔUSE 1	pGEM7-wt DSE ΔUSE 1	—	*Afl* III
ΔUSE 2	pGEM7-wt DSE ΔUSE 2	—	*Afl* III
ΔUSE 3	pGEM7-wt DSE ΔUSE 3	—	*Afl* III
USE 1-5'mt	pGEM7-USE 1 Rmt-DSE	—	*Afl* III
USE 1-3'mt	pGEM7-USE 1 Lmt-DSE	—	*Afl* III
USE 1	pGEM7-USE 1	—	*Cla* I
USE 2	pGEM7-USE 2	—	*Cla* I
USE 3	pGEM7-USE 3	—	*Cla* I
USE 1-(a)	pGEM7-left pyr	—	*Cla* I
USE 1-(b)	pGEM7-middle pyr	—	*Cla* I
USE 1-(c)	pGEM7-righr pyr	—	*Cla* I
wtpA1-DSE	pGEM7-wt DSE	—	*Kpn* I
wtpA2	pGEM7-wtpA2	—	*Kpn* I
wtpA1-pA2	pGEM7-*polo* wtpA1-pA2	—	*Kpn* I
Ad-L3	Ad-L3	—	*Bam* HI
Ad-L3MA	Ad-L3MA	—	*Bam* HI
pGEM7	pGEM®-7Zf(+)	—	*Hind* III (Fig.12A) *Cla* I (Fig. 15B)

Table AVI. DNA constructs used in the synthesis of each pre-mRNA. Whether if the pre-mRNAs were to be used in cleavage, polyadenylation or UV. cross-linking reactions, the DNA constructs were linearized with appropriate restriction enzyme.

Abbreviations

A: adenine (deoxy)nucleotide

α-CP2*:* α-complex protein 2

***A.gambiae*:** *Anopheles gambiae*

***A. mellifera*:** *Apis mellifera*

AMP: adenosine monophosphate

APF: after pupa formation

ATP: adenosine triphosphate

***awh*:** *arrowhead*

bp: base pair

BSA: bovine serum albumin

C: cytosine (deoxy)nucleotide

CacyBP: calcyclin binding protein

Cdc: cell division cycle

Cdk: cyclin-dependent kinase

cDNA: complementary DNA

CF: cleavage factor

Ci: curie

CMV: cauliflower mosaic virus

CP: creatine phosphate

***cox-2*:** *cyclooxygenase-2*

cps: counts per minute

CPSF: cleavage/polyadenylation specificity factor

CSF: cleavage/specificity factor

CstF: cleavage stimulation factor

C-terminal: carboxy-terminal

CTD: carboxy-terminal domain

CTP: cytidine triphosphate

***CyO*:** *curly of oster*

DAPI: 4', 6-Diamidino-2-phenylindole dihydrochloride

dATP: deoxyadenosine triphosphate

dCTP: deoxycytidine triphosphate

DEPC: Diethyl pyrocarbonate

***Df*:** deficiency

dGTP: deoxyguanosine triphospahte

DNA: deoxyribonucleic acid

DNase: deoxyribonuclease

dNTPs: deoxynucleoside-triphosphate mix

dsDNA: double strand DNA

DSE: downstream sequence element

DTT: Dithiothreitol

dTTP: deoxythymidine triphosphate

ECL: enhanced chemiluminescence

***E. coli*:** *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

EGTA: Ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid

EMS: ethylmethanesulfonate

***esg*:** *escargot*

EST: expressed sequence tag

***e(r)*:** *enhancer of rudimentary*

FBS: fetal bovine serum

G: guanine (deoxy)nucleotide

g: gram

GFP: green fluorescent protein

GTP: guanosine triphosphate

h: hour

HAT: half a TPR

HBS: Hepes-buffered saline

***hdc*:** *headcase*

Hepes: N-(Hydroxyethyl)piperazine-N′-(2-ethanesulfonicacid)

hFip1: human homologue of Factor interacting with Pap1

hnRNP: heterogeneous nuclear ribonucleoprotein

HRP: horseradish peroxidase

HSF1: heat shock transcription factor 1

***hsp70*:** *heat shock protein 70*

***If*:** *irregular facets*

IgM: immunoglobulin M

IL-10: interleukin-10

kb: kilobase(s)

kcal/mol: kilocalories per mole

kDa: kilodalton(s)

L: litter

LB: Luria-Bertani culture medium

LEC: larva epidermal cell

LPS: lipopolysaccharide

M: Molar

mAB: monoclonal antibody

mCi: millicurie

μCi: microcurie

MEARA/G: methionine-glutamic acid-alanine-arginine-alanine/glycine

mg: milligram

μg: microgram

min: minute(s)

ml: milliliter

μl: microliter

mM: millimolar

MOPS: 3-(N-Morpholino)propanesulfonic acid

mRNA: messenger RNA

mt: mutant

MW: molecular weight markers

N: normality

NE: nuclear extract(s)

NLS: nuclear localization signal

NP-40: nonidet™ P 40

nt: nucleotide(s)

N-terminal: amino-terminal

OD: optic density

ON: overnight

ORF: open reading frame

pA: poly(A) signal

PABPN1: poly(A) binding protein nuclear 1

PAGE: polyacrylamide gel electrophoresis

PAP: poly(A) polymerase

PBS: phosphate-buffered saline

pBs: BlueScribe cloning vector

PBT: PBS/Triton X-100

PCBP: poly(C) binding protein

PCR: polymerase chain reaction

Plk: Polo-like kinase

PIPES: piperazine-N,N′-bis(2-ethanesulfonic acid)

pmol: picomoles

PMSF: phenylmethylsulfonyl fluoride

PolII: polymerase II

Poly(A): polyadenylation or polyadenylic acid

Poly(C): polycytidylic acid

Poly(rC): polyribocytidylic acid

pre-mRNA: precursor messenger RNA

PSF: PTB-associated-splicing factor

PTB: polypyrimidine tract binding protein

PVA: poly (vinyl alcohol)

PVP: polyvinylpyrrolidone

Py: pyrimidine(s)

RAM: rabbit anti-mouse

RBD: RNA binding domain

RNA: ribonucleic acid

RNase: ribonuclease

rPAP: recombinant PAP

rpm: rotations per minute

RT: room temperature

s: second(s)

***Sco*:** *scutoid*

SDS: sodium dodecyl sulfate

SELEX: systematic evolution of ligands by exponential enrichment

SL2: Schneider's *Drosophila* Line 2

***SM6*:** *Second Multiple 6*

snRNA: small nuclear RNA

snRNP: small nuclear ribonucleoprotein

ss: splice site

SSC: saline sodium citrate buffer solution

ssDNA: single strand DNA

SSPE: saline sodium phosphate EDTA buffer

st error: standard error

SUMO: small ubiquitin-like modifier

SV40E: simian virus 40 early

SV40L: simian virus 40 late

sxl: *sex-lethal*

T: thymine deoxynucleotide

TAE: tris-acetate-EDTA

TBE: tris-borate-EDTA

TCA: trichloroacetic acid

***T. castaneum*:** *Tribolium castaneum*

***TM6*:** *Third Multiple 6*

TPR: tetratrico-peptide repeat(s)

Tris: tris(hydroxymethyl)aminomethane

tRNA: transfer RNA

TY: tryptone yeast culture medium

T4 PNK: T4 polynucleotide kinase

U: uracil nucleotide

UCSC: University of California Santa Cruz

USE: upstream sequence element

UTP: uridine triphosphate

UTR: unstranslated region

UV.: ultra-violet

V: volts

v/v: volume/volume

W: watts

WD repeats: tryptophane-aspartate repeats

wt: wild type

w/v: weight/volume

PART VI

REFERENCES

References

Atwater, J.A., Wisdom, R., and Verma, I.M. 1990. Regulated mRNA stability. *Annu Rev Genet* **24**: 519-541.

Audibert, A., Juge, F., and Simonelig, M. 1998. The suppressor of forked protein of Drosophila, a homologue of the human 77K protein required for mRNA 3'-end formation, accumulates in mitotically-active cells. *Mech Dev* **72**(1-2): 53-63.

Audibert, A. and Simonelig, M. 1998. Autoregulation at the level of mRNA 3' end formation of the suppressor of forked gene of Drosophila melanogaster is conserved in Drosophila virilis. *Proc Natl Acad Sci U S A* **95**(24): 14302-14307.

Audibert, A. and Simonelig, M. 1999. The suppressor of forked gene of Drosophila, which encodes a homologue of human CstF-77K involved in mRNA 3'-end processing, is required for progression through mitosis. *Mech Dev* **82**(1-2): 41-50.

Bainbridge, S.P. and Bownes, M. 1981. Staging the metamorphosis of Drosophila melanogaster. *J Embryol Exp Morphol* **66**: 57-80.

Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* **11**(13): 1703-1716.

Barreau, C., Paillard, L., and Osborne, H.B. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res* **33**(22): 7138-7150.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**(7): 1001-1010.

Ben-Dov, C., Hartmann, B., Lundgren, J., and Valcarcel, J. 2007. Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem*.

Benoit, B., Juge, F., Iral, F., Audibert, A., and Simonelig, M. 2002. Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. *Proc Natl Acad Sci U S A* **99**(16): 10593-10598.

Benoit, B., Nemeth, A., Aulner, N., Kuhn, U., Simonelig, M., Wahle, E., and Bourbon, H.M. 1999. The Drosophila poly(A)-binding protein II is ubiquitous throughout Drosophila development and has the same function in mRNA polyadenylation as its bovine homolog in vitro. *Nucleic Acids Res* **27**(19): 3771-3778.

Bentley, D.L. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* **17**(3): 251-256.

Bienroth, S., Keller, W., and Wahle, E. 1993. Assembly of a processive messenger RNA polyadenylation complex. *Embo J* **12**(2): 585-594.

Bienroth, S., Wahle, E., Suter-Crazzolara, C., and Keller, W. 1991. Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J Biol Chem* **266**(29): 19768-19776.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291-336.

Brackenridge, S. and Proudfoot, N.J. 2000. Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. *Mol Cell Biol* **20**(8): 2660-2669.

Brown, K.M. and Gilmartin, G.M. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* **12**(6): 1467-1476.

Brown, P.H., Tiley, L.S., and Cullen, B.R. 1991. Efficient polyadenylation within the human immunodeficiency virus type 1 long terminal repeat requires flanking U3-specific sequences. *J Virol* **65**(6): 3340-3343.

Bruce Alberts, A.J., Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. 2002. *Molecular Biology of the Cell*. Garland Science Textbooks.

Bruce, S.R., Dingle, R.W., and Peterson, M.L. 2003. B-cell and plasma-cell splicing differences: a potential role in regulated immunoglobulin RNA processing. *Rna* **9**(10): 1264-1273.

Buratowski, S. 2005. Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* **17**(3): 257-261.

Callebaut, I., Moshous, D., Mornon, J.P., and de Villartay, J.P. 2002. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30**(16): 3592-3601.

Calvo, O. and Manley, J.L. 2001. Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol Cell* **7**(5): 1013-1023.

Carswell, S. and Alwine, J.C. 1989. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol Cell Biol* **9**(10): 4248-4258.

Castelo-Branco, P., Furger, A., Wollerton, M., Smith, C., Moreira, A., and Proudfoot, N. 2004. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* **24**(10): 4174-4183.

Castrillon, D.H., Gonczy, P., Alexander, S., Rawson, R., Eberhart, C.G., Viswanathan, S., DiNardo, S., and Wasserman, S.A. 1993. Toward a molecular genetic analysis of spermatogenesis in Drosophila melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis. *Genetics* **135**(2): 489-505.

Chen, F., MacDonald, C.C., and Wilusz, J. 1995. Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res* **23**(14): 2614-2620.

Chen, Z., Li, Y., and Krug, R.M. 1999. Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *Embo J* **18**(8): 2273-2283.

Cherrington, J. and Ganem, D. 1992. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *Embo J* **11**(4): 1513-1524.

Chkheidze, A.N., Lyakhov, D.L., Makeyev, A.V., Morales, J., Kong, J., and Liebhaber, S.A. 1999. Assembly of the alpha-globin mRNA stability complex reflects binary interaction between the pyrimidine-rich 3' untranslated region determinant and poly(C) binding protein alphaCP. *Mol Cell Biol* **19**(7): 4572-4581.

Choi, Y.D. and Dreyfuss, G. 1984. Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. *J Cell Biol* **99**(6): 1997-2004.

Chou, Z.F., Chen, F., and Wilusz, J. 1994. Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals. *Nucleic Acids Res* **22**(13): 2525-2531.

Christofori, G. and Keller, W. 1988. 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. *Cell* **54**(6): 875-889.

Christofori, G. and Keller, W. 1989. Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. *Mol Cell Biol* **9**(1): 193-203.

Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**(21): 2755-2766.

Coller, J. and Parker, R. 2004. Eukaryotic mRNA decapping. *Annu Rev Biochem* **73**: 861-890.

Conne, B., Stutz, A., and Vassalli, J.D. 2000. The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med* **6**(6): 637-641.

Curtiss, J. and Heilig, J.S. 1995. Establishment of Drosophila imaginal precursor cells is controlled by the Arrowhead gene. *Development* **121**(11): 3819-3828.

Curtiss, J. and Heilig, J.S. 1997. Arrowhead encodes a LIM homeodomain protein that distinguishes subsets of Drosophila imaginal cells. *Dev Biol* **190**(1): 129-141.

Danckwardt, S., Gehring, N.H., Neu-Yilik, G., Hundsdoerfer, P., Pforsich, M., Frede, U., Hentze, M.W., and Kulozik, A.E. 2004. The prothrombin 3'end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. *Blood* **104**(2): 428-435.

Danckwardt, S., Hartmann, K., Katz, B., Hentze, M.W., Levy, Y., Eichele, R., Deutsch, V., Kulozik, A.E., and Ben-Tal, O. 2006. The prothrombin 20209 C-->T mutation in Jewish-Moroccan Caucasians: molecular analysis of gain-of-function of 3' end processing. *J Thromb Haemost* **4**(5): 1078-1085.

Danckwardt, S., Kaufmann, I., Gentzel, M., Foerstner, K.U., Gantzert, A.S., Gehring, N.H., Neu-Yilik, G., Bork, P., Keller, W., Wilm, M., Hentze, M.W., and Kulozik, A.E. 2007. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *Embo J* **26**(11): 2658-2669.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. 1997. Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389**(6649): 399-402.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. 2000. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *Embo J* **19**(21): 5895-5904.

Deak, P., Omar, M.M., Saunders, R.D., Pal, M., Komonyi, O., Szidonya, J., Maroy, P., Zhang, Y., Ashburner, M., Benos, P., Savakis, C., Siden-Kiamos, I., Louis, C., Bolshakov, V.N., Kafatos, F.C., Madueno, E., Modolell, J., and Glover, D.M. 1997. P-element insertion alleles of essential genes on the third chromosome of Drosophila melanogaster: correlation of physical and cytogenetic maps in chromosomal region 86E-87F. *Genetics* **147**(4): 1697-1722.

DeZazzo, J.D. and Imperiale, M.J. 1989. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol Cell Biol* **9**(11): 4951-4961.

DeZazzo, J.D., Kilpatrick, J.E., and Imperiale, M.J. 1991. Involvement of long terminal repeat U3 sequences overlapping the transcription control region in human immunodeficiency virus type 1 mRNA 3' end formation. *Mol Cell Biol* **11**(3): 1624-1630.

Donaldson, M.M., Tavares, A.A., Ohkura, H., Deak, P., and Glover, D.M. 2001. Metaphase arrest with centromere separation in polo mutants of Drosophila. *J Cell Biol* **153**(4): 663-676.

Edmonds, M. 2002. A history of poly A sequences: from formation to factors to function. *Prog Nucleic Acid Res Mol Biol* **71**: 285-389.

Edwalds-Gilbert, G., Veraldi, K.L., and Milcarek, C. 1997. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* **25**(13): 2547-2561.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C. 2002. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**(5563): 2258-2261.

Funk, C.D., Funk, L.B., Kennedy, M.E., Pong, A.S., and Fitzgerald, G.A. 1991. Human platelet/erythroleukemia cell prostaglandin G/H synthase: cDNA cloning, expression, and gene chromosomal assignment. *Faseb J* **5**(9): 2304-2312.

Fuse, N., Hirose, S., and Hayashi, S. 1994. Diploidy of Drosophila imaginal cells is maintained by a transcriptional repressor encoded by escargot. *Genes Dev* **8**(19): 2270-2281.

Gaiano, N., Amsterdam, A., Kawakami, K., Allende, M., Becker, T., and Hopkins, N. 1996. Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383**(6603): 829-832.

Galli, G., Guise, J., Tucker, P.W., and Nevins, J.R. 1988. Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain RNAs. *Proc Natl Acad Sci U S A* **85**(8): 2439-2443.

Galli, G., Guise, J.W., McDevitt, M.A., Tucker, P.W., and Nevins, J.R. 1987. Relative position and strengths of poly(A) sites as well as transcription termination are critical to membrane versus secreted mu-chain expression during B-cell development. *Genes Dev* **1**(5): 471-481.

Garcia-Bellido, A. and Merriam, J.R. 1971. Clonal parameters of tergite development in Drosophila. *Dev Biol* **26**(2): 264-276.

Gawande, B., Robida, M.D., Rahn, A., and Singh, R. 2006. Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline. *Embo J* **25**(6): 1263-1272.

Gehring, N.H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M.W., and Kulozik, A.E. 2001. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat Genet* **28**(4): 389-392.

Gilbert, S.F., ed. 2000. *Developmental Biology*. Sinauer Associates, Inc.

Gilmartin, G.M., Fleming, E.S., Oetjen, J., and Graveley, B.R. 1995. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev* **9**(1): 72-83.

Gilmartin, G.M. and Nevins, J.R. 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes Dev* **3**(12B): 2180-2190.

Gilmartin, G.M. and Nevins, J.R. 1991. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. *Mol Cell Biol* **11**(5): 2432-2438.

Glover, D.M. 2005. Polo kinase and progression through M phase in Drosophila: a perspective from the spindle poles. *Oncogene* **24**(2): 230-237.

Gompel, N. and Carroll, S.B. 2003. Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* **424**(6951): 931-935.

Gorlach, M., Burd, C.G., and Dreyfuss, G. 1994. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem* **269**(37): 23074-23078.

Graber, J.H., Cantor, C.R., Mohr, S.C., and Smith, T.F. 1999. In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proc Natl Acad Sci U S A* **96**(24): 14055-14060.

Grosveld, G.C., de Boer, E., Shewmaker, C.K., and Flavell, R.A. 1982. DNA sequences necessary for transcription of the rabbit beta-globin gene in vivo. *Nature* **295**(5845): 120-126.

Guerra, M., Postlethwait, J.H., and Schneiderman, H.A. 1973. The development of the imaginal abdomen of Drosophila melanogaster. *Dev Biol* **32**(2): 361-372.

Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C., and Mattaj, L.W. 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**(3): 531-541.

Hall-Pogar, T., Liang, S., Hague, L.K., and Lutz, C.S. 2007. Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR. *Rna* **13**(7): 1103-1115.

Hall-Pogar, T., Zhang, H., Tian, B., and Lutz, C.S. 2005. Alternative polyadenylation of cyclooxygenase-2. *Nucleic Acids Res* **33**(8): 2565-2579.

Hamilton, B.J., Nagy, E., Malter, J.S., Arrick, B.A., and Rigby, W.F. 1993. Association of heterogeneous nuclear ribonucleoprotein A1 and C proteins with reiterated AUUUA sequences. *J Biol Chem* **268**(12): 8881-8887.

Hartenstein, V. 1993. *Atlas of Drosophila Development* Cold Spring Harbor Laboratory Press.

Hayashi, S. 1996. A Cdc2 dependent checkpoint maintains diploidy in Drosophila. *Development* **122**(4): 1051-1058.

Hayashi, S., Hirose, S., Metcalfe, T., and Shirras, A.D. 1993. Control of imaginal cell development by the escargot gene of Drosophila. *Development* **118**(1): 105-115.

Hayashi, S. and Yamaguchi, M. 1999. Kinase-independent activity of Cdc2/cyclin A prevents the S phase in the Drosophila cell cycle. *Genes Cells* **4**(2): 111-122.

Herrmann, S., Amorim, I., and Sunkel, C.E. 1998. The POLO kinase is required at multiple stages during spermatogenesis in Drosophila melanogaster. *Chromosoma* **107**(6-7): 440-451.

Higgs, D.R., Goodbourn, S.E., Lamb, J., Clegg, J.B., Weatherall, D.J., and Proudfoot, N.J. 1983. Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* **306**(5941): 398-400.

Hirose, Y. and Manley, J.L. 1998. RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**(6697): 93-96.

Hirose, Y. and Manley, J.L. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**(12): 1415-1429.

Hla, T. and Neilson, K. 1992. Human cyclooxygenase-2 cDNA. *Proc Natl Acad Sci U S A* **89**(16): 7384-7388.

Hofmann, I., Schnolzer, M., Kaufmann, I., and Franke, W.W. 2002. Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes. *Mol Biol Cell* **13**(5): 1665-1676.

Jacobs, H.W., Knoblich, J.A., and Lehner, C.F. 1998. Drosophila Cyclin B3 is required for female fertility and is dispensable for mitosis like Cyclin B. *Genes Dev* **12**(23): 3741-3751.

Jenny, A., Hauri, H.P., and Keller, W. 1994. Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. *Mol Cell Biol* **14**(12): 8183-8190.

Jenny, A., Minvielle-Sebastia, L., Preker, P.J., and Keller, W. 1996. Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* **274**(5292): 1514-1517.

Juge, F., Audibert, A., Benoit, B., and Simonelig, M. 2000. Tissue-specific autoregulation of Drosophila suppressor of forked by alternative poly(A) site utilization leads to accumulation of the suppressor of forked protein in mitotically active cells. *Rna* **6**(11): 1529-1538.

Kambach, C. and Mattaj, I.W. 1992. Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. *J Cell Biol* **118**(1): 11-21.

Kaminski, A., Hunt, S.L., Patton, J.G., and Jackson, R.J. 1995. Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *Rna* **1**(9): 924-938.

Kapp, L.D. and Lorsch, J.R. 2004. The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **73**: 657-704.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., Weber, R.J., Haussler, D., and Kent, W.J. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res* **31**(1): 51-54.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., and Keller, W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *Embo J* **23**(3): 616-626.

Keller, W., Bienroth, S., Lang, K.M., and Christofori, G. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *Embo J* **10**(13): 4241-4249.

Kleiman, F.E. and Manley, J.L. 2001. The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* **104**(5): 743-753.

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., and Handa, H. 2006. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res* **34**(21): 6264-6271.

Kuhn, U. and Wahle, E. 2004. Structure and function of poly(A) binding proteins. *Biochim Biophys Acta* **1678**(2-3): 67-84.

Kujovich, J.L. 2006. Prothrombin Thrombophilia. In *GeneReviews* (ed. R.A. Pagon). University of Washington, Seattle.

Kumagai, A. and Dunphy, W.G. 1996. Purification and molecular cloning of Plx1, a Cdc25-regulatory kinase from Xenopus egg extracts. *Science* **273**(5280): 1377-1380.

Lake, R.J. and Jelinek, W.R. 1993. Cell cycle- and terminal differentiation-associated regulation of the mouse mRNA encoding a conserved mitotic protein kinase. *Mol Cell Biol* **13**(12): 7793-7801.

Legendre, M. and Gautheret, D. 2003. Sequence determinants in human polyadenylation site selection. *BMC Genomics* **4**(1): 7.

Lewin, B., ed. 1997. *Genes VI*. Oxford University Press.

Liang, S. and Lutz, C.S. 2006. p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *Rna* **12**(1): 111-121.

Llamazares, S., Moreira, A., Tavares, A., Girdham, C., Spruce, B.A., Gonzalez, C., Karess, R.E., Glover, D.M., and Sunkel, C.E. 1991. polo encodes a protein kinase homolog required for mitosis in Drosophila. *Genes Dev* **5**(12A): 2153-2165.

Lukiw, W.J. and Bazan, N.G. 1997. Cyclooxygenase 2 RNA message abundance, stability, and hypervariability in sporadic Alzheimer neocortex. *J Neurosci Res* **50**(6): 937-945.

Lutz, C.S. and Alwine, J.C. 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes Dev* **8**(5): 576-586.

Lutz, C.S., Cooke, C., O'Connor, J.P., Kobayashi, R., and Alwine, J.C. 1998. The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. *Rna* **4**(12): 1493-1499.

Lutz, C.S., Murthy, K.G., Schek, N., O'Connor, J.P., Manley, J.L., and Alwine, J.C. 1996. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev* **10**(3): 325-337.

Ma, J., Gunderson, S.I., and Phillips, C. 2006. Non-snRNP U1A levels decrease during mammalian B-cell differentiation and release the IgM secretory poly(A) site from repression. *Rna* **12**(1): 122-132.

MacDonald, C.C. and Redondo, J.L. 2002. Reexamining the polyadenylation signal: were we wrong about AAUAAA? *Mol Cell Endocrinol* **190**(1-2): 1-8.

MacDonald, C.C., Wilusz, J., and Shenk, T. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* **14**(10): 6647-6654.

Madhavan, M.M. and Madhavan, K. 1980. Morphogenesis of the epidermis of adult abdomen of Drosophila. *J Embryol Exp Morphol* **60**: 1-31.

Madhavan, M.M. and Schneiderman, H.A. 1977. Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development of Drosophila melanogaster. *Wilhelm Roux's archives of developmental biology* **186**: 269-305.

Makarova, O., Kamberov, E., and Margolis, B. 2000. Generation of deletion and point mutations with one primer in a single cloning step. *Biotechniques* **29**(5): 970-972.

Makeyev, A.V. and Liebhaber, S.A. 2002. The poly(C)-binding proteins: a multiplicity of functions and a search for mechanisms. *Rna* **8**(3): 265-278.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**(7121): 953-956.

Manley, J.L. 1983. Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate. *Cell* **33**(2): 595-605.

Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M.T., and Milcarek, C. 1998. Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. *Proc Natl Acad Sci U S A* **95**(19): 11095-11100.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**(5): 911-940.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614): 357-361.

Mifflin, R.C. and Kellems, R.E. 1991. Coupled transcription-polyadenylation in a cell-free system. *J Biol Chem* **266**(29): 19593-19598.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. 2006. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *Embo J* **25**(20): 4854-4864.

Mirouse, V., Formstecher, E., and Couderc, J.L. 2006. Interaction between Polo and BicD proteins links oocyte determination and meiosis control in Drosophila. *Development* **133**(20): 4005-4013.

Mitchelson, A., Simonelig, M., Williams, C., and O'Hare, K. 1993. Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability. *Genes Dev* **7**(2): 241-249.

Moore, C.L., Chen, J., and Whoriskey, J. 1988. Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. *Embo J* **7**(10): 3159-3169.

Moore, C.L. and Sharp, P.A. 1984. Site-specific polyadenylation in a cell-free reaction. *Cell* **36**(3): 581-591.

Moore, C.L. and Sharp, P.A. 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell* **41**(3): 845-855.

Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L., and Proudfoot, N.J. 1998. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* **12**(16): 2522-2534.

Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N.J. 1995. Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *Embo J* **14**(15): 3809-3819.

Mount, S.M. and Salz, H.K. 2000. Pre-messenger RNA processing factors in the Drosophila genome. *J Cell Biol* **150**(2): F37-44.

Moutinho-Santos, T., Sampaio, P., Amorim, I., Costa, M., and Sunkel, C.E. 1999. In vivo localisation of the mitotic POLO kinase shows a highly dynamic association with the mitotic apparatus during early embryogenesis in Drosophila. *Biol Cell* **91**(8): 585-596.

Murata, T., Nagaso, H., Kashiwabara, S., Baba, T., Okano, H., and Yokoyama, K.K. 2001. The hiiragi gene encodes a poly(A) polymerase, which controls the formation of the wing margin in Drosophila melanogaster. *Dev Biol* **233**(1): 137-147.

Murthy, K.G. and Manley, J.L. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* **267**(21): 14804-14811.

Murthy, K.G. and Manley, J.L. 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* **9**(21): 2672-2683.

Nakielny, S., Fischer, U., Michael, W.M., and Dreyfuss, G. 1997. RNA transport. *Annu Rev Neurosci* **20**: 269-301.

Natalizio, B.J., Muniz, L.C., Arhin, G.K., Wilusz, J., and Lutz, C.S. 2002. Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. *J Biol Chem* **277**(45): 42733-42740.

Ninov, N., Chiarelli, D.A., and Martin-Blanco, E. 2007. Extrinsic and intrinsic mechanisms directing epithelial cell sheet replacement during Drosophila metamorphosis. *Development* **134**(2): 367-379.

Pan, Z., Zhang, H., Hague, L.K., Lee, J.Y., Lutz, C.S., and Tian, B. 2006. An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism. *Gene* **366**(2): 325-334.

Pearson, J., Godinho, S.A., Tavares, A., and Glover, D.M. 2006. Heterologous expression of mammalian Plk1 in Drosophila reveals divergence from Polo during late mitosis. *Exp Cell Res* **312**(6): 770-781.

Penalva, L.O. and Sanchez, L. 2003. RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. *Microbiol Mol Biol Rev* **67**(3): 343-359, table of contents.

Perez, I., Lin, C.H., McAfee, J.G., and Patton, J.G. 1997. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *Rna* **3**(7): 764-778.

Peterson, M.L. 1994. Regulated immunoglobulin (Ig) RNA processing does not require specific cis-acting sequences: non-Ig RNA can be alternatively processed in B cells and plasma cells. *Mol Cell Biol* **14**(12): 7891-7898.

Peterson, M.L., Bingham, G.L., and Cowan, C. 2006. Multiple features contribute to the use of the immunoglobulin M secretion-specific poly(A) signal but are not required for developmental regulation. *Mol Cell Biol* **26**(18): 6762-6771.

Peterson, M.L., Gimmi, E.R., and Perry, R.P. 1991. The developmentally regulated shift from membrane to secreted mu mRNA production is accompanied by an increase in cleavage-polyadenylation efficiency but no measurable change in splicing efficiency. *Mol Cell Biol* **11**(4): 2324-2327.

Peterson, M.L. and Perry, R.P. 1989. The regulated production of mu m and mu s mRNA is dependent on the relative efficiencies of mu s poly(A) site usage and the c mu 4-to-M1 splice. *Mol Cell Biol* **9**(2): 726-738.

Phillips, C., Jung, S., and Gunderson, S.I. 2001. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. *Embo J* **20**(22): 6443-6452.

Phillips, C., Pachikara, N., and Gunderson, S.I. 2004. U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. *Mol Cell Biol* **24**(14): 6162-6171.

Pogge von Strandmann, E., Senkel, S., and Ryffel, G.U. 2001. ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. *Biol Chem* **382**(9): 1379-1385.

Portman, D.S. and Dreyfuss, G. 1994. RNA annealing activities in HeLa nuclei. *Embo J* **13**(1): 213-221.

Preker, P.J. and Keller, W. 1998. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci* **23**(1): 15-16.

Preker, P.J., Lingner, J., Minvielle-Sebastia, L., and Keller, W. 1995. The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell* **81**(3): 379-389.

Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**(3): 272-278.

Proudfoot, N.J. and Brownlee, G.G. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**(5574): 211-214.

Proudfoot, N.J., Furger, A., and Dye, M.J. 2002. Integrating mRNA processing with transcription. *Cell* **108**(4): 501-512.

Qian, Y.W., Erikson, E., Li, C., and Maller, J.L. 1998. Activated polo-like kinase Plx1 is required at multiple points during mitosis in Xenopus laevis. *Mol Cell Biol* **18**(7): 4262-4271.

Retelska, D., Iseli, C., Bucher, P., Jongeneel, C.V., and Naef, F. 2006. Similarities and differences of polyadenylation signals in human and fly. *BMC Genomics* **7**: 176.

Roberts, D.B., ed. 1986. *Drosophila: a Practical Approach*. IRL Press at Oxford University Press.

Roseland, C.R. and Schneiderman, H.A. 1979. Regulation and Metamorphosis of the Abdominal Histoblasts of Drosophila melanogaster. *Wilhelm Roux's archives of developmental biology* **186**: 235-265.

Rothnie, H.M., Reid, J., and Hohn, T. 1994. The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. *Embo J* **13**(9): 2200-2210.

Ruegsegger, U., Beyer, K., and Keller, W. 1996. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* **271**(11): 6107-6113.

Ruegsegger, U., Blank, D., and Keller, W. 1998. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell* **1**(2): 243-253.

Russnak, R. and Ganem, D. 1990. Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes Dev* **4**(5): 764-776.

Russnak, R.H. 1991. Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive. *Nucleic Acids Res* **19**(23): 6449-6456.

Ryan, K., Calvo, O., and Manley, J.L. 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *Rna* **10**(4): 565-573.

Sambrook, J. and Russell, D., ed. 2001. *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press.

Sanfacon, H., Brodmann, P., and Hohn, T. 1991. A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes Dev* **5**(1): 141-149.

Schek, N., Cooke, C., and Alwine, J.C. 1992. Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Mol Cell Biol* **12**(12): 5386-5393.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res* **18**(19): 5799-5805.

Sheets, M.D. and Wickens, M. 1989. Two phases in the addition of a poly(A) tail. *Genes Dev* **3**(9): 1401-1412.

Shell, S.A., Hesse, C., Morris, S.M., Jr., and Milcarek, C. 2005. Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection. *J Biol Chem* **280**(48): 39950-39961.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**(8): 1034-1050.

Singh, R., Valcarcel, J., and Green, M.R. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**(5214): 1173-1176.

Sisodia, S.S., Sollner-Webb, B., and Cleveland, D.W. 1987. Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Mol Cell Biol* **7**(10): 3602-3612.

Smale, S.T. and Tjian, R. 1985. Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Mol Cell Biol* **5**(2): 352-362.

Sullivan, W., Ashburner, M., and Hawley, R.S., ed. 2000. *Drosophila Protocols*. Cold Spring Harbor Laboratory Press.

Sunkel, C.E. and Glover, D.M. 1988. polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. *J Cell Sci* **89 (Pt 1)**: 25-38.

Swanson, M.S. and Dreyfuss, G. 1988. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol* **8**(5): 2237-2241.

Takagaki, Y., MacDonald, C.C., Shenk, T., and Manley, J.L. 1992. The human 64-kDa polyadenylylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs. *Proc Natl Acad Sci U S A* **89**(4): 1403-1407.

Takagaki, Y. and Manley, J.L. 1992. A human polyadenylation factor is a G protein beta-subunit homologue. *J Biol Chem* **267**(33): 23471-23474.

Takagaki, Y. and Manley, J.L. 1994. A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. *Nature* **372**(6505): 471-474.

Takagaki, Y. and Manley, J.L. 1997. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* **17**(7): 3907-3914.

Takagaki, Y. and Manley, J.L. 1998. Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation. *Mol Cell* **2**(6): 761-771.

Takagaki, Y. and Manley, J.L. 2000. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* **20**(5): 1515-1525.

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., and Shenk, T. 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* **4**(12A): 2112-2120.

Takagaki, Y., Ryner, L.C., and Manley, J.L. 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52**(5): 731-742.

Takagaki, Y., Ryner, L.C., and Manley, J.L. 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev* **3**(11): 1711-1724.

Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**(5): 941-952.

Tavares, A.A., Glover, D.M., and Sunkel, C.E. 1996. The conserved mitotic kinase polo is regulated by phosphorylation and has preferred microtubule-associated substrates in Drosophila embryo extracts. *Embo J* **15**(18): 4873-4883.

Thisted, T., Lyakhov, D.L., and Liebhaber, S.A. 2001. Optimized RNA targets of two closely related triple KH domain proteins, heterogeneous nuclear ribonucleoprotein K and alphaCP-2KL, suggest Distinct modes of RNA recognition. *J Biol Chem* **276**(20): 17484-17496.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33**(1): 201-212.

Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J.C. 1991. The human immunodeficiency virus type 1 polyadenylylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylylation. *Proc Natl Acad Sci U S A* **88**(6): 2108-2112.

van de Weerdt, B.C. and Medema, R.H. 2006. Polo-like kinases: a team in control of the division. *Cell Cycle* **5**(8): 853-864.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19**(11): 1315-1327.

Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J., and Milcarek, C. 2001. hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol Cell Biol* **21**(4): 1228-1238.

Vethantham, V., Rao, N., and Manley, J.L. 2007. Sumoylation modulates the assembly and activity of the pre-mRNA 3' processing complex. *Mol Cell Biol* **27**(24): 8848-8858.

Wahle, E. 1991a. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* **66**(4): 759-768.

Wahle, E. 1991b. Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* **266**(5): 3131-3139.

Wahle, E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* **270**(6): 2800-2808.

Wahle, E., Lustig, A., Jeno, P., and Maurer, P. 1993. Mammalian poly(A)-binding protein II. Physical properties and binding to polynucleotides. *J Biol Chem* **268**(4): 2937-2945.

Weaver, T.A. and White, R.A. 1995. headcase, an imaginal specific gene required for adult morphogenesis in Drosophila melanogaster. *Development* **121**(12): 4149-4160.

Weiss, E.A., Gilmartin, G.M., and Nevins, J.R. 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. *Embo J* **10**(1): 215-219.

Whitelaw, E. and Proudfoot, N. 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *Embo J* **5**(11): 2915-2922.

Wickens, M. and Stephenson, P. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. *Science* **226**(4678): 1045-1051.

Williams, C.J. and O'Hare, K. 1996. Elimination of introns at the Drosophila suppressor-of-forked locus by P-element-mediated gene conversion shows that an RNA lacking a stop codon is dispensable. *Genetics* **143**(1): 345-351.

Wilusz, J., Pettine, S.M., and Shenk, T. 1989. Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. *Nucleic Acids Res* **17**(10): 3899-3908.

Wilusz, J. and Shenk, T. 1988. A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell* **52**(2): 221-228.

Wilusz, J. and Shenk, T. 1990. A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. *Mol Cell Biol* **10**(12): 6397-6407.

Wilusz, J., Shenk, T., Takagaki, Y., and Manley, J.L. 1990. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol Cell Biol* **10**(3): 1244-1248.

Wojcik, E., Murphy, A.M., Fares, H., Dang-Vu, K., and Tsubota, S.I. 1994. Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. *Genetics* **138**(4): 1163-1170.

Xing, H., Mayhew, C.N., Cullen, K.E., Park-Sarge, O.K., and Sarge, K.D. 2004. HSF1 modulation of Hsp70 mRNA polyadenylation via interaction with symplekin. *J Biol Chem* **279**(11): 10551-10555.

Yan, D.H., Weiss, E.A., and Nevins, J.R. 1995. Identification of an activity in B-cell extracts that selectively impairs the formation of an immunoglobulin mu s poly(A) site processing complex. *Mol Cell Biol* **15**(4): 1901-1906.

Yan, J. and Marr, T.G. 2005. Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Res* **15**(3): 369-375.

Yokoyama, C. and Tanabe, T. 1989. Cloning of human gene encoding prostaglandin endoperoxide synthase and primary structure of the enzyme. *Biochem Biophys Res Commun* **165**(2): 888-894.

Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., and Hovorun, D.M. 2003. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res* **31**(5): 1375-1386.

Zhao, J., Hyman, L., and Moore, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**(2): 405-445.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**(13): 3406-3415.