The I factor 5' UTR: physical properties and possible role in epigenetic control

Laura Bisoni

A thesis presented for the degree of PhD

University of Edinburgh

1999

Dedication

Dedico questa Tesi di Laurea a coloro che mi sono piu cari: i miei genitori Fausto e Giovanna, mia sorella Antonella e la sua famiglia, Nicola & Carmine, e James con tanto affetto.

Declaration

This thesis has been composed by myself and the work presented herein is my own, except where stated.

Laura Bisoni

September 1999

Acknowledgements

First and foremost I wish to thank my supervisor Professor David J. Finnegan for taking me up and for providing the Academic framework within which this project could develop. I am most grateful to him now for strongly encouraging me to finish what I had started and for keeping me on the straight and narrow. I also wish to express my enormous gratitude to Dr Jeff D. Bond, my second supervisor and mentor, for his unfailing help, advice, support, rapid rescue service and teachings from day one (literally) of my arrival at this University.

My thanks for all their support go to all my colleagues/friends (past & present) at the ICMB: they know who they are! I also wish to thank all members of staff at the Institute: media and glassware providers, the engineers, the servitors. Special regards to the photography team: Sir Frank, Graham and David for putting up with me and willingly giving me hours of their time (esp computer time). I hope the final result matches their efforts.

For their help with Thesis matters I thank the following: Prof. Adrian Bird and his group, for providing the bending ideas and the necessary plasmid; Prof, Paul Englund for sending me the bent Kinetoplast DNA; Professor Renato Paro for his excellent anti-Pc antibodies and for general advice on procedures; Professor James Birchler and his colleagues for emailing their invaluable protocol for fluorescence immunostaining. I also wish to thank Dr Marie Louise Loupart, Neville and Sharon for their patient help with the fluorescent microscope and Dr Margarete Heck for help and advice. Moreover I thank all the *I* factor colleagues in France for exciting Garchy meetings. In particular I am grateful to Drs Isabelle Busseau and Marie-Carmen Seleme for useful hot-off-the press email exchanges, and Dr Chantal Vaury for sending me some of her unpublished work.

I am particularly indebted to the following people: James McCormack for everything: endless patience, love, support and for keeping me sane (I hope)!; Dr Marie-Christine Chaboissier who introduced me to the joys of fly pushing and micromanipulation and, who, in the midst of it all, has become a very dear friend. Her excellent work on the I factor forms a crucial part of this thesis and her long distance friendship has been of great support to me; Dr David Blesa: I am delighted he joined our group even if only for one year, bringing fresh ideas (not least the "telomerase model") and invaluable expertise! I thank him for showing me the

secrets of the perfect Spanish chromosome squashes, for his friendship, for his Nature consultation and reference providing services and much more....(not least that wonderful walnut wine!); Heather Precious a dear and close friend since the first time we met! Her faith in me has kept me afloat; Helen Hibbert to whom I express my gratitude for offering me the hospitality of her home in this last difficult year.

To all those who helped me by reading drafts of the manuscripts go my most heartfelt thanks. All mistakes in this Thesis are entirely due to my ability to ignore good advice.

Finally, these acknowledgements would not be complete if I failed to honour the memory of the late Miss Enid Dickie to whom I dedicate the joys and tribulations of completing this Thesis.

Abbreviations

A	Adenosine
bp	base pairs
C	Cytosine
°C	degrees Celsius
CAT	Chloramphenicol acetyl transferase
Ci	Curie
dNTPs	doexynucleotides triphosphate
dH_20	de-ionised water
DAPI	4",6-diamino-2-phenylindole
DNA	deoxyribonucleic acid
DNase	deoxyribonulcease
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetra-acetic acid
G	guanosine
LB	Luria-Bertani (bacterial medium)
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeats
m	milli ($x10^{-3}$)
M	Molar
min	minutes

μ	micro ($x10^{-6}$)
n	nano ($x10^{-9}$)
OD	optical density
p	pico ($x10^{-12}$)
PA	polyacrylamide
PcG	polycomb group
PCR	polymerase chain reaction
PREs	polycomb response elements
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
RNase	ribonuclease
SDS	sodium dodecyl sulphate
T	thymine
TEMED	N,N,N'-tetramethylethylenediamine
Tris	tris (hydroxymethyl) aminomethane
TrxG	Trithorax group
tsds	target site duplications
UTR	untranslated transcribed region
UV	unltraviolet
v/v	volume pr volume
w/v	weight per volume

The I factor 5' UTR: physical properties and possible role in epigenetic control

Abstract

The I factor is a Long Interspersed Nuclear Element of *Drosophila melanogaster*. Transposition of the I factor occurs via an RNA intermediate and is restricted to the germline of hybrid females that have received one half of their chromosome complement from a parent devoid of active I factors. If the I factor devoid chromosomes are of maternal origin, I factor transposition is associated with lethality of the F2 embryos and high frequency of lethal mutations in the F3 generation. In contrast to its restricted pattern of transposition, I factor expression is detectable in somatic tissues of all flies that carry active I factors. Accumulation of I factor copies over the generations is associated with cessation of transposition and a decrease in expression in the female germline. As previously shown, the I factor 5' Untranslated Transcribed Region (UTR) can impose on a reporter gene a pattern of expression similar to that observed for complete I factors, i.e. high levels of expression in the germline of permissive females and lower yet significant levels in other tissues and in males. Furthermore, the I factor 5'UTR is sensitive to the presence of complete I factors and multiple copies of this sequence significantly reduce of the F2 lethality associated with I factor transposition. This work investigates the role of the 5' UTR in determining parental effects and in mediating processes of Polycomb-dependent I factor silencing. In addition, the possibility that the high Adenine-Thymine content of the 5'UTR might impose on this sequence an unusual conformation that could be responsible for some of the biological properties of the I factor is also investigated.

Table of contents

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1. The I factor and its host: a complex interaction

1.1.1. The hybrid dysgenesis syndrome in *Drosophila melanogaster*

The scientific literature of the 1970s is thick with reports of a novel phenomenon in *Drosophila* genetics. Since the previous decade, investigators from different laboratories had been investigating what, if any, would be the effects of individual chromosomes from naturally occurring populations of fruit flies on the fitness of laboratory stocks (reviewed in Engels, 1989). In order to extract "wild" chromosomes, they crossed newly caught flies with specimens from designated laboratory stocks. The striking finding was evidence of recombination in the F1 male hybrids (Hiraizumi, 1971). Even more remarkable came the later realisation that male recombination was only one symptom of a more complex syndrome of germline abnormalities that included: temperature-sensitive sterility, high frequency of mutations, chromosome aberrations and transmission ratio distortions (Kidwell *et al.*, 1977).

A hallmark of hybrid dysgenesis (Sved, 1976; Kidwell and Kidwell, 1976), as this new genetic disease became known , is its nonreciprocity. Affected F1 hybrids come mostly from crosses between fathers from natural populations and mothers from laboratory stocks (Kidwell, *et al.*, 1977). The progeny of the reciprocal crosses are usually unaffected and in rare cases show only mild symptoms. The factors responsible for this complex genetic dysfunction appeared to be scattered on the chromosmes of the paternally contributing strains (Engels, 1979b; Engels, 1980).

With hindsight it is remarkable to realise that the conditions that made these observations possible were set some 50 years earlier by the unsuspecting pioneers of *Drosophila* research, who had captured flies from the wild in different parts of the world and established them in the laboratory. Little did they know that this

forced isolation of fly populations would have contributed to the discovery of transposable elements in *Drosophila* and played such a crucial role in our understanding of the complex relationship with their host. Following on from the early observations, it soon became clear that hybrid dysgenesis was the result of complex chromosome-cytoplasm interactions (Engels, 1979a). Parallels were being drawn between the dysgenesis phenomenon and the much older observations on the mutator effects in *Drosophila* and hypotheses put forward about a possible relationship between the causative agents of hybrid dysgenesis and IS elements in bacteria (Green, 1973; Green, 1977).

The details of the complex syndrome affecting the progeny of a dysgenic cross were clearly pointing to the existance of at least two non-interacting forms of hybrid dysgenesis each with its specific set of manifestations and responses to modifiers, implying the involvmemnt of two different causative agents (Kidwell, 1979). To date three independent systems of hybrid dysgenesis in *Drosophila melanogaster* have been described: the P-M system (Engels, 1989), the *I-R* system (Finnegan, 1989a) and the H-E system (Blackman and Gelbart, 1989). In each case elements from a distinct family of transposable sequences, P elements, *I* factors and hobo elements respectively, are mobilized *en mass* with measurable consequences on the phenotype of the hosts. The P-M hybrid dysgenesis system is by far the best understood of the three. What has been learnt about it as well as the way in which it has been elucidated have set the pace for a whole new era in *Drosophila* molecular genetics providing indispensable tools for transgenesis (Spradling and Rubin, 1982; Rubin *et al.*, 1982). Moreover P-M dysgenesis has put forward a paradygm for the investigators of related phenomena, not least the *I-R* hybrid dysgeneis system.

1.1.2. The P-M hybrid dysgenesis system as a paradygm

In the P-M system, a cross between M females (from a Maternally contributing strain) and P males (from a Paternally contributing strain) results in progeny of both sexes suffering hybrid dysgenesis (Bregliano and Kidwell, 1983; Engels, 1989). The effects of hybrid dysgenesis are most severe if the F1 hybrids are placed at 29° C during late embryonic to early larval stages, which coincide with crucial events in germline development. Upon dissection, the adult gonads appear rudimentary and atrophic with little scope for successful reproduction, a condition known as gonadal dysgenesis (GD). A cross between P females and M males almost always results in healthy progeny even at the restrictive temperature. The surviving F2 individuals, descendants of the dysgenic cross, bear the signs of additional genetic abnormalities

associated with P-M hybrid dysgenesis, namely male recombination, mutations, and gross chromosomal rearrangements. The key manifestations of P-M hybrid dysgenesis were used to assign known fly strains to the M or P category following inter-strain crosses: almost invariably M strains descend from old laboratory stocks established before the 60's, whereas P strains comprise all the most recent laboratory stocks and all flies newly caught from the wild (Kidwell, *et al.*, 1977).

1.1.3. I-R hybrid dysgenesis

A distinct form of hybrid dysgenesis was first reported in 1971 which exclusively affected the female progeny of crosses between certain fly strains (Picard and L'Heritier, 1971; Picard *et al.*, 1972; Kidwell, 1979; Bregliano *et al.*, 1980; Bregliano and Kidwell, 1983; Finnegan, 1989a). The mothers in these dysgenic crosses come from a relatively small number of strains, denominated R for Reactive, descended from old laboratory stocks established no later than the 1940's. All remaining laboratory stocks, as well as natural populations world wide, provide the interacting fathers and are called *I* for Inducer. Affected F1 hybrid females, called SF from the French *stérilité femelle*, behave as if they carried maternal-effect embryonic lethal mutations: despite having normal ovaries showing no signs of GD, a variable percentage of the F2 embryos die after only a few nuclear divisions (Picard and L'Heritier, 1971; Picard, *et al.*, 1972).

Detailed examination of the dead embryos reveals that most of them have arrested at an ill defined preblastoderm stage and display a number of cytological abnormalities including chromosome fragmentation and enlarged spindles and asters (Picard *et al.*, 1977). Similar abnormalities associated with impaired nuclear divisions and preblastoderm arrest are also observed in unfertilised eggs undergoing parthenogenic development (Picard, *et al.*, 1977; Lavige, 1986). The number of nuclei in F2 embryos, however, is higher on average than that of parthenogenetic eggs (Lavige, 1986)and immunostaining of sperm tail markers indicates that about 94 % of the SF eggs are fertilised (Brigitte De Saint Phalle, personal communications). The exact causes of *I-R* associated embryonic lethality remain unclear. The surviving F2 embryos go on to develop normally and there is no measurable increase in mortality at subsequent larval and pupal stages (Picard, *et al.*, 1977). The adults that finally eclose appear phenotypicaly normal, yet further sib matings reveal that they harbour mutations, chromosomal rearrangements and suffer chromosome loss all at high frequencies (Picard *et al.*, 1978). The female progeny of the reciprocal cross (Inducer female x Reactive male), were named RSF

and are usually unaffected although they sometimes show mild dysgenic symptoms. Male siblings are unaffected in all cases (Picard and L'Heritier, 1971).

The probability of survival of the F2 embryos is increased whenever their SF mothers are placed at 30°C one to two days prior to egg laying (Picard, *et al.*, 1977). When the SF females are restored to 20°C, F2 survival returns to the original levels. However, before the original survival values are re established, the number of surviving embryos often falls to levels below those observed before the temperature shift. This decrease in embryonic survival reflects, in its severity and duration, the length of the thermic treatment. What seems to be critical is a change in temperature rather than the temperature itself: if SF females are raised at 29 °C throughout their development, there is no effect on the survival of their F2 progeny (Bucheton, 1979). It was also observed that as the SF mothers age, the percentage of surviving F2 embryos increases: in extreme cases, all the embryos descended from a 2-3 day old SF female might die, yet by the time the same SF female is 20 days old , 50 % or more of its progeny will survive (Picard, *et al.*, 1977).

Interestingly, embryonic lethality, also expressed as percentage hatchability, is completely independent of the genotype of the embryo (Picard, *et al.*, 1977). This observation reinforces the idea that an egg maturation abnormality is associated with *I-R* dysgenesis. Whatever is causing embryonic lethality acts prezygotically: the embryos die before significant expression levels of genes from the zygotic genome is achieved. Ageing and temperature changes must also be acting maternally so as to affect the oogenesis dysfunction brought about by the *I-R* interaction. This conclusion is further supported by the observation that placing F2 embryos at 30°C for a period of 24 hours, does not detectably change their probability of survival.

There is an interesting dimorphism in the genetic properties of the F2 survivors (Picard, 1978b). The F2 males have inducer behaviour when mated with reactive females. Their female siblings instead retain the SF characteristics of their mothers for up to 5 generations after the original dysgenic cross. The genotype of the mates of SF females seems to play an important role in determining the severity of dysgenesis in their daughters. The percentage hatchability of the F3 embryos is lowest if the SF grandmothers are mated with reactive males from the maternal strain in the original dysgeneic cross. In contrast F3 hatchability is increased if the mates of SF females are any of the following: 1) a sibling, 2) an inducer male, or 3) even a male from a strain that classifies as weakly reactive within the *I-R* system.

As a rule, mating SF females with inducer males not only decreases the severity of dysgenesis in their daughters, but also reduces the number of generations required to obtain inducer females among their descendants.

Another interesting phenomenon associated with *I-R* hybrid dysgenesis is the irreversible acquisition by reactive chromosomes of inducer properties after a single passage through the germline of a hybrid female. This process, known as chromosomal contamination (Picard, 1976), occurs at very high frequency in the germ-line of SF females and at lower frequency in the RSF germline (Picard, 1978b). A similar phenomenon was only rarely reported in the case of P-M dysgenesis (Engels, 1980; Bregliano and Kidwell, 1983), a difference that, in the light of what we now know, can be attributed to the distinct molecular biologies of the two transposable elements involved.

1.1.4. Reactivity

The severity of the dysgenesis syndrome suffered by SF females and measured as percentage hatchability of F2 embryos, varies from cross to cross (Bucheton *et al.*, 1976; Picard, *et al.*, 1978; Bregliano, *et al.*, 1980; Bregliano and Kidwell, 1983; Finnegan, 1989a). A small proportion of this variation is attributable to minor strain specific variations in the inducer levels of the fathers. With the present understanding, this small yet measurable variation in inducer levels can be attributed to the effect of genetic background and genomic location on the activity of the *I* factor. The bulk of variation in F2 embryo hatchabilities between SF females from different crosses can, however, be traced back to significant quantitative variability in a cellular state of their mothers, called reactivity. If the age of the SF females is held constant at the time of egg laying, high hatchabilities, close to wild type (80 % or more), conventionally indicate low reactivity levels of their R mothers, whilst low hatchabilities (50% or less) denote high reactivity levels of the R mothers. One complication is that, females within a given strain often appear heterogeneous with respect to their reactivity levels.

A great deal of the intra-strain variation in reactivity, seems to be caused by differences in the age of the females at the time of egg laying. Sublines comprising females of low or high reactivity can be selected from a given strain by simply breeding from older or younger females respectively for several generations. Surprisingly, the consequences of such breeding schemes are detectable in the hatchabilities of embryos from SF daughters as many as five generations after age selection along the maternal line has been relaxed.

Whatever the physiological environments of young versus older ovaries might be, they persist many generations after the age constraints that determine them have been removed. Reactivity levels appear to affect only the probability of F2 embryo survival but not the phenotype at the time of death (Lavige, 1986). Comparative electron microscopy of F2 embryos, whose grandmothers had different reactivity levels, reveals that although arrested embryos from weakly reactive females might have undergone a greater number of nuclear divisions, they are otherwise indistiguisheable from arrested embryos with a strong reactive background.

Temporary shifts in the temperature of rearing of reactive females reduce reactivity levels and therefore increase F2 hatchabilites in a similar heritable fashion as observed for ageing (Picard, *et al.*, 1977). The effects of age selection and thermic treatment on reactivity levels are, therefore, heritable through the female germline and entirely reversible: in both cases a number of generations elapses before the application or removal of modifiers have measureable effects.

1.1.5. Genetic determinants of reactivity

Experiments aimed at identifying the genetic determinants of reactivity revealed a complex interaction of cytoplasmic and chromosomal factors (Picard, 1978c)(Picard, 1978c; 1982,Bucheton). When reactive chromosomes from a strong reactive stock are progressively introduced into a weak reactive cytoplasm, by mating weak R females with males from a strong R stock, the resulting daughters at first display weak reactivity levels similar to those of their mothers. Over a number of generations, however, reactivity levels increase steadily until they reach those associated with the strong reactive chromosomes. Reactivity levels will display similar dynamics but in the opposite direction if the reciprocal cross (strong R female x male from weak R stock), is performed. Thus, in the short term, inheritance of reactivity is cytoplasmic through the maternal line. In the long term, however, reactivity levels are determined by multiple genetic factors with additive effects of the chromosomes. This duplicity of determinants could explain the reversibility of the effects of age selection and thermic treatment on reactivity levels.

A number of scenarios can be hypothesised where temperature and ageing cause quantitative and/or qualitative alterations of molecular factors directly or indirectly involved in determining reactivity. From the time modifier action ceases, a few generations are required to restore original reactivity levels as the pools of such factors are replenished and/or altered factors replaced by newly synthesised ones. In

any case the process of gradual yet random substituion of components provides the stochastic elements that can account for the complex dynamics of reactivity within populations and through the generations. An example of the latter is the observation that dysgenesis does not affect all the female offspring of SF (and RSF) females: whilst F2 daughters fit within relatively narrow distributions of hatchabilities, there is clear segregation of a few unaffected female siblings in what appears to be a stochastic manner (Picard, 1978c).

Mating schemes aimed at determining the inheritance of dysgenesis by the descendants of RSF females uncovered another surprising property of reactivity (Picard, 1978c). When unaffected RSF females are mated to reactive males, the resulting F2 females suffer dysgenesis. The severity of the dysgenesis is greatest if the reactive males come from the paternal strain. This observation contradicts the assumption that has so far been implicit: that reactivity represents a permissive and therefore recessive state whereas the inducer character is dominant. RSF females, however, have received inducer cytoplasm and chromosomes from their mothers, which implies that, they should behave as inducer females if reactivity were a recessive state. There are at least two possible explanations for the observed dysgenesis in the daughters of RSF females. One possibility is that the paternal reactive chromosomes are able to start a reactive cytoplasm "*de novo*" (Picard, 1978c). In this case reactivity is seen as codominant and is expressed if enough reactive chromosomes are introduced in an inducer background.

An alternative possibility is that mating RSF females with reactive males further reduces (despite the occurrance chromosomal contamination in the germline of these females (Picard, 1976)) the number of inducer chromosomes, disrupting a critical cytoplasm/chromosome ratio of inducer determinants (Finnegan, 1989a). In the female germline this would permit interaction between inducer factors and host reactivity which results in F3 embryonic lethality. According to this second view, reactivity coexists with the inducer state at all times and would be no longer detectable in inducer strains because the very process by which it is measured, *I* factor transposition and its consequences on embryonic viability, does not occur in inducer backgrounds. The early genetic evidence lends some support to the second hypothesis: ageing and thermic treatment of SF and RSF females affect dysgenesis in their daughters in the same direction as observed for the females in the initial dysgenic cross (Picard, 1978c). This means that, despite the presence of inducer chromosomes, reactivity is still detectable through its responses to modifiers.

In the P-M system, the M state or cytotype represents the absence of P element products and is therefore both necessary and sufficient to P element expression (Ronsseray *et al.*, 1993). However, reactivity is not simply the absence of *I* factors: it is necessary but not sufficient per se for *I* factor transposition. The latter can only occur in the few generations that it takes to establish the inducer state, after which *I* factor transposition stops despite the persistance of reactivity (Pelisson and Bregliano, 1987). Moreover, unlike the M cytotyope, reactivity can be modified *a priori* many generations before interaction with the *I* factors, by changing the breeding conditions of reactive stocks.

The effects of ageing and temperature on reactivity levels have, for almost thirty years now, been regarded as hallmark properties of reactivity (Bregliano, *et al.*, 1980). There is, however, evidence showing that ageing and temperature might be universal modifiers and at least in *Drosophila* have measurable directional effects on a large number of physiological processes (Marinkovic and Bajraktari, 1988; Zwaan *et al.*, 1992; Rogina and Helfand, 1995; Huey *et al.*, 1995; James *et al.*, 1997; Imasheva *et al.*, 1998). That is to say responsiveness to ageing and temperature does not uniquely identify reactivity but rather represents a general and probably adaptive response of the whole organism to changing internal and external stimuli. The prediction is that the effects of such modifiers should be detectable in many independent metabolic processes.

Speculations on the nature of reactivity have yielded similar hypotheses invoking the role of a DNA repair process in modulating the severity of embryonic lethality and the frequency of lethal mutations (Proust *et al.*, 1992; Bregliano *et al.*, 1995). The only recent attempt to test such hypotheses has been mostly concerned with measuring the modulating effects of DNA mutagens and ageing on *I-R* induced embryonic lethality (Laurencon and Bregliano, 1995; Bregliano, *et al.*, 1995; Laurencon *et al.*, 1997). The model being tested relies on the relationship between reactivity and a hypothetical *Drosophila* DNA repair/recombination system, paralleling the bacterial inducible stress response. *I* factor transposition would activate the inducible components of the host DNA repair/recombination machinery just like DNA mutagens, such as inhibitors of nucleotide sythensis and gamma rays, do in bacteria. The different reactivity levels are identified with strain and possible age dependent differences in the inducibility potential of DNA repair/recombination components.

The main problem with the experimental design used by Bregliano, *et al.*, 1995, Laurencon *et al.*, 1995 and Laurencon *et al.*, 1997 is that the responses to mutagens, that should represent the level of induction of the DNA repair/recombination pathway, are measured in a dysgenic background. In other words, the lack of independent tests for reactivity and induction of the DNA repair/recombination system and the absence of appropriate controls result in failure to establish a convincing connection between these two processes. With one exception, all strains used differed in reactivity levels because of age effects, making it difficult to distinguish between general pleiotropic effects of ageing on independent processes and true identity between them. Even when age independent polymorphisms in reactivity levels were considered, only one strain per reactivity level was tested which is insufficient to draw definite conclusions.

Previous studies have shown that recombination frequencies in SF females with different reactive backgrounds, do not differ significantly from non dysgenic controls over a number of genetic intervals on all three major chromosomes (Picard, *et al.*, 1977). Unfortunately the experiments were not designed to investigate the effects of reactivity on recombination levels and therefore different intervals were analysed in different crosses. It is significant, however, that recombination levels in each interval considered were not affected by the simultaneous occurrence of dysgenesis, which clearly contradicts predictions from the repair/recombination model proposed above.

More recently, however, experiments involving a recombinant *I* factor, carrying the *yellow* gene in place of its coding sequences, have uncovered a strong association between the occurrence of *I-R* dysgenesis and an increase in recombination frequency in genetic intervals on the X and II chromosomes (Chaboissier *et al.*, 1995). Recombination frequencies in the dysgenic germ line were found to be significantly higher than in non dysgenic controls, despite the fact that one of the two chromatids involved in the exchange in each case was from a balancer chromosome.

In both the 1977 and 1995 studies, recombinants were scored in the progeny of SF females. The contrasting results could be partly explained by taking into account the effect of the different genetic backgrounds of the strains involved. More significant, however, could be the fact that in the earlier study the SF females were mated with inducer males from the paternal strain, whereas in the later study SF females were mated with reactive males. The former mating scheme, however, does not appear to

affect the severity of SF dysgenesis because Picard, *et al.*, 1977 report a high degree of embryonic lethality amongst the F2 progeny which, nevertheless, yielded normal numbers of recombinants for each interval considered. One possible explanation is that the association between dysgenesis and altered recombination frequencies, if a genuine one, does not hold true for all genetic intervals (Laurencon, *et al.*, 1997). Furthermore, the identity of reactivity with an inducible DNA repair/recombination process make it difficult to explain the fact that expression of transgenes driven by the control region of the *I* factor detect reactivity levels in the absence of any DNA damage or recombinogenic effect (Lachaume and Pinon, 1993; McLean *et al.*, 1993).

In summary, reactivity behaves like a quantitative trait, not unlike body size for example, with reactiviy values distributed along a continuum from neutral or extremely weak reactive to very strong reactive strains. The strong non genetic component of this trait becomes evident as reactivity levels change daily with age and breeding temperature. The egg maturation process in *Drosophila*, may render the effects of environmental and physiological changes heritable, perhaps through the accumulation in the maturing egg of age/temperature-modified components.

The nature of reactivity, however, remains a mystery. It probably represents a complex physiological state associated with the ovarian environment and partly determined by the molecular components of many metabolic processes, including perhaps DNA repair and recombination, yet restricted to none of them in particular. According to this view, reactivity exists in all flies, both males and females, of all strains, reactive and inducer, yet it cannot be detected, at least for the time being, because the only test available, i.e. *I* factor transposition, is a process restricted to the germline of *I -R* hybrid females.

1.2. Molecular and evolutionary biology of the I factor

1.2.1. The I factor is a LINE-like retrotransposable element

The determinants of the inducer state, termed *I* factors, appeared as discrete genetic entities distributed along the chromosome arms (Picard, 1976). In inducer backgrounds, *I* factors are inherited in a Mendelian fashion and are separable by

recombination. Individual chromosomes differ in their inducer potential and some chromosomes devoid of inducer properties have been isolated from inducer strains (Picard and Pelisson, 1979). The process of chromosomal contamination (Picard, 1978b), represented the most important clue to the mobile nature of *I* factors (Pelisson, 1981; Proust and Prudhommeau, 1982).

Newly contaminated reactive chromosomes, together with ther inducer properties, often also carry novel mutations that can be isolated in the descendants of dysgenic flies. It was plausible that at least some of these mutations were a direct consequence of *I* factor insertions at the affected locus (Pelisson, 1981). The problem was to isolate one such mutation within a gene of known sequence. This same approach had successfully led to the isolation of the first P element sequences inserted into the newly characterised *white* gene (Rubin, *et al.*, 1982). A similar approach was taken to attempt isolation of putative *I* factors (Bucheton *et al.*, 1984). Two independent mutations affecting the *white* locus and isolated in the descendants of *I-R* dysgenic crosses were first characterised genetically and shown to be closely linked to inducer activity (Pelisson, 1981). Restriction analysis of the two alleles, called w^{IR1} and w^{IR3} respectively, revealed that the main difference betweem them and the wild type *white* locus is the presence of a 5.4 kb insertion at overlapping positions in each of the mutant alleles (Bucheton, *et al.*, 1984).

The inducer properties associated with the w^{IR1} insertion suggested it represents an active element and it was therefore sequenced in its entirety together with the ends of five other independent insertions in the *white* locus (Sang *et al.*, 1984; Fawcett *et al.*, 1986) also recovered from dysgenic individuals. The *I* factor insertion responsible for the wIR3 allele, which is identical to the wIR1 *I* sequence, was cloned into a P element transformation vector and shown to be capable of autonomous transposition (Pritchard *et al.*, 1988). The structure of a full length *I* factor (Abad *et al.*, 1989; Busseau *et al.*, 1994). is shown in figure 1.2.1-1.

I factors are 5359 bp long plus a variable number of TAA repeats, usually between 4 and 8, at their 3' end. From left to right the main features of the *I* factor sequence are: a 186 bp 5'UTR followed by 1287 bp long ORF1, a 53 bp linker region that joins ORF1 to the larger 3675 bp ORF2, and, finally, a 167 bp long 3'UTR separates the end of ORF2 from the first TAA repeat. Sequence comparison of the 5' ends of different *I* factors reveals that they are highly conserved except for a common polymorphism at position 3 where G or T can be found.

Structure of the I factor

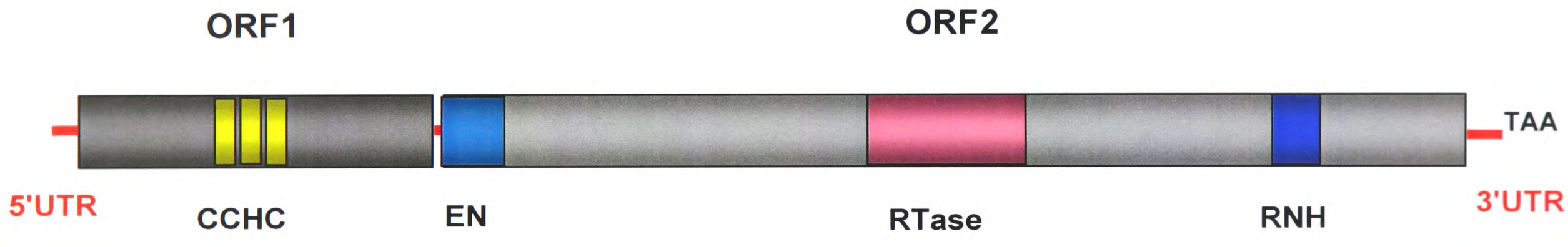


Fig. 1.2.1-1: Structure of the I factor. The open reading frames are indicated as bars of different shades of grey. The three yellow bars in ORF1 symbolise the conserved cysteine rich motif (CCHC). The main catalytic domains in ORF2 are as follows: EN-endonuclease; RTase- reverse transcriptase; RNH- RNase H. The red line represents the untranslated regions at the 5' and 3' ends of the element respectively and the short spacer region separating the two ORFs. Only one TAA repeat is shown.

Despite the variability in number, the TAA repeat unit is also highly conserved with only one case, so far, of a longer unit of the type (TAAA) being found at the 3' end of an *I* factor inserted at the *yellow* gene. Target site duplications (tsds) of variable length (7 to 14 bp) and sequence flank *I* factor insertions at their genomic locations. Even *I* factors that have inserted at identical position within the *white* gene, as is the case for w^{IR1} and w^{IR3}, have different length tsds. Despite the lack of a clear consensus, tsds are, in general, AT rich and in 6 our of 19 tsds compared so far, the first three positions from the 5' end are occupied by TAA residues.

The properties of *I* factors, deduced from the genetics of the *I-R* interaction, postulate that transposition and regulatory functions should be, at least in part, self encoded (Fawcett, *et al.*, 1986). Although the products of the two ORFs have not yet been characterised biochemically, sequence analysis has revealed significant homolgy with protein domains of known function which are relevant to the biological activity of *I* factors (Fawcett, *et al.*, 1986; Di Nocera and Casari, 1987; Finnegan, 1989a; Bucheton, 1990). ORF1 has the potential to encode a protein with three cystein rich motifs of the type CX2CX4HX4C, where X is any aminoacid. This type of motif identifies the RNA binding domain of viral *gag* protein that packages viral RNA into infectious particles. The ORF1 product has been shown to form large multiprotein aggregates and to bind both DNA and RNA non specifically *in vitro* (Dawson *et al.*, 1997).

The product of ORF2 shares homology with the two catalytic domains, namely Reverse transcriptase (RTase) and RNase H (RNH), found in RNA-dependent DNA polymerases encoded by the *pol* gene of retroviruses and other retroelements (Fawcett, *et al.*, 1986). The putative RTase domain spans a region of about 250 amino acid residues almost in the middle of the ORF2 protein. The RNH domain, required for degradation of the RNA template after synthesis of the complementary DNA strand by the RTase, resides within approximately 120 residues between the RTase domain and the C terminus of ORF2. The first 200 aminoacid residues of the putative ORF2 protein share homology with the apurinic/apyrimidinic endonuclease (AP1), an enzyme that plays a critical role in DNA repair (Martin *et al.*, 1995; Feng *et al.*, 1996). In close proximity to the C terminus there is one cysteine rich sequence (not shown in figure 1.2.1-1), with potential for DNA binding. Finally a putative nuclear localisation signal lies at aminoacid position 330 of the ORF2 product (Seleme *et al.*, 1999). Both open reading frames lack putative consensus splice sites, and experimental evidence

suggests that they are translated independently from *I* factor derived bicistronic transcripts (Bouhidel *et al.*, 1994).

I factors have little in common with the "classic" transposable elements epitomised by IS sequences in bacteria, controlling elements in maize, and, more recently, P elements in *Drosophila*. Mobile elements of this type are known to transpose by a cut-and-paste mechanism and constitute class II transposable elements according to the conventional classification system (Finnegan, 1989b; Finnegan, 1990). On the other hand *I* factors lack the Long Terminal Repeats (LTRs) found in retroelements more closely related to retroviruses (Fawcett, *et al.*, 1986). LTR elements, as their name implies, transpose via an RNA intermediate and aspects of their biology recall stages of the viral life cycle. They are grouped together to form class I.1 retrotransposable elements.

The structural and predicted functional features of the *I* factors, place them amongst class I.2 retrotransposable elements, together with other members of the Long Interspersed Nuclear Element (LINE) or non-LTR superfamily (Fawcett, *et al.*, 1986). In particular the two ORFs with homology to *gag* and *pol* retroviral sequences share the greatest similaritiy with mammalian LINEs rather than the retroviral counterparts (Fawcett, *et al.*, 1986; Di Nocera and Casari, 1987). The A rich sequence at the 3' end, which in different LINEs may or may not be the result of polyadenylation, is also distinctive of LINEs and of related sequences of the processed pseudo gene type such as SINEs (Singer, 1992). The structural homology of *I* factors to LINE retroelements has also important implications in terms of *I* factor evolution and makes clear predictions with respect to transposition mechanisms.

1.2.2. Genomic distribution

The use of *I* factor derived probes for Southern analysis revealed the presence of *I* factor sequences not only in inducer strains, as expected, but, surprisingly, also in reactive ones (Bucheton, *et al.*, 1984). Hybridisation patterns of *I* factor probes to genomic DNA digests is almost invariant between different reactive strains. Probes derived from certain parts of the *I* factor, particularly those spanning the ORF2 sequence, hybridise poorly to reactive genomic DNA. In contrast, in inducer strains, superimposing on a hybridisation pattern indistinguishable from that of reactive

stocks, additional bands hybridise strongly to probes derived from all parts of the *I* factor.

In situ hybridisation of similar *I* factor probes to polytene chromosomes from salivary glands, showed a clustering of *I* factor sequences at the chromocentre of both reactive and inducer strains (Bucheton, *et al.*, 1984). In particular the pericentromeric *I* factors from both strain types appear embedded in beta-heterochromatin (Vaury *et al.*, 1989) which consists of complex arrangements of middle repetitive sequences and is moderately under replicated in polytene chromosomes (Gabor Miklos and Cotsell, 1990). Detailed analysis revealed the complex repetitive nature of the sequences flanking the putative heterochromatic *I* factor insertions, thus confirming their genomic location (Crozatier *et al.*, 1988; Vaury, *et al.*, 1989).

I factor sequences have been mapped also to mitotic heterochromatin (Carmena and Gonzalez, 1995). Hybridisation signals to *I* factor sequences, appear in discrete blocks associated with the heterochromatin along one arm of chromosome 2 and at the distal ends of chromosomes 3 and 4. *I* factor sequences also hybridise strongly to the heterochromatin on mini chromosomes derived from rearrangements of the X chromosome and yet hybridisation is not detected on either of the wild type sex chromosomes. The association with heterochromatin blocks is not unique to the *I* factor but applies also to other mobile sequences (Carmena and Gonzalez, 1995), suggesting the possibility that transposable elements actively participate in chromosome dynamics and in particular could play a role in heterochromatisation (Steinemann *et al.*, 1993).

The reasons for the apparent accumulation of *I* factor sequences and other transposable elemens in *D. melanogaster* heterochromatin remain uncertain (Pimpinelli *et al.*, 1995; Dimitri and Junakovic, 1999). One possibility is that transposable elements are actively and preferentially targeted to heterochromatin by transposition functions or are recruited there by host encoded factors. The main driving force behind this transposition bias could be the increased likelihood that insertions in these regions of low gene density are neutral, which would be to the mutual advantage of host and genomic parasite. Alternatively, rates of transposable element insertions might be relatively uniform throughout hetero- and euchromatic regions of the genome: transposable element insertions would accumulate and persist for longer evolutionary times in heterochromatin because recombination

events, that contribute to the loss of repetitive sequences, occur at reduced frequencies in heterochromatin.

Recent experimental evidence shows that, following *I-R* dysgensis, heterochromatic genes on chromosome 2 are about 20-fold more mutable than euchromatic genes on the same chromosome (Dimitri *et al.*, 1997). The findings take into account the fact that heterochromatin has an estimated 150 times less genes than euchromatin. The association of *I* factor heterochromatic transposition events with lethal mutuations (Dimitri, *et al.*, 1997), however, contradicts the earlier assumption that heterochromatin represents neutral ground for insertion of mobile sequences, but these findings need to be confirmed for the heterochromatic regions of the other chromosomes. Whatever the mechanism might be, the selective forces behind the increased transposition rate of *I* factors to heterochromatin remain to be elucidated.

Hybridisation of *I* factor sequences to the chromosome arms is only detectable in inducer strains (Bucheton, *et al.*, 1984; Vaury, *et al.*, 1989). The only notable exceptions are conserved euchromatic bands detectable in the pericentromeric regions of chromosomes 2 and 4 of both strain types (Vaury, *et al.*, 1989). Approximately 15 bands can be counted along the chromosome arms of inducer strains (Bucheton, *et al.*, 1984; Biemont, 1986) which is in agreement with the densitometry estimate from Southern blots of 10 to 15 copies of *I* factor per haploid genome (Bucheton, *et al.*, 1984).

Large scale population studies, looking at the distribution of *I* factor sequences along the chromosome arms, have made a number of important revelations regarding the dynamics of active *I* factors within and between strains of *Drosophila melanogaster*. The data from two independent studies clearly show that *I* factors transpose within natural and laboratory inducer populations (Biemont, 1986; Leigh-Brown and Moss, 1987). *I* factor sequences have in fact been found associated with "spontaneous" mutations that arose *de novo* within inducer populations (Peifer and Bender, 1986; Mount *et al.*, 1988). *I* factors can, therefore, mobilise at detectable rates even in the absence of hybrid dysgenesis. In particular, *I* factors were found to transpose with a frequency of about 10^{-3} per element per generation in two inbred laboratory strains (Harada *et al.*, 1990). The extent of *I* factor insertion polymorphism observed in individuals of an inbred line, supports a possible causal relationship between inbreeding, that can happen also in peripheral natural populations, and spontaneous mobility of *I* factors (Di Franco *et*

al., 1992). Mobilisation of transposable elements could, in fact, contribute to the high genetic variance observed in certain inbred fly populations.

The extremely low excision rates measured for the *I* factor (Harada, *et al.*, 1990), fit well within replicative models of *I* factor transposition, whereby new copies are generated in the absence of excision. This is also in agreement with the failure to detect reversions of *I* factor induced alleles (Pelisson, 1981). Interestingly, the lowest mean copy number of *I* factor insertion was found on the X chromosome (Biemont, 1986). This observation agrees with non neutral models of transposable element insertion, that predict active selection against sex-linked deleterious transposon insertions for which 50% of the population would be hemizygous (Langley *et al.*, 1983).

The distribution of *I* factor sequences along the chromosome arms of different strains shows great variability, with most sites being of low occupancy (Prudhommeau and Proust, 1990; Harada, *et al.*, 1990). High occupancy sites have also been identified in certain strains and these have been explained either as founder *I* factor insertions (Harada, *et al.*, 1990), or as evidence of preferential insertion sites, especially when cytological high occupancy sites coincide with the known location of genes that are often affected by *I-R* dysgenesis-induced mutations (Sang, *et al.*, 1984; Prudhommeau and Proust, 1990; Proust, *et al.*, 1992). Overall the evidence available favours a model where the properties and the evolutionary history of individual strains determine to a certain extent *I* factor distribution patterns and rates of spontaneous transposition (Biemont, 1992).

Comparative restriction mapping and heteroduplex studies of random *I* factor clones from inducer and reactive backgrounds, strongly suggested that heterochromatic *I* factors differ from euchromatic copies because of multiple complex rearrangements (Crozatier, *et al.*, 1988). More detailed sequence analysis of individual heterochromatic *I* factors (Vaury *et al.*, 1990) identifies the rearrangements as base pair substitutions, deletions, insertions and duplications, all of which compromise the coding potential of the two ORFs and make it unlikely that any of the elements characterised encode functional polypeptides. Of the two ORFs the second is the one most often affected by large deletions, which can account for the hybridisation patterns reported above. Hot spots for rearrangements, particularly small deletions, can be identified as bounded by short direct repeats in the sequence of complete *I* factors.

Most of the heterochromatic elements characterised are not flanked by tsds (Crozatier, *et al.*, 1988) although exceptions have been reported(Sezutsu *et al.*, 1995). In contrast, the presence of tsds seems to be an almost constant feature of euchromatic *I* factors, even amongst those that have suffered 5' truncation (Crozatier, *et al.*, 1988). The 3' end A-rich repeat of heterochromatic *I* sequences has a different repeat unit of the type TAA(TAAA)n which may have implications in terms of the evolutionary origin of these elements.

Percentage similarity of heterochromatic *I* sequences with each other and with euchromatic *I* factors, is on average close to 94 % which means that the first are as divergent from each other as they are from the latter (Vaury, *et al.*, 1990). This is in contrast to the high degree of sequence conservation (greater than 99%) between euchromatic *I* sequences. The complex nature and multiplicity of the rearrangements separating heterochromatic *I* sequences from euchromatic ones, make a recent evolutionary relationship between the two categories of *I* sequences highly improbable (Vaury, *et al.*, 1990). It invalidates, for example, an earlier hypothesis for the recent origin of active *I* factor sequences, that envisaged them as the fortuitous products of recombination events between heterochromatic *I* sequences (Bucheton, *et al.*, 1984). Similar reasons render highly unlikely the inverse evolutionary relationship, that sees heterochromatic *I* sequences as recent derivatives of active *I* factors.

The available evidence, therefore, suggests that heterochromatic *I* factor sequences have been in the *Drosophila melanogaster* genome long enough to diverge to a considerable extent from each other and from the active *I* factors of present day inducer strains, to which they are, however, clearly related. The peculiar distribution of *I* factor sequences in reactive and inducer strains of *Drosophila melanogaster* and the apparent copy number restriction of active elements in inducer strains, have important implications for *I* factor evolution and regulation.

1.2.3. The I factor in the melanogaster group

Throughout the 70's and 80's, laboratory and natural Drosophila *melanogaster* populations were routinely tested for *I -R* hybrid dysgenesis (Bucheton, *et al.*, 1976; Kidwell, 1983). Reactive strains were only found amongst stocks that had been bred under laboratory conditions for 1000 or more generations, the equivalent of the 50 or more years that had elapsed since their capture in the wild. Inducer

strains, on the other hand, were found to be ubiquitous amongst early and recent laboratory stocks and in wild populations around the globe even from the most remote areas. Thus, whilst inducer strains constitute the majority of *Drosophila melanogaster* populations both in captivity and in the wild, reactive strains appear confined to a small minority of old laboratory stocks.

How is it possible to explain such disproportion of inducer vs. reactive strains and the apparent absence of the latter from present day wild population? One possibility is to assume that *I* factors were ubiquitous amongst natural *Drosophila melanogaster* populations, but were rapidly lost from the genome of the early laboratory stocks, perhaps through recombination (Bucheton, *et al.*, 1976). This process of "stochastic loss", may have been facilitated by the drastic reduction in the effective population size imposed by the artificial breeding conditions (Engels, 1980). Present day inducer stocks in the laboratory represent on going tests of this hypothesis, albeit circumstances favourable to the random loss of moderately repetitive sequences are likely to occur also in nature. So far, however, there is little experimental evidence supporting stochastic loss, although, in one instance, three reactive lines appear to have descended from a common inducer progenitor after only 400 generations of sib matings (Ytterborn, 1985).

The alternative possibility is to assume that most natural *Drosophila melanogaster* populations were of the reactive type early this century and that *I* factor sequences then spread rapidly through them over the following decades (Bregliano and Kidwell, 1983; Kidwell, 1983). Notably, this hypothetical wave of invading *I* factor sequences coincides with the spreading of *Drosophila melanogaster* and other related species around the world, a process that was undoubtedly facilitated by international fruit exports. One source of contention is for the origin of *I* factor sequences in this rapid invasion hypothesis (Kidwell, 1983). Having excluded the heterochromatic *I* factor sequences as potential progenitors (Vaury, *et al.*, 1990; Bucheton *et al.*, 1992) and given the lack of evidence for an infectious nature of the *I* factor itself (Engels, 1980), two possibilities remain: 1) active *I* factors originated from a small isolated population of *Drosophila melanogaster* which as a consequence of human intervention spread around the globe acting as a vector for this transposable element by vertical transmission; 2) active *I* factors existed in a closely related species from which they spread to *Drosophila melanogaster* by horizontal transfer.

Bucheton, *et al.*, 1992, suggest that, following horizontal transfer to *D melanogaster, I* factors may have initially remained at low frequency or were perhaps confined to peripheral populations. After the big world wide expansion in fruit fly populations in the 1930s, the *I* factor spread rapidly as inducer strains came into contact with reactive ones. A recent survey of transposable element disributions in *D. melanogaster* and in the closely related *D. simulans* species, supports this view of low initial frequency of a given element, followed by a rapid increase in copy number (Vieira *et al.*, 1999). This "awakening" of transposable elements would coincide with world wide colonisation by the host species. Rapid invasion by horizontal transfer, mediated by a semiparasitic mite, has been proposed to explain distribution of P elements in laboratory strains and wild populations(Kidwell, 1983; Houck *et al.*, 1991).

Processes of stochastic loss and rapid invasion by mobile element sequences need not be mutually exclusive (Kidwell, 1983) and in fact it is likely that they both operate in nature and their relative rates might change in response to changing environmental conditions. This implies that mobile sequences can sweep through fly populations and be lost, either by excision/recombination or by mutation, in relatively short evolutionary times. Rapid cycles of invasion and loss (by mutation) can account for the existence of relics of previous invasions such as the heterochromatic *I* factors (Vaury, *et al.*, 1990; Bucheton, *et al.*, 1992; Sezutsu, *et al.*, 1995). The latter may have been retained because they had inserted in genomic regions of low recombination.

Looking at the distribution of *I* factor homologues in other species of the genus *Drosophila*, can shed some light on the evolutionary history of this mobile element within its host. The classification within the *Drosophila* genus is shown in figure 1.2.3-1 (Bucheton, *et al.*, 1992).

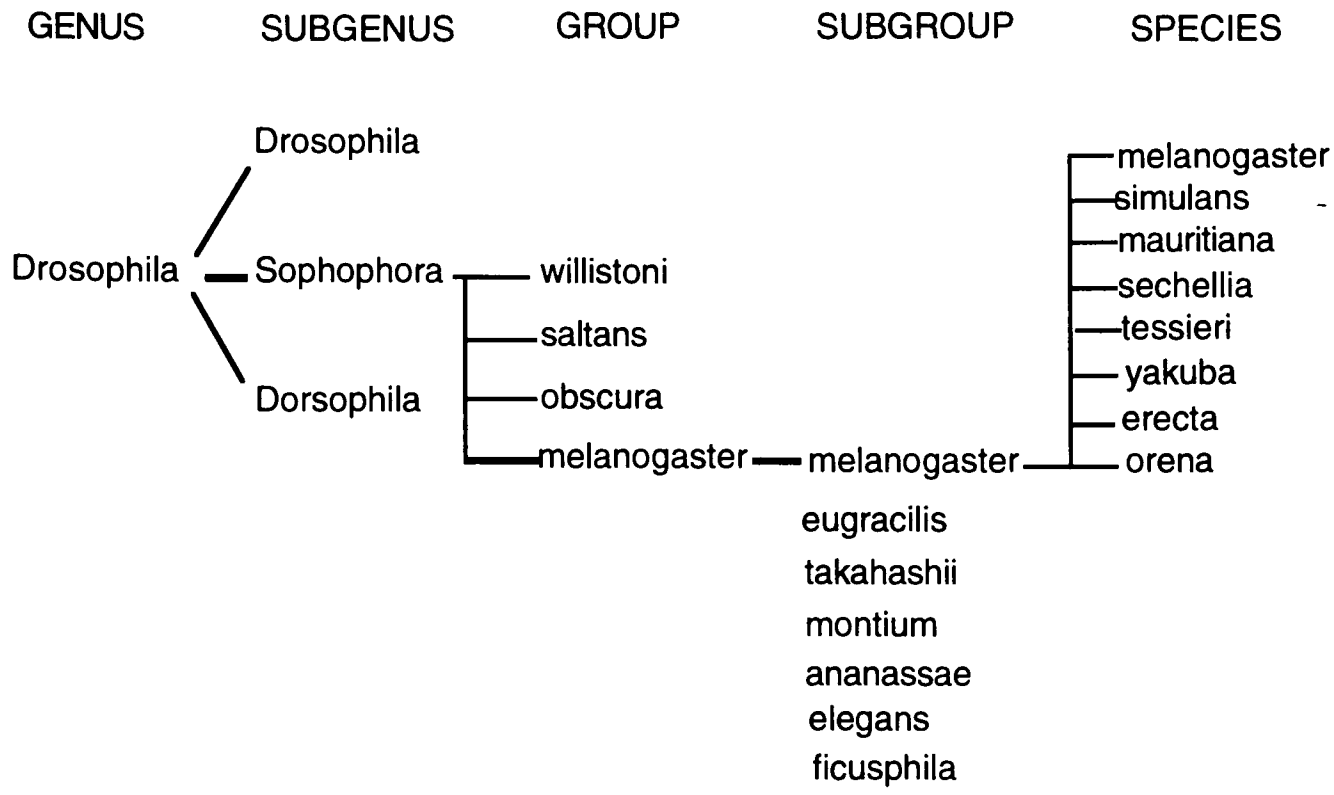


Fig. 1.2.3-1: Classification within the Drosophila genus.The members of the melanogaster species are listed in order (top to bottom) of increasing evolutionary divergence from D.melanogaster.

Sequences with homology to the *I* factor have been found in 20 out of 21 species within the *melanogaster* subgroup by Southern analysis (Stacey *et al.*, 1986; Bucheton *et al.*, 1986). *In situ* hybridisation to salivary gland chromosomes from larvae of different species, reveal that distribution of *I* factor homologues is similar to that observed in inducer strains of *Drosophila melanogaster* (Bucheton, *et al.*, 1986; Simonelig *et al.*, 1988): all stocks of the species analysed have sequences with homology to both heterochromatic and euchromatic *I* factors of *D.melanogaster*, and no equivalent to reactive strains was reported.

The use of either *I* factor probes from fragments diagnostic for inducer strains (Bucheton, *et al.*, 1986) or of different stringency conditions for probe hybridisation (Stacey, *et al.*, 1986), reveals that the degree of homology between the *I* factors of different species, correlates well with the accepted phylogenetic relationships within the *melanogaster* subgroup. This has been confirmed by random cloning of *I* factor homologues from four additional species within the *melanogaster* subgroup (Sezutsu, *et al.*, 1995).

There is greater similarity between active *I* factors from *D. melanogaster* and their putative counterparts in the two most closely related species, *D. simulans* and *D.*

mauritiana, than there is between active and inactive copies within individual species. The putative active *I* factor from *D. simulans*, for example, shares 99.5 % nucleotide similarity with a clone from an active *I* factor of *D. melanogaster* and the percentage similarity of the latter sequence with its counterpart in *D. mauritiana* is only slightly less striking at 98.1 % (Sezutsu, *et al.*, 1995). Furthermore, a complete copy of the *I* factor homologue from *D. tessieri*, which has 85 % similarity to the *melanogaster* *I* factor, is not only capable of transposition when inserted in the genome of reactive *D. melanogaster* strains, but can also confer an inducer phenotype on the transgenic flies (Abad, *et al.*, 1989; Vaury *et al.*, 1993).

The phylogenetic distribution of euchromatic *I* factor sequences within the *melanogaster* species subgroup is in agreement with a model of rapid invasion by horizontal transfer from a closely related species (Bucheton, *et al.*, 1992; Sezutsu, *et al.*, 1995). This might in part reflect the limitations of the horizontal transfer mechanism for the *I* factor that would require close kinship of the donor and recipient species. The near identity of *I* factor sequences in *D.melanogster* and *D.simulans*, suggests that the latter may have served as the source of this mobile element (Bucheton, *et al.*, 1986; Sezutsu, *et al.*, 1995). In contrast, P elements are believed to have entered the *D.melanogaster* species directly from *D.willistoni*, a member of a different species group (Daniels *et al.*, 1990).

Under the above assumptions, the existence of *D.melanogaster* strains devoid of active *I* factors would be due to the genetic isolation in the laboratory of the early fly specimens, the capture of which must predate or be contemporary with the onset of *I* factor invasion (Kidwell, 1979). The apparent absence of equivalent reactive strains in other species (Bucheton, *et al.*, 1986; Simonelig, *et al.*, 1988; Sezutsu, *et al.*, 1995), could be tentatively interpreted as evidence that *D.melanogaster* was the last species in the subgroup to receive the *I* factor. On the other hand, there is some evidence that most of the stocks from species other than *D.melanogaster*, were established with flies caught after the accepted date for the *I* factor invasion (Simonelig, *et al.*, 1988) and it is also plausible that surveys of these species may not have been so exhaustive as to completely rule out the existence of reactive-like strains.

The fact that *I* factor sequences are not well represented outwith the *melanogaster* species group, strongly suggests that the spread of this element took place after speciation in the *Sophophora* subgenus (Stacey, *et al.*, 1986; Sezutsu, *et al.*, 1995). This offers a much more satisfactory explanation than having to postulate the

existence of *I* factor sequences in the ancestor to the *Drosophila* genus, followed by loss after speciation to give the present distribution. A process of stochastic loss must, however, be invoked to explain the evolutionary history deduced for heterochromatic *I* sequences, whereby the active ancestors must have been lost prior to speciation in the *melanogaster* group(Bucheton, *et al.*, 1986; Sezutsu, *et al.*, 1995). Similar processes of rapid acquisition and associated loss have been suggested for other transposable elements of *Drosophila* (Robertson and Lampe, 1995).

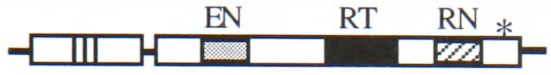
1.2.4. Other non-LTR elements in Drosophila melanogaster

LINEs (Singer and Skowronski, 1985), better known as non-LTR elements (Eickbush, 1992) , are widespread and probably very old components of eukaryotic genomes (Malik *et al.*, 1999). In mammals, where they were first discovered, LINEs make up between 15 and 20 % of the total genome (Hutchison III *et al.*, 1989). The large number (10^4-10^5) of copies of these elements found in mammalian genomes come from a few active master copies that are closely related and belong to one of two families commonly found in each host species.

In contrast, the *I* factor of *D.melanogaster* is one amongst several distinct families of non-LTR elements that coexist within the genome of this organims. The names and the main structural features of those that have been better characterised so far are listed in the table below (adapted from (Di Nocera *et al.*, 1994; Blesa and Martinez-

Sebastian,1997)

Table 1.2.4-I: non-LTR retroelements in D. melanogaster

Element	Structure	Copy number/ genome	Genomic location
I factor		~30	Hetero/Euchromatic Dispersed
F		50-100	Hetero/Euchromatic Dispersed
Doc		>30	Hetero/Euchromatic Dispersed
G		?	Tandem; interspersed
Jockey		~50	Hetero/Euchromatic Dispersed
R1		?	rRNA genes; site specific
R2		?	rRNA genes; site specific
TART		?	Telomeres; tandem interspersed
Het-A		?	Telomeres; tandem interspersed

The structures of the elements are scaled;white boxes represent coding regions; stippled, black and striped boxes, denote the endonuclease(EN), RTase (RT) and RNase H (RN) domains respectively; thick horizontal lines represent non coding 5' and 3'UTRs and spacer sequences; three closely spaced vertical lines indicate the relative position (where known) of the ORF1 cystein rich domains; a single vertical line marks the separation between ORFs where appropriate; an asterisk indicates the approximate position of cystein rich domains in ORF2, where present.

The *F* (Di Nocera *et al.*, 1983), *jockey* (Mizrohki *et al.*, 1985) and *Doc* (Bender *et al.*, 1983; Driver *et al.*, 1989) elements, have similar genomic distribution to the *I* factor with sites both on the chromosome arms and in the heterochromatin. G elements, in contrast, are organised in tandem repeats mostly restricted ot the chromocentre, but copies are also found in the the non-translated spacer of non-nucleolar rDNA units (Di Nocera *et al.*, 1986). R1 and R2 elements insert at

specific distinct sites in the non-translated spacer of nucleolar rDNA (Jakubczak *et al.*, 1991). Fifty percent or more of rDNA genes may contain R2 and R1 insertions and the relative orientation of these retroelements is compatible with cotranscritpion with the rDNA genes.

Multiple copies of the HeT-A (Biessman *et al.*, 1990) and TART (Levis *et al.*, 1993) elements colocalise at the ends of Drosophila chromosomes, where they appear to be important and active components of the *Drosophila* telomeres (Pardue *et al.*, 1996). HeT-A is truly unique amongst *Drosophila* LINEs, not only because it does not encode products with recogniseable RTase function, but also because its transcription requires head-to-tail arrangement of HeT-A sequences: within the array, transcription of a downstream element is driven by regulatory sequences present within the 3' end of the element just upstream (Danilevskaya *et al.*, 1997). There are additional, less well known LINEs in the genome of *D.melanogaster* such as the BS element (Udomkit *et al.*, 1995), and Helena, a *D.virilis* LINE, found to occur at low frequencies also in the *D.melamogaster* genome (Vieira, *et al.*, 1999).

A recent model describes the evolutionary relationship between LINEs in the *Drosophila* genus (Blesa and Martinez-Sebastian, 1997). Based on both DNA and protein similarity, this evolutionary hypothesis suggests that the *I* factor is probably very similar to the ancestral element of most *Drosophila* LINEs in having retained endonuclease, RTase and RNase H domains in its ORF2 and Cysteine rich domains in both ORFs and in being devoid of an oligo(A) sequence at its 3' end. According to this view, *jockey*, *F*, *G*, *Doc*, BS and TART derive from a variant of the common *I* factor-like ancestor, which lost RNase H and cysteine rich domains (and possibly RTase in the case of HeT-A), acquired longer 5' and 3' UTRs and oligo (A) stretches at its 3' end.

1.3. I factor transposition and regulation

1.3.1. Mechanism of transposition

The structural similarity of the *I* factor to LINE elements makes clear predictions about a general mechanism of transposition involving an RNA intermediate (Fawcett, *et al.*, 1986). A putative RNA intermediate of *I* factor transposition was identified by Northern analysis (Chaboissier *et al.*, 1990). This 5.4 kb transcript is most abundant is SF females where *I* factor transposition levels are highest. A transcript of the same size is also present in smaller amounts in RSF females. This is in agreement with the level of *I* factor transposition estimated, from the frequency of chromosomal contamination in the ovaries of these females, to be five fold less than in SF ovaries. On the threshold of detectability, similar full length transcripts were reported also in inducer flies, which is compatible with the evidence of spontaneous transposition of *I* factors in inducer background.

S1 mapping experiments, confirmed the *I* factor transcripts to be full length within the detection limits of the technique. The fact that transcripts appear to begin at position 1 of the *I* factor supported the hypothesis of *I* factor transcription starting from an internal promoter (Fawcett, *et al.*, 1986). Moreover, the observation that no sequences are probably removed upon maturation of *I* factor transcripts, has important consequences for the translation and stoichiometry of the two ORFs products (Chaboissier, *et al.*, 1990; Bouhidel, *et al.*, 1994).

Experimental evidence showing that *I* factor transcripts are indeed used as substrates for transposition was obtained independently using two related approaches that ultimately involved the insertion within the *I* factor of intron-carrying sequences and the subsequent monitoring of intron-loss associated with transposition events. In the first approach (Chaboissier, *et al.*, 1990), approximately 1kb from the P element transposon, including the second intron and part of the flanking exons, were inserted in the 3' end of the *I* factor ORF2, downstream of the putative RNase H domain. The *I* factor construct carrying P element sequences, was found to be incapable of autonomous transposition but when introduced in SF females, where reverse transcriptase and other self-encoded functions can be supplied in trans, its transposition frequency was estimated at $<10^{-2}$. In the second approach (Jensen and Heidmann, 1991), a construct including the neomycin resistance gene, interrupted by an artificial intron, was inserted in the *I*

factor 3' UTR in reverse orientation with respect to the *I* factor direction of transcription. Once introduced in a reactive background, this recombinant *I* factor was capable of autonomous transposition at a frequency of 10^{-4}.

Thus, as found by both Chaboissier, et al., 1990 and the Jensen et al., 1991, the presence of additional heterologous sequences had dramatic effects on transposition frequencies, reducing them by several orders of magnitude with respect to estimated transposition rates in dysgenic ovaries (Chaboissier, *et al.*, 1990; Busseau, *et al.*, 1994) and those of a cloned *I* factor (Pritchard, *et al.*, 1988). The recovery of marked I factors that lacked the intron sequences (Chaboissier, *et al.*, 1990; Jensen and Heidmann, 1991) has provided definitive proof that *I* factor transcripts, after being processed according to the splicing signals provided, served as substrates for reverse transcription and genomic integration.

Little is known so far about the details of the transposition process of non-LTR retroelements. However, studies on the retrotransposition of Cin4, a non-LTR element from *Z.mays* , has resulted in a model that in its general outline might be applicable also to other non-LTR retroelements (Schwarz-Sommer *et al.*, 1987). The principal points of this model are as follows: (i) it gives an identity to the main retrotransposition intermediate which would consist primarily of the association between the element RNA and the RTase; (ii) reverse transcription occurs in the nucleus, at the site of insertion; (iii) integration is not a separate process but rather the end product of *in situ* reverse transcription; (iv) the priming function is provided by one of the free 3' hydroxyls generated upon double stranded nicking of genomic DNA.

Experimental evidence supporting *in situ* DNA-primed reverse transcription has been gathered from studies on the transposition of the *Bombyx mori* non-LTR retroelement R2Bm. Biochemical evidence for the existance of an endonuclease function was first obtained for the protein product encoded by this element, which cuts DNA in a sequences specific manner (Xiong and Eickbush, 1988). Sequence analysis of PCR amplified reverse transcription products of R2Bm RTase confirmed DNA-priming of reverse transcription by the free 3'OH created to the left of the first nick at the site of integration (Luan *et al.*, 1993).

The tagging of functional *I* factors has allowed molecular characterisation of the ends of three newly transposed copies: all are flanked by tsds, but in two cases, part of the 5' end is missing (Jensen and Heidmann, 1991). Similar 5' truncated *I* factor sequences had been isolated from *I-R* dysgenesis induced alleles of the *white*

and *yellow* genes (Busseau *et al.*, 1989a), and had also been reported amongst random *I* factor clones from inducer strains (Crozatier, *et al.*, 1988). Variable truncation of the 5' end is believed to be one of the most common causes of size polymorphism in other non-LTR retroelements (Eickbush, 1992), particularly mammalian LINEs (Hutchison III, *et al.*, 1989). As in the case of the *I* factor, 5' truncated elements are unlikely to represent faithful copies of equally truncated genomic insertions.

Important transcription regulatory functions are believed to be associated with sequences at the 5' end of non-LTR retroelments, and the absence of these sequences in the truncated elements implies that the latter represent inactive members of a given non-LTR family. The common occurrence of 5' truncation in the *I* factor and other non-LTR elements, however, suggests that sequences at the 5' end, although likely to be necessary for transcription of a given element, are not required for completion of the transposition process (Finnegan, 1989a). In contrast, retroviruses and the closely related classI.1 retroviral retrotransposons, seem to require sequences at both ends to generate the full length cDNA copy of their respective genomes [(Eickbush, 1992; Finnegan, 1993; Finnegan, 1997)].

Sequence analysis of a number of 5' truncated copies of Cin4 revealed the existence of a variable degree of homology between sequences within and upstrem of the tsd and those that would lie just 5' to the truncation breakpoint in non-truncated copies of the element. To explain these findings, the authors suggest a model whereby random and probably imperfect pairing at any point along the element RNA and the genomic sequences 5' to the staggered cut, could cause premature arrest of reverse transcription. Subsequent RNA degradation and gap filling (by RTase or host encoded DNA repair enzymes) would result in a truncated copy of the element flanked by tsds.

The sequence analysis of newly transposed, tagged *I* factors, suggests that 5' truncation, as in the case of Cin4, must occur during transposition. Sequence homology between the *I* factor and the genomic DNA flanking the sites of insertion of 5' trunacted copies, is not so convincing as in the case of the Cin4 element (Busseau, *et al.*, 1989a). This suggests that, in the case of the *I* factor, different events must be responsible for early arrest of the reverse transcriptase during first cDNA strand synthesis (Busseau, *et al.*, 1994). One possibility is the premature integration of the *I* factor RNA at the 3' end overhang of the genomic target site upstream of the insertion. This event would promote early template switching by

the RTase from RNA to flanking genomic DNA thus generating a 5' truncated template for second strand synthesis (Busseau, *et al.*, 1989a). Another possibility, that has been suggested to explain 5' truncation in R2 elements, concerns the role that secondary structure formation in LINE transcripts may have in promoting pausing of the RTase thereby increasing the chance of RNA to DNA strand switching (George and Eickbush, 1999). The ability to use both RNA and DNA as templates is an intrinsic catalytic ability of the retroviral RTases and is essential to ensure synthesis of a complete DNA copy of the retroviral genome.

Template switching can best explain also the rearrangements observed at the 5' end of Drosophila R2 elements at their genomic location. These rearrangements present themselves as variable size deletions and/or duplications of both R2 and genomic sequences at the 5' end of the insertions (George and Eickbush, 1999). The target-site specificity of R2 elements, makes the detection of such events relatively easy. In the case of the *I* factor, only one instance of 5' truncation of a newly transposed tagged copy has been found associated with what appears to be a rearrangement of the flanking sequences (Jensen *et al.*, 1994). This 5' truncated *I* factor copy is flanked by unrelated sequences, which appear to result from symmetrical deletions of the tsds that would normally flank the 5' and 3' ends of the element in its genomic location. These events are more easily accounted for if *I* factors insert at staggered cuts with 5' protruding ends, because this would permit displacement of these short complementary flaps of genomic DNA so that some or all of their residues are not copied during reverse transcription (Jensen, *et al.*, 1994). A smilar process is believed to explain the precise deletion of two bp from the genomic target site found on either side of all R2 insertions. The two bp long 5' overhangs are generated by the site specific R2 encoded endonuclease but are never recovered after R2 insertion (Luan, *et al.*, 1993).

If the overhang displacement model were correct, then the RTase would have to be capable of symmetric processing of the overhangs at any given site. An alternative explanation to the problem of symmetry of tsd deletion events is insertion of a retroelement at staggered cuts with 3' overhangs, which have been processed by cellular exonucleases. Symmetric displacement of 5' overhangs, on the other hand, could explain the observed variability in the size of tsds flanking *I* factor insertions so that it would be no longer necessary to attribute such variation to incosistencies in the nicking activity of the *I* factor's own endonuclease. Sequence data from transposition of tagged elements are required to confirm that tsd size variation is associated with deletions of corresponding length in the intact target site. Little is

known so far about the endonucleolytic activity of *I* factor ORF 2 product although studies aimed at its characterisation are currently in progress (Seleme, *et al.*, 1999).

Studies of R2Bm reverse transcription have shown that initiation is specific for R2Bm RNA and that it requires a 250 nucleotides long sequence transcribed from the 3' UTR of the element (Luan and Eickbush, 1995). R2Bm transcripts whose 3'ends are exact copies of the corresponding region in the genomic R2Bm sequences, are used most efficiently as templates for reverse transcription but additional and possibly non templated nucleotides appear at the junction with the flanking genomic DNA. Addition of even a few rATPs seems to interfere with initiation, and a similar decrease in efficiency is observed if the last 3 to 6 rNTPs are removed from the 3' end of R2Bm transcripts . Precise initiation of reverse trancription at the last residue of R2Bm RNA, without additional editing, is apparently ensured if a few nucleotides from the conserved insertion site are added at the 3'end of the R2Bm transcripts (Luan and Eickbush, 1996). The postulated ability of RTases to add nucleotides *de novo* without the use of a template, has been suggested as a possible explanation for the origin of poly A tracts or repeat units found at the 3' end of LINE-like elements, including the *I* factor (Luan and Eickbush, 1995).

The role of the *I* factor 3' end in transposition has recently been investigated in greater detail (Chaboissier *et al.*, 1999). In particular the effects of removal or substitution of the TAA repeats on the efficiency of reverse transcription have been analysed. Full length, transposition competent copies of the *I* factor were engineered at the DNA level to have either a poly (A) tract of 13 residues in place of the 3' end TAA repeats or to have the TAA repeats simply removed with no sequences to replace them. Transgenic lines carrying the modified *I* factors were establisehd after P element mediated germline transformation.

Sequence analysis of newly transposed copies of poly(A) *I* factors reveals that they have acquired a considerable number of additional A residues at their 3'end but are otherwise indistiguisheable from the progenitor poly(A) *I* factor sequence, apart from their novel genomic locations. Genomic copies of retrotransposed TAA deleted *I* factors are of two distinct types: type 1 appear to have retained at least one nucleotide from the original tsd and in addition have aquired "*de novo* " poly(A) tails between 25 and 54 residues long; type 2 have no poly(A) sequences, but have retained 11 nucleotides from the flanking sequence of the progenitor *I* factor.

Closer inspection of the sequences 3' of type 2 transposition products, reveals that 7 out of the 11 nucleotides from the original tsd have been partially duplicated giving rise to 7 bp long head to tail repeats. These intriguing findings have prompted a new model that attributes telomerase-like properties to the *I* factor RTase (see fig 1.3.1.-1)(Chaboissier, *et al.*, 1999).

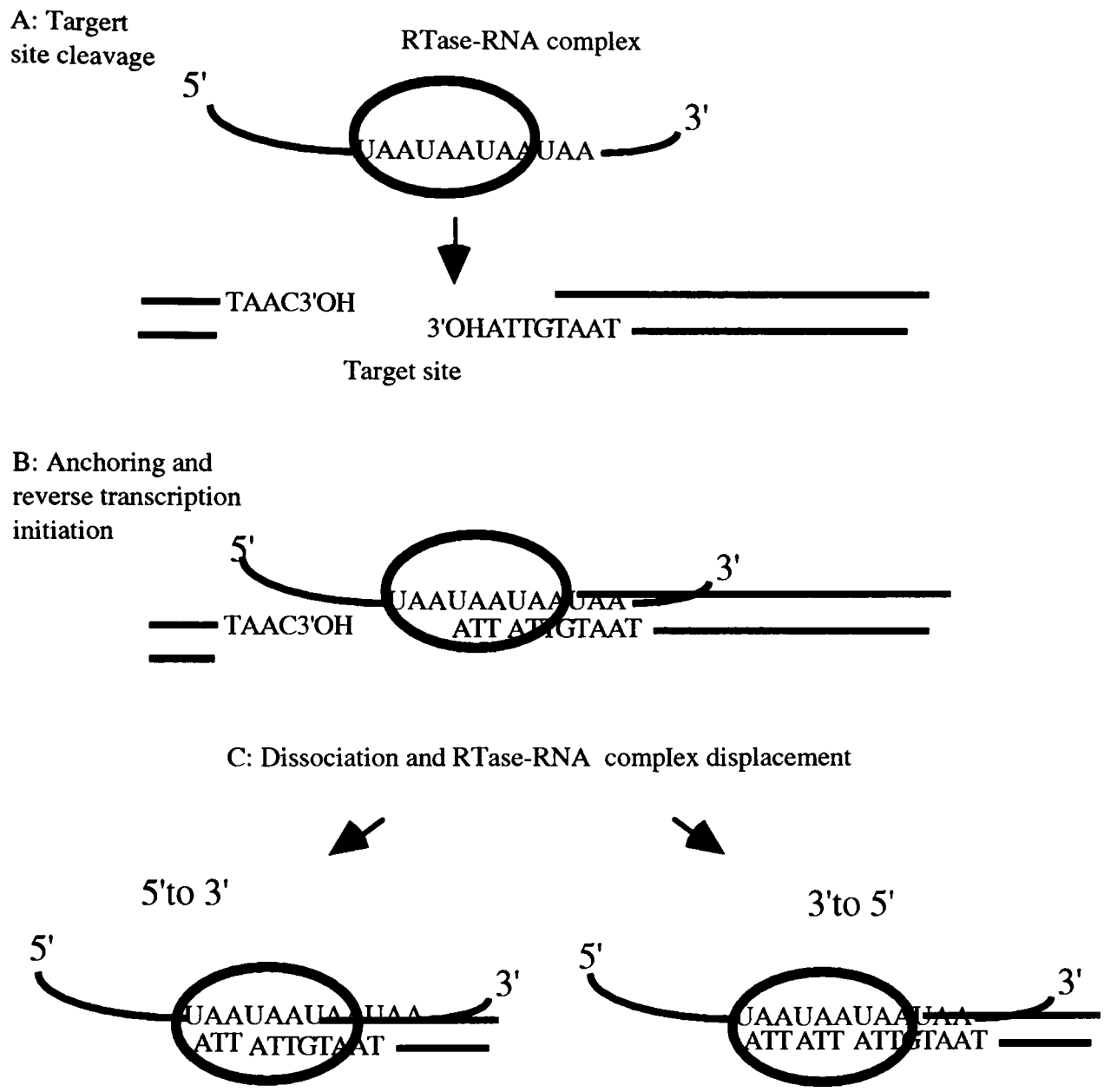


Fig 1.3.1-1:Telomerase-like activity of I factor RTase. The RTase is represented by the red circle; putative read-through sequences transcribed from the donor site are in green whilst the first nucleotides of the I factor coding strand are in blue; the black curved line represents the I factor transcript sequences upstream of the TAA repeat whilst the black straight line represents undetermined genomic sequences flanking the target site.

According to this view, the 3' end of the *I* factor RNA would have a templating role not dissimilar from that of the telomerase own RNA component. The repeats observed in type 2 reverse transcription products become easily explained if one assumes that a region of the *I* factor RNA transcript close to but not exactly at the 3'end is held in the catalytic site of the RTase and serves as a template for initiation of cDNA synthesis. Part of this sequence or residues immediately downstream

might even partially base pair to complementary residues present by chance to the left of the genomic 3'OH, thereby anchoring the RTase-RNA complex to the new insertion site.

Extension of the cDNA copy past the initial residues would require 5' to 3' displacement of the RNA template relative to the RTase catalytic site. However, in order to explain the results obtained by Chaboissier, *et al.*, 1999 it must be assumed that the RTase-RNA complex as a whole dissociates from the genomic target. This dissociation may occur repeatedly at the time of reverse transcription initiation: the number of "jumps" undertaken by the RTase-RNA complex prior to commencing the extension phase of cDNA synthesis, will then determine the number of TAA repeats recovered in the finished product.

The relationship between RTases and telomerases is not an entirely novel concept (Eickbush, 1997; Pardue *et al.*, 1997), but it is particularly exciting to find supporting experimental evidence in *Drosophila*, where telomeres appear in fact to be maintained by LINE-like elements and their associated functions (Pardue, *et al.*, 1996). It is rather more difficult by any model to explain the large expansion of poly(A) tracts either when associated with pre-existing oligo A stretches or appearing *de novo* as in the case of the type 1 retrotransposition products described above.

Lengthening of short poly(A) tracts after transposition has also been reported for human LINE1 (L1Hs) elements (Moran *et al.*, 1999), but in this case it appears associated with the normal process of polyadenylation of L1Hs sequences. Given the lack of evidence in favour of polyadenylation of *I* factor transcripts (Chaboissier, *et al.*, 1990; Chaboissier, *et al.*, 1999), however, it cannot be excluded that the RTase itself might be responsible for the lengthening of L1Hs poly (A) tract. The extent of RTase mediated poly(A) lengthening could also reflect abnormal physical properties of this type of homopolymeric sequence.

The work on the *I* factor 3' end also uncovers additional events associated with *I* factor transcription: the association of newly transposed copies with genomic sequences that flank the progenitor *I* factor implies a frequent transcriptional read through to include non-*I* factor downstream sequences (Chaboissier, *et al.*, 1999). This read through phenomenon, far from being restricted to the modified *I* factors has also been clearly demonstrated for wild type *I* factors carrying the TAA repeats, but in the latter case the additional nucleotides are not recovered at the new insertion site and can only be detected at the RNA level.

The absence of read through products from the sequences flanking newly transposed *I* factors implies that reverse transcription is not initiated at the last nucleotide of the *I* factor RNA (Chaboissier, *et al.*, 1999). The telomerase model of RTase function can accommodate for this by postulating that residues at the extreme 3'end of the *I* factor might rarely if ever be at the RTase catalytic site but could either play a role in the anchoring (Feng *et al.*, 1998) and/or might help binding and correct positioning of the RTase on the *I* factor RNA as deduced for R2Bm (Luan and Eickbush, 1996). In addition to these roles, the UAA repeats of wild type *I* factors transcripts would also serve as template for reverse transcription initiation thus ensuring that nucleotides upstream of the repeat are always included in the cDNA. The AT richness of tsds is compatible with a degree of preference for integration sites that can provide stabilising base pairing. In the absence of UAA repeats, read through sequences, that would normally lie outside the RTase catalytic site, can take over the role of templates for initiation of reverse transcription.

Whatever its role in transposition, read through is the result of the interaction between the transcriptional machinery and termination signals at the end of the *I* factor element. Read through events have been recently documented also for human L1Hs elements and they appear to be favoured by the presence of polyadenylation signals downstream of the one native to the element (Moran, *et al.*, 1999). The fortuitous insertion of L1Hs elements upstream of a gene polyadenylation signal would result in the production of read through transcripts from the newly transposed elements. The read through sequences can then insert at new genomic locations along with the retrotransposon, contributing in the long term to large scale genomic rearrangements (Moran, *et al.*, 1999).

The exact mechanism of *I* factor transposition remains to be elucidated and will ultimately require purification of *I* factor encoded products and establishment of an *in vitro* transposition assay. Many clues, however, have come from careful observation of transposition products and from experimental evidence obtained with other non-LTR elements. We now have a better idea of the substrates and logistics of reverse transcription at least in its initial stages. The A-rich repeats at the 3'end of the *I* factor have been shown to play an important role in the early steps of reverse transcription and, by providing the initial template, they might buffer the upstream sequences against loss due to imprecise initiation. A similar role could be attributed to the A-rich sequences found at the 3' ends of other non-LTR elements.

The experimental evidence supports a model of telomerase-like function for the *I* factor RTase-RNA complex, strengthening the evolutionary links between the catalytic activities of cellular telomerases and "parasitic" RTases. Moreover, read through phenomena, reported for *I* factors and human LINEs, have implications for both transcription termination steps and initiation of reverse transcription as well as providing the first evidence of sequence transduction mediated by retroelements.

1.3.2. Mutagenic effects of I factor insertions

The molecular changes associated with *I* factor transposition, can not only give better insights into some of the steps of the transposition process, but can also have important effects on the expression and regulatory potential of the sequences they affect. Even though a given insertion event may be neutral *per se*, recombination events between dispersed *I* factor sequences and recombinogenic intermediates generated at the time of reverse transcription can have detectable effects on the phenotype of the host.

Chromosome rearrangements provided the first detectable evidence of the molecular rearrangements postulated to be the cause of the recessive lethal mutations associate with *I-R* hybrid dysgenesis (Picard, *et al.*, 1978). Inversions and deletions are amongst the most common rearrangements recovered on polytene chromosomes of *I-R* dysgenic hybrids and these are found associated with as much as 30 % of *I-R* induced lethals (Prudhommeau and Proust, 1990). *I* factor sequences have been detected at the break points of many of the rearrangements characterised (Busseau *et al.*, 1989; Proust, *et al.*, 1992). Ectopic recombination between *I* factor sequences present in otherwise non-homologous regions of the genome, can explain most of the rearrangements observed. The latter are recovered directly in the progeny of SF females, i.e. in the absence of pre-existing *I* factor insertions at those sites on the parental chromosomes. This means that if ectopic recombination is the right explanation, genetic exchange between *I* factors at non-homologous locations would have to occur immediately after transposition (Busseau, *et al.*, 1989; Proust, *et al.*, 1992).

An interesting alternative to ectopic recombination, that can account also for the timing of the chromosomal events, is the possibility of direct target exchange during the transposition process (Busseau, *et al.*, 1989; Proust, *et al.*, 1992). Current models of *in situ* DNA-primed reverse transcription do not impose any restriction

on the source of free DNA ends available to complete cDNA synthesis. It is plausible, in fact, that a large number of free ends are generated by the simultaneous endonuclease activity of *I* factor encoded RTase during *I-R* dysgenesis. For this model to work, however, one would have to postulate that RTase generates double strand breaks and that there is no mechanism to ensure chromosome integrity by anchoring the two DNA ends. Alternatively, it is also possible that, the *I* factor RTase may have used pre-existing DNA ends to initate reverse transcriptionin in those instances where rearrangements are observed.

Both ectopic recombination and target exchange processes may be operating during hybrid dysgenesis. Ectopic exchange between dispersed *I* factor sequences, however, can also occur in the absence of *I* factor transposition, contributing to the long term mutagenic effect of *I* factor insertions that might otherwise be neutral. Incidentally, either of the above processes could account for the finding of some newly transposed *I* factor copies not flanked by tsds (Jensen, *et al.*, 1994; Chaboissier, *et al.*, 1999).

Genetic and molecular investigations of alleles generated during *I-R* hybrid dysgenesis, have provided a better picture of the numerous ways in which the presence of *I* factor sequences can affect gene function. Depending on the relative orientation of *I* factor sequences with respect to the affected gene, additional intiation and termination signals might become available to the transcription machinery. Cotranscription of host gene and *I* factor sequences can result in chimaeric transcripts with effects on the splicing pattern and protein translation due to activation of cryptic splice sites, the creation of frameshifts, the presence of premature stop codons and polyadenylation signals as well as the potential to generate protein fusion products. The effects of different *I* factor insertions recovered at the *w* locus have been studied in detail. In particular efforts have been made to correlate site and relative orientation of *I* factor insertions with the observed effects on eye phenotype and transcription products. A total of 8 *w* alleles, named w^{IR1} to 8, have been isolated and their main features are summarised in the table 1.3.2-1.

3737Table 1.3.2-I: List of *w* alleles caused by I factor insertions [a].

	Insertion site	Orientation †	Eye phenotype
w^{IR1}	intron 5	antiparallel	ts§ red-brown
w^{IR2*}	5' UTR	antiparallel	red-brown
w^{IR3}	intron 5	antiparallel	ts§ red-brown
w^{IR4}	intron 5	antiparallel	ts§ red-brown
w^{IR5}	3' UTR	parallel	red-brown
w^{IR6}	intron 1	parallel	red-brown
w^{IR7}	DELETION------	----------	bleached-white
w^{IR8}	DELETION------	----------	bleached-white

a compiled from data in (Sang, *et al.*, 1984; Lajoinie *et al.*, 1995) ;
† Parallel = in the same orientation as *w* transcription; Antiparallel = in the opposite orientation to *w* transcription;
§ where indicated ts stands for temperature sensitive: eye colour is lighter at 25°C than at 20° C;
* This insertion represents phenotypic reversion of w^{1} a bleached-white allele caused by a Doc element insertion.

Alleles w^{IR1} to $w^{IR\,6}$ are associated with full length insertions of *I* factor sequences. In particular, alleles $w^{IR1,3,4}$ have *I* factor insertions at the same site within the fifth intron of the w gene. The transcription products of the two intronic insertion alleles, w^{IR1} and $w^{IR6,}$ were characterised by Northern analysis (Lajoinie, *et al.*, 1995). In addition to the wild type transcript, the w^{IR1} allele gives rise to two additional transcripts that are not present in wild type controls: the longest mRNA results from retention of the fifth intron and rises the question of why the splicing inhibitory effect of *I* factor insertion is not all-or-nothing and a large proportion of the chimaeric pre mRNAs are successfully spliced; the shorter transcript appears to be due to early transcriptional termination within the *I* factor.

Transcription products from the w^{IR6} allele are also of three kinds: a wild type mRNA where all the introns are successfully spliced, a chimaeric mRNA where the first intron is retained and finally an unusual chimaeric mRNA where cryptic splice sites are activated to include a small exon derived from the 5'UTR of the *I* factor. A premature stop codon within this exon prevents translation of a sizeable fusion

protein product. Interestingly, some of the chimaeric transcripts show tissue and developmental specificity and are absent from the testis where the w gene is known to be expressed.

Overall it appears that the alternative transcription and RNA processing signals present in the *I* factor sequences are often ignored by the transcription machinery which ensures synthesis of wild type product. Since alleles w^{IR1} to w^{IR6} show dosage compensation, it is possible to explain the eye phenotype simply in terms of a decreased dosage of wild type transcripts as transcription functions are diverted towards synthesis of the chimaeric RNAs. Although this has not been investigated, it is possible that different temperatures could affect the ability of the transcription and pre mRNA processing machinery to choose between *w* gene and *I* factor signals thus changing the ratio of wild type to chimaeric transcripts with visible consequences on the eye phenotype.

Further insights in the interactions between *I* factor *cis*- and *trans*-acting components and those of the affected gene can be gained by looking for unlinked mutations that modify the *I* factor induced phenotype. When looking for modifiers of transposon induced mutations it is important to distinguish between effects due to a specific interaction of the modifier with the gene and those effect resulting from interaction of the modifier with the transposable element. Once a candidate modifier gene has been isolated, dissection of the two types of effects requires testing the modifier against a number distinct transposon-dependent and transposon-independent alleles.

Examples of locus specific modifiers are the *Lighten up* (*Lip*) and *zeste* genes which seem to be specific for *w* mutant alleles. Mutant *Lip* alleles modify *w* alleles induced by insertions of various retrotransposons, including the *I* factor. *Lip* mutation enhance (i.e. give a darker eye colour) all the known *I* factor induced *w* alleles, independently of their site of insertion and relative orientation (Csink *et al.*, 1994). The mutant zeste allele, z^1, exerts paring dependent regulation of wild type *w* sequences, a phenomenon known as transvection (Jack and Judd, 1979; Hazelrigg and Petersen, 1992). In z^1 flies, heterozygous for the w^{IR1} allele, the typical z^1 induced bright yellow-orange eye colur, becomes lighter (Pelisson, 1981). In addition to *I* factor induced *w* alleles, z^1 has been found to affect a number of different insertional alleles at the *w* locus and to interact with transposon specific modifiers (Csink *et al.*, 1994a).

The only modifier, so far, that appears to be specific for *I* factor induced alleles is the *Mod(w^{IR6})* allele of the M6+ locus. which maps to the right arm of chromosome 3 (Vaury *et al.*,). As its name implies, *Mod(w^{IR6})* was isolated following P element mediated mutagenesis because it further reduces eye color in w^{IR6} mutants. *Mod(w^{IR6})* appear to act specifically on parallel *I* factor insertions in the *w* gene and has not effect on *w* allels due to point mutations or to the insertion of unrelated transposable elements.

Reporter gene studies show that *Mod(w^{IR6})* does not appear to exert its effect on the regulatoy sequences of the *I* factor and it is possible that its action requires sequences from the main body of the element. *Mod(w^{IR6})* exerts a slight but detectable effect on the ratio of transcription products associated with the w^{IR6} allele by increasing the amount of the longest chimaeric mRNA, a change that could explain the lighter eye colour phenotype. Using well established tests, *Mod(w^{IR6})* has been shown to act as an enhancer of position effect variegation (PEV) but does not appear to interact with components of the silencing multiprotein complex encoded by the Polycomb group genes. Modifiers of *I* factor-independent w alleles also appear to have a role in chromatin structure and have effects on PEV, suggesting that this apparent bias might be due to the genetic context rather than to a specific interaction with the *I* factor. The specificity of *Mod(w^{IR6})* action makes it unlikely to exert more general effects on *I* factor transposition although this possibility has not been investigated.

1.3.3. An internal RNA polymerase II promoter

Several of the structural and functional properties of the first *I* factor sequences characterised, strongly suggested that the complete element should include all the regulatory sequences required for its own expression (Fawcett, *et al.*, 1986). Reliance on chance regulatory elements provided by host flanking sequences, would prove rather inefficient and could not account for the ability of *I* factors to transpose independently of their genomic location and their relative orientation. Moreover, the requirement for heterologous regulatory sequences, could not explain the tissue specificity and timing of *I* factor expression, and different *I* factors would be expected to display expression patterns likely to reflect those of the nearest gene.

Self-encoded regulatory sequences like those hypothesised for the *I* factor and other non-LTR elements, must be able to direct their own transcription and, thus, they

must be recognised by the host transcription machinery. Consensus sequences normally recognised by the RNA pol II transcription machinery, such as the TATA box, are not found at the 5' end of non-LTR elements, where promoter sequences are most likely to be located (Fawcett, *et al.*, 1986). Furthermore, the runs of oligo dT found on the non coding strand of the 5'UTR of many non-LTR elements would act as strong terminators for RNA pol III (Bogenhagen and Brown, 1981; McLean, *et al.*, 1993), rendering these sequences unsuitable substrates for transcription by this polymerase.

Promoters of non-LTR elements are not unprecedented because since the early 1980's, an incresasing number of pol II eukaryotic genes have been found to posses no obvious TATA box. The TATA box is an AT rich sequence found 20-30 bp upstream of the transcription start site of many structural genes (Zawel and Reinberg, 1993; Novina and Roy, 1996). Recent evidence suggests that the TATA consensus is not essential for transcription initiation but rather stimulates promoter strength by increasing the frequeny of re-initiation of transcription (Yean and Gralla, 1997).

Promoters lacking the TATA box, fall into two classes (Smale and Baltimore, 1989; Smale, 1997). The first class includes GC rich promoters typical of many housekeeping genes: they are characterised by multiple starting sites scattered over a relatively large region with multiple binding sites for the transciption factor Sp1. The second class includes the remaining promoter sequences that have neither TATA nor GC rich regions and are particularly prevalent amongst developmentally regulated genes: transcription from these promoters starts at a well defined position, usually from a single nucleotide. By the end of the 80's, examples of genes controlled by promoter elements of this second type included a number of homeotic genes in *Drosophila* and genes regulated during immunodifferentaition in mammals (Novina and Roy, 1996; Smale, 1997).

The quest to define the promoter element that acts in the absence of TATA to guide transcription from a single start site led to the discovery of the Initiator (Inr) element and to a novel perception of promoters as having modular structure. Inr was first characterised in the TATA-less promoter of the murine lymphocite specific terminal deoxynucleotidyltransferase (TdT) gene (Smale and Baltimore, 1989) and has been shown to consist of a relatively variable short sequence which can by itself direct a low level of transcription initiation from a single internal nucelotide (Weis and Reinberg, 1992; Arkhipova, 1995). Since then Initiator modules have been

shown to be essential in both TATA-containing and TATA-less promoters, where they provide independent recognition sites for components of the RNA pol II transcriptional machinery (Weis and Reinberg, 1992; Gill, 1994; Novina and Roy, 1996; Smale, 1997).

RNA pol II was first shown to participate in transcription of LINE elements in the mouse and in the parasitic protozoa *T.brucei* (Jackson *et al.*, 1985; Murphy *et al.*, 1987). Conclusive experimental evidence for the existence of an internal RNA pol II promoter in non-LTR retroelements came from studies on the 5'UTR of the *jockey* element in *Drosophila* (Mizrokhi *et al.*, 1988). The promoter activity associated with the first 13 nucleotides at the 5' end of the element, was shown to be sensitive to standard doses of alpha-amanitin, a fungal toxin that inhibits RNA pol II activity. Primer extention studies also demonstrated that jockey transcripts begin at the first nucleotide of the element, suggesting that the sequences at the extreme 5' end of *jockey* might have an Inr-like function.

The 5' UTR of the *I* factor is able to drive germline specific expression of reporter genes in reactive and dysgenic ovaries, while only background levels are observed in the corrresponding tissue of RSF and inducer ovaries and in male and female somatic tissues (Lachaume *et al.*, 1992; McLean, *et al.*, 1993). This 186 bp sequence from the *I* factor 5' end is therefore sufficient to confer upon a heterologous gene, an expression pattern that is consistent with the spatial and temporal distribution of full length *I* factor transcripts (Chaboissier, *et al.*, 1990) as well as with the known genetic properties of the *I* factor. As in the case of the *jockey* element, sensitivity to alpha-amanitin, identified RNA pol II as the main transcription function, while primer extension confirmed that *I* factor transcripts begin at position +1 (McLean, *et al.*, 1993).

Analysis of a large *Drosophila* promoter data set has contributed to better define the conserved consensus sequences believed to play a critical role in transcription from TATA-less promoters (Arkhipova, 1995). This computer based approach has been complemented with in vitro footprinting analyses of the binding of transcription factors to TATA-less promoter sequences (Burke and Kadonaga, 1996). A TFIID footprint has permitted to refine the *jockey* Inr consensus and has provided the first evidence for the existance in LINEs, of a novel conserved promoter module called the Downstream Promoter Element (DPE) (Burke and Kadonaga, 1996).

TFIID is a multiprotein complex that directs transcription preinitiation assembly on both TATA-containing and TATA-less promoters (Hoffmann *et al.*, 1997). Certain

TFIID subunits are believed to make direct DNA contacts with the major promoter modules including TATA (where present), Inr and DPE (Verrijzer and Tjian, 1996). The TATA-binding protein (TBP) subunit of TFIID is known to specifically bind the AT rich region associated with the TATA box, although the role of TBP in initiation at TATA-less promoters remains unclear (Zawel and Reinberg, 1993; Roeder, 1996; Smale, 1997).

The *I* factor putative Inr has been shown to be essential for *I* factor promoter activity although it can tolerate a great deal of sequence variation, at least in culture cells (McLean, *et al.*, 1993). Interestingly the naturally occurring CAGT variant considerably stimulates *I* factor promoter function with respect to the CATT variant. The two variants of the *I* factor promoter may have arisen in *D.melanogaster* during the invasion or might represent two pre-existing variants selected for in the original host species that entered the genome of *D.melanogaster* separately. It is not known whether these sequence polymorphisms also have an effect on *I* factor control in the now predominant inducer state and, therefore, whether or not selection forces are acting on them.

The *I* factor DPE and sequences immediately flanking it , have been shown to be essential for efficient intiation at the Inr in culture cells (Minchiotti *et al.*, 1997). In the same study, DPEs from the *I* factor and two other Drosophila LINEs, namely the F and Doc elements, are found to be interchangeable with respect to their ability to drive transcription from the respective heterologous Inrs. Moreover the DPE from all three elements were shown to interact with sequences within the hsp70 promoter in a distance dependent manner. Although this has been interpreted as representing synergistic interactions between the hsp 70 TATA box and the DPEs of the respective retroelements (Minchiotti, *et al.*, 1997), controls involving TATA-less hsp 70 sequences were not performed and therefore it cannot be excluded that other sequences in this promoter are responsible or participate in the interaction with the DPEs. Experiments of these kind are of limited value when the main interest is to investigate promoter functions under physiological conditions. It could be argued, for example, that given the germline specificity of *I* factor expression, *Drosophila* cell lines, that are of somatic origin, do not offer a valid cellular context against which to study aspects of *I* factor regulation, although the *I* factor has been shown to transpose in culture cell (Jensen *et al.*, 1994a).

Evaluating the importance of individual promoter modules is complicated by the fact that contrasting results are, indeed, obtained depending on whether tests are

carried out in culture cells or in whole flies. In tissue culture cells, the first 40 nucleotides of the *I* factor promoter, which include the DPE, result in 4 fold higher reporter gene expression than the one measured for the entire 5' UTR, which has been interpreted as due to the removal of inhibitory sequences present downstream of the DPE (McLean, *et al.*, 1993). When the same 1-40 construct is tested in transgenic flies, however, reporter expression is specifically reduced in the germline by as much as 50 fold, suggesting that a tissue specific enhancer resides between nucleotide 100 and 186 (Udomkit *et al.*, 1996). It is evident in this case how quite opposite conclusions can be arrived at by interpreting results from different and non equivalent systems.

The important result is that the first 40 nucleotides of the *I* factor 5' UTR, that contain all the conserved promoter modules, cannot, on their own, impose germline specific expression on a reporter gene. In contrast, the 100 to 186 sequence of the *I* factor 5' UTR is capable of driving germline specific reporter gene expression when placed upstream of a heterologous TATA-containing promoter in an orientation independent manner, as would be expected of an enhancer sequence (Udomkit, *et al.*, 1996).

The involvement of sequences in addition to the conserved promoter modules is far from surprising and, in fact, has to be postulated to explain the unique expression patterns of individual LINEs and cellular genes that would otherwise rely exclusively on closely related sequences to fulfil unique expression requirements. Sequence variation within the conserved promoter modules can only serve to modulate the efficiency of transcription but is unlikely to change promoter specificity. The ability of DPEs and Inrs from different LINEs to interact with each other (Minchiotti, *et al.*, 1997) supports this view although testing the Inr/DPE chimaeric constructs in whole flies is likely to be more informative when it comes to investigate possible changes in and/or loss of tissue specificity.

In TATA-containing promoters, activator proteins are believed to bind upstream enhancer elements in response to specific cellular signals (Verrijzer and Tjian, 1996; Hoffmann, *et al.*, 1997). The interaction of enhancer-bound activators with the subunits of the promoter-bound TFIID has been suggested to result in a conformational change in the TFIID complex that would be recognised by the RNA pol II enzyme as a signal to initiate transcription. Thus specific activators and their cognate target sequences are invoked to confer response dependent activation upon a given gene. In the case of internal promoters such as those associated with LINEs,

enhancer like function has to depend on sequences downstream of the main promoter modules. It is interesting in this respect that sequences downstream of the DPE in TATA-less promoters, show little sequence homology (Arkhipova, 1995), which is in agreement with the expectation of them serving as mediators of promoter specificity.

In the case of the *I* factor, nuclear extracts, prepared from female and male tissues and *Drosophila* culture cells, contain one or more functions able to bind specifically to *I* factor sequences within the 100-186 putative enhancer (Udomkit, *et al.*, 1996). *In vitro* footprinting has identified a single protected site extending from nucleotides 137 to 150. This short sequence, called site 1, acts as a specific competitor for the binding of the ubiquitous nuclear function(s) *in vitro*. Deletion of site 1 from the 5'UTR of the *I* factor decreases expression of a reporter gene by up to 1000-fold in ovaries and up to 100-fold in female somatic tissue. However, it should be noted here that the results obtained with site 1 deleted constructs are currently under revision (I. Clark, pers. comm.). A concatemer composed of a slightly longer version of site 1 (residues 138-157) is specifically bound in vitro by the *Drosophila* Adult Enhancer Factor 1 (Aef-1) (*I.* Clark, unpublished). Aef-1 is a zinc finger protein originally identified as a specific repressor of the *Drosophila* Alcohol dehydrogenase (Adh) gene in the adult fat body (Falb and Maniatis, 1992). The consensus sequence recognised by Aef-1 at the *Drosophila* Adh enhancer, contains the a CAACAA motif also present in site 1. The relationship between Aef-1 and the site 1 binding function(s) in *Drosophila* nuclear extracts is still unclear (*I.*Clark, pers. comm.).The biological significance of Aef-1 binding to the *I* factor promoter in vivo, is subject of current investigation (*I.* Clark, unpublished).

1.3.4. Multistep regulation

The available experimental evidence regarding control of *I* factor expression and transposition contribute to a very complex picture, with many contradictions and unresolved issues. Transposition of the *I* factor, detected as the occurrence of chromosomal contamination, appears strictly limited to the germline of SF and RSF females (Picard, 1976), and is only a rare, sporadic event in the germline of inducer females. The distribution of *I* factor full length transcripts, as detected by Northern analysis, was found to correlate well with that predicted from the genetics of the element, although the sentitivity of the technique used was greatly reduced by RNA

fractionation procedures that resulted in the loss of about 50% of the *I* factor transcripts (Chaboissier, *et al.*, 1990).

Early reporter gene studies using the bacterial *Beta-galactosidase* gene and *in situ* staining of transgenic ovaries revealed a pattern of *I* factor driven expression that overlapped with that deduced for the full length transcripts (Lachaume, *et al.*, 1992; Lachaume and Pinon, 1993). Reporter expression was limited to the germline cells within the ovaries of reactive, SF and to a lesser extent RSF females, whilst only control levels were detected in inducer ovaries. CAT-based reporter genes also detected a reactive ovary specific expression, but, given the greater sensitivity of the detection system, measureable levels of CAT activity were also reported in the somatic tissues of both R and *I* females (Udomkit, *et al.*, 1996). Sequences within the *I* factor 5' UTR responded also to temperature and age induced variations in the reactivity of the recipient strains (Lachaume, *et al.*, 1992; Lachaume and Pinon, 1993).

The temporal and cytological localisation of the *I* factor ORF 1 product is in good agreement with the expected pattern of *I* factor expression (Seleme, *et al.*, 1999). In SF females, ORF1 signal is detected from stage 1 of oocyte developmenet, which takes place in the posterior 50 % of the germarium. The apparent absence of ORF 1 product from the tip of the germarium, where germline stem cells are located, agrees with previous reporter gene studies and can account well for the pattern of recovery of F2 mutants which are very rarely found to occur in clusters (Tatout *et al.*, 1994). ORF1 signal appears in the nurse cell cytoplasm from where it is then transported into the oocyte: here the signal remains cytoplasmic, localising in the immediate surroundings of the oocyte nucleus. These findings contrast with earlier reports regarding a fusion product between the bacterial *Beta-galactosidase* gene and a deletion derivative of ORF1, which was found to localise mostly to the nuclei of nurse cells and oocyte (Lachaume, *et al.*, 1992; Lachaume and Pinon, 1993; Tatout, *et al.*, 1994). It is possible that the fortuitous creation or activation of a cryptic nuclear localisation signal might be responsible for the aberrant localisation pattern of this fusion protein.

The cytoplasmic ORF1 product then follows the antero-dorsal migration of the oocyte nucleus from its original location at the posterior end of the oocyte (Seleme, *et al.*, 1999). Finally, ORF1 signal disappears progressively from stage 10 of oocyte development. In old SF ovaries (>20 days), ORF1 expression is more patchy, but, where present, its pattern is indistiguisheable from that found in young SF ovaries.

In RSF ovaries, ORF1 staining is much weaker than in young SF ovaries and does not appear to be specific for the oocyte cytoplasm. A few RSF egg chambers, however, show ORF1 signal similar in strength and localisation to that found in young SF ovaries. One of the main features of ORF1 expression is that it is discontinuous, giving stained ovarioles a variegated appearance: non-staining egg chambers can be found flanked on both sides by staining ones and vice versa. Although observed also in young SF ovaries, variegation of ovarioles with respect to distribution of ORF1 product is much more pronounced in the ovaries of RSF and old SF females. Similar variegation of ovarioles had been previously observed also with *I* factor driven reporter genes and could explain why each egg is unique in terms of its probability of surviving *I-R* dysgenesis (Lachaume, *et al.*, 1992; Tatout, *et al.*, 1994).

Taken together, the data obtained from studies on transcript localisation, reporter gene expression and ORF1 protein localisation, suggest a good correlation between *I* factor expression and transposition. Thus, a relatively simple model of germline specific transcriptional control could account for the *I* factor expression pattern. In its simplest form, this transcriptional regulation model predicts that a host encoded protein, able to stimulate transcription at the *I* factor promoter, would be present (or active) in the female germline but would be absent (or inactive) in other female tissues and in males.

The high levels of *I* factor expression and transposition in the germline of *I-R* hybrid females represent a transient event both within the lifetime of a given female and through the generations. Ageing and the associated decline in reactivity levels is believed to be responsible for the progressive reduction in *I* factor expression within the germline of individual females. The kinetics of *I* factor expression through the generations, on the other hand, appear to be affected to a certain extent by the *I* factors themselves. The activity of a reporter gene driven by the *I* factor 5' UTR, is not detectable in inducer background and, most importantly, the *I* factor driven transgene appears to be silenced after 10 or so generations of sib matings between the descendants of a dysgenic cross (Lachaume and Pinon, 1993). Although no attempt has been made to detect reporter transcript levels, transcriptional silencing due to effects on the *I* factor promoter have been postulated in all cases.

Interestingly, the dynamics of *I* factor copy number are inversely correlated with reporter gene activity: the number of *I* factor insertions has been shown to increase

progressively for up to fifteen generations after a dysgenic cross until it reaches a total of between 20 and 30 copies per genome, by which time also the manifestations of hybrid dysgenesis have subsided (Pelisson and Bregliano, 1987). *I* factor copy number is highly conserved in inducer strains, independently of whether they come from the wild or were generated in the laboratory and does not increase even under selection schemes that artificially prolong the *I-R* dysgenesis conditions (Pelisson and Bregliano, 1987). Furthermore, an *I* factor homologue from *D.tessieri* transposes in reactive strains of *D.melanogaster* until it reaches 10 to 15 copies per haploid genome, even though in its original host its copy number is only 4 copies per haploid genome (Vaury, *et al.*, 1993). This observation suggests that host specific factors, must also influence the extent of transposition of mobile elements.

Given the extremely low excision rates of the *I* factor, it is improbable that this maximum copy number is maintained as a dynamic equilibrium of novel insertions and counterbalancing excisions and it is more likely to be the result of cessation of transposition (Pelisson and Bregliano, 1987). There might, therefore, be a correlation between the *I* factor reaching its steady state copy number and the apparent silencing of *I* factors and of *I* f actor-driven transgenes. Similar phenomena of dosage dependent gene silencing were first observed in plants. In these organisms, the introduction of transgenes homologous to endogenous genes often fails to increase product dosage and in many cases results in both types of sequences being inactivated (Flavell, 1994).

Two types of this dosage-sensitive gene silencing, also called cosuppression, have been described in plants and are known as transcriptional gene silencing (TGS) and posttranscriptional gene silencing (PTGS) respectively (Flavell, 1994; Matzke and Matzke, 1995; Vaucheret *et al.*, 1998) (Chapter 4). Both rely on the existence of homology at the DNA and/or RNA level among the sequences involved, hence this silencing phenomenon is also referred to as homology-dependent gene silencing (Matzke and Matzke, 1995). As the nomenclature implies, TGS involves silencing at the promoter so that transcription of the silenced locus is inhibited (Vaucheret, *et al.*, 1998). In one variant of TGS, called *trans*-silencing, the silenced state of one promoter is transmitted to unlinked homologous promoter sequences, presumably via DNA-DNA association (Mette *et al.*, 1999). Recently, however, it has been shown that RNA transcripts containing promoter sequences, can act in trans to silence homologous promoters (Mette, *et al.*, 1999). In the case of PTGS, transcription of the silenced gene continues but the mRNA turnover rate is greatly increased (Depicker and Van Montagu, 1997; Grant, 1999). In both cases, the end

result is detected as greately reduced or variegated expression of the affected genes and transgenes.

A titration model of *I* factor silencing has been suggested, according to which transcriptional inactivation of the *I* factor over the generations, is due to the progressive alteration of the ratio between the number of *I* factor regulatory sequences and the amount of a soluble protein factor that binds to them (Finnegan, 1989a; Udomkit, *et al.*, 1996). This soluble factor might be a host encoded transcriptional activator that is present in limiting amounts in the female germline. *I* factor activation would require binding of two or more activator moieties to sites of decreasing affinity within any given *I* factor sequence. As the *I* factor copy number increases the activator molecules would scatter between high affinity sites on different *I* factors sequences, so that the activator becomes depleated and can no longer bind individual *I* factor promoters with the right stoichiometry to trigger activation.

The site 1 binding function(s) could be a possible candidate for the role of titrateable transcriptional activator (Udomkit, *et al.*, 1996). Titration processes seem to be used in nature to ensure precise timing of gene expression. Titration of the transcriptional repressor *Fushi tarazu*, due to gradual change in nucleocytoplasmic ratio, is believed to play an important role in activation of patterning genes during early embryonic development in *Drosophila* (Pritchard and Schubiger, 1996). Alternatively, *I* factors might be silenced by a self encoded product that accumulates in the female germline and needs to reach a certain level before it can exert its silencing effect (Busseau, *et al.*, 1994).

To test the hypothesis of *I* factor silencing by titration of a transcriptional activator, plasmid constructs were designed to artificially increase the copy number of the *I* factor 5'UTR (Chaboissier *et al.*, 1998). These constructs carry two or three copies of the 1-186 sequence from the *I* factor in the form of head-to-tail tandem repeats (Chapter4). P element mediated transgenesis followed by matings between transformed lines and selection of recombinants resulted in the creation of a dosage series, so that individual homozygous lines carried 2, 4, 6, 8, 12 or 22 copies of the *I* factor promoter per genome. Females from individual lines were then mated independently with (i) males from a transgenic stock carrying the CAT reproter driven by a single copy of the *I* factor 5'UTR and with (ii) males from special inducer stocks that harbour only one copy of the active *I* factor.

Reporter gene activity was measured in the female progeny from the (i) crosses whilst in the (ii) crosses, estimates were made of the hatchabilites of the resulting "SF" females. Reporter gene expression, measured as the acitivity of the CAT enzyme, decreased gradually along the increasing 5'UTR dosage series, whereas the hatchabilities values showed a threshold-like effect with a sharp increase at the maximum 5'UTR dosage tested. A similar dosage series comprised transgenic lines carrying 4, 6 or 10 copies per genome of the *I* factor 5'UTR, harbouring a deletion of site 1. Contrary to expectations, deletion of site 1 decreased only slightly the ability of the *I* factor 5'UTR to reduce transgene expression, although the effect on hatchability of these site-1 deleted constructs was not tested. This suggests that site 1 binding function(s) are unlikely to be the transcriptional activators being titrated out but does not exclude the possibility that other as yet unidentified factors, binding elsewhere along the *I* factor 5'UTR, might play this role.

There are, however, a number of problems with the experimental design used. Firstly, the *I* factor "silencing" sequences are arranged in tandem rather than being dispersed as in the case of naturally occuring *I* factors. There have been reports in *Drosophila* of a phenomenon termed repeat induced gene silencing (RIGS), whereby heterochromatin forms *de novo* in correspondence of regions carrying tandemly repeated sequences (Sabl and Henikoff, 1996). Heterochromatisation results in variegated silencing of the repeated genes and can "spill-over" also to non repeated flanking sequences. Although no variegation of the closely linked mini *white* and *yellow* marker genes was detected in the case of the 5'UTR tandem repeat contructs, it remains difficult to determine to what extent RIGS might have influenced the outcome of these experiments.

Secondly, if RIGS is not operating, then there is a high probability that the *I* factor promoters are indeed transcribed, even in their tandem configuration. The presence of promoter derived transcripts means that RNA mediated *trans*-silencing, rather than the proposed titration mechanism, might have been responsible for the observed silencing. A similar process of RNA mediated cosuppression has been demonstrated for the *Drosophila Adh* locus (Pal-Bhadra *et al.*, 1997). Moreover, the results obtained with the site 1 deleted constructs, do not invalidate this possibility because the early data, regarding the strong depressing effect of site 1 deletion on transgene expression, are currently under revision (*I.* Clark, pers. comm.).

The view of transcriptional regulation as the main form of control over *I* factor transposition, has recently been challenged by the outcome of a number of independent experiments that imply the existence of posttranscriptional mechanisms of *I* factor control. An RT-PCR based approach, that made use of primers specific for transcripts from euchromatic *I* factors, detected comparable amounts of *I* factor transcripts in the soma and ovaries of SF, RSF and inducer females and in male siblings from all three categories (de La Roche Saint Andre and Bregliano, 1998). This contrasts with the large amount of genetic evidence suggesting that the *I* factor does not transpose detectably in the female soma and in males (Finnegan, 1989a; Bucheton, 1990). At the molecular level, mapping of euchromatic *I* factor insertions on polytene chromosomes of inducer and dysgenic origin, failed to detect positional polymorphisms in nuclei from individual flies (Biemont, 1986; Pelisson and Bregliano, 1987). Thus, the existence of *I* factor transcripts in flies and tissues where transposition has never been reported, implies that the presence or absence of factors downstream of transcription is responsible for this failure of transposition.

In agreement with previous findings (Chaboissier, *et al.*, 1990), up to five times more *I* factor transcripts were found in the ovaries of young SF females descended from strongly reactive mothers. This excess of *I* factor transcripts appears to be strain specific (de La Roche Saint Andre and Bregliano, 1998), and correlates with the manifestations of hybrid dysgenesis: as hatchability increases with the age of these SF females, *I* factor transcripts in their ovaries decrease to the levels found in all other categories of flies. In this regard it should be noticed that *Drosophila* ovaries are a composite tissue that includes cells from both germline and somatic lineages (King, 1970). The RT-PCR substrate in this study, was total RNA from whole ovaries, which means that *I* factor RNA from cells of somatic origin would make a substantial contribution to the amount of final product. This possibility is supported by the recent finding that ORF1 product is present in large amounts in the follicular somatic cells in the ovaries of both SF and RSF females, probably as a result of active *I* factor transcription in these cells (Seleme, *et al.*, 1999). It is possible, for example, that the basal *I* factor transcription levels detected in old SF ovaries are due to this somatic contribution, whereas the apparent excess reported in the ovaries of young SF females from strong reactive background is due to *de novo* *I* factor transcription in their germline cells.

Another point apparently overlooked in the above study, is the finding that two sets of primers specific for the 5' end and ORF2 sequences of *I* factor transcripts respectively, do not detect equimolar amounts. In particular the ORF2 specific set,

albeit able to detect overexpression in SF ovaries, gives overall reduced amounts of product in all categories of flies tested. This product imbalance, if not explaineable by primer-dependent variation in amplification efficiencies, must reflect the existence of a previously unreported population of *I* factor transcripts that lack sequences from essential coding regions. These apparently 3' deleted transcripts might be degradation products and could also play a role in posttranscriptional control of *I* factor transposition.

Recent experimental evidence suggests that *I* factor encoded regulation in the dysgenic female germline might, in fact, rely on PTGS mechanisms of cosuppression. Constructs carrying the first 969 bp from ORF2 downstream of the hsp70 promoter significantly reduce embryonic lethality in the progeny of SF females whose mothers carried the transgene (Jensen *et al.*, 1999). This effect becomes progressively stronger as the number of copies of the transgene increases and it takes a number of generations after transgenesis to become fully established. When less than 3 copies per haploid genome are present in the orignal transgenic line, it may take up to 10 generations before any effect is observed. As it has been pointed out (Birchler *et al.*, 1999), the generational kinetics of *I* factor silencing are unique among the cases of cosuppression described to date. Reduction of embryonic lethality requires transcription but not translation of the *I* factor transgene and can persist, albeit significantly reduced, for at least one generation after the transgene has been removed (Jensen, *et al.*, 1999). Moreover the phenomenon is not restricted to the sequence tested as sequences from other parts of the *I* factor give comparable results. There is also a strong parental effect, whereby embryonic lethality returns to dysgenic values if, in the generation immediately preceding the dysgenic cross, the transgene was inherited through the father rather than the mother.

One question not addressed in the above study, is the fact that significant amounts of transcripts, derived from heterochromatic *I* factor sequences, have been detected in R, *I*, SF, RSF females and in their male siblings (Chaboissier, *et al.*, 1990). As previously suggested (Chaboissier, *et al.*, 1990), it is plausible to assume that these abundant homologous transcripts might be contributing to the silencing effect. If transcripts from heterochromatic *I* are present in the reactive female germline then one might ask why they are not sufficient to silence incoming active *I* factors. One possible answer is that, as already observed, *I* factor silencing requires a number of generations to become established independently of whether the silencing RNA were already present or are synthesized *de novo*. Alternatively, it is possible that the extent of sequence divergence is such that these transcripts of heterochromatic

origin are ineffective as silencing mediators. This last possibility is quite unlikely, however, given the evidence that cosuppression, at least in plants, is in fact stimulated by the presence of heterologous sequences or other forms of structural distortion in the silencing RNAs (Depicker and Van Montagu, 1997; Grant, 1999).

Another problem with the above study is that the reduction in embryonic lethality is taken as a direct measure of *I* factor expression and therefore transposition. As discussed above, *I* factor expression does not necessarily correlate with transposition even in the dysgenic germline. Moreover *I* factor transposition has never been directly linked to the causes of embryonic lethality. The frequency of *I* factor transposition in the RSF germline, for example, as estimated from the frequency of chromosomal contamination, is only five fold less than that measured in SF germline (Picard, 1976).

Despite this relatively small difference in transposition levels, embryonic lethality and *I* factor associated mutations occur at high frequency only in the progeny of SF females. The critical question seems to be not why SF and RSF females are different, but rather, why RSF females behave more like *I* females, despite active *I* factor transposition (de La Roche Saint Andre and Bregliano, 1998). Moreover, how should these well established observation be correlated with the recent studies on the distribution of ORF1 in SF and RSF ovaries (Seleme, *et al.*, 1999)? Further evidence in favour of an uncoupling between *I* factor transposition and the main manifestations of *I-R* dysgenesis, comes from experiments where low hatchabilities were selected for a number of generations in the female descendants of dysgenic crosses (Pelisson and Bregliano, 1987). As judged from estimates of the *I* factor copy number, transposition had apparently ceased, and yet the dysgenesis symptoms suffered by these flies were still strong. This led to the hypothesis that embryonic lethality can occur in the absence of transposition and therefore "cannot simply result from the mutator activity of *I-R* dysgenesis".

Thus, evidence so far, suggests a model whereby host and/or *I* factor specific regulatory mechanisms are operating in different tissues. Control in the female soma and in male tissues, appear to be independent of *I* factor dosage. Hence, cellular factors (or their absence) must be responsible for the posttranscriptional regulation postulated to explain the absence of transposition events despite the presence of *I* factor transcripts in these tissues. It may be relevant at this point to recall that the insertion of full length *I* factors within the transcribed region of the *white* gene, did not appear to alter the turnover rate of *white* derived transcripts

(Lajoinie, *et al.*, 1995), suggesting that RNA mediated silencing mechanisms may not be operating in these tissues. In contrast to events in the female soma and in males, the severity of the dysgenesis syndrome in the female germline, can be mocdulated by *I* factor copy number, suggesting that some *I* factor-dependent, dosage-sensitive mechanism could be operating in this tissue.

In conclusion, the evidence available so far, although incomplete and contradictory, favours a complex model where *I* factor control is exerted at multiple levels in different tissues and involves interactions between host- and self-encoded products, the latter being RNA and probably proteins. The mechanisms invoked vary from simple on/off transcriptional control, to titration of transcriptional regulators and other dosage sensitive processes such as TGS and PTGS. Posttranslational regulation of *I* factor products as well as modulation of the various steps leading to transposition, e.g. transport to the nucleus, target site cleavage, reverse transcription/integration, have not been considered but remain a possibility open to investigation.

1.4. Objectives of the present work

The experimental evidence reviewed in sections 1.3.3. and 1.3.4 implicates the *I* factor 5' UTR in key aspects of regulation. These can be summarised as follows: 1) The I factor 5' UTR can impose on a linked reporter gene, an expression pattern that reproduces the one of length I factors; 2) this sequence contains conserved promoter motifs, including an Inr and a DPE, that are associated with a number of internally transcribed promoters in *Drosophila*; 3) in addition, this sequence also contains a tissue specific enhancer that, when linked to a heterologous promoter, drives high levels of reporter gene expression in the female germline of transgenic flies; 4) a conserved CAACAA motif within this enhancer is specifically bound in vitro by the *Drosophila* encoded transcriptional regulator Aef-1; 5) deletion of the enhancer decreases transcription only in the ovaries, whilst basal levels in female somatic and male tissues remain unchanged; 6) expression of transgenes driven by I factor 5' UTR is greatly reduced in inducer background; 7) multiple copies of the 5'UTR downregulate unlinked transgenes driven by the same sequence and also decrease the lethality of F2 embryos associated with *I-R* dysgenesis.

The work described here focuses on different and partly complementary properties of the I factor 5' UTR. In Chapter 3 attempts have been made to correlate the

unusual A-T richness of this sequence with biologically relevant physical properties such as the presence of static bends. This was done using a combination of experimental approaches and computer based prediction algorithms. The results are then critically assessed in relation to known examples of intrinsically bent molecules and their interactions with protein factors. The role of the 5'UTR in mediating preferential maternal inheritance of transmissible silenced states is investigated in Chapter 4. The findings are then discussed in the light of epigenetic silencing phenomena in *Drosophila* and other organisms. Direct interactions between the I factor 5' UTR and the *Drosophila* Polycomb protein, implicated in processes of epigenetic inheritance of silenced states are investigated in Chapter 5. The results are discussed taking in to account models of Polycomb-mediated gene silencing. Finally, future prospects and possible directions of investigation are outlined.

Chapter 2

Materials and methods

2. Materials and methods

2.1. Materials

2.1.1. Antibodies

Affinity purified monoclonal anti-Polycomb antibodies from rabbit, a gift from Prof. Renato Paro, Heidelberg, Germany.

Anti-dioxigenin-Alkaline-phosophatase antibody conjugates from sheep for non-radioactive detection of dioxigenin labeled DNA probes, Boehringer-Mannheim.

Anti-rabbit secondary Horseradish peroxidase antibody conjugates from goat for non-fluorescent immunostaining, Promega.

Anti-rabbit biotin antibody conjugates, from goat for fluorescent immunostaining, Harlan Sera Ltd.

Avidin-Texas Red conjugates for fluorescent immunostaining, Vector Laboratories Ltd.

2.1.2. Bacterial strains and culture conditions

Strain: XL1BLUE- Δ (mcrCB-hsdSMR-mrr). lac[F' proAB, lacIqZ ΔM15] Su-.

Luria-Bertani broth: Bacto tryptone (Difco), 10 g; Bacto yeast extract (Difco), 5g; NaCl, 5g; per litre adjusted to pH 7.2.

Luria-Bertani agar: Luria-Bertani broth with 15 g/ l Bacto agar (Difco).

Ampicillin (Penbritin, Beecham Research) was added as required to a final concentration of 100 μg/ml

2.1.3. Fly strains and maintenance

137 (w^k ; R ; I promoter::CAT in pCATw8) (McLean *et al.*, 1993).

169 (w^k; I; I promoter::CAT in pCATw8) (McLean, *et al.*, 1993).

192 (w^k ; R ; CAT in pCATw8) Carol McLean, PhD thesis 1992.

Ja (R; *y* ;*w*) and derivative:

380 (Ja; R; *y*+ ; *w*+) (Chaboissier *et al.*, 1998 and Chapter 4).

"French" media: Oxoid No.3 agar, 7.5g; polenta, 55g; dried yeast pellets, 550g; nipagen (150mg/ml made up in 95% Ethanol), 10ml; dH_2O, 100 ml.

Apple agar: Bacto agar (Difco), 2.5g per 100 ml pure apple juice.

All flies were kept at constant temperature (25 ° C)

2.1.4. Oligonucleotides

Name	Sequence
From I factor sequences:	
5' X-I	5' CCG**CTCGAG**CATTACCACTTCAACCTCCG 3'
5'41	5' CCG**CTCGAG**CAGTCTAAAGCCTCG 3'
3'99	5' CCG**CTCGAG**GTTTATCAAGATTTTGC 3'
3'186	5' CCG**CTCGAG**GATTGTTGGTTAAGGGC 3'
431T	5' **CTCGAG**CAAATATCAACCACAAAGAG 3'
From MboI fragment of L.tarentolae k DNA (see fig 3.1.3)	
Knp 5'	5' GATCTAGACTAGACGC 3'
Knp3'	5' GCTCTAGAGATCCAAATCTCCCAC 3'
From *yellow* gene	
yw5'	5' GATTCCGGCCACTCTG 3' (from exon 2 position 3998 (+) strans)
yw3'	5' GTTAACCTTGATGCTG 3' (from exon 2 position 5385 (-) strand)
From *white* gene	
wh 5'	5' CCAGCGAAGCCGTCGAC 3' (from exon 2 position 11855 (+) strand)
wh3'	5' GCCACCAGAGTGACCAG 3' (from exon 2 position 12890 (-) strand)

2.1.5. Plasmids

Name	Description	Reference/Source
pGEM-T	TA cloning vector	Promega
p186T.1	Used to amplify I factor 5'UTR.	(McLean, *et al.*, 1993)
C42PIY	Used to amplify *yellow* marker gene.	(Chaboissier, *et al.*, 1998)
pW8.186.CAT	Used to amplify *white* marker gene.	(McLean, *et al.*, 1993)
pBend	Used for bending studies.	(Kim *et al.*, 1989)

2.1.6. Probes

Yellow probe was labeled by random priming using digoxygenin-dUTP (alkali labile) following manufacturer instructions (Boheringer Mannheim). The fragment for labeling was prepared by PCR amplifcation of the *yellow* gene cloned in C42PIY plasimd using yw5' and yw3' primers.

White probe was labeled by random priming using digoxygenin-dUTP (alkali labile) following manufaturer's instructions (Boheringer Mannheim). The fragment for labeling was prepared by PCR amplifcation of *white* gene cloned in pW8.186.CAT plasimd using wh5' and wh3' primers.

2.1.7. General solutions

The main chemical suppliers were SIGMA, Fisons, BDH and Boheringer Mannheim.

4 x gel loading buffer :

20% glycerol (v/v); 0.05 % bormophenol blue in TE.

Ficoll loading buffer: 0.25% Bromophenol Blue; 0.25% % xylene cynol; 15% Ficoll (type 400, Pharmacia)

5 x TBE gel running buffer: 54 g Tris; 27.5 g Boric Acid; 20 ml 0.5 M EDTA.

PBS : 10x for 1 litre 75.97g NaCl (130mM); 12.46g $Na_2HPO_4.2H_2O$ (7mM); 4.8g $NaH_2PO_4.2H_2O$ (3mM).Adjust pH to 7.5 and sterilize by autoclaving. PBT PBS pH7.5, 0.1 % BSA, 0.2 % Tween 20

Ringers solution: 6.5g NaCl; 0.14g KCl, 0.2 g $NaHCO_3$; 0.01g $NaH_2PO_4.2H_2O$; 0.12 g $CaCl_2.2H_2O$

Solutions used for specific protocols are given in the appropriate Methods sections. The composition of solutions that are supplied by manufacturers with specific enzymes are not given.

2.2. Methods

2.2.1. Agarose gel electrophoresis

DNA samples were electrophoresed on 0.5 to 1.5 % MP agarose (Boehringer Mannheim) in 1x TBE. Ethidium Bromide (EtBr) (working concentration 0.5 mg /ml) was usually added to the running buffer, except when gel were used for detection of bending. In this latter case an EtBr stain was prepared in 1x TBE and the gels were soaked in it for 30 mins and then detained for 20 min in d H_2O, before visualising under UV on a transilluminator.

2.2.2. Bacterial transformations

XL1BLUE bacterial cells from a fresh overnight culture were diluted 1:20 in prewarmed LB broth and grown for approximately 1 hr or until the OD_{595} reached a reading of 0.5. At this point cells were spun at 5000 rpms for 5 min on a bench centrifuge to sediment the cells. The supernantant was discarded and replaced with 5 ml of ice cold 0.1 M CaCl. Cells were resuspended by voprtexing and incubated on ice for up to 1 hr. The bacterial suspention was then spun as before and the supernatant discarded and replaced with 1 ml of ice cold 0.1 M CaCl and the cells resupsended. 100 to 200 μl of this cell suspension was added to an appropriate dilution of the plasmid DNA to be transformed and mixed well. Cells and DNA were incubated on ice of up to 30 min, then heat shocked at 37 °C for 90 seconds and finally returned to ice for 2 min. 0.5 ml of LB broth were then added and the cell left to recover at 37 °C (with gentle shaking) for up to 1hr. After this time the recovered

cells were spun on abench cetrifuge and 0.4ml of the supernatant removed. The cells were resuspended in the remining supenantant and spread on fresh LB+ Ampicillin plates using sterile technique. The plates were incubated overnight at 37°C and the number of colonies were checked in the morning agaisnt the appropriate controls.

2.2.3. Bradford assay

Protein samples from fly extracts were diluted to 1 ml dH20 and 1 ml Coomassie Protein Reagent (Pierce). Absorbance avlues at 595 nm were then determined and protein concentrations estimated against a standard curve obtained from measurements of bovine serum albumine standard dilutions prepared under identical conditions. For each extract two differnt dilution were prepared in duplicate and mean values calculated.

2.2.4. CAT - ELISA assay

The CAT-ELISA assay (Boehringer Mannheim) was performed folowing amnufacturer instructions except that dilution of CAT standard and adjusting of sample volumes was done using a 0.25 M Tris (pH 7.8) solution.

2.2.5. Cloning

I factor 5' UTR fragments in the pBend vector were cloned cloned in the pGEM T vector (Promega) following amnufaturer's instructions. After gel purification of DNA fragments, cloning into teh pBend vector was carried out using T4 DNA ligase and buffer (New England Biolabs) following manufaturer instructions.

2.2.6. Chromosome squashes preparation

For hybridisation to DNA probes

Larvae were grown on French fly food with added dried yeast and were kept at 25 °C. Salivary glands of third instar larvae were dissected in Ringer's solution. Excess fat was removed with twizers. A pair of glands was fixed for about 30 seconds in 45

% Acetic Acid. The glands were then placed on a siliconsed cover slip and a microdrop of 1:2:3 fixer solution was added. Looking under the dissecting microscope the glands were destroyed with the twizers, separating individual cells. A 10 μl droplet of the 1:2:3 fixer solution was added and the cells spread over a larger area of the cover slip. The whole process of gland destruction and spreading lasted about 3 minutes. The cover slip was then recovered by surface tension touching the drop of 1:2:3 fixer on the cover slip with a subbed glass slide. Cells and nuclei were broken by repeadetly dropping a sharp needle or scalpel on the cover slip and then moving the cover slip from side to side and monitoring the process under a face contrast microscope. Squashing required wrapping of a strip of tissue paper around the slide so as to immobilise the cover slip. With the coverslip face down a rubber cap was placed on the slide and the slide was hit with a rubber hammer 2-3 times and then continuous pressure was applied for about a minute by pressing with the hand on the rubber cap. Squashing was also monitored under face contrast and judged successful when the chromosomes appeared pale with no dark birefrangent bands. After squashing the chromosome preparation was placed in liquid nitrogen for 20 seconds or until sizzling stopped. The cover slip was removed quickly and the slide plunged in 70 % EtOH for at least 5 min and then in 96 % EtOH for 5 min. After the final EtOH wast the slide was air dried and rehydrated with dH2O for examination under the microscope and quality assessment. Good slides where the chomosome arms lay flat and are well spread were dehydrated again and stored at 4°C till ready for hybridisation.

Siliconising cover slips: Individual cover slips were dipped in a 1% Silicone solution [SIGMA] in Chloroform (Fisons) and then allowed to dry overnight in a dust free environment.

Subbing solution (helps biological preparation to stick better to glass surfaces):1% gelatin (SIGMA), 0.1% Chromium alum (SIGMA). The mixture was heated to 60°C taking care it did not boil and used after cooling. Unused subbing solution was stored at 4°C.

Subbing of slides: Microscope slides were washed in a mild detergent, rinsed many times in ditstilled water with a final rinse in 50 EtOH and then air dried.The dried slides were dipped for a few seconds in subbing solution. Subbed slides were left to dry in a dust free environemnet and were stored in racks in closed containers at room temperature.

1:2:3 Fixer solution: Prepared by mixing 1 part Lactic acid (Fisons) with 2 parts d H2O and 3 parts Acetic Acid (Fisons). Note: both fixing solutions were prepared fresh just before use.

For Immunostaining (adapted from (Zink *et al.*, 1991)

For high polyteny larvae were grown at 18 °C on medium with live yeast. Salivary glands were dissected in PBS pH 7.5 + Triton X-100. For fixation two pairs of glands were placed into a droplet of SOLUTION 1 for 10-20 sec. The glands were then immediately removed to a 35 μl droplet of SOLUTION 2 on a coverslip for 2-3 min. The cover slip was then picked up with a Poly-L-lysine coated glass slide that was then gently but extensively tapped and moved using the tip and the eraser end of a pencil. The slide was then turned over onto blotting paper and thumb pressurewas applied. The slide was submerged in liquid Nitrogen and the cover quickly slipped off slip with a razor blade. The slides were immediately placed in PBS and washed twice 15 min each. After washes, the slides were placed in PBT and washed twice 10 min each agitating. 0.5 μl of affinity purified Rabbit-Anti-Pc per slide were then added (from primary Abs 1:200 stock) diluted in PBT. Preparations were covered with coverslip and keept in humid box for 1hour at RT. Slides were then washed in PBT twice for 10 min each agitating. The biotinylated goat-anti-rabbit secondary Abs (preabsorbed to wild-type embryos) were then added (diluted in PBT) + 2% of normal goat serum and preparations incubated 45 min at RT. Slides were then washed twice in PBT for 10 min agitating. The avidin-biotin-HRP complex pre incubated in PBT was then added and the chromosome prearations incubated 45 min at RT. The slides were then washed again in PBT twice for 10 min at RT, and incubated with DAB solution (5mg/ml in PBT) for 5 min at RT. The DAB-only solutions was then repalced with DAB + 0.01% H2O2 and the staining reaction let proceed 1-10 min checking under the microscope. Reactions by washing slides in PBS. Chromosomes were counterstaining with Giemsa (1:130 dilution in 10mM $NaPO_4$ buffer pH 6.8) for 20 s then rinsed briefly in dH_2O. Slides were washed and mounted in 90 % glycerol and examined under the microscope immediately (staining faded within hours).

PBT: PBS pH7.5, 0.1 % BSA, 0.2 % Tween 20

SOLUTION 1 : 3.7% Formaldehyde, 1% Triton X-100, PBS pH 7.5

SOLUTION 2 : 3.7% Formaldehyde, 50 % Acetic acid .

For Fluorescence Immunostaining

Materials

Liquid nitrogen

0.7% NaCl

siliconized 22 mm and non-siliconised 22 x 40 mm coverslips

EtOH cleaned slides

PBS pH 7.2

PBT = PBS + 1% BSA + 0.2% Triton X-100 + [0.02% azide]

Solutions: (always prepared on the day)

1. 5X formaldehyde = 18.5% paraformaldehyde

Heat 5ml of water and then add 0.925g of paraformaldehyde were added to 5 ml of pre heated dH20 under a fume hood and the mixture vortexed. 35μl of 1N KOH were added and the solution vortexed again till well until in solution.

2. **Solution I**: 0.2 ml of 5X formaldehyde
0.01 ml Triton X-100
0.79 ml PBS pH 7.2

3. **Solution II**: 0.2 ml 5X formaldehyde
0.3 ml water
0.5 ml acetic acid

The solutions were kept on ice and replaced with freshely made solutions every hour.

Mature third instar larvae were dissected in 0.7% NaCl. The larval salivary glands were fixed in 40 μl of Solution I on a siliconised coverslip for 30 seconds to 1 minute (maximum). Preparations were fixed in 40 μl of Solution II for 2to 5 minutes on the coverslip by pouring off the Sol I. The coverslip was picked up with a slide and inverted. Preparations were squashed by tracing back and forth, and then holding the coverslip and tapping the glands. The slide was then wrapped in tissues and pressure applied to flatten. The slide was then placed in liquid nitrogen and kept there until all squashes were ready. All incubation steps from now on were done at room temperature unless otherwise noted. SLides were washed two times (10 minutes each) in PBS and then blocked in PBT azide for 30 minutes. After a breif washing in PBS the slides were incubated with anti Pc antibody (the equivalent of 1 μl per slide of a 1:200 dilution in 40 ul of PBT) under a 22 x 40 mm coverslip in a humidified

dish. Incubation was left to proceed overnight. In the morning slides were washed twice (10 minutes each) in PBS. Slides were then blocked in PBT plus 2% goat serum for 30 minutes and then incubated with secondary biotinylated Ab (the equivalent of 1μl per slide of a 1:200 dilution stock in 40 μl PBT) under a coverslip in a humidified dish. Incubation with secondary antibodies was left to proceed for up to 3 hours in a humid box. Slides were then washed twice (10 min each) in PBS. All following steps were carried out in darkness. Slides were incubated for up to 5 hours with streptavidin linked Texas Red (VECTOR laboratories Ltd) in a humid box wrapped in foil to protect from light. The slides were then washed again in PBS twice (10 min each) and the chromosome preparations were stained with DAPI for up to 5 min and washed again twice in PBS followed by air drying. Finally slides were mounted in a glycerol based solution containing the antibleaching agent Mowiol (CALBIOCHEM®). Slides were analysed under an Olympus microscope.

2.2.7. DNA sequencing

I factor and Kinetoplast DNA fragments cloned within the pGEMT and pBend 2 vectors were sequenced using the corresponding primers. Appropriate amounts of primer and template plasmid DNA were supsended in a final volume of 10 μl in dH_20 and boiled together for 4 minutes. The denatured mixture was then snap frozen on dry ice of liquid nitrogen and thowed briefly at 37°C just before use.Sequencing reactions were set up using the T7 Sequenase version 2.0 DNA sequencing kit following manufacturer instructions.

2.2.8. Embryo collection for extract preparation

As a positive control to test specificity of affinity purified anti-Pc anitbodies, extracts were prepared from 18-22 hour old embryos, a time of maximum *Polycomb* expression. Staged embryos were collected on apple agar, dechorionated in 50 % bleach and then rinsed briefly in dH_20 and then transferred to an eppendorf. The dechorinated embryos were then washed in 10 mM Tris pH7.2, 300 mM NaCl and 0.5 % Triton X-100 (SIGMA) and spun briefly on abench centrifuge (12000 rpm) to pellet. Excess supernatant was removed and the preparation was then snap frozen in liquid nitrogen. 300 μl of sample buffer (0.0625 M Tris-HCl pH 6.8, 2 % SDS, 10 %

glycerol, 5 % Beta-mercaptoethanol and 0.001 % Bromo Phenol Blue) were then added and the sample homgenised. The homogenate was then spun briefly on a bench centrifuge (12000 rpm) to pellet debris and 15 μl of supernatant were loaded on an SDS PAGE.

2.2.9. End labeling

End labelling of I factor 5' UTR fragments, PCR amplified kinetoplast DNA and 123 bp ladder for non-denturing gel electrophoresis was carried out using T4 polynucleotide kinase (New England Biolabs) following manufaturer instructions.

2.2.10. Extraction of DNA from agarose gel

Several μg of plasmid DNA was digested to completion with with approprite enzyme (New ENgland Biolabs) then the products were run on high percentage agarose gel (1.5%) until fragment were resolved. Troughs were then cut into the gel just below the position of the samples and and voltage reapplied until the samples reached the edge of the trough. From this point on electrophoresis was interrupted every 20/30 seconds and approx 200μl factions were collected in individual eppendorfs each time, till the most of the sample had been reomoved. The fractions were then visualised under a UV transilluminator and thhose giving strongest staining were pooled and extracted once with phenol (SIGMA), once with phenol:chlorophorm:isoamyl alcohol 24:24:1 (SIGMA; Fisons) and once with chlorophorm: isoamyl 24:1. The precipitated in 2 volumes ethanol and 1/10 volume 3M NaAcetate pH5.2 and finally washed in 75% EtOH and resuspended in appropriate volume of 10 mM Tris pH 8.0. Identical procedures were used to gel purify PCR products.

2.2.11. Fly dissection and extract preparation

100 pairs of ovaries were dissected from the rest of the female carcass in Ringer's solution under a Zeiss dissection microscope. The tissue fractions were then spun on a bench top centrifuge at 12000 rpms for 5 min to pellet and the excess Ringer's pipetted out and replaced with 500μl of 0.25 M Tris-HCl pH 7.8. Intact males were

subjected to identical treatment. Tissues were homogenised throuroughly and then put through six cycles of freeze thawing. After the fianl thawing the samples were incubated for 5 min at 65 °C and then spun for 5 min at 12000rpms. The supernatants were placed in fresh tubes and then snap-frozen in liquid nitrogen till required for Bradford assay or CAT -ELISA.

2.2.12. Molecular weight markers

DNA molecular weight markers for agarose or polyacrylamide gel electrophoresis were:

- ØX174 phage DNA digested with Hae III (1000μg/ml) (Kramel Biotech)

- 123bp ladder from mouse prolactin gene (1μg/ μl) (SIGMA).

2.2.13. Mapping of polytene chromosomes

Identification of chromosome arms and polytene banding patterns was carried out by direct comparison with the polytene chromosome maps provided in (Lefevre Jr.,).

2.2.14. Plasmid DNA extraction

Plasmid DNA was prepared from fresh overnight bacterial cultures (XL1BLUE) using the Quiagen mini or midi prep kit following manufacturer instructions.

2.2.15. Polyacrylamide gel electrophoresis

Detection of DNA bending.

The I factor 5'UTR (in plasmid p186.T) and the MboI fragment (in plasmid pPE103) were first sequenced, to confirm their identities (see Chp 2). Specific primers were designed and used to amplify each fragment of interest. The PCR products were purified and their ? ends labelled with the P^{32} radioisotope. After further purification to remove unicorporated radiolabel the fragments were loaded on a native 12% PA gel that had been equilibrated at 4° C overnight and then pre run for >1hr till a constant temperature of 18° C was reached. The temeperature was checked periodically by placing a thermometer in the buffer tank immediately above the gel and a reading takin when the meniscus had stabilised. gels were run at 1200 Vs for up to 16 hours. As a size marker the 123 bp mouse ladder was loaded together with

the test samples. All samples were mixed with an appropriate volume of Ficoll loading buffer (see SOLUTIONS) to give a final volume of 10 μl.

For DNA sequencing

6% Polyacrylamide gels + 6M urea in 1x TBE were pre-warmed to 50-55 °C (19 Amps- 1500 Volts for 1 hour) and then samples for sequencing were loaded and resolved as appropriate.

For SDS-PAGE

After pouring resolvng and stacking gels, samples were loaded and gel run for approprite length of time at 100 Volts. For immediate visualisation of protein, gels were rescued and placed in stain solution for 45 min and then washed in destain solution usually overnight. Destained gels could then be inspected directly or dried for long term storage. For Western blotting, transfer of proteins to PVDF membrane (Boehringer-Mannheim) was carried out from unstained gels under an applied electric field (40-43 volts) in tranfer buffer at 4 °C for 1.5 hours. After transfer was complete the PVDF memebrane was processed for non-radioative detection of the bound protein with the Boehringer-Mannheing kit (using 40μl of 1:200 primary anti Pc antibodies) following manufacturer instructions.

Resolving gel 5 ml of 10%:
(% of acrylamide in the resolving gel depends on size of protein!)
Resolving gel buffer .....1.2 ml
30 % Acrylamide.....1.6 ml
Dist. water.....2.2 ml
25% APS.....30μl
TEMED.....3μl
Adjust vols of acrylamide & water for different % gel)

Stacking gel 2.5 ml of 5 %:
Stacking gel buffer .....0.6 ml
30 % Acrylamide.....0.4 ml
Dist water.....1.4 ml
25 % APS .....15 μl
TEMED.....1.5 μl

Stacking gel buffer
5.1g Tris (0.42 M)
0.4g SDS (0.4%)
pH to 6.7 with HCl
Sterilize by autoclaving

Resolving gel buffer 100ml:
18.15g Tris (1.5M)
0.4g SDS (0.4%)
pH to 8.9 with HCl
Sterilise by autoclaving

10 x Running gel buffer (TGS= TRIS, GLYCINE,SDS) for 1L
30g Tris (250 mM)
144g Glycine (1.92 M)
10g SDS (1 %)

Stain 550ml:
1g Coomassie brilliant blue R-250
dissolve in 250 ml methanol
then add 250 ml distilled water
followed by 50 ml acetic acid

Destain 500 ml:
400 ml distilled water
50 ml methanol (10% v/v)
50 ml acetic acid (10% v/v)

2.2.16. Polymerase chain reaction conditions

All PCR reactions, unless otherwise stated were carried out using the following concentrations of reagents:

Template.....200ng
primers.....20pmol
Deoxyribonulceotides 25μM
$MgCl_2$ .....1.5 mM

In all reactions the equivalent of 1 unit of Taq polymerase (Boheringer Mannheim) and were used and 1x working strength of the corresponding reaction buffer (Boheringer Mannheim) . All reactions were carried out in a final volume of 100 μl. The thermal cycler used was the Hybaid Omne. Amplification conditions are specified below for individual primer sets.

Amplification of I factor derived fragments for detection of bending:

Primer set	Expected product
5' X-I / 3'186	1-186
5' 41 / 3'99	41-99
431T / 3'186	99-186

For all three sets of primers the following conditions were used

Function	Temperature	Time (min)
Denaturing	93 °C	5
Annealing	40° C	2
Extention	72° C	1
Number of cycles 3. Followed by		
Deanturing	93° C	1.5
Annealing	40° C	1.5
Extention	72° C	1

Number of cycles 27. Total cycles 30

Amplification of Kinetoplast *MboI* framgment

Prmier set	Product
Knp5' / Knp 3'	*MboI* fragment

Function	Temeperature	Time(min)
Denaturing	95° C	5
Annealing	49° C	3
Extention	72° C	2
Number of cycles 3.		
Denaturing	95° C	1
Annealing	49° C	2
Extention	72° C	2

Number of cycles 17.

Amplification of yellow gene exon 2 sequence

yw5' / ywp 3'	exon 2 fragment (1387bp)	
Function	Temeperature	Time(min)
Denaturing	95° C	5
Annealing	48° C	2
Extention	72° C	3
Number of cycles 1.		
Denaturing	95° C	2
Annealing	48° C	2
Extention	72° C	3
number of cycles 29.		

Amplification of white gene exon 2 sequences

wh5' / wh3'	exon 2 fragment (1035bp)	
Function	Temeperature	Time(min)
Denaturing	95° C	5
Annealing	58° C	2
Extention	72° C	2
Number of cycles 1.		
Denaturing	95° C	2
Annealing	58° C	2
Extention	72° C	2
number of cycles 29.		

2.2.17. Protein extracts from Pc expressing *E.coli* cells

E.coli cells contaning Pc-encoding plasmid were grown at 37 °C in selective medium till they reached an OD og 0.5 as measured at 595 nm. At this point PC expression was induced by addition of 10 mM IPTG. After induction cells were incubated for 2 hours at 37°C and then harvested by spinning for 5 min at 5000 rpm on a bench top centrifuge. Non-induced controls were treated under identical conditions. Excess supernatant was removed and the pelleted cells were resuspended by pipetting in boiling buffer (below) and then incubated at 100 °C for 5-10 min. Samples were then spun for 5 min at 12000 rpm and 20 μl supernatant were loaded in each well as appropriate.

6 x Sample boiling buffer :
1.0 ml stacking gel buffer
0.8 ml 25 % SDS (25 % SDS)
0.5 ml Beta-mercaptoethanol
1.0 ml glycerol
0.05% (w/v) BBP dye

2.2.18. Radioisotopes

alpha-32P-dCTP (3000 Ci/mM) for end labeling and alpha-35S-dATP (400 Ci/mM) for sequencing were supplied by Amersham.

Chapter3

Physical properties of the I factor promoter

3. Physical properties of the I factor promoter

3.1. Intrinsic bending of DNA

3.1.1. Introduction

The aims of the work described in this chapter were to investigate the possibility that the *I* factor 5'UTR can adopt an unusually bent conformation in solution, and to establish a possible correlation between this structural feature and some of the biological properties of the *I* factor. The bending hypothesis of the *I* factor 5'UTR was based on considerations of the significant dA-dT richness of this sequence. This base composition has been found associated with an unusual helical conformation that can result in DNA bending (Hagerman, 1990; Crothers *et al.*, 1990).

Theoretical models describing intrinsically curved DNA molecules were prompted in the late '70s by increasing knowledge of the extent of DNA packaging in the eukaryotic nucleus and in viral capsids (Crothers, *et al.*, 1990). Although in many cases the folding of DNA is associated with bound protein, protein-induced deflections of the double helix were seen as reflecting a propensity to deformation of the DNA molecule that is determined entirely by the local base composition.

The first experimental evidence in favour of naturally occurring intrinsically bent DNA molecules, came from observations of the unusual electrophoretic behaviour of a short DNA fragment derived from the *Leishmania tarentolae* mitochondrial genome. In this parasitic protozoan, like in other tryponosomatidiae, the mitochodrial (kinetoplast) DNA has an unusual organisation (Shapiro, 1993; Shapiro and Englund, 1995): it consists of a complex network of some 10000 minicircles and approximately 50 so-called maxicircles that are interlocked and packaged within the mitochondrial matrix. Minicircle sizes among species range between 0.5 kb and 2.9 kb, whereas maxicircles can be anything between 20 and 40 kb in length. Maxicircles are considered to be the equivalent of the mitochondrial genome found in other eukaryotes because they encode ribosomal RNAs as well as several proteins most of which are involved in mitochondrial energy transduction. Most of the

transcripts from maxicircle genes undergo the unusual process of RNA editing where nucleotides are inserted or deleted at specific positions to form the mRNA. The minicircles within a network, vary in base composition and code for a variety of essential guide RNAs that direct RNA editing.

A 414 bp *MboI* restriction fragment from *L.tarentolae* kinetoplast DNA (kDNA) (Marini and Englund, 1983) displayed a curious electrophoretic anomaly: migration of this fragment through polyacrylamide (PA) gels was greatly reduced with respect to its mobility on agarose gels (Simpson, 1979; Marini *et al.*, 1982a). Moreover, the migration anomaly of the *MboI* fragment grew larger with increasing acrylamide concentrations, so that the size of this fragment estimated from its mobility on a 12% PA gel was more than twice the size estimate obtained on a 0.5% agarose gel, the latter being reasonably close to its real size in bp (Marini, *et al.*, 1982a). The average pore size is significantly smaller (1 to 8 nm) in PA gels than in agarose gels (42 to 390 nm) and, in the former case, it is also very close to the diameter of the double helix (2 nm) (Levene and Zimm, 1989; Diekmann, 1992). Since pore size is an approximation of inter fiber spacing within the gel matrix, increasing gel concentration decreases the average size of the gel pores that provide passageways for DNA molecules migrating along the applied electric field (Lumpkin and Zimm, 1982; Lerman and Frisch, 1982). It was, therefore, plausible that PA gels were detecting departures of the *MboI* fragment from the rod-like shape adopted by DNA molecules in acqueous and gel solutions (Marini, *et al.*, 1982a).

Further characterisation revealed that a number of variables had measurable effects on the electrophoretic mobility of the *MboI* fragment (Marini *et al.*, 1984). At high voltage, for example, when rod-shaped molecules are predicted to migrate faster, the mobility of the *MboI* fragment was reduced further. A similar enhancement of the migration anomaly was seen if electrophoresis was carried out at low temperature ($\leq$ 5 °C). In contrast, the migration anomaly of *MboI* was greatly alleviated if ethanol (5%-25%) or urea were added to the electrophoresis buffer. A similar effect was seen at high temperatures, so that at 63°C and above the *MboI* fragment migrated as expected for a molecule of its size. Little or no effect on *MboI* mobility was observed after varying the ionic strength of the electrophoresis buffer (Marini, *et al.*, 1984), although later investigation have shown that millimolar amounts of Mg^{2+} are required to detect electrophoretic abnormalities (Brukner *et al.*, 1995). Following denaturation (by boiling), each of the two strands of the *MboI* fragment migrated as expected for their size, but the mobility abnormality reappeared after one cycle of renaturation under controlled conditions. This observation strongly suggested that

the electrophoretic behaviour of *MboI* depends on properties of the double helical structure which are not preserved upon separation of the two strands.

In gel filtration experiments, the *MboI* fragment eluted like a molecule with an apparent size of 375 bp, that is to say about 10 % smaller than its real size in base pairs. This methodology is very sensitive to the overall shape of a molecule, in so far that elution profiles depend on the speed with which molecules can diffuse through the gel beads that pack the column. Since, compact molecules have faster elution times than long rod-like ones, the result obtained for the *MboI* fragment suggested that the latter is more compact than predicted from its size in bp. Transient electric dichroism (TED), relies on the ability of charged molecules to orient within an electric field. When the field is abruptly removed, the molecules will gradually lose the alignment. This "decay" can be detected as changes in the absorbance of polarised light that is shone through the sample. The time that it takes for this decay to be completed is called rotational relaxation time and in a way it is a reflection of the frictional resistance of the molecule to rotation within the solution, which in turn depends on its shape and size. The rotational relaxation time of the *MboI* fragment was smaller than the value measured for a 309 bp control molecule. Independent studies using a variant of TED called transient electric birefringence (TEB), which uses two beams of polarised light at right angles to each other were carried out on dimerised subfragments of the *MboI* molecule (Hagerman, 1984). These studies confirmed that those *MboI* subfragments showing the largest electrophoretic abnormality also had TEB diffusion coefficients typical of molecules shorter than their size in bp.

The hypothesis put forward to explain the physical properties of the *MboI* fragment was that this molecule could be bent in solution (Marini, *et al.*, 1982a; Marini, *et al.*, 1984). Bending would bring the ends of the molecule closer together and, therefore, give it an overall more compact shape, which could account for its behaviour in gel filtration, TED and TEB experiments. Bending could also explain the observed electrophoretic anomaly of the *MboI* fragment. Existing theoretical models that describe DNA migration through a gel (Lumpkin and Zimm, 1982; Lerman and Frisch, 1982), predicted that departures from a rod-like shape would result in decreased mobility because of increased friction between the distorted molecule and the gel matrix. Accounting for the complex electrophoretic behaviour of the *MboI* fregment required also the introduction of parameters that express the elastic interactions between the gel and the bent DNA molecule (Levene and Zimm, 1989).

The intrinsic bending hypothesis had to be tested against several possible alternatives that could equally well explain the experimental results (Marini, *et al.*, 1984). Specific aggregation of the *MboI* fragments due to unusual stickiness, for example, was ruled out by calculation of their sedimentation velocity on sucrose gradients, which gave values compatible with those of a monomer. The presence of chemical moieties, e.g. methyl groups, due to *in vivo* modification could be easily excluded because the *MboI* fragment has identical electrophoretic behaviour whether it has been isolated directly from the kinetoplast or after propagation through methylation competent and incompetent *E.coli* strains. Tightly bound protein not removed by standard DNA purification procedures could also account for the electrophoretic behaviour the *MboI* fragment. This possibility was invalidated by the fact that extensive protein kinase treatment as well as repeated phenol extractions had no effect on *MboI* mobility. Finally, certain forms of secondary structure, such as cruciforms, also result in contour length reduction and anomalous migration through a gel. Disruption of interstrand interactions due to secondary structure formation, such as the loops in the hairpins of a cruciform, often generates single stranded regions that are sensitive to attack by single strand specific nucleases. Prolonged treatment of the *MboI* fragment with large amounts of S1 nuclease did not in any way diminish its electrophoretic abnormality. It was therefore concluded that the unusual properties of the *MboI* fragment are determined by a sequence-induced bend in the doule helix (Wu and Crothers, 1984; Hagerman, 1984; Marini, *et al.*, 1984).

3.1.2. Models of DNA bending

The natural variation in base composition along a DNA molecule creates anisotropy, that is to say, different stretches of DNA are non-equivalent and have unique sequence-determined structural features. Precise stacking of the aromatic rings of the bases, for example, depends on the steric interaction between adjacent nucleotides. Dinucleotide combinations that prevent parallel stacking of the base rings can cause local distortions in the double helix by introducing a wedge. If such distortion-inducing dinucleotides occur with a precise periodicity that matches the helical turn (10.5 bp), the net result is unidirectional bending of the DNA axis in a single plane. This is because matching sequence and helical periodicities, place the wedges at exactly the same position on the same side of the double helix: in this configuration, all the distortions will add up constructively to impose an overall curvature on the DNA molecule. These assumptions formed the basis of the early "wedge" models of

DNA bending (Trifonov and Sussman, 1980; Crothers, *et al.*, 1990). Wedge models have since evolved into a number of structure prediction algorithms under the generalised name of nearest-neighbour models. The main differences among them lies in the source of the structural parameters and in whether calculations are applied to di-, tri-, or tetranucleotide neighbours (Goodsell and Dickerson, 1994).

Positional autocorrelation analysis can be used to determine the periodicity of a sequence element (e.g. a dinucleotide) along a DNA molecule. This procedure involves examining the frequency of occurrence of the same sequence element at different distances from each other. Periodicity is then detected as regularly spaced peaks of occurrence of that sequence element on a background of random distributions. Application of this methodology to a set of 32 eukaryotic sequences yielded clear periodicities of dinucleotide combinations corresponding to the DNA helical turn in solution (10.3-10.4 bp/turn) (Trifonov and Sussman, 1980). Of the 16 possible dinucleotide combinations, the ApA doublet and its complement TpT, were singled out as those with the stronger periodicities. Together with the fact that such periodicities were not readily detectable in prokaryotic sequences, this was interpreted as evidence that dinucleotide distributions in eukaryotic genomes have been selected for to facilitate the wrapping of DNA around the core histones in the first step of DNA packaging (Trifonov and Sussman, 1980).

Positional autocorrelation analysis of the bent kDNA fragment showed a clear periodicity of ApA and TpT dinucleotides along its sequence (Marini, *et al.*, 1982a). The presence of polyd(A)-polyd(T) segments (A-tracts) with a specific phase along the *MboI* fragment was confirmed in a series of experiments aimed at mapping the centre (or focus) of bending of this molecule (Wu and Crothers, 1984). Head-to-tail dimers of a 240bp subfragment of the *MboI* molecule were generated by ligation. A number of restriction enzymes were chosen so that each had a unique restriction site within the 240bp monomer and would therefore cut twice within the dimer. The fragments produced by digestion with the selected enzymes were identical but circularly permutated versions of each other. The restriction products were then run on PA gels under conditions favourable for detection of bending. The hypothesis behind the experimental design was that the fragments carrying the bend region in their centre would be slow migrating whilst fragments carrying the bend at their ends would be fast migrating. Plotting relative mobility versus position (in bp) allowed to extrapolate with a good degree of certainty the centre of the bend from the values obtained for the fastest migrating species. This bending focus resided almost exactly in the centre of the 240 bp fragment and was flanked on either side by short A tracts

with the consensus ($CA_{5-6}T$) and with a periodicity of half a helical turn (~ 5 bp)(see fig 3.1.4-1).

Preliminary data from physical studies of synthetic polyd(A)-polyd(T) oligos revealed an unusual diffraction pattern that could be explained by model building if the two strands were not in the same helicalconfiguration (Arnott *et al.*, 1983). The best fit with the data was obtained if the polyd(A) chain underwent a conformational readjustment from the B to the A form whilst the poly d(T) chain remained B-like. The main difference between A and B forms of DNA is in the puckering conformation of the deoxyribose ring which pertains to the relative angles between carbon to carbon bonds in the ring (Dickerson *et al.*, 1982). Different conformations of the sugar ring have consequences for the structure of the whole helix and the result is a shorter and thicker helix in the case of the A form with about 11 residues per turn, and a very deep major groove and shallow minor groove, whilst the B form is a slimmer helix with about 10 bp per turn and comparable depth of major and minor grooves. The A and B forms represent transient states of the double helix that can interconvert as a consequence of changes in hydration and temperature, although base composition also plays a role in determining the stability of a given conformation and the ease of transition.

The unusual structure assumed by polyd(A)-polyd(T) oligos, where the A and B forms coexist in the same helix was named heteronomous DNA (H-DNA) (Arnott, *et al.*, 1983). The hypothesis put forward to explain kDNA bending was that short polyd(A)-polyd(T) tracts retain their non-B conformation even when intercalated with B-form DNA (Wu and Crothers, 1984). Model building predicted that energetically favourable stacking of base rings at the junction between H and B DNA segments should result in an abrupt bend of the DNA at that point because the helical axes of the two segments are not parallel (Selsing *et al.*, 1979; Arnott *et al.*, 1982; Arnott, *et al.*, 1983). This junction-bend hypothesis, offered an important alternative to nearest neighbour models as a possible explanation of intrinsic DNA bending (Arnott, *et al.*, 1983; Wu and Crothers, 1984; Koo and Crothers, 1987; Crothers, *et al.*, 1990) Moreover, it could explain bending of the *MboI* fragment in relation to the precise phasing of the ($CA_{5-6}T$) consensus (Crothers, *et al.*, 1990): the half helical turn periodicity meant that a H-DNA to B-DNA junction occurred every 10 or so base pairs. This meany that successive H/B junctions were placed on the same side of the helix and any distortions occurring at this point would be additive.

The crucial assumption made by the junction model regarding bending at phased A-tracts, is that the 5' and 3' junctions of the A-tracts are not equivalent, and, in fact, most of the bending (in the direction of the minor groove) appears to be contributed by the 3' junction (Koo *et al.*, 1986). The asymmmetry* of A-tract junctions, was suggested by the unexpected finding of a large mobility anomaly for an oligomer with a repeat unit of the type $(A_9N_1)_n$. If 5' and 3' junctions were equivalent, this molecule would be expected to behave as if straight because the two junctions are almost one helical turn apart and their contributions to curvature should cancel out. This junction asymmetry not only accounts for the anomalous behaviour of $(A_9N_1)_n$ containing molecules, but also explains bending of the *MboI* fragment. The phasing of the $(CA_{5\text{-}6}T)$ repeat within the *MboI* fragment is such that successive 3' junctions are separated from each other by almost exactly one helical turn, which places them on the same side of the helix. Moreover, the 5' junctions are half a helical turn away, which means that their effects on the overall helical axis is therefore expected to be additive and to result in curvature of the *MboI* molecule. Hydroxyl radical cleveage experiments support the non-equivalence of 5' and 3' junctions of A tract (Burkhoff and Tullius, 1987). The hydroxyl radical agent attacks the sugar backbone of a DNA molecule with equal probability along the entire length of the helix, but experimental applications of this technique to known bent sequences have shown this agent to be sensitive to distortions in the DNA helix that render certain part of the sugar backbone less accessible. Hydroxyl radical cleavage of molecules containing phased A-tracts gives a sinusoidal pattern with the lowest cutting frequencies observed at the A tracts. In particular, the cleavage frequency decreases smoothly from the first (5') A through the last (3') A where minima are observed. The reduced clevage frequencies at A-tracts are attributed to decreased accessibility of the minor groove due to its bend-induced compression (Burkhoff and Tullius, 1987).

The main difference between nearest-neighbours and junction models of DNA bending resides in their respective geometrical representations of the double helix

---

* Asymmetry here refers to the amplitude of the helical deflection angle, which is postulated to be larger at the 3' than at the 5' junction. However, in order to explain the outcome of ligation ladder experiments and the bending of the kDNA fragment, 5' and 3' junctions must also be asymmetric with respect to the *direction* of bending. In other words, B-DNA to A-tract transitions must bend the helical axis in the opposite direction with respect to A-tract to B-DNA transitions. This is not clearly stated in Koo et al., 1986 but is implied in the digram shown in fig 1 of Crothers et al., 1990.

(Crothers, *et al.*, 1990). According to the first, adjacent dinucleotides (steps) are viewed as geometrically independent from each other because each defines its own helical axis. Departures from parallel stacking of the bases in the dinucleotide step result in local distortion of the helix. The sum of the wedges contributed by individual dinucleotide steps causes smooth bending of the DNA molecule, without significant departures from the B conformation. Estimates of wedge angle amplitudes (which reflect the extent of non parallelism between base pairs in a step) have shown that AA/TT dinucleotide steps are one among five dinucleotide combinations where adjacent base rings are non-parallel with the largest wedge angle being associated with the AG/CT dinucleotide (Bolshoy *et al.*, 1991). In contrast, the junction model defines individual straight axes for conformationally distinct segments of DNA and the bending is predicted to occur at the junction between the segments because their respective axes are not parallel to each other. Nearest-neighbour models are prediction tools that aim at determining the rotational parameters of successive base pairs from which global structures are deduced. The junction model can therefore be viewed as a very special application of nearest-neighbour predictions, where global structure is deduced for parameter calculations of DNA segments longer than two base pairs, and where significant departure from standard B-DNA is postulated.

Both nearest-neighbour and junction models can explain the same DNA structure and both require precise phasing of the wedges or junction bends with the helical turn so that deflections of the helical axis can add constructively. Each, however, makes distinct predictions in terms of the effects on the bending of sequence changes. Since, according to the nearest-neighbour models, the AA/TT steps in the A-tract are contributing local helical distortions that are additive with respect to those contributed by the mixed sequence tracts, the prediction is that base changes in the intervening sequences should have significant effects on the bending. The junction model, on the other hand, predicts that the most dramatic effects on the bending will be obtained with base changes within the A-tract, whilst altering nucleotides in the regions of B-DNA should only have small modulating effects. The interruption of A-tracts is expected to lower the free energy value associated with the A-tract conformation (Koo and Crothers, 1987). Thermodynamic considerations, predict that A-tracts should be at least 3 bases long before they can adopt a stable non-B conformation. This is because the favourable free energy generated by the structural switch of shorter (<3 bases) A-tracts to a less constrained non-B form cannot offset the unfavourable free energy generated at the junction with B DNA.

The available experimental evidence, offers some support to the junction model of DNA bending, in so far that the properties of A-tract containing molecules do not appear to be easily described from the parameters of individual dinucleotide steps. Crystallographic data have confirmed the unusual conformation of polyd(A)-polyd(T) oligos, although they invalidate the earlier H-DNA hypothesis (Alexeev *et al.*, 1987). Diffraction patterns are consistent with both strands in a polyd(A)-polyd(T) oligo having the same B-like conformation. There are, however, some significant departures from the parameters defined for standard B-DNA, sufficient to require the B' notation for this non-canonical B-conformation of the double helix. The most salient feature of a B'-form helix is the fact that the bases in each A-T pair are rotated relative to each other along their longitudinal axis, generating what is known as a propeller twist (Nelson *et al.*, 1987). The main consequence of this propeller twist is that the nitrogen at position 6 in the Adenine ring can now hydrogen bond to oxygens at position 4 in the rings of two successive Thymines. This generates a series of bifurcating hydrogen bonds that are predicted to confer an unusual rigidity upon polyd(A)-polyd(T) oligos. This increased stiffness might explain why A homopolymers exhibit faster mobilites of PA gels than mixed sequence DNA molecules (Anderson, 1986). Rigid rod-like molecules are, in fact, expected to experience minimal friction as they pass through the tight pores of PA gels. Furthermore, A homopolymers have consistently failed to form nucleosomes in vitro (Rhodes, 1979; Simpson and Kunzler, 1979; Drew and Travers, 1985), which again supports the hypothesis of increased rigidity for DNA molecules with this base composition.

Another important piece of information to emerge from the crystallographic data is the fact that the AA/TT steps within the A tract, do not appear to have the necessary roll and tilt components that would contribute to the wedge effect (Nelson, *et al.*, 1987). This finding, therefore, seemed to exclude a wedge contribution from the AA/TT steps within the tract, although more careful evaluation warrants some caution in interpreting DNA crystallographic data for strucuture prediciton purposes (Haron *et al.*, 1994). In several cases, in fact, the conformation of the crystallised double helix had to be attributed to its forced packaging within the crystal unit cell and therefore did not necessarily reflect conformations adopted in solution* (Di

---

*Goodsell et al., 1994 find that the CCATTAATGG decamer assumes different conformations depending on the type (trigonal or orthogonal) of crystal in which it is packaged. The authors interpret

Gabriele *et al.*, 1989; Lipanov *et al.*, 1993; Goodsell *et al.*, 1994). Thus, the structural data so far cannot unambiguously distinguish between abrupt changes at the junction of conformationally distinct DNA segments and gradual wedge contributions.

Studies based on ligation ladders of synthetic A-tract containing oligonucleotides have also confirmed some of the sequence predictions of the junction model (Crothers, *et al.*, 1990). They have shown, for example, a requirement for three or more Adenine residues in the A repeat before a bending effect is observed (Hagerman, 1985; Koo, *et al.*, 1986). Although base substitutions in the region of B-DNA have only small modulating effects on the electrophoretic mobility of the oligomers, the identity of the bases immediately adjacent to the A-tracts, appeared to be important (Koo, *et al.*, 1986). Maximum bending for A-tracts of a given length is observed when the 5' and 3' positions are occupied by C and T residues repsectively. Moreover, Guanidine residues cannot substitute for Adenines within a bending inducing repeat, suggesting that homopolymeric tracts of the two purines have rather different properties, and bending of the double helix is uniquely associated with precisely phased polyd(A) segments. Experimental evidence exists, however, showing DNA bending induced by G-tracts (Brukner *et al.*, 1993). Although molecules containing phased G-tracts usually display normal electrophoretic

---

this as evidence against both wedge and junction models of DNA bending and suggest that it supports their hypothesis of non-A-tract bending. According to this view crystal structure data should be interpreted in a "dynamic" fashion as implying that A-T steps provide flexible hinges about which DNA can assume a number of conformations including bending towards the major groove, as deduced from crystal structure analysis. The non-A-tract bending model, therefore, denies the existance of stable "static" curves in the DNA, and suggests that sequence variation only imparts various degrees of flexibility on the DNA helix. As correctly pointed out by Haron et al., however, i) the crystallographic data mentioned above can be interpreted also as evidence that the type of crystal lattice determines to a certain extent the confromation of crystallised oligos which therefore do not necessarily reflect their preferrred conformations in solutions and ii) bending of A tracts towards the major groove contraddicts a number of experimental observations. Moreover, the junction and wedge models fully recognise the polymorphic nature of A-tract containing molecules in solution as implied by well documented temperature and solvent effects on the electrophoretic detection of bending. Finally as pointed out by Crothers et al., 1990, the idea of a flexible hinge is rendered obsolete by the requirement for a precise phasing of A-tracts and of other bend-inducing motifs which has been clearly shown experimentally.

behaviour, when G tracts are ligated in phase with bent A-tract containing molecules, the result is elimination of the elctrophoretic abnormality. In contrast, G-tracts ligated exactly out of phase with bending inducing A-tracts, can accentuate the electrophoretic anomaly. These obseravtion strongly suggest that G tracts must be favouring bending in the direction opposite to that of A-tracts. Furthermore, the effect of phased G-tracts on the electrophoretic anomaly of A-tract containing bent molecules, supports the hypothesis that macroscopic distortions of the double helix depend on interactions well beyond the dinucleotide step.

The relative simplicity of the junction model, permits a reinterpretation of the effects of temperature and solvent concentration on the electrophoretic mobility of the *MboI* fragment in terms of their predicted effects on the stability of the B' conformation adopted by A tracts (Crothers, *et al.*, 1990). Temperatures above 23 °C or ethanol concentrations above 5% (v/v) could destabilise the unusual helical structure of B'-DNA, by disrupting the energetically favourable hydration network predicted to be associated with this non-canonical B conformation. Once this chain of water* molecules has been disturbed, the energetically unfavourable configuration of A-T base pairs within A-tracts is converted to a less strained and more B-like conformation which reduces the deflection of the A-tract helical axes and therefore alleviates junction bends.

3.1.3. Biological significance of DNA bending

The bent structure of kinetoplast DNA raised questions regarding its possible biological role. Early speculations, suggested that the bent structure of kDNA minicircles might be essential to ensure correct packaging of the minicircles into the mitochondrial matrix (Marini *et al.*, 1982b). This hypothesis predates the discovery of a coding potential of the minicircles and therefore reflects the purely structural role attributed to whole minicircles that were seen as simple "link" components within the kDNA network. The detailes of the kDNA replication process (Ryan *et al.*, 1988), however, suggested an important functional role for the bent motif of minicircles (Linial and Shlomai, 1987a; Linial and Shlomai, 1987b). During network

---

* Goodsell et al., 1994, have obtained crystallographic data of short oligos that are compatible with the existance of such a hydration spine along one side of the B-form DNA helix.

replication, individual minicircles are released from the network and replicated as monomers. Daughter minicircles reattach to the network periphery and this process is repeated until all minicircles have been replicated. Once the entire network has doubled in mass, it splits into two daughter networks. One problem facing the replication machinery is how to distinguish between newly replicated and unreplicted minicircles. Newly replicated minicircles carry nicks and single stranded gaps, which probably play a role in permitting discrimination of replicated vs unreplicated minicircles. A nicking function purified from the trypanosomatid *Crithidia fasciculata*, has been shown to specifically bind to and nick the bent region of the *C.fasciculata* minicircle *in vitro* (Linial and Shlomai, 1987a). The binding efficiency of the nicking enzyme to the bend in *C. fasciculata* minicircle DNA is independent of the nucleotide sequence but appears to be very sensitive to the degree of bending of the target sequence (Linial and Shlomai, 1987b). Furthermore, the nicked circles are poor substrates for purified *C. fasciculata* topoisomerase, an enzyme believed to be involved in decatenation of unreplicated minicircles from the rest of the network. This experimental evidence suggests that the kDNA bend might represent an important structural feature of kinetoplast minicircles that is specifically recognised by enzymes involved in network DNA replication.

Apart from being a disitinctive feature of Trypanosomatid kinetoplast DNA, naturally occurring bent DNA motifs have been found associated with replication relevant sequences in bacteriophage lambda (Daniels *et al.*, 1980), the yeast *Saccharomyces cerevisiae* (Snyder *et al.*, 1986; Anderson, 1986), Simian virus 40 (SV40) (Ryder *et al.*, 1986), Adenovirus (Anderson, 1986) and in a number of other prokaryotic and eukaryotic replication systems (Hagerman, 1990).

The single origin of replication of bacteriophage lambda (ORI) displays unusual electrophoretic mobility on PA gels and it consists of short A-tracts repeated with the helical turn (Zahn and Blattner, 1985). This region is also the binding site for the phage encoded O replication protein. The O protein has four binding sites, called iterons (numbered *I*-IV), within the ORI sequence and circular permutation studies showed that the putative focus of bending is within iteron IV (Zahn and Blattner, 1985). Ring closure experiments, that measure the frequency with which the intramolecular ends of a DNA molecule become substrates to a DNA ligase, have also shown that binding of the O protein requires A-tracts and accentuates ORI bending (Zahn and Blattner, 1987).

The autonomously replicating region 1 (ARS1) of *S.cerevisiae* seems to rely on a bent helical structure in the vicinity of a conserved replication consensus in order to increase efficiency of replication initiation. The ARS1-associated bent motif is also the binding site for a yeast protein factor and deletion studies have shown this site to be an essential determinant of plasmid stability (Snyder, *et al.*, 1986). Experiments where a synthetic bent DNA sequence could substitute for the ARS1 associated bent motif, demonstrate more conclusively the role of the bend in increasing efficiency of ARS1 recognition by yeast replication enzymes (Williams *et al.*, 1988).

The SV40 bent motif within region *I* of the viral origin of replication, overlaps an essential binding site for the virally encoded large T antigen (Tag) protein (Ryder, *et al.*, 1986). Experimental evidence shows that the single A_6-tract within region *I* is responsible for the intrinsic bend and that base substitutions within the tract visibly decrease the bending as detected on PA gels. The Tag protein does not bind directly the A-tract but rather two Tag moieties, either as monomers or as part of a dimer, bind to the flanking GC rich sequences. On the basis of these findings, it has been suggested that the bending at the A_6-tract may help coordinate the binding of Tag subunits to their target sites in region *I* (Ryder, *et al.*, 1986).

Within the Adenovirus genome, bending is detected exclusively for a 158 bp fragment located in the vicinity of sequences that are essential not only for viral replication but also for transcription and packaging of the viral genome (Anderson, 1986). Interestingly, binding competition and mobility shift electrophoresis analyses have shown that the *C.fasciculata* nicking enzyme is able to bind all the bent motifs described above (Linial and Shlomai, 1988). Given the extent of sequence divergence among the bent motifs used in this study, this finding provides further evidence that the *C.fasciculata* nicking function is truly recognising a DNA structure rather than a specific sequence.

In the case of the *I* factor, the A-rich 5'UTR is a substrate to transcription processes. There is a large amount of evidence supporting a role of bent DNA motifs in transcription processes. The application of computer based analytical methods that permit the detection of patterns in DNA sequences, have uncovered subsets of *Escherichia coli* promoters that carry potentially bend-inducing A-tracts in proximity of conserved promoter motifs (Galas *et al.*, 1985). Programs designed to detect runs of phased A-tracts have identified A-rich sequences in association with strong *E.coli* promoters (Plaskon and Wartell, 1987).

In *Salmonella typhimurium* a region of the $tRNA^{his}$ promoter located about 70 bp upstream of the transcription start site contains phased A tracts and was found to display abnormal electrophoretic mobility (Bossi and Smith, 1984). This region came to attention because a three bp deletion overlapping position -70 resulted in a significant reduction (50%) in transciption from the $tRNA^{his}$ promoter and also relieved the electrophoretic abnormality associated with this DNA sequence. Interestrngly the 3 bp deletion in this upstream region has comparable effects whether the mutant promoter is studied *in vivo* or *in vitro* with purified protein factors. This suggested that the role of the -70 region was not that of binding a trascription factor but rather this sequence might be influencing RNA polymerase binding. Furthermore, substitution of the entire -47 to -70 region with random DNA has the same effect on transcription as the original three bp deletion. This finding suggests that the exact sequence around the -70 region is not essential for transcription from the $tRNA^{his}$ promoter, but rather that it acts as an enhancer, increasing efficiency of transcription. It is interesting to note that the promoter is very strong but is also able to respond rapidly to changes in metabolite concentration. A promoter with comparable properties which has also been found associated with a bent motif, belongs to the human beta-actin gene (Kawamoto *et al.*, 1989). The beta-actin promoter belongs to a class of inducible promoters that respond to external stimuli and become rapidly but transiently transcribed. An A-rich region was identified upstream of the -40 CAAT motif and the bending focus mapped to positions -60 to -68 by circular permutation experiments. Moreover, linker scanning experiments showed that an unidentified protein factor present in cell extracts binds within the bending focus and that disruption of this sequence reduced binding. Thus, as in the case of the *S.typhimurium* $tRNA^{his}$ promoter, the bent motif appears to be associated with an enhancer-like function.

In *Drosophila melanogster,* a region of DNA bending has been identified in the UTR of the retrovirus *gypsy* by hydroxy radical cleavage (Spana and Corces, 1990). This 340 bp region contains 12 conserved motifs with similarity to the mammalian octamer motif which is recognised by homeobox containing proteins. These octamer-like motifs in the *gypsy* retrovirus are bound by the *Drosophila* protein called *suppressor of hairy wing* (su(Hw)). The su(Hw) protein contains 12 zinc finger motifs and binds the 340 bp region in the gypsy UTR with great affinity. Interestingly, each of the octamer-like motifs is flanked by A-tracts in inverse orientation. Hydroxy radical cleavage shows the bending to be associated with these A-rich sequences, and in combination with results from methylation protection

studies, reveals that although not directly bound by the su(Hw) protein, these A rich sequences are important for efficient su(Hw) binding. Mutations that disrupt the A-tracts, reduce su(Hw) binding and relieve *gypsy* induced mutations in vivo.

One feature shared by both transcription and replication processes, that could shed some light as to the role of bent motifs associated with replication origins and promters, is their requirement for unwinding of the DNA double helix. Free energy considerations lead to speculations that in the case of intrinsically bent sequences, the favourable energy associated with the wrapping of the already bent DNA around the (DNA or RNA) polymerase or other protein factor, may render the subsequent unwinding step less favourable (Travers, 1989). This is in contrast to protein binding to unbent region where the energy expended in bending the DNA around the protein, would be converted into tortional stress that renders DNA unwinding energetically favourable. Studies on the effects of synthetic bent DNA sequences on transcription from the *E.coli lac* promoter, do indicate that binding of the RNA polymerase is stimulated up to 10 fold by the presence of artificially introduced bends in the promoter region (Gartenberg and Crothers, 1991). These studies, however, did not look at subsequent transcription steps. Previous work, however, had already shown that synthetic curved sequence can substitute for protein induced bends at the *E.coli gal* operon and, in some cases, give higher promoter activities than wild type (Bracco *et al.*, 1989). Thus, at least in prokaryotes, bent motifs can truly act as transcriptional enhancers, even though this effect is not easily predicted from free energy considerations. Recent evidence suggests that in eukaryotes, the TATA box motif is strongly bent in one direction and that binding of the TBP protein forces the DNA to bend in the opposite direction (Davis *et al.*, 1999). This event should therefore change the energy balance in favour of DNA unwinding, at least in the case of TATA-containing promoters. Furthermore, intrinsically bent sequences placed upstream of simple artificial TATA box contaning promoters, have been shown to potentiate transcription.

Bent DNA motifs are not confined to origins of replication origins and promoters. Using two-dimensional PA gel electrophoresis has permitted the isolation of "libraries" of bent DNA from both *E.coli* and *S.cerevisiae* (Mizuno, 1987; Mizuno and Itoh, 1988). Curved DNA sequences have been found associated with the centromere consensus sequence of all 16 chromosomes of *S.cerevisae* and have been shown, at least in the case of chromosomes 12 and 16, to be important for chromosome segregation (Murhpy *et al.*, 1991; Bechert *et al.*, 1999). Interestingly, the centromeric bent motif showed maximum electrophoretic abnormality at 30° C

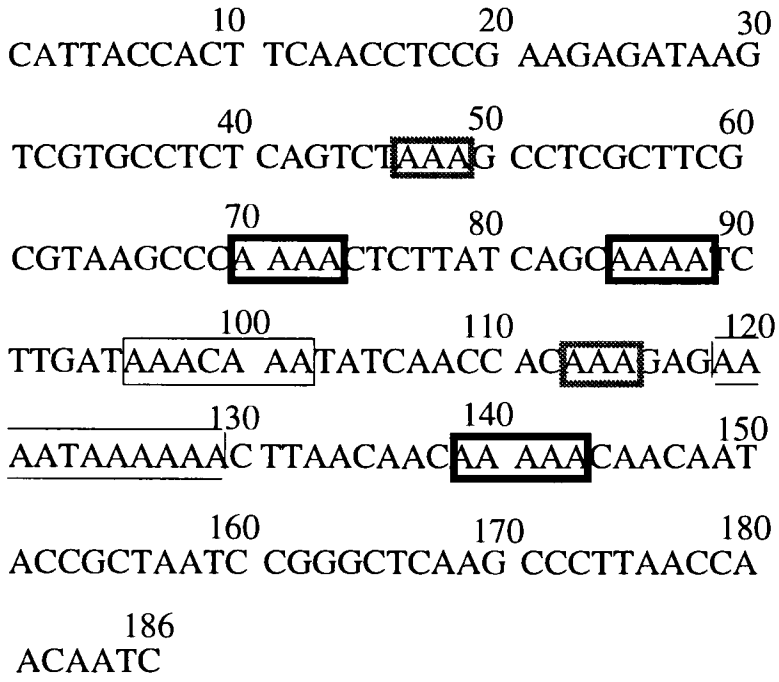
which is also the optimal temperature for growth of yeast (Bechert, *et al.*, 1999). Bent DNA motifs have also been found to be a major component of satellite DNA in a number of organisms (Radic *et al.*, 1987; Pasero *et al.*, 1993; Fitzgerald *et al.*, 1994; Ohyama *et al.*, 1998). The possible role of DNA curvature in mediating chromatin condensation was first suggested by experiments that made use of drugs, such as Distamycin A, that are know to relieve DNA bending: the effect of this drug *in vivo* was to cause satellite heterochromatic decondensation.

Overall the evidence available suggests that sequence induced distortions of the double helix can play an important role in determining protein-DNA interactions. Such interactions can be seen as part of a continuum where one extreme is occupied by sequence specific proteins and the other extreme by proteins that only recognise structural features. Moreover, the *in vitro* sensitivity of DNA bends to a number of environmental factors, such as temperature, ion concentration, hydration etc., suggests a possible role of these DNA motifs as environmental sensors (Gabrielian *et al.*, 1997) with repercussions on vital cellular processes such as replication, gene expression and chromosome segregation.

3.1.4. The I factor 5'UTR contains A-tracts

Structure predictions using the nearest-neighbour models, require relatively complex calculations of rotational parameters between groups of 2, 3 or 4 base pairs. In contrast, the junction model, defines only two main parameters, namely an A rich consensus and its phasing with the helical turn. Structural predictions based on this model can therefore be made simply by visual inspection. A direct sequence comparison was, therefore, carried out between the *I* factor 5' UTR and the kDNA *MboI* fragment. Both sequences are shown in figure 3.1.4.1 and the relevant A-repeats are highlighted.

Sequences on either side of the main bending focus in the kDNA fragment, can be represented as an $(A_{5\text{-}6}N_{6\text{-}5})_n$ repeat. This repeat unit places successive 3' junctions almost exactly every helical turn, which places successive bends on the same side of the double helix. In experiments involving synthetic oligonucleotide concatemers, a repeat unit of the type $(A_5N_5)_n$, has been associated with the largest electrophoretic abnormality when compared to similar repeats that differ only by the number of N



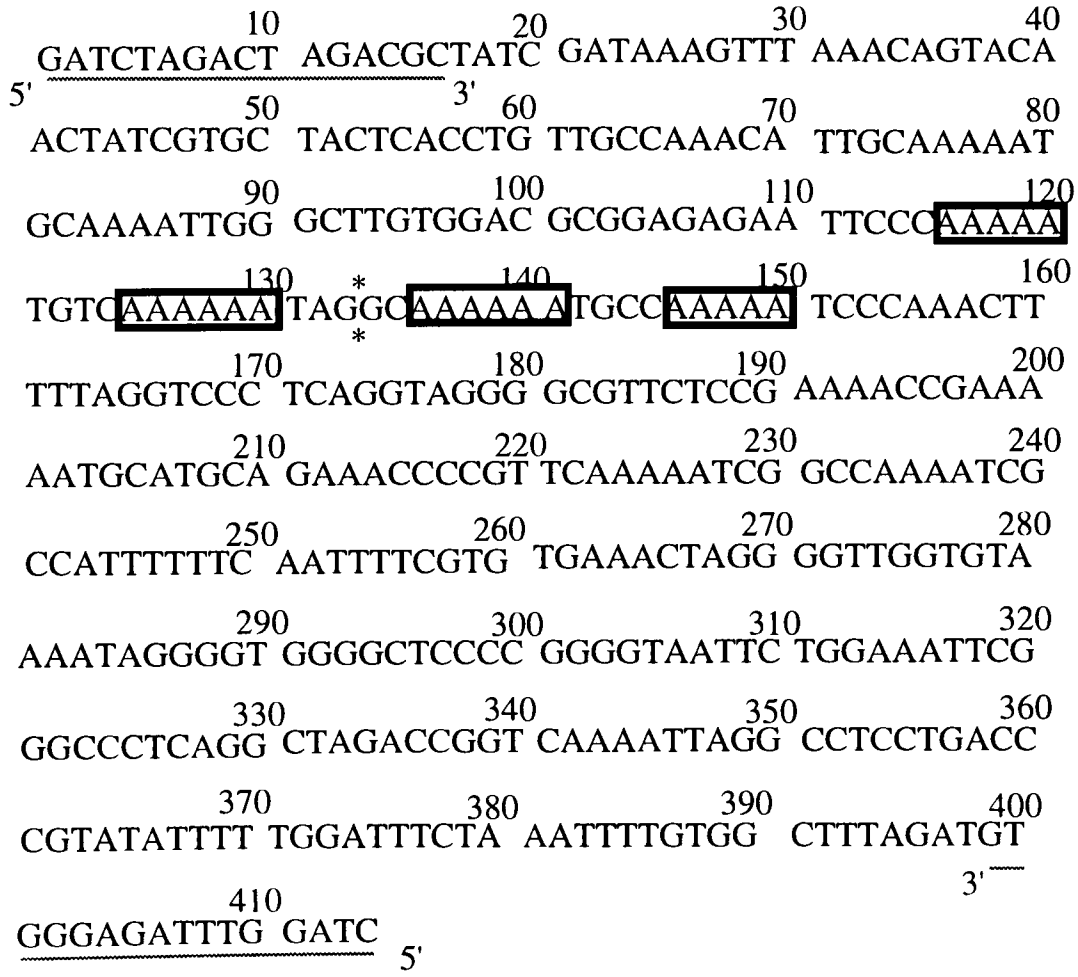


Fig. 3.1.4.1 ; The I factor 5'UTR (Fawcett et al.,1984) is given above. A tracts are framed by thick black lines when continuous, in grey when three residues long and by thin lines when interrupted (interruptions are only considered when three or more A residues flank the interrupting nucleotide). The sequence of the Kinetoplast MboI fragment (Marini et al., 1983) is represented below. The A tracts that compose the CA5-6T repeat are framed in black. The two asterisks mark the focus of bending as determined by Crothers & Wu, 1984. The thin grey lines indicate the primer orientation used to amplify the fragment in the present work.

residues they contain (Koo, *et al.*, 1986). The junction model attributes the electrophoretic behaviour of this type of sequence to its global curvature contributed by the precisely phased bends.

The A-tracts within the *I* factor 5' UTR vary in length and have an apparently irregular distribution. Closer inspection reveals that starting with the A-tract at position 70, five 3' junctions can be identified, whose periodicity can be expressed by the funcion $(A_n+N_n=14)_5$. The phasing of 3' junctions within this unit is of approximately 1.5 helical turns, which places successive bends on opposite sides of the DNA helix. The periodicity of 3' junctions within the *I* factor 5'UTR is thus very similar to that predicted for oligomers with a repeat unit of the type $(A_5N_{10})n$, that have been shown to migrate normally through PA gels (Koo, *et al.*, 1986). The junction model explains the normal electrophoretic behaviour of this type of A-tract containing molecule because the effects of out-of-phase bends cancel out and do not produce overall curvature (Koo, *et al.*, 1986; Crawford *et al.*, 1999). However, the junction bends still act locally to distort the axis at alternating positions along the DNA helix, so that $(A_5N_{10})n$ oligomers are expected to have a "zig-zag" configuration (Koo, *et al.*, 1986).

Theoretical considerations suggest that the helical axis of molecules with out-of-phase distortions, will follow the path of a regular superhelix, the diameter of which is expected to be a few times that of an ordinary DNA rod (Calladine *et al.*, 1988). Electrophoretic theory predicts that DNA molecules in a superhelical configuration will migrate more slowly in high percentage PA gel, than regular rod-shaped molecules. This is only expected to be true, however, if the length of the molecule equals or exceeds one superhelical turn (200 bp). The diameter of molecules shorter than one superhelical turn does not significantly differ from that of the double helix so that no electrophoretic anomaly should be detected. A similar superhelical configuration determined by local out of phase bends is therefore predicted also for the *I* factor 5'UTR. The next steps consisted in applying nearest-neighbour predictions to both the *I* factor 5'UTR and the kDNA fragment as well as determining their electrophoretic mobilities under indetical conditions. The outcomes of the theoretical considerations are compared and discussed in the light of the experimental results.

3.1.5. Prediction algorithms of DNA bending

Despite the apparent monotony of the DNA sequence, long range structure prediction is a non trivial and partly unresolved problem (Goodsell and Dickerson, 1994). Stated simply, the main obstacle to correct structure prediction of DNA, is the fact that many local molecular arrangments can produce identical global properties, and, *vice versa*, a given global conformation can be produced by many different local arrangements. The smaller the structure prediction unit, the more accentuated the problem. This is true of the early dinucleotide based algorithms. Theoretical studies of sequence dependent conformation of B-DNA have demonstrated that the propensity to bending of any of the 16 dinucleotide combinations, far from being limited by the energetically favourable spacial arrangements of the constituent base pairs, actually depends significantly on the larger sequencial context (Poncin *et al.*, 1992; Hunter, 1993). Hence the accuracy of long range structure prediction is greater, the more sequence context information is included. This awariness has made trinucleotide based algorithms the preferred option for structure prediction.

The most user-friendly publicly available program for prediction of sequence dependent DNA curvature is the bend.it program that runs on a WWW server (Gabrielian *et al.*, 1996; Gabrielian, *et al.*, 1997; Munteanu *et al.*, 1998). This program is based on the so called sequence-dependent anisotropic-bendability (SDAB) model (Munteanu, *et al.*, 1998). This model is a sophistication of simple elastic models where DNA is represented as a straight cylindrical rod that is deformable in all directions in a sequence independent manner.

The SDAB representation of DNA is that of a segmented rod, where each segment (a cylinder cross section in three dimensional representation) is a base pair. The SDAB segmented rod, in contrast to the respresentation of simple elastic models, is anisotropically deformable i.e. it has preferred directions of flexibility. This assumption has no experimental basis (Munteanu, *et al.*, 1998) but relies on early theoretical calculations of the potential energy of tetranucleotides (Zhurkin *et al.*, 1979) according to which the DNA double helix may have preferred directions of bending towards the major and minor grooves (what is known as symmetrical anisotropy). The SDAB model simplifies this in terms of asymmetric anisotropy, i.e. the DNA rod will bend preferntially in one direction, that corresponds to the major groove. The unidirectional-bending simplification made by the SDAB model is justified in terms of the source of the trinucleotide bendability parameters that were derived from the statistical determination of cutting rates by the bovine pancreatic

deoxyribonuclease *I* (DNase *I*) enzyme (Brukner, *et al.*, 1995; Munteanu, *et al.*, 1998).

DNase *I* cuts DNA with rates that vary by 100 fold or more along the length of the molecule. Differential cleavage by DNase *I* is not due to sequence specifity but rather to its sensitivity to sequence dependent structural variations in the double helix. The idea of using DNase *I* as a probe of sequence dependent DNA bending, came from the outcome of high resolution studies of DNase *I*-DNA complexes (Brukner, *et al.*, 1995). The crystallographic data revealed the DNase *I* bound to a stretch of 6 base pairs on the surface of the minor groove, whilst the helical axis is bent away from the enzyme towards the major groove. Assuming that sequence-dependent flexibility of a DNA molecule towards the major groove is the main factor determining substrate preference by DNase *I*, it follows that experimental determination of DNase *I* cleavege frequencies for a large set of sequences, should give estimates of the bending propensity of the sequences tested. Such measurements have been obtained for a set of 32 out of the 64 possible trinucleotide combinations (allowing for strand symmetry) and the data was cenverted into a bendability scale where high positive values (expressed in arbitrary units) denote high bendability whilst large negative values denote rigidity. In this context, the term bendability describes the bending propensity of a given sequence and incorporates the dynamic flexibility component and the static curvature component, but cannot distiguish between them (Gabrielian, *et al.*, 1996).

The bend.it program uses two types of bendability scales (matrices), one derived from the DNase *I* parameters, and one that represents the average (consensus) between the DNaseI bendability values for a given trinucleotide and the corresponding values obtained from nucleosome positioning experiments (Gabrielian, *et al.*, 1996). The consensus matrix was devised to render the bendability profiles of trinucleotide sequences independent of the method used to determine them. Both DNase *I* and consensus matrices express the propensity of a given sequence to bend at the position where maximum values are observed, but neither can provide a quantitative measure of curvature (Gabrielian, *et al.*, 1996). The latter requires specification of the helical geometry parameters* that describe the angles of rotational displacement between successive base pairs. X-ray

---

* The parameters used are based on definitions of base pair geometry ratified in the EMBO workshop on DNA curvature and bending (Diekmann, 1989).

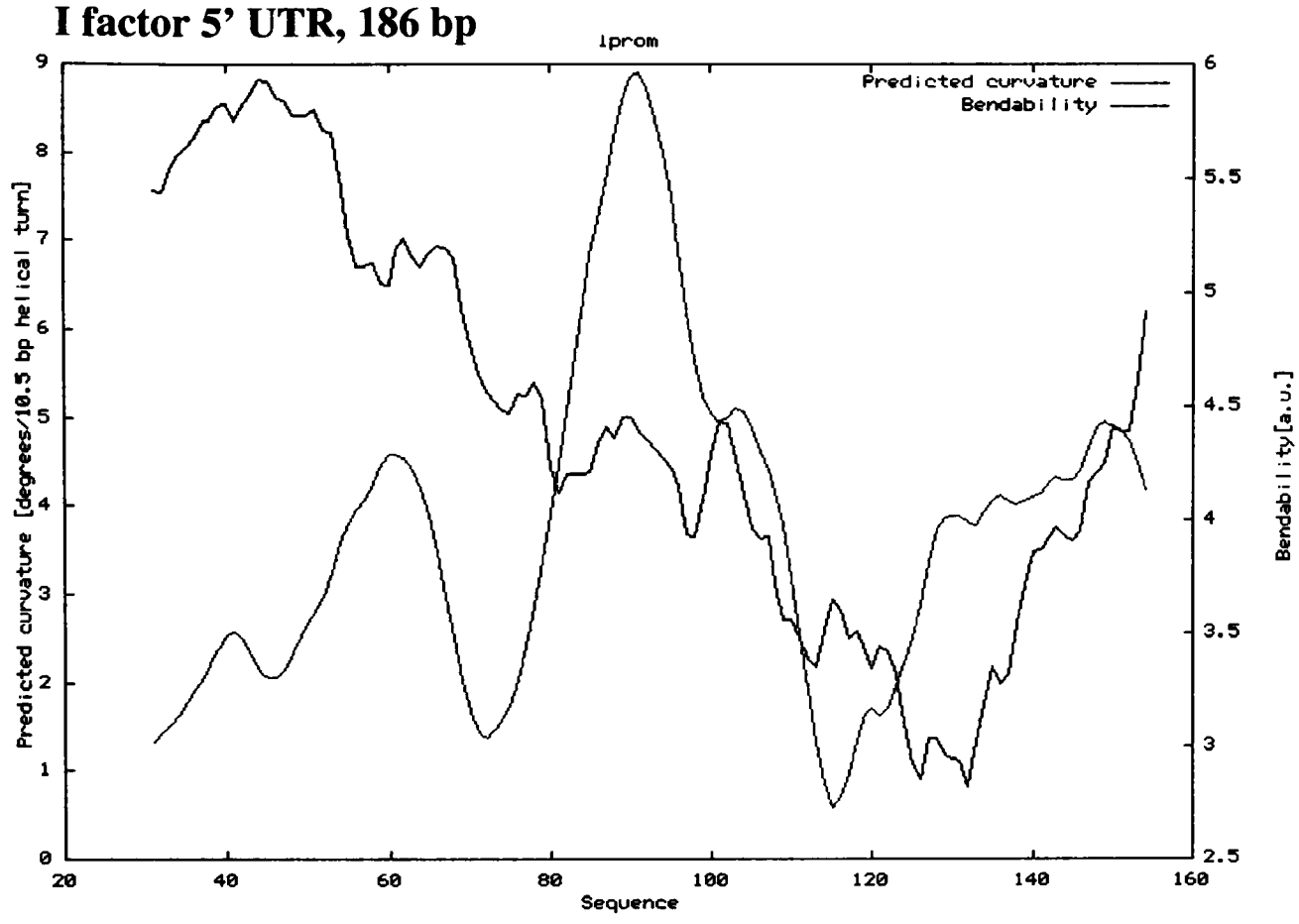
crystalography, Nuclear magnetic resonance (NMR), nucleosome positioning, gel mobility analysis, and theoretical calculations can all provide estimates of the angular amplitudes describing relative displacement of base-pair steps (Gabrielian, *et al.*, 1997). However, when measurements from each source were compared for their ability to discriminate between known bent and straight molecules, the NMR based predictions proved to be the most accurate (Gabrielian, *et al.*, 1996). The reason behind the greater accuracy of NMR helical parameter estimates may be the fact that they represent time averages of relative atomic distances calculated for DNA molecules in solution and are therefore more likely to 'capture' the preferred confomartion adopted by the double helix in solution (Ulyanov and James, 1995).

The bend.it program estimates curvature from the average, over approximately 3 helical turns (10.5 bp / per turn = 31 bp window size), of the deflection angle per helical turn (Gabrielian, *et al.*, 1996). These deflection angles were derived by applying the BEND program (Goodsell and Dickerson, 1994), to helical parameters derived from the matrix data (Gabrielian, *et al.*, 1997). One important simplification implicit in the bend.it program, is that apart from the roll, twist and tilt angles, which, as said before, are specified by the particular matrix used, all other paramenters are assumed to be equal to those of standard B-DNA (Goodsell and Dickerson, 1994; Munteanu, *et al.*, 1998). It was therefore of particular interest to see how accurately this program would predict curvature of a DNA molecule such as the kinetoplast DNA for which departures from standard B-form configuration had been postulated as the main cause of bending (Wu and Crothers, 1984).

Figures 3.1.5-1, 3.1.5-2 and 3.1.5-3 show the curvature profiles of the *I* factor 5' UTR and the kDNA *MboI* fragments against bendability and G+C contents. In figure 3.1.5-1 and 3.1.5-2, the consensus matrix is used to calculate the values of curvature angles, expressed in degrees per helical turn (10.5 bp; Y axis right) for the *I* factor 5'UTR (Iprom; top) and the *MboI* fragment (kDNA; bottom). In figure 3.1.5-1 the curvature angles are plotted against bendability (expressed in arbitrary units;Y axis left) whilst in figure 3.1.5-2 the curvature angles are plotted against G+C content (expressed as a percentage; Y axis left). In figure 3.1.5-3 the DNase *I* matrix is used to calculate curvature angles and these are plotted against bendability and the notations are the same as in the previous two figures. Plots of curvature versus G+C content using the DNase *I* matrix are omitted because the relevant information is already provided from the equivalent plots obtained with the consensus matrix in figure 3.1.5-2. In all cases the X axis indicates the positions along the DNA molecule to which the measurements on the Y axis apply. Because of the window size (31 bp)

Fig3.1.5-1 . bend.it prediction. Matrix: Consensus; curvature versus bendability

I factor 5' UTR, 186 bp



***L.tarentolae* kDNA MboI fragment, 141 bp**

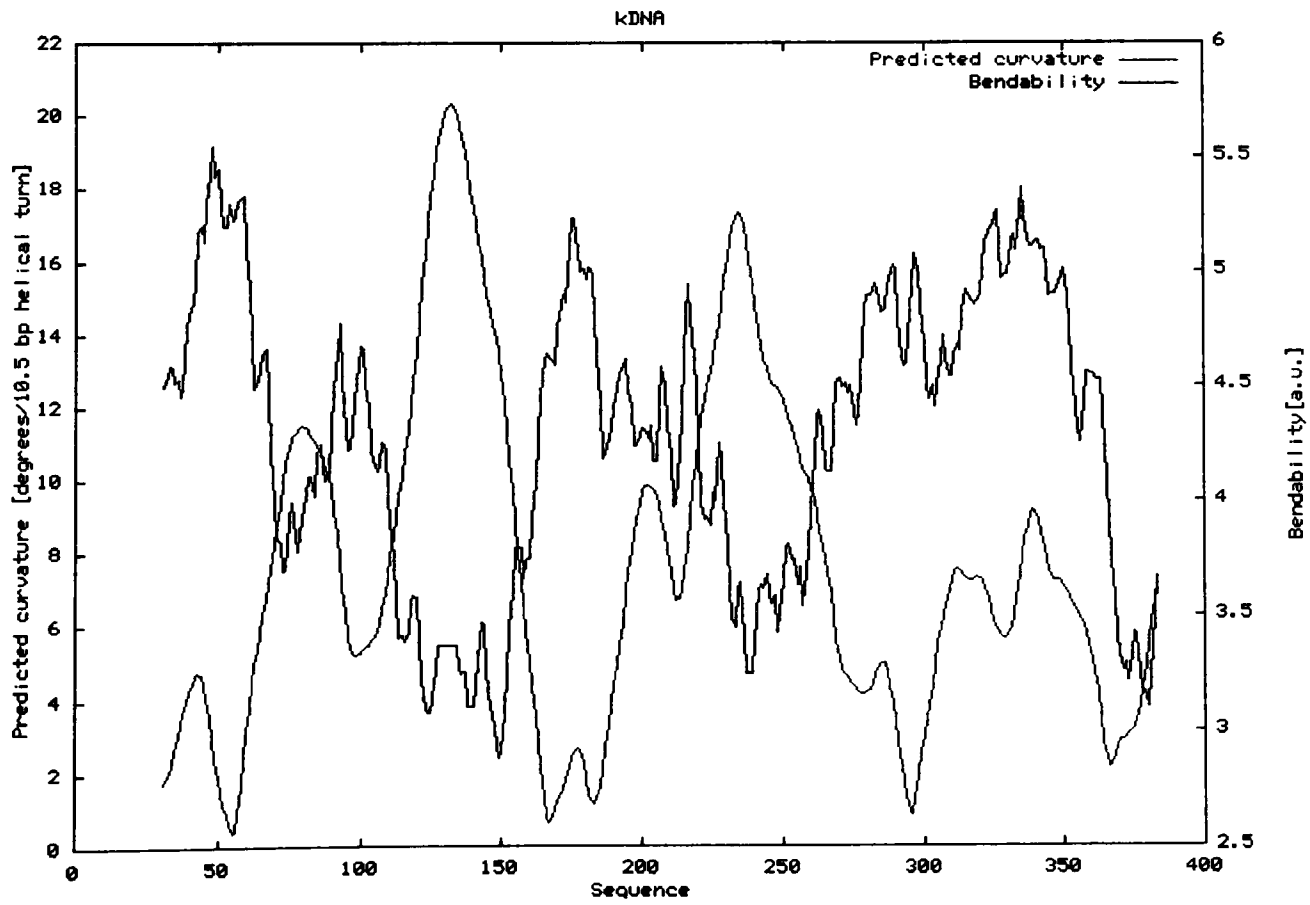
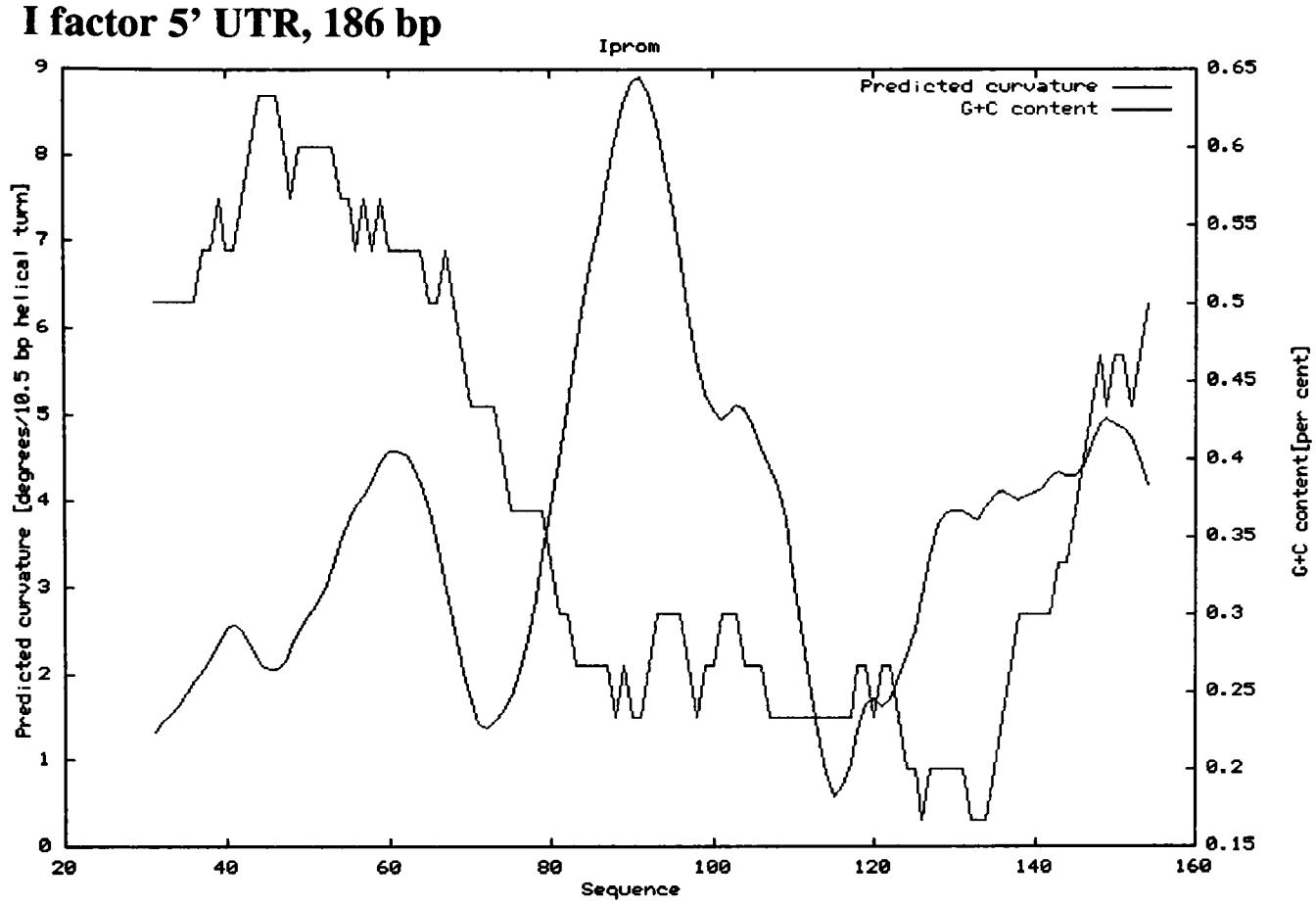


Fig 3.1.5-2. bend.it predictions. Matrix: Consensus. Curvature vs G+C content

I factor 5' UTR, 186 bp



***L.tarentolae* k DNA MboI fragment, 414 bp**

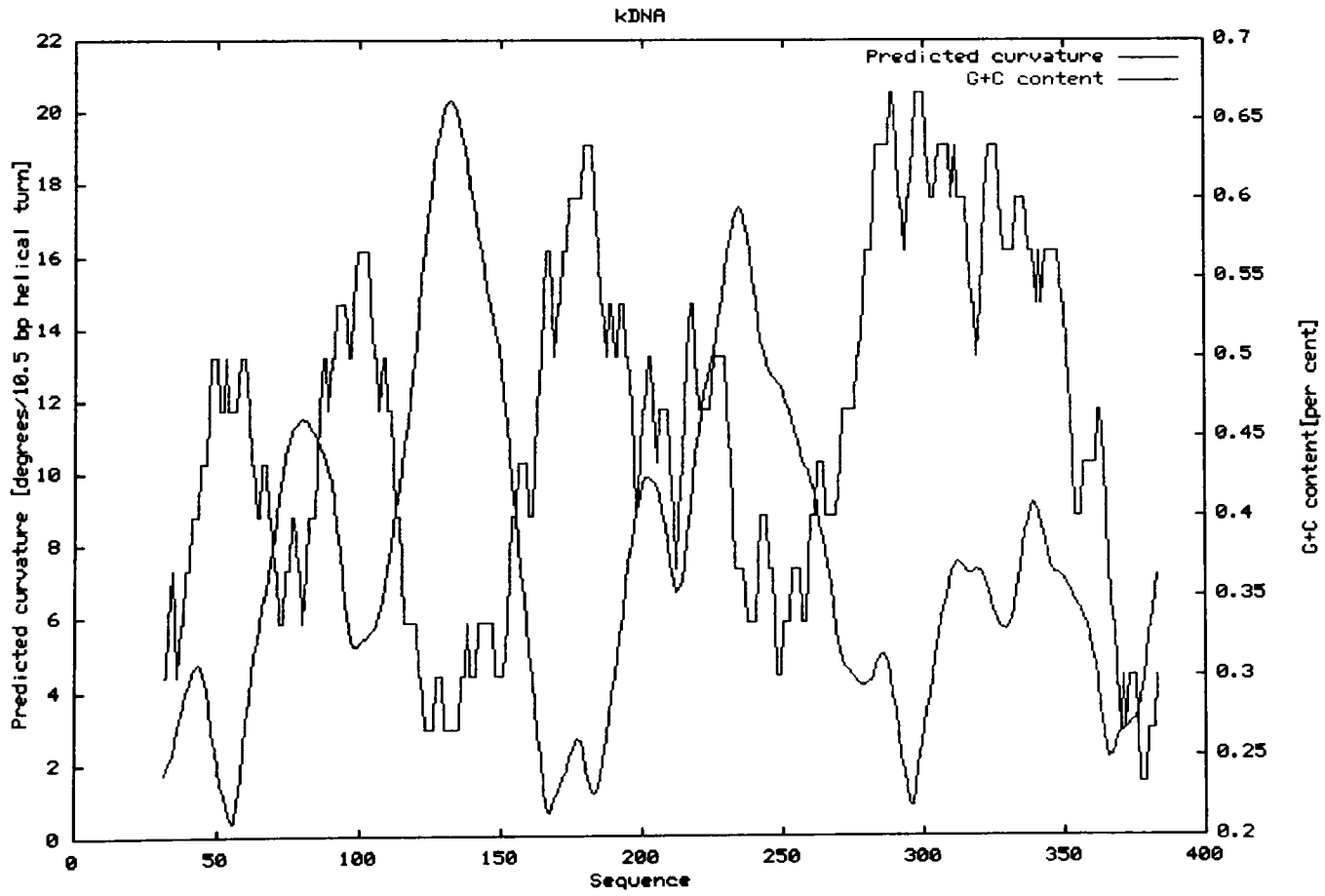
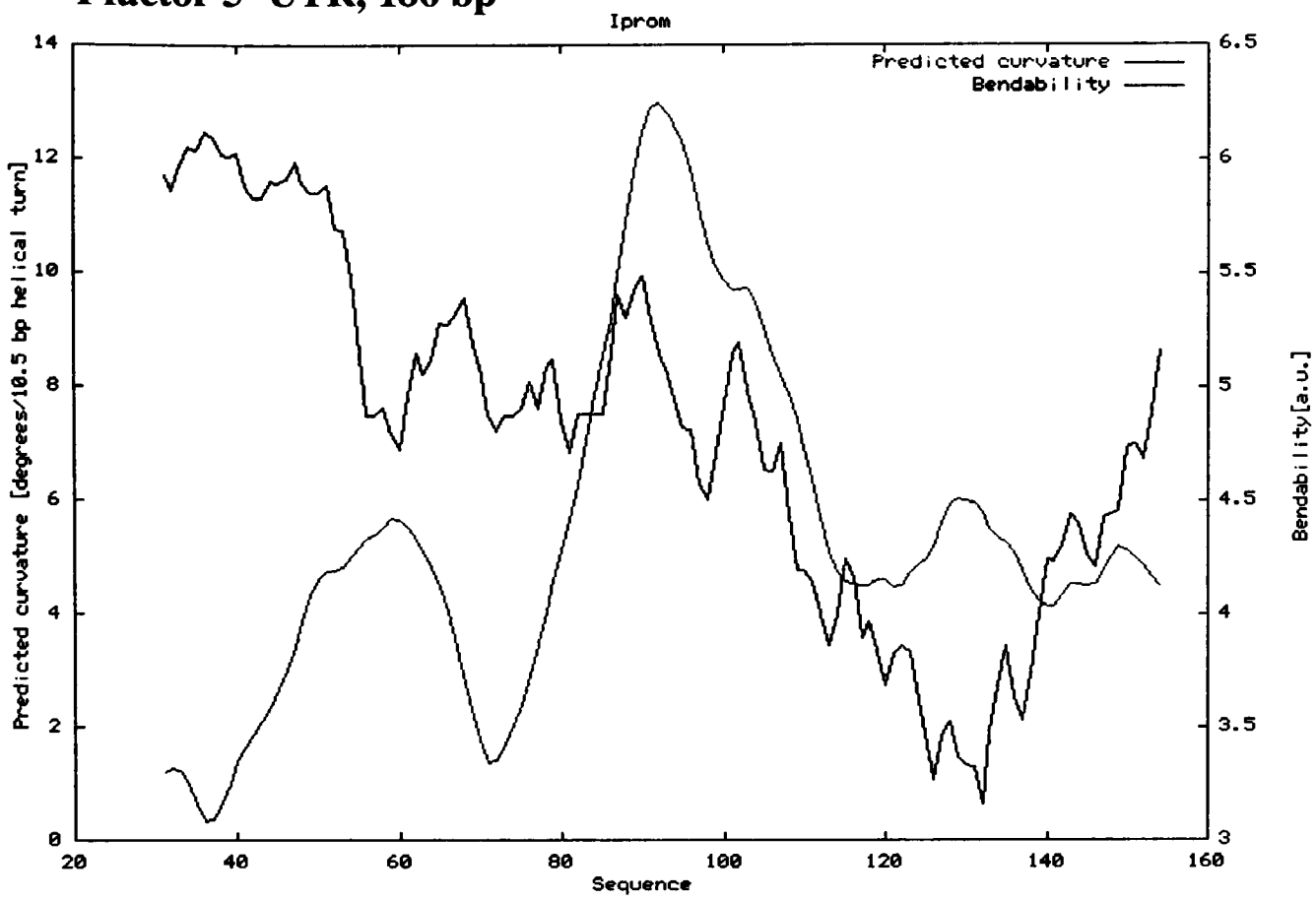
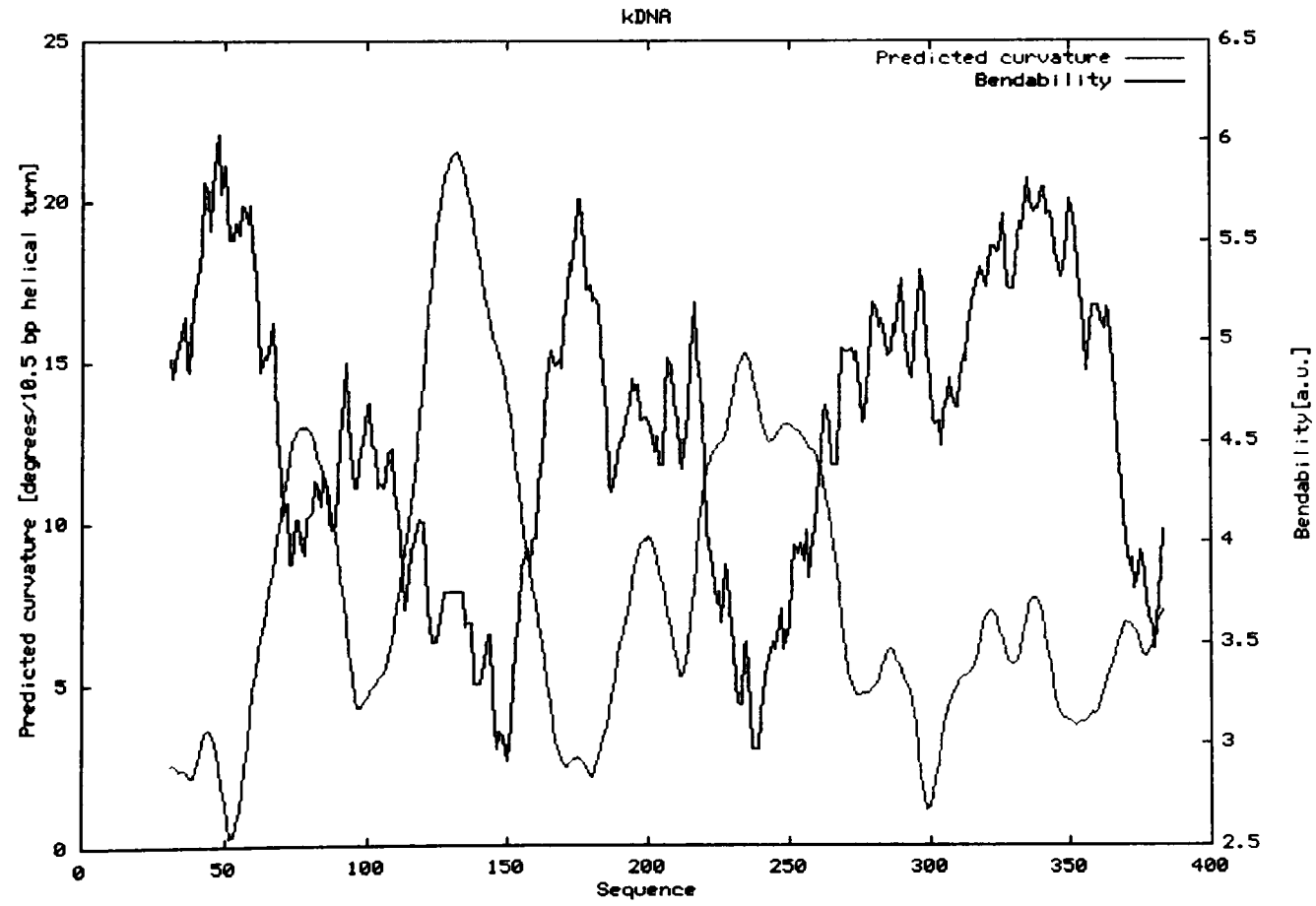


Fig 3.1.5-3. bend.it predictions. Matrix: DNaseI; curvature vs bendability

I factor 5' UTR, 186 bp



***L.tarentolae* kDNA MboI fragment, 414 bp**



for the determination of average curvature angles, the calculations do not include the first and last 31 bp of both sequences.

The curvature versus bendability plots for the kDNA *MboI* fragment show a major peak of curvature with an angular amplitude of approximately 21°. This maximum of curvature falls very close to the 140 bp position that coincides with the bending focus as predicted by circular permutation experiments (Wu and Crothers, 1984). There is also a second peak of potential curvature that falls about 100 bp downstream of the main one and is particularly conspicuous in the plots obtained using the consensus matrix whilst it appears much broader if the DNaseI matrix is used. The approximate angle of this second peak is between 16° and 17°. The existance of a second focus of bending was suggested by Wu & Crothers (1984) as a possible explanation for the unusual shape of the curve of electrophoretic mobility versus position in bp obtained in their circular permutation experiments. The bend.it curvature predictions for the kDNA *MboI* fragment, therefore, are in good agreement with the electrophoretic data.

The *I* factor plots show a single main peak of curvature almost exactly in the centre of the 5'UTR, around position 90 bp. The curvature angle has an amplitude of about 8.9° on consensus plots and 13° on DNaseI plots, both values being considerably smaller than the value representing the angular amplitude of the kDNA main bending focus. The discrepancy between the *I* factor 5'UTR curvature angles predicted using the DNaseI and consensus matrices is rather striking in the light of the fact that there is little disagreement between the DNaseI and consensus estimates of curvature angles for the kDNA fragment. The smaller angle obtained for the *I* factor 5'UTR using the consensus matrix, must result from disagreements between the DNaseI and nucleosome wrapping bendability scales when applied to the *I* factor 5'UTR sequence context. In contrast the two bendability scales must give very similar values for the kDNA sequence.

Another noteworthy feature of the bend.it profiles is the inverse reciprocity between curvature and bendability in each graph. This pattern is all the more conspicuous in the regions of maximum curvature observed in the kDNA graphs. The data, therefore, imply that the most curved segments of DNA are also those of least bendability. This is not a paradox, but rather it is in good agreement with the rigid properties attributed to static bends of the type postulated for curved molecules such as the kDNA fragment. Peaks of curvature and corresponding troughs of bendability have been observed in many bend.it predictions of known bent molecules that have

conspicuous electrophoretic abnormalities (Sondor Pongor, pers. comm.). It may, therefore, be significant that the *I* factor curvature peak is not associated with a trough in bendability and the implications of this observations will be discussed below in the light of the elctrophoretic data.

The graphs showing curvature versus G+C content, indicate that in both the kDNA and the *I* factor 5'UTR the peaks of curvature lie in regions of relatively low G+ C content (25 to 30 %). It is interesting to note that the G+C and the bendability plots are almost superimposable, which supports the idea of a greater general flexibility for DNA regions of mixed sequence that differentiates them from regions whereparticular bases predominate (e.g. A-tracts).

The amplitudes of curvature angles and the bendability values extrapolated from the graphs are summarised in table 3.1.5-I, where ranges of control values from sets of known bent and straight molecules have been included for comparison. As a consequence of the discrepancy remarked above in the *I* factor 5'UTR curvature estimates obtained with the DNaseI and consensus matrices, the *I* factor curvature angle falls either just within the the range specified for the control curved molecules (DNaseI) or just within the range defined for straight molecules (consensus). Although not evident from the values in table 3.1.5-I, the remarkable consistency in the angle of curvature and bendability of the kDNA, estimated using DNaseI or consensus matrices, nor the inconsistencies of the *I* factor 5' UTR, are unprecedented. Discrepancies of angle amplitudes of less than 0.1° or of more than 9° between DNaseI and consensus estimates are observed, with the average DNaseI-consensus discrepancy being of 2.3° for curved molecules and 2.6° for straight molecules. Moreover, the discrepancy is bidirectional in both the curved and straight molecule sets, i.e. it may either reflect higher DNaseI angle amplitudes vs lower consensus angle amplitudes, as observed in the case of the *I* factor 5"UTR, or it may result from the inverse relationship.

Three synthetic bent oligos included in the curved ranges, have curvature vs bendability values very similar to those of the *I* factor 5'UTR (~13° and ~5.1 a.u. respectively), the main difference being that the DNaseI-consensus discrepancy is considerably smaller than that of the *I* factor 5'UTR (the highest discrepancy value for the three sequences considered is of 3°). Taken independently of the curvature angles, however, the bendability of the *I* factor is one of the highest amongst those recorded for either curved or straight molecules. Finally, it should be noted that the inverse correlation between curvature and bendability breaks down completely for

Table 3.1.5-I

	MATRICES			
	DNase I		Consensus	
DNA sequences	Curvature^	Bendability ~	Curvature	Bendability
Iprom	13°	5.3	8.9°	4.4
kDNA*	21° (16°)	3.7 (3.5)	21° (17°)	3.4 (3.6)
Curved **	13.3° - 27.6° (21.5°)	2.2 - 5.0 (3.9)	10.8° - 20.5° (20.5°)	2.3 - 4.7 (3.5)
Straight***	0.002° - 12.8°	0.1 - 5.1	0.002° - 9.6°	0.06 - 4.6

^ Curvature is measured in degrees per helical turn

~ Bendability is measured in arbitrary units (a.u.) The Iprom and kDNA bendability values are extrapolated directly by determining where a straight line drawn from the point of maximum curvature meets the bendability curve.

* The values in parenthesis indicate a second possible focus of bending at position 240 bp of the MboI fragment, the existance of which was suggested by Wu & Crothers, 1984, to explain the unusual curve obtained for the MboI fragment when plotting mobility versus position in circular permutation experiments.

** and *** refer to a set of11 and 7 curved and straight DNA molecules respectively. These estimates were obtained by the authors (Gabrielan, A., Vlahovicek, K., Pongor, S., 1998 www2.icgeb.trieste.it) using the bend.it program and only the extreme values are shown here. The ** range include also the kDNA values for the main bending focus obtained by the authors and these are shown in parenthesis immediately underneath the range values. Note that the curvature and bendability ranges although referred to the same set of molecules, do not imply any strict correspondence between extreme bendability and curvature values, e.g. the molecule with curvature 13.3° has a bendability of 5.0 and not of 2.2. The only exception are the lowest values observed for the *** ranges which do refer to the same molecule, the A homopolymer.

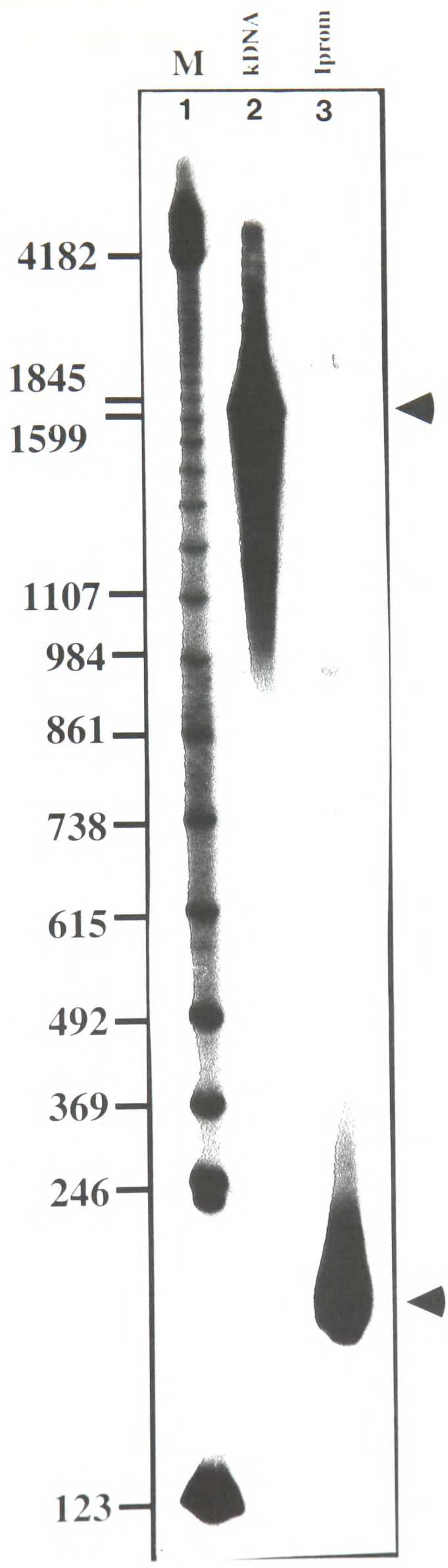
the most extreme cases, such as the rigid unbent A-homopolymer, where to extremely small (yet measurable) angles of curvature (0.002°) correspond extremely low bendabilities (0.1 and 0.06 a.u.) (see table 3.1.5-I).

3.1.6. Detection of DNA bending by polyacrylamide (PA) gel electrophoresis

In order to detect whether the A-tracts within the *I* factor 5' UTR cause a helical distortion detectable as electrophoretic abnormality, this sequence was subjected to electrophoretic analysis on a non-denaturing high percentage PA gel. The *MboI* fragment from the *L.tarentoalae* kinetoplast DNA was included to test whether the conditions used were suitable for detection of DNA bending. The test fragments and the size marker, prepared and labeled as described in Chapter 2, were loaded on a native 12 % PA gel. The results are shown in figure 3.1.6-1. Lane 1 contains the size marker, whilst the PCR amplified *MboI* fragment is in lane 2, and the PCR amplified *I* factor 5'UTR is in lane 3. The *I* factor 5'UTR migrates well within the limits of its size in bp as judged from comparison with the molecular weight marker. In contrast, the kDNA fragment displays a conspicuous electrophoretic abnormality: its size as estimated from the molecular weight marker correspond to that of a fragment approximately 1800 bp long. This is considerably larger than the values reported in the literature and obtained under comparable conditions.

Marini *et al.*, 1984 found that on a 12% gel at 23°C the *MboI* has an apparent size of 1380 bp as measured against *Hae III* fragments of phiX174 or *AvaII* fragments of SV40 DNA. It is unlikely that the difference in the markers used can account for the discrepancy in the apparent size of the *MboI* fragment observed here and that previously reported (Marini, *et al.*, 1982a). It is more plausible that this discrepancy is caused by differences in the gel porosity due to the chemical composition of the gel. The acrylamide to bisacrylamide ratio used in the present experiment was 19:1. This means approximately 1/3 less acrylamide to bisacrylamide than used by Marini *et al*, 1982, to determine the electrophoretic mobility of the *MboI* fragment (29:1 acrylamide to bisacrylamide). The porosity of acrylamide gels depends on both the length of the polyacrylamide chains and the degree of cross linking between them: the first is determined by the percentage of acrylamide whereas the second depends on the relative concentration of the divalent agent bisacrylamide which actually forms the crosslinks (Sambrook *et al.*, 1989). Thus for any given acrylamide percentage, a decrease of the proportion of acrylamide relative to bisacrylamide results in formation of more cross links per acrylamide chain and, therefore, in a decrease of gel porosity. The latter may, thus, be responsible for the observed enhancement of the electrophoretic abnormality of the *MboI* fragment.

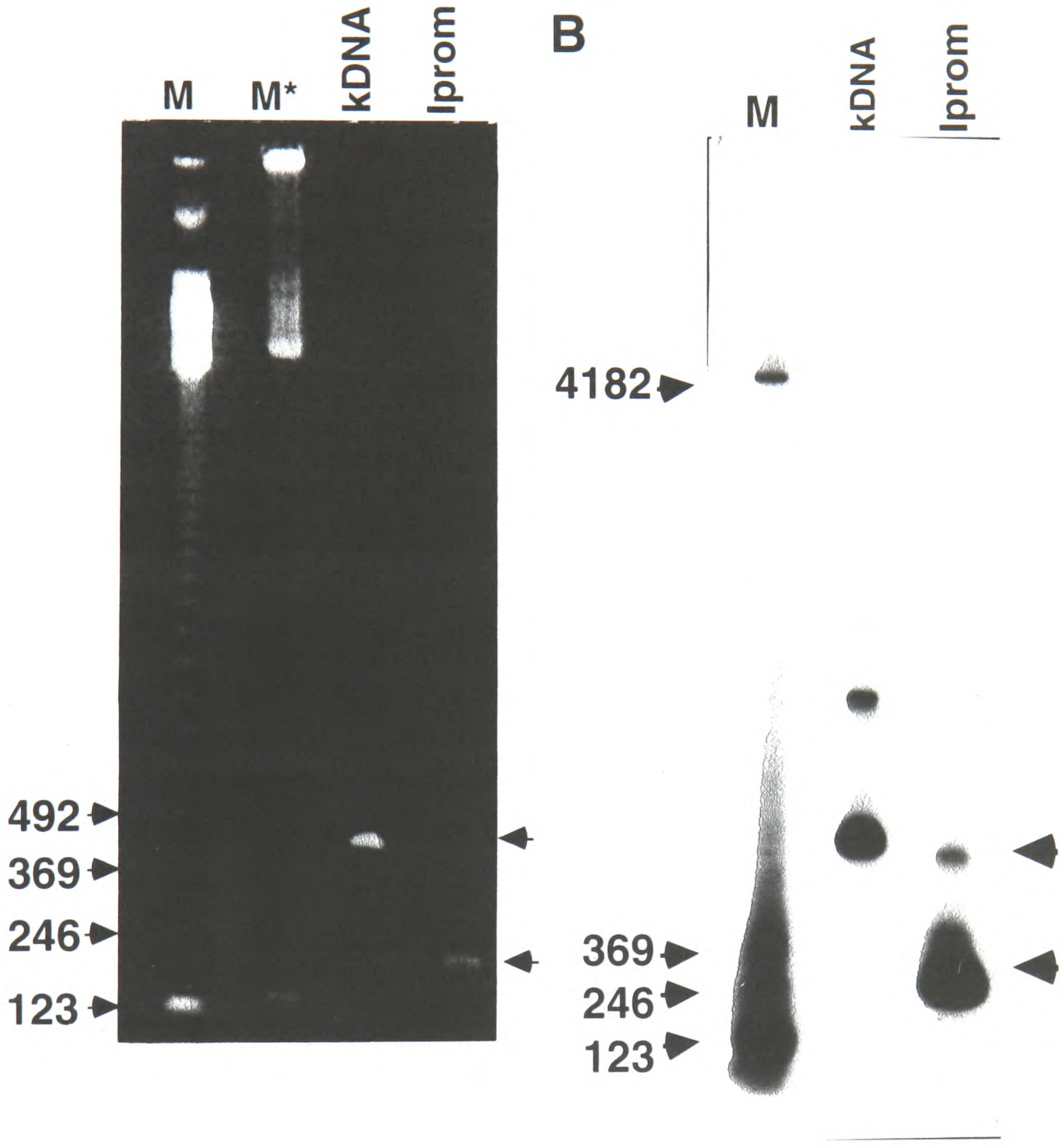
Figure 3.1.6-1 : Analysis of the I factor 5' UTR on a non-denaturing 12% polyacrylamide gel. Lane 1-M: Molecular weight marker, mouse 123 bp ladder; lane 2- kDNA: "bent" kinetoplast DNA from *Leishmania tarentolae;* lane 3- Iprom: I factor 5' UTR.


M
kDNA
Iprom
1
2
3
4182
1845
1599
1107
984
861
738
615
492
369
246
123

In order to test whether the enhanced electrophoretic anomaly of the *MboI* fragment is due to the particular gel composition, samples from the same isotope-labeled preparations used for PA gel electrophoresis were run on a 1.5% agarose gel and then subjected to Southern transfer (gel B figure 3.1.6.-2). In addition, samples from unlabeled preparations of the same fragments, were electrophoresed on a 1.5% agarose gel under conditions identical to these of the labeled fragments. After electrophoresis, the gel was stained with ethidium bromide and visualised under a UV lamp (gel A figure 3.1.6-2). The ethidium bromide was not added to the samples during electrophoresis, because this intercalating agent has been shown to completely relieve electrophoretic abnormalities attributed to bending at standard concentrations (1µg/ml) (Diekmann and Lilley, 1987).

Under electrophoretic conditions similar to those described above, Marini *et al.*, 1994, report an apparent size of 450 bp for the *MboI* fragment. Although transfer of the labeled marker was not very efficient, figure 3.1.6-2 B shows that the *MboI* fragment migrates as a molecule of apparent size 450 bp which is in agreement with the size estimates reported in the literature. This can be seen more clearly on the ethidium bromide stained gel (gel A). The apparent size of the *I* factor 5'UTR on agarose gel appears unchanged from estimates obtained on PA gels and remains very close its actual size in base pairs. Thus the results obtained so far suggest that under conditions suitable for the detection of bending, the *I* factor 5'UTR shows no detectable electrophoretic abnormality.

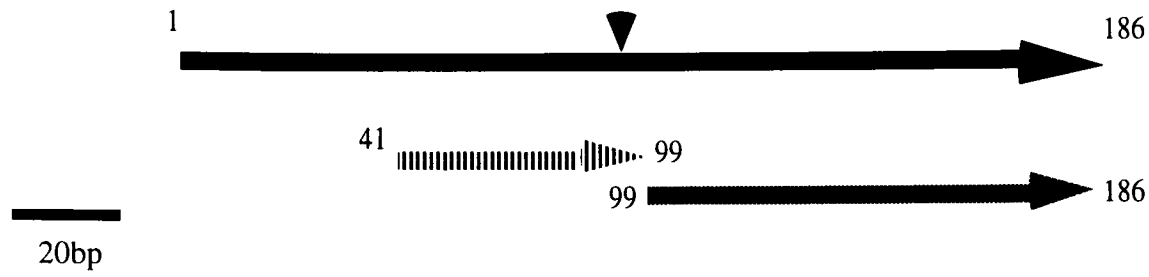
Figure 3.1.6-2: Analysis of the I factor 5' UTR on an Ethidium Bromide stained 1.5 % agarose gel. A:Lane M: molecular weight marker: mouse 123 bp ladder; M*: as lane M except that marker was passed through purification column, to show loss of intermediate bands; lane kDNA: kinetoplast DNA fragment; lane Iprom: I factor 5' UTR. B: Gel run under the same conditions as in A except that DNA fragments in each sample were end labelled using d ATP-P-^{32}P and T4 polynucleotide kinase. Sample M* was not loaded on gel B. Only relevant marker sizes are indicated.


A
M
M*
kDNA
Iprom
492
369
246
123
B
M
kDNA
Iprom
4182
369
246
123

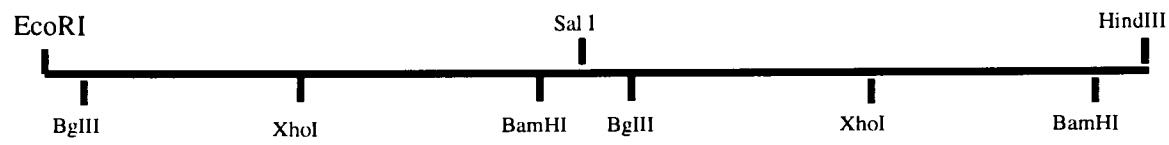
The helical distortions associated with the *I* factor 5' UTR might, however, be too mild for detection even under conditions that considerably enhance the electrophoretic abnormality of a known bent molecule. Experimental evidence has shown that addition of normally migrating arms to both ends of a curved fragment can considerably accentuate the migration anomaly (Diekmann and Wang, 1985). Moreover, although the region of maximum curvature is almost exactly in the centre of the *I* factor 5'UTR, which is an ideal location for electrophoretic detection (Wu and Crothers, 1984), it is plausible that features of the flanking sequences may attenuate the effects of bending to an extent that they become undetectable (Brukner, *et al.*, 1993; Ohyama, *et al.*, 1998). In particular, Ohyama et al., 1998, isolated fragments of bovine satellite DNA that migrate unusually rapidly through PA gels. This rapid migration property was believed to be responsible for concealing the slow migration properties of flanking bent satellite DNA. To investigate these possibilities, the entire sequence of the *I* factor 5'UTR as well as two subfragments were cloned in the pBend2 vector (Kim *et al.*, 1989), a plasmid especially designed for detection of both intrinsic and protein induced bending. Figure 3.1.6.-3 A shows the a map of the *I* factor 5'UTR and the subfragmets used in this experiment. A simplified map of the the pBend 2 polylinker, where only the relevant restriction sites are indicated, is shown in part B. The restriction products containing the *I* factor derived sequences used for the electrophoretic analysis are given together with their names andsizes in bp in part C. The *I* factor 5'UTR and the subfragments were cloned in to the unique SalI site of the pBend vector from PCR products (Chapter 2).

The pBend 2 vector polylinker has been designed to have restriction sites arranged in equidistant pairs, permitting the generation of molecules of identical length that are circularly permutated. The unbent polylinker arms can, on the one hand, enhance the electrophoretic detection of bending, presumably by amplifying the effects of a centrally located bend, and, on the other hand, can provide a different sequence context, that permits to test the effect of flanking sequences on the electrophoretic detection of bending. The restriction products in figure 3.1.6-3 C, were run on a 12% non-denaturing PA gel under conditions identical to those described above. The results are shown in figure 3.1.6.-4.

A: I factor 5'UTR and subfragments

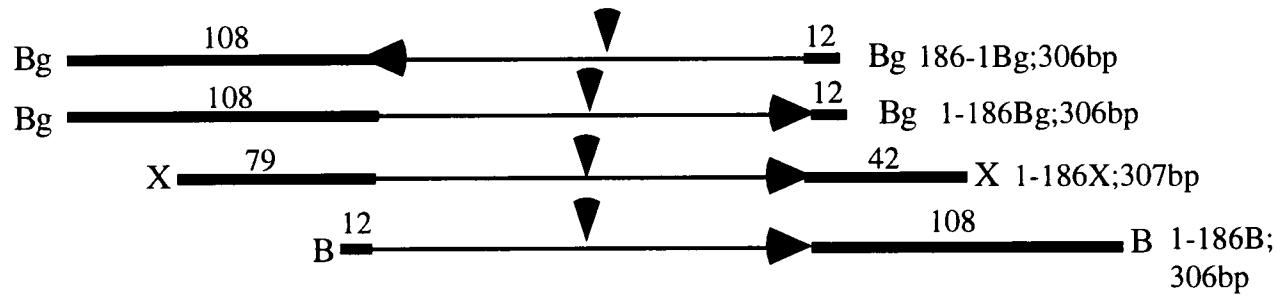


B: Relevant restriction sites in pBend vector polylinker

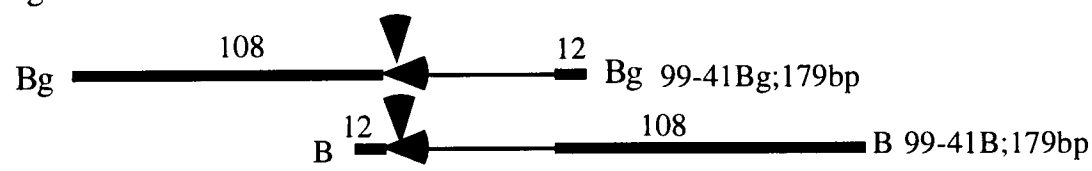


C: Restriction products used for electrophoretic analysis

I factor 5'UTR 1-186



41-99 subfragment



99-186 subfragment

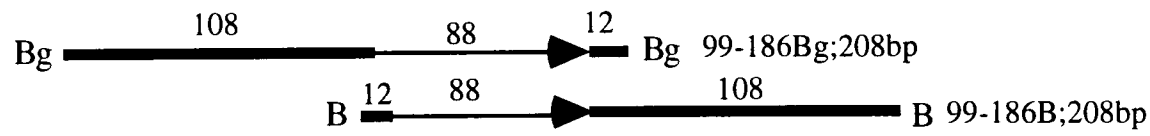
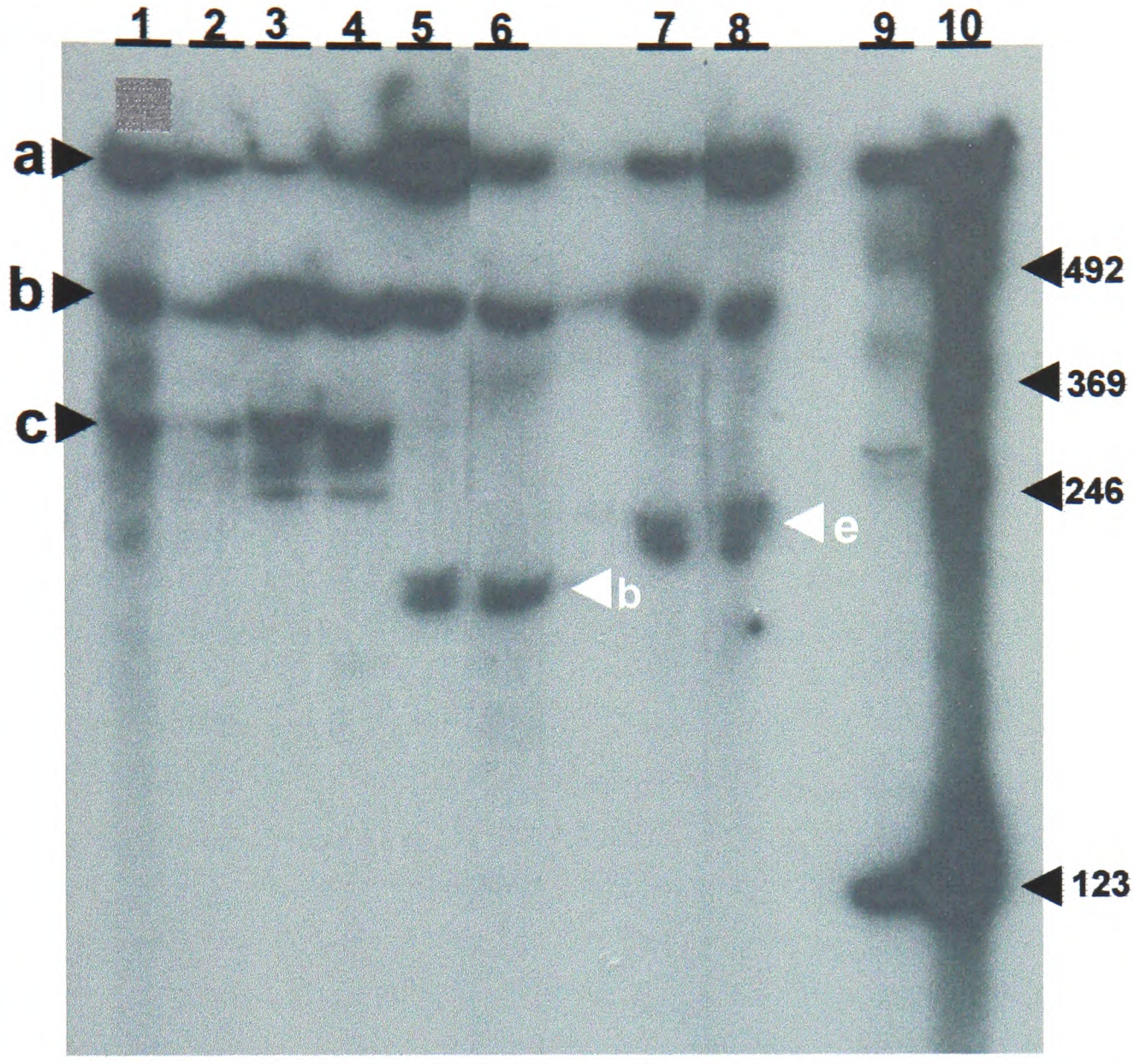


Figure 3.1.6-3. A.: The I factor 5'UTR nucleotides 1-186 and the relative position and orientation of the subfragments used in this experiment. The position of the putative focus of bending as predicted by the bend.it program is indicated by a downward pointing arrow head, almost exactly in the centre of the I factor 5'UTR. B.: A simplified restriction map of the pBend 2 vector polylinker. Unique restriction sites and the paired sites used to release the fragments for electrophoretic analysis are shown above and below the horizontal line representing the pBend 2 polylinker respectively. All fragments from the I factor 5' UTR were cloned in the unique SalI site. C.: The restriction products from the pBend2 vector polylinker (thick lines) containing the I factor derived fragments (thin lines; the arrowheads indicate relative orientation). The putative focus of bending is indicated by a downward pointing arrow head. Sizes in bp are given above each fragment and omitted when uncertainties arise because the exact location of the predicted bending focus is not known. Total sizes in bp of each restriction product and their names are indicated to the right. The restriction enzyme abbreviations are: B=BamHI; Bg= BglII; X= XhoI.

Figure 3.1.6-3; Analysis of the I factor 5' UTR and derivative fragments on a non-denaturing 12% polyacrylamide gel. The I factor 5' UTR and PCR amplified fragments were cloned into the pBend 2 vector and then digested with appropriate enzymes in order to give circularly permutated molecules. The restriction products were end labelled using dATP-P- ^{32}P and T4 polynucleotide kinase. Samples are as follows: Lane 1: 186-1 Bg; lane 2; 1-186 Bg; lane 3: 1-186 X; lane 4: 1-186 B, all these fragments have expected sizes between 306 and 307 bp; lane 5: 99-41 Bg; lane 6: 99-41 B, these fragments have expected sizes of 179bp; lane7 :99-186 Bg; lane 8: 99-186B, these fragments have expected sizes of 208 bp. Lanes 9 and 10 : molecular weight marker: mouse 123 bp ladder, relevant sizes in bp are indicated. Arrow heads are as follows: a: loading wells; b: vector backbone; c : restriction fragments from the 1-186 pBend 2 clones; d: restriction fragments from the 99-41 pBend2 clones; e: restriction fragments from the 99-186 pBend2 clones.


1
2
3
4
5
6
7
8
9
10
a
b
c
492
369
246
123
e
b

Two independent clones carrying the full length *I* factor 5'UTR in opposite orientations were isolated and used in this study. Assuming no effect on electrophoretic mobility attributable to the polylinker sequences of the pBend 2 vector, the186-1Bg and 1-186Bg products should have identical mobilities. Moreover, if there are no helical distortions associated with the ends of the 5'UTR, which escaped detection because of the size of the curvature prediction window used by the bend.it program, the 1-186Bg product should have a mobility indistinguisheable from that of the 186-1Bg and 1-186B fragments. The 1-186X fragment adds 79 bp and 42 bp arms to the left and right sides respectively of the *I* factor 5'UTR. If the presence of these additional sequences enhances the effects of the previously undetectable putative bend present at position 90 bp of the *I* factor 5'UTR, then the 1-186X fragment should be the slowest migrating of this set. As it can be seen in the figure 3.1.6-4, lanes 1 to 4, all four restriction fragments migrate with identical or nearly identical mobilities as expected for unbent molecules.

From extrapolation of the relative position of the *I* factor 5' UTR bending focus, the 99-41 subfragment should contain this region of helical distortion very close to its 3' end. However, since it was not possible to determine the exact position of the putative bending focus, it is uncertain whether or not the 99-186 subfragment carries this region or part of it close to its 5' end. In both sets of restriction products, the Bg fragments are expected to be the slowest migrating. The B fragments, are expected to show normal mobility unless their removal from the longer sequence context of the entire 5'UTR, allows detection of previously unuspected bends. As it can be seen from figure 3.1.6-4, lanes 5 to 6 and to 7 to 8, the fragments within each set display identical mobilities. Moreover, size estimates for all three sets of restriction products, determined by compariosn with the molecular weight marker in lanes 9 and 10, are in good agreement with their known sizes in bp. Thus it can be concluded that under conditions suitable for electrophoretic detection of DNA bending, the *I* factor 5'UTR shows no electrophoretic abnormality attributable to the presence of a bend. Moreover, two subfragments overlapping or adjacent to a region of predicted bending show no detectable electrophoretic abnormality.

3.1.7. Discussion

The *I* factor 5'UTR has been shown experimentally to contain the basic promoter elements (Inr and DPE) between nucleotides 1 and 40 and a tissue specific enhancer between nucleotides 41 and 186 (Chapter 1 section 1.3.3). This sequence has a very

high A-T content (70 to 80%) in a region comprised between nucleotides 80 and 140, which overlaps at least in part, the putative tissue specific enhancer. Visual inspection of the *I* factor 5' UTR reveals the organisation of the A rich sequences into variable length repeats. A-rich sequences have been found associated with unusual double helix conformations that affect protein binding and result in abnormal electrophoretic behaviour on high percentage PA gels (see sections 3.1.1 and 3.1.4).

Electrophoretic anaylsis of the *I* factor 5' UTR failed to detect any migration anomaly in this sequence. Under identical conditions, a migration anomaly was detected for the intrinsically bent control *MboI* fragment from *L. tarentolae* minicircle DNA. The apparent size of the *MboI* fragment was approximately 500 bp larger than that previously reported (Marini, *et al.*, 1984). This enhancement of electrophoretic anomaly was attributed to the lower porosity of the PA gel used in the present experiment. The *I* factor 5' UTR showed normal electrophoretic behaviour also when circularly permutated molecules containing this sequence were analysed. Two subfragments from the 5'UTR which overlapped or contained sequences immediately flanking the predicted focus of bending (see below) were also analysed as circularly permutated molecules and found to display normal electrophoretic behaviour. It was therefore concluded that the A rich sequence found in the *I* factor 5' UTR does not confer upon this sequence helical distortions that are detectable by electrophoretic analysis. Furthermore it was possible to exclude that the properties of sequences immediately 3' of the region predicted to have maximum curvature had properties that attenuate the effects of existing bends.

According to the junction model, the length of the A-tracts and of the spacer region in the *I* factor 5' UTR combine in such a way as to place successive bend inducing 3' junctions on opposite sides of the DNA helix. The junction model predicts that this configuration should not result in global curvature of this sequence. The staggered bends on opposite sides of the double helix, however, are expected to introduce 'kinks' that should favour a superhelical configuration in solution. Superhelicity can result in altered electrophoretic mobilities only if the DNA molecule is long enough to form at least one superhelical turn. This critical length depends on the properties of the particular sequence and has to be determined experimentally. If the junction model correctly predicts the conformation of the *I* factor 5'UTR then its normal electrophoretic behaviour must result because this molecule is shorter than one superhelical turn and therefore it migrates through the gel mesh as a rod with the diameter of the double helix and should not experience major frictional forces.

Nearest neighbour predictions using the bend-it program place a region of curvature almost exactly in the centre of the *I* factor 5' UTR at around position 90bp with an angular amplitude of between 8.9° and 13°. Comparison with values obtained for known bent and unbent molecules suggests that the bend in the *I* factor 5' UTR is relatively shallow although potentially still within a range of predicted curvatures that are detectable by PA electrophoresis. However, most bent molecules display also low bendability values in correspondence of their regions of high curvature which is in agreement with the relative rigidity of static curvature. The predicted curvature for the *I* factor 5' UTR is associated with relatively high bendability with values between 5.3 and 4.4 a.u.. Molecules with curvature vs bendability profiles similar to those of the *I* factor 5'UTR, display low migration rates on polyacrylamide gels typical of bent molecules. Thus, high bendability in correspondence of regions of maximum curvature is not distinctive of either bent or unbent molecules.

The only significant difference between the *I* factor 5'UTR and molecules with comparable curvature-bendability profiles is the conspicuous discrepancy between curvature amplitudes estimated using DNaseI and those obtained with the consensus bendability scale. Although large discrepancies are observed also in the case of known bent molecules, they are associated with larger curvature angles and correspondingly lower bendability values. Thus, the bend.it program does not clearly place the *I* factor 5'UTR in either the bent or unbent category, whereas the electrophoretic data strongly suggest that the *I* factor 5' UTR behaves as an unbent molecule. It therefore appears that for molecules with similar curvature vs bendability profiles, the critical factor that permits discrimination between bent and unbent conformations may be the extent of agreement between predictions obtained with DNaseI and consensus bendability scales.

The bend.it program predicts a relatively high bendability value for the central region of the *I* factor 5' UTR. Bendability describes a propensity of a given sequence to bending and therefore makes no predictions as to whether this propensity can, under certain conditions, result in static curvature that is detectable by electrophoresis. Bendability, however, gives an idea of the flexibility of a given DNA sequence. The location of the high bendability region within the *I* factor 5' UTR suggets that it might serve as a flexible hinge between two functionally distinct parts of the *I* factor regulatory region namely its promoter and the ovary specific enhancer. The two regions and the regulatory proteins bound to them might therefore be brought into contact perhaps through the binding of another protein within the flexible hinge

sequence. Many of the bent motifs described in section 3.1.4 were also sites of protein binding an event which often enhanced the effects of bending.

The *I* factor 5' UTR is specifically bound by a protein factor or factors present in *Drosophila* nuclear extracts. This factor(s) recognise the site 1 sequence which lies betwen residues 138 and 157 and is therefore within the A-rich region. Moreover the transcriptional silencer Aef-1 has been shown to bind the same sequence in vitro. It would therefore be interesting to test the possibility that binding of the nuclear factor(s) and or of the Aef-1 protein induces a bend in this region of the *I* factor 5' UTR. An additional role for the base composition of the *I* factor 5'UTR is suggested by the finding that G-tracts within the 5S RNA promoter of *X. laevi* promote DNA structures that resemble double stranded RNA (Fairall *et al.*, 1989). This sequence-imposed structure is believed to explain the ability of the TFIIIA transcription factor to stimulate transcription from the 5S RNA promoter and to bind the RNA product to form a 7S storage particle. Thus, it is plausible that the base composition of the *I* factor 5' UTR might be important also at the RNA level by e.g. stabilising particular secondary structure motifs that may be important for recognition and processing of the *I* factor transcripts.

Chapter 4

Parental effects of I factor mediated silencing

4. Parental effects of I factor mediated silencing

4.1. Introduction

A number of observations suggest that transposition of the *I* factor within the female germline is subjected to a form of copy number control that is partly determined by species specific host factors (Chapter 1, section 1.3.4). Firstly, all inducer strains which have been analysed, whether they come from natural populations or have been bred in the laboratory from crosses between reactive and inducer stocks, carry between 10 and 15 copies of euchromatic *I* factor sequences per haploid genome. Secondly, artificially prolonging the conditions of hybrid dysgenesis, does not increase *I* factor copy number above this 'inducer' level. Thirdly, the *I* factor homolog from *D.tessieri*, transposes within the germline of *D.melanogaster* females until it reaches a copy number similar to that of the endogenous *I* factor, that is three times higher than that in its original host species.

The *I* factor 5' UTR 'responds' to the presence of other *I* factor sequences within the same genome in germline cells. Expression of a reporter gene, driven by the entire 5'UTR, is reduced by approximately 30 fold in inducer ovaries compared to reactive ovaries (Udomkit *et al.*, 1996). This reactive/inducer ratio is reduced by 10 fold when only nucleotides 1 to 100 of the 5'UTR are present in the reporter construct. This was interpreted as evidence that, within the 5'UTR, sequences between nucleotides 101 and 186 might mediate most of this copy number effect (Udomkit, *et al.*, 1996). Multiple copies of the *I* factor 5'UTR have been shown to downregulate expression of an unlinked reporter gene that is driven by a single copy of this sequence (Chaboissier *et al.*, 1998). Reporter gene downregulation is proportional to 5' UTR copy number. Moreover, when flies carrying multiple copies of the *I* factor 5' UTR are crossed with inducer flies, embryonic lethality in the progeny of the resulting SF females is significantly reduced with respect to controls (Chaboissier, *et al.*, 1998). This reduction in embryonic lethality is again proportional to the number of copies of the *I* factor 5' UTR. The capacity to contribute to the copy number effect is not exclusive to the 5' UTR, since other sequences taken from the main body of the *I* factor and placed downstream of a heterologous promoter can mediate similar effects on the severity of *I*-R hybrid dysgenesis (Jensen *et al.*, 1999). In the absence

of the *I* factor's own regulatory sequences, this copy number effect seems to require transcription (but not translation) of the *I* factor sequence under investigation.

One intriguing aspect of the *I* factor mediated copy number effect is its dependance on the parental origin of the sequences involved, in a fashion that mirrors the directionality of the *I*-R dysgenic cross. This parental effect was first reported for a transposition-defective *I* factor derivative that carried a large deletion within ORF2 but was otherwise identical to active *I* factors (Jensen *et al.*, 1995). This construct was found to alleviate the effects of hybrid dysgenesis (measured as embryonic lethality) to a much greater extent when inherited from the mother than when inherited from the father (see fig. 4.1.1-1). Interestingly, as can be seen from the scheme in figure 4.1.1-1, inheritance from the mother is also associated with a persistance of the effects of the transgene for at least one generation after the transgene has been removed through mating. This apparent "memory" for the presence of the transgene was not observed when the transgene was transmitted paternally. The original interpretation of these findings was that the defective *I* factor transgene was acting as a modifier of reactivity the effects of which were heritable and reversible as previously reported for temperature and ageing (Jensen, *et al.*, 1995). Subsequent experiments involvng a transgene that carried a fragment from the *I* factor ORF2 downstream of the hsp70 promoter (Jensen, *et al.*, 1999), showed that alleviation of the hybrid dysgenesis symptoms and the parental effect were both positively correlated with the number of copies of the transgene present (Cahpter 1, section 1.3.4). Moreover, for a given copy number, embryonic lethality would decrease as the number of generations since transgenesis increased. Thus it appears that *I* factor derived sequences can immunise the ovaries of flies that carry them in a dosage dependent manner. This immunising effect is stronger when the *I* factor containing transgenes are inherited maternally.

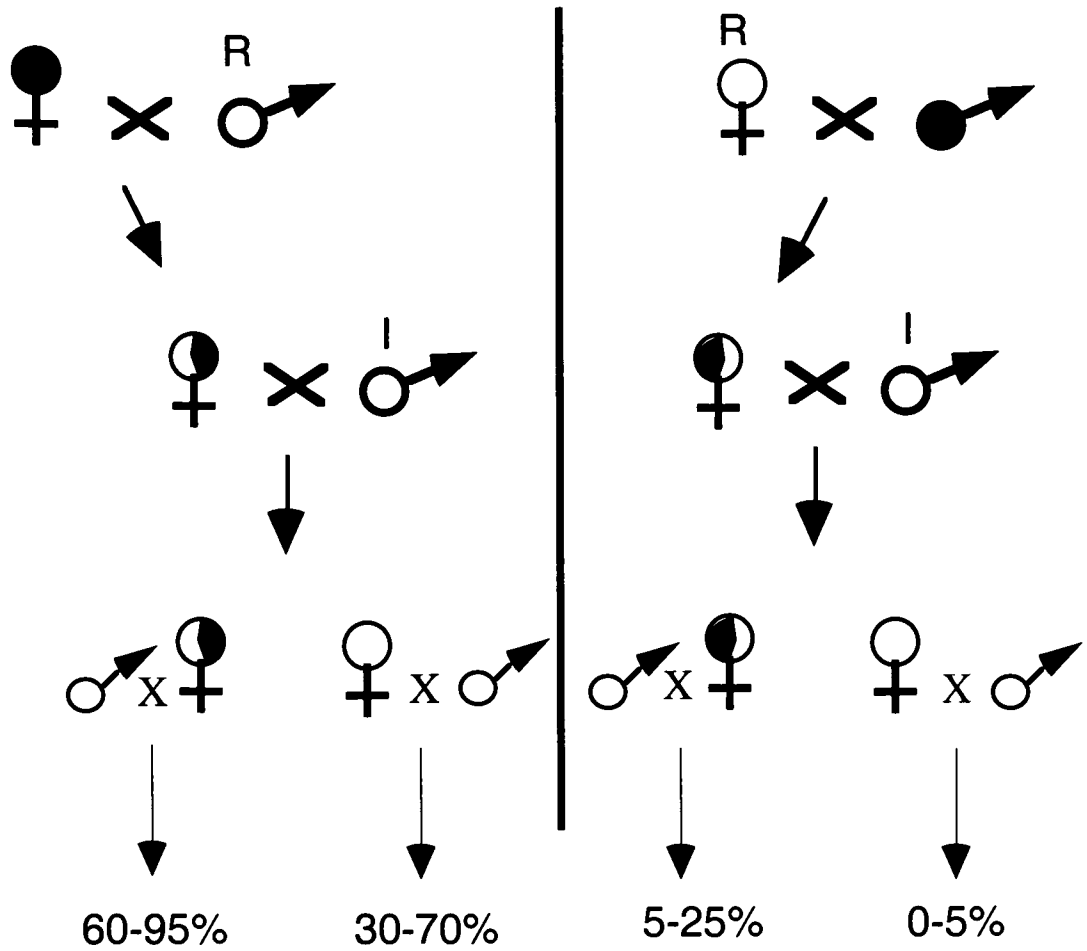


Fig 4.1.1-1. Mating scheme showing the parental effect on reduction of embryonic lethality observed with an I factor containing transgene (data from Jensen et al., 1995).

4.1.1. Copy number-dependent gene silencing in Drosophila

The dosage-dependent silencing of the *I* factor in the female germline might represent a host encoded defence mechanism to limit the spreading of invading DNA sequences or it might reflect the *I* factor's own strategy to maximise its survival by minimising fitness damage to the host. Alternatively, the establishment of the inducer state might result from a combination of both processes operating simultaneously. Little is known so far of the strategies of non-LTR retroelements to regulate their copy number within their hosts, mostly because *D.melanogaster* represents the only known system where the existance of reactive strains permits the study of events immediately following *de novo I* factor invasion. In contrast to the scarcity of knowledge regarding retroelement self limiting strategies, there are an increasing number of examples of *Drosophila* dosage sensistive gene silencing mechanisms.

One of the first examples of a dosage sensitive gene inactivation phenomenon in *Drosphila* was reported in connection with the silencing of mini white transgene arrays (Sabl and Henikoff, 1996; Dorer, 1997). Eye pigmentation was not proportional to transgene copy number and in fact tended to decrease with increasing size of the array. For a given array size, silencing was stronger for insertion sites in the proximity of centromeric heterochromatin. The effects of genome location on the strength of transgene silencing is reminescent of position effect variegation (PEV), a phenomenon whereby euchromatic genes placed in the proximity of constituitive heterochromatin acquire variegated expression (Karpen, 1994; Eissenberg *et al.*, 1995; Henikoff, 1995; Wallarth, 1998). The impaired expression experienced by genes subjected to PEV is believed to be caused by spreading of silencing heterochromatin to the gene sequences, thus resulting in decreased accesibility of the promoter to transcription factors. A possible relationship between trasgene array silencing and PEV is supported also by the fact that known genetic modifiers of PEValso affect transgene silencing (Dorer and Henikoff, 1994).

The principal determinants of the strength of transgene silencing, however, are the properties intrinsic to the array iteself, mainly copy number and relative orientation of the transgenes withn the array (Dorer and Henikoff, 1994; Sabl and Henikoff, 1996). This is particularly evident in the observation that a single mini white insertion in a region of boundary between hetero- and euchroamtin resulted in

normal transgene expression whereas variegation was observed following duplication of the transgene at the same site (Dorer and Henikoff, 1994).

Phenomena of transgene array silencing have been observed in many systems ameneable to genetic manipulation. Recent work in mice has unambiguosly shown that at least one example of transgene array silencing in this organism is also dictated mainly by the size of the array and not by postion effects (Garrick *et al.*, 1998). Using the *Cre-lox* P recombination system, it was possible to trigger controlled excision of transgenes from an array at a given genomic location. Invariably it was found that reduction in transgene copy number resulted in increased transgene expression as well as decreased levels of chromatin condensation and methylation at the site of array insertion. In *Arabidopsis thaliana*, the equivalent phenomenon has been given the name of repeat induced gene silencing (RIGS) (Assaad *et al.*, 1993). In general, transgene array silencing in eukaryotes, seems to rely on heterochromatisation, in response to increased gene dosge (Dorer, 1997).

Recent experimental evidence has uncovered instances of dosage dependent gene silencing in *Drosophila* that may be more relevant to *I* factor copy number control, because they involve transgene copies at different genomic locations rather than in arrays. Insertion of multiple unlinked copies of the heterologous transgene *w-Adh*, consisting of the *white* (*w*) gene control region and the Alcohol dehydrogenase (*Adh*) structural gene, results in downregulation of the *w-Adh* transgenes themselves, and of the endogenous *Adh* gene (Pal-Bhadra *et al.*, 1997). Down regulation, is clearly proportional to transgene copy number, resulting in transcript reduction by 2 to 5 fold, and is maintained stably from larval stages to the adult. Suppression of the Adh gene also results in decreased levels of alcohol tolerance in the transgenic flies, implying that the lower transcripts levels do in fact have phenotypic consesequences. Interestingly, at any given *w*-Adh dosage, suppression effects were always stronger in females then in males. This sex specific difference in suppression effects remains unchanged within a dosage series and therefore does not appear to be associated with the suppression phenomenon itself, but, rather it is more likely to reflect differential expression from the *w* promoter in the two sexes.

At any given copy *w-Adh* dosage, downregulation is greatest whenever transgene copies are homozygous, suggesting direct interactions between transgene insertions that may be facilitated if they are located on homologos chromosomes. There is circumstantial evidence that somatic pairing between homologues is retained in adult cell lineages of Dipterans like *Drosophila* and represents a remnant of polyteny in

larval tissues (Hiraoka *et al.*, 1993; Henikoff and Comai, 1998). The close interaction beween homologous sequences, facilitated by the organisation of homologous chromosomes in Drosophila nuclei, might be at the basis of a whole host of gene silencing phenomena that rely on close direct interaction between the sequences involved. These so-called *trans*-sensing processes (Henikoff, 1995; Henikoff and Comai, 1998) are not, however, limited to related sequences on homologous chromosomes, but have been reported to occurr also between homologous dispersed sequences (Henikoff, 1997). Transvection is a classic example of trans-sensing phenomena in Drosophila that pertains to the interaction between disticnt allelic forms of a gene(Pirrotta, 1990; Tartof and Henikoff, 1991; Hazelrigg and Petersen, 1992). The product of the *zeste* gene is a known mediator of transvection and is known to interact directly with loci that experience transvection (Jack and Judd, 1979). Mutations affecting the Zeste protein inhibit transvection. Despite the enahcement of *w-Adh*-mediated cosuppression by homologous paring, mutations in the *zeste* gene, known to interfere with transvection, had no effect on *w-Adh* downregulation (Pal-Bhadra, *et al.*, 1997).

In contrast, the products of Polycomb group (PcG) genes, were found to play a role in w-Adh cosuppression. PcG proteins play an essential role in early *Drosophila* development , by maintaining the repressed state of a number of homeotic genes in specific cell lineages (Eissenberg, *et al.*, 1995; Kennison, 1997; Breiling and Paro, 1997). Mutations in the PcG loci Polycomb (Pc) and Polycomb-like (Pcl) reduced *w-Adh* suppression by 37 % in females and 50 % in males (Pal-Bhadra, *et al.*, 1997). The sex specific effects of the PcG mutations on Adh mediated suppression, although not investigated further, could reflect true sex dependent differences in the action of the two PcG genes. Further evidence of PcG involvement in Adh dosage-dependent supression, came from immunostaining of polytene chromosomes with antibodies against Pc which identified novel bands at the site of insertion of transgenes (Pal-Bhadra, *et al.*, 1997). Most importantly, Pc recruitment was specific for suppressed transgenes, as shown by the fact that Pc bands were not observed in the absence of suppression at the sites of transgene insertion. Similar results were obtained with antibodies against polyhomeotic (ph), another PcG protein (Pal-Bhadra, *et al.*, 1997), in agreement with current theories of PcG protein action, which is believed to involve recruitment of large multiprotein complexes that include different members of the PcG (Pirrotta, 1998).

Dosage dependent suppression is not a special property of the *w-Adh* transgene but can be detected, albeit to a slightly reduced level, also by simply increasing the

number of copies of a wild type *Adh* gene (Pal-Bhadra, *et al.*, 1997). In contrast, increased dosage of the *w* promoter as part of the *w*-Adh transgene failed to result in detectable effects on the endogenous *w* gene expression. One possibility could be that in contrast to the *I* factor promoter, the *w* regulatory sequence is not transcribed and transcription might be a prerequisite for copy number-dependent silencing as it appears to be the case for the *I* factor (Jensen, *et al.*, 1999). However, even increasing the dosage of a wild type *w* gene did not result in suppression effects: in fact, *w* gene expression appears to increase in proportion to gene dosage (Pal-Bhadra, *et al.*, 1997). It is interesting in this instance to notice that whilst the *w* gene is susceptible to copy-number dependent silencing when in transgene arrays, dosage detection does not seem to function when copies of this gene are scattered at different positions around the genome.

Phenomena of dosage dependent transgene silencing are well documented in plants and fungi and go under the collective name of cosuppression (Vaucheret *et al.*, 1998; Grant, 1999). Dosage dependent downregulation of the *Drosophila* Adh gene represents the first example of cosuppression phenomena in animals (Bingham, 1997). In plants, cosuppression phenomena are widespread and are believed to represent an evolutionarily conserved defense mechanism against invading multicopy parasites such as transposable elements and retroviruses. Systematic inactivation of viral genomes by cosuppression permits plants to cure themselves, a process known as recovery. A recovered plant retains the viral genome but acquires immunity to further infection by sequence related viruses. Although the mechanisms of copy number detection remain unclear in all cases, there are important parallels between the requirements for gene silencing in plants and in *Drosophila*. The latest evidence, attributes an important role to pairing between the transgenes and the endogenous homologous sequences in plants (Grant, 1999), as has been suggested for dosage dependent silencing events in *Drosophila* (Pal-Bhadra, *et al.*, 1997). This DNA-DNA interaction may trigger silencing directly or might promote production of aberrant RNA transcripts (Flavell, 1994; Grant, 1999). These abnormal transcripts could act as diffusible agents that can promote silencing by a variety of mechanisms that go under the general name of post transcriptional gene silencing (PTGS). Some PTGS mechanisms envisage that the aberrant transcripts act via RNA-RNA interactions by stimulating degradation of homolgous transcripts in the cytoplasm. Alternatively, PTGS could take place via RNA-DNA interactions (*trans*-silencing), whereby the silencing transcripts interact directly with homologous genomic sequences in the nucleus.

Cosuppression by RNA degradation was one of the first mechanisms proposed to explain posttranscriptional cosuppression in plants (Flavell, 1994; Vaucheret, *et al.*, 1998; Grant, 1999). Although there are many ways in which an abnormal RNA could mediate degradation of homologous transcripts, almost all models proposed so far postulate a requirement for the synthesis of short complementary copies of the RNA by cellular RNA-dependent RNA polymerases (RdRPs). Support for a possible role of RdRPs in gene silencing phenomena comes from the finding that introduction of double stranded RNA in C.elegans (Fire *et al.*, 1998) or *Drosophila* (Kennerdell and Carthew, 1998) results in suppression of homologous endogenous genes, a phenomenon known as RNA interference.

Until recently, RdRPs had never been directly associated with cosuppression phenomena. The first direct implication of one such enzymatic function in transgene silencing comes from studies in the Ascomycete fungus *Neurospora crassa*, where the process is called quelling (Cogoni and Macino, 1999). In an ingenious experiment, a silencing transgene was used as the mutagenic agent to allow isolation of genes that suppressed quelling. Isolation of the gene (called *qde1*) associated with restoration of transgene expression revealed a conserved protein sequence with good homolgy to a known RdRP from tomato. Moreover, data base searches identified putative homologues also in *Schizosaccahromyces pombe*, in *Arabidopsis thaliana* and in *Caenorhaditis elegans*. No RdRP homologue has as yet been found in *D.melanogaster*.

Trans-silencing mechanisms of cosuppression were hypothesised to explain systemic transgene silencing in plants, whereby the entire plant can acquire immunity to subsequent infection even though the original site of viral entry was restricted to a few cells (Grant, 1999). Trans-silencing models invoke the existence of a diffusible silencing agent that interacts with the transgene sequences in the nuclei of all cells in the plant. Aberrant "silencing" RNAs derived from the original transgene are the prime candidates for a role as diffusible and self perpetuating silencing agents. They are believed to exert their silencing role by entering the nucleus and interacting directly with homologous genomic sequences. Interaction of the aberrant transcripts with related genomic sequences would result in production of more aberrant transcripts at the expense of the normal transcription product thus serving two functions: disrupting gene expression and ensuring continued production of the silencing agent. Recent work in *Antirrhinum majus* has shown that transcripts derived from a normally untranscribed promoter region can induce silencing of genes carrying homologous sequences in trans (Mette *et al.*, 1999). This provides strong

evidence supporting an important regulatory role for the interaction between genes and homologous RNA molecules. RNA-protein complexes are known to regulate gene expression. In mammals the non-coding RNA product of the Xist gene is expressed only by the one of the two X chromosomes in female somatic nuclei that has been chosen for inactivation. The Xist RNA acts in *cis* by coating and inactivating the X chromosome from which it is expressed (Panning and Jaenisch, 1998). In Drosophila, the role of an RNA-proten complex is the exact mirror image of that observed in mammals. The non-coding RNA transcript from the X-linked roX gene interacts with an autosomally encoded protein and then coates the male X chromosome. The result is hypertranscription as part of the process of male dosage compensation (Willard and Saltz, 1997; Meller *et al.*, 1997; Amrein and Axel, 1997).

4.1.2. Inheritance of silenced states

Gene silencing phenomena are inherited mitotically as can be seen, for example, in the patches of colourless ommatidia in the eyes of flies carrying silenced transgene arrays or in the stability of Adh cosuppression from larvae to adults. Cosuppression in plants and quelling in fungi can also be inherited meiotically (Hollick *et al.*, 1997; Selker, 1997; Bingham, 1997; Vaucheret, *et al.*, 1998). In animal systems the evidence so far suggests that silenced states of genes are maintained mitotically but are usually unstable through meiosis. One possible reason for this difference is that whilst in animals the germ cells differentiate very early in development, in plants the future gametes are formed after extensive growth as part of the somatic lineage (Hollick, *et al.*, 1997). This could ensure that plant gametes retain transcriptional imprints that were originally imposed somatically.

Meiotic inheritance of silenced states of genes is particularly well documented in plants with respect to the phenomenon of paramutation (Hollick, *et al.*, 1997). Paramutation has been studied genetically for over 40 years and can be defined as a meiotically heritable change in the expression of one allele invoked by another specific allele. Such allelic interactions represent violations of Mendel's first law which states the independent segregation of alleles. Alleles that are sensitive to paramutation, are called 'paramutable' whilst alleles that induce paramutation are called 'paramutagenic'. The expression level of a paramutable allele is invariably reduced through interaction with a paramutagenic allele. Following meiosis, the paramutable allele retains this reduced expression state and becomes itself paramutagenic, whilst the original paramutagenic allele emerges unchanged from

this interaction. In most cases the paramutagenic allele imposes its repressed state on the paramutable allele, although, there are instances when the repressed transcriptional state acquired by the the paramutable allele does not reflect the expression level of the paramutagenic allele. In all cases, however, the decreased expression of paramutated alleles is associated with a decrease in steady state RNA levels. Most importantly, paramutational changes in expression levels are purely epigenetic in nature that is to say they are not associated with changes in the nucleotide sequence. The occurrrence of paramutation is not limited to endogenous alleles as it has also been reported to occur between unlinked transgenes (Meyer *et al.*, 1993) in a manner reminescent of dosage sensitive silencing phenomena in other experimental systems including *Drosophila* (Hollick, *et al.*, 1997).

Despite its meiotic stability, paramutation is not associated with parental effects: the outcome of a cross between gametes containing paramutable and paramutagenic alleles is invariant regardless of which allele is transmitted through the male or the female (Hollick, *et al.*, 1997). There is however another silencing phenomenon, namely genomic imprinting, that is characterised by parent specifc gene silencing. Genomic imprinting was first described in insects almost 40 years ago (Crouse, 1960), but it is imprinting manifestations in mammals that have been the object of the most intense investigations over the past two decades (Efstratiadis, 1994). In its original definition (Crouse, 1960), genomic imprinting refers to a process of gene silencing that is determined by the sex of the parent. The parent specific imprint is transmitted faithfully through the somatic lineage of the progeny so that within the same somatic cell nucleus, identical alleles coexist that have opposite transcriptional states depending on their parental origin. In this respect, genomic imprinting differs from paramutation and other trans-sensing effects because the transcriptional state of one allele is not transmitted to the other, despite their simultneous presence within the same nucelus. In humans, for example, it is essential that only one of the two alleles is expressed: aberrant imprinting or aberrant transmission of imprinted regions have often dramatic and life threatening consequences (Efstratiadis, 1994). In the germline lineage, the genomic imprint is erased, so that the gametes will carry alleles with an imprint specific for the sex of the individual that produced them. Imprinting phenomena that fulfill the above definition have been described also in *Drosophila*, although in this Dipteran, as well as in other insect orders where impriting has also been reported, biallelic expression or silencing of an imprinted gene is of little or no consequence.

Imprinting of a *white* transgene inserted at the tip of the Y chromosome is one of the first clear examples of paternal imprinting in *Drosophila* (Golic *et al.*, 1998). By producing a stock in which both males and females carried the marked Y chromosome, it was possible to investigate the effect of parental origin on transgene expression. The original phenotype associated with the transgene was one of variegated expression giving a mosaic of patches of red and colourless ommatidia. This phenotype could be attributed to the heterochromatic location of the transgene that was, therefore, believed to a consequence of PEV. Levels of transgene expression were significantly higher in both males and females which had inherited the Y chromosome from their mothers rather than from their fathers. This effect remained detectable in different genetic backgrounds, thus excluding a possible role of strain specific PEV modifiers. Moreover *w* transgene expression was truly determined by parent specific imprints and was not the results of a maternal effect.

Similar imprinting phenomena have been recently reported for three *Drosophila* X-linked loci, *narrow abdomen*, *tiny* and *garnet* (Lloyd *et al.*, 1999). The three loci were brought in close proximity to the pericentromeric heterochromatin of the X chromosome by a complex rearrangement involving both an inversion and a large deletion. As in the previous case, paternally transmitted alleles of all three genes show strongly variegated expression whilst maternally transmitted alleles give nearly wild type phenotypes. As before "hidden" PEV modifiers and maternal effects were excluded suggesting that the observed differential expression of the three loci was due to the establishment of genuine parent-specific imprints. In this instance, however, the effect of known PEV modifiers on the expression of the imprinted gene was investigated. Suppressors of PEV completely eliminated the effects of imprinting on paternally transmitted alleles but only when present in the progeny. The same modifiers had no effect when they were introduced in the parents. This has led to the hypothesis that establishment of the imprint in the germline and its somatic perpetuation in the progeny might occur by different mechanisms.

In summary, three are a number of gene silencing mechanisms in *Drosophila* with parallels in other eukaryotic systems, from plants to humans, that can help understand the mechanisms of *I* factor control. Their relevance to *I* factor copy number control and maternal effects will be discussed below.

4.1.3. Effectors of gene silencing

Most of the silencing phenomena in plants, fungi and mammals are associated with increased methylation levels (Eden and Cedar, 1994; Martienssen and Richards, 1995). Although the cause/effect relationship between methylation and silencing is still a matter of debate, methylation of selected bases has great potential as a heritable imprint specifying the transcriptional state of a gene. In contrast, there is no evidence for DNA methylation in the genome of *D.melanogaster* (Urieli-Shoval *et al.*, 1982), so that other components of the chromatin assembly must take up the role of epigenetic imprints. Increasing evidence suggests that methylation is not simply a substitute for higher order chromatin structures, but rather that it represents an additional level of information that can affect and direct changes in chromatin structure (Ng and Bird, 1999). There are, however, instances of gene silencing phenomena in methylation proficient organisms when changes in methylation do not appear to play a role. In plants, some paramutable and paramutagenic alleles, for instance, do not differ significantly in methylation levels (Hollick, *et al.*, 1997)., despite the conspicuous difference in expression levels. These observations have lead to models of paramutational interactions that resemble those envisaged for transvection and other trans-sensing phenomena in *Drosophila*. These models invoke complex DNA-DNA interactions mediated and stabilised by protein (and perhaps RNA?) components (Hollick, *et al.*, 1997). It is the protein components that are believed to render such interactions and their effects on gene expression heritable.

In *Drosophila*, PcG and Trithorax group (TxG) proteins are prime candidates for having a role as mediators of epigenetic states of gene expression (Eissenberg, *et al.*, 1995; Breiling and Paro, 1997; Kennison, 1997; Pirrotta, 1998). Apart from their role in development and cosuppression, PcG proteins have been implicated also in transvection-like gene silencing interactions that are reminiscent of paramutation (Gindhart and Kaufman, 1995; Pirrotta, 1997). *white* transgenes, carrying fragments derived from the control region of the homeotic gene *sex combs reduced* (*scr*), mediate silencing of unlinked transgenes that carried related, but not identical sequences derived from the *scr* locus. Homeotic genes are the main developmental targets of PcG (and TxG) proteins and carry target sequences called Polycomb response elements (PREs) that mark sites of multiprotein complex assembly for members of the PcG. The *scr*-derived sequences that mediate transgene silencing were believed to carry PREs because the silencing effects were PcG dependent.

4.2. Results

4.2.1. The I factor 5' UTR mediates parental effects

Parental effects of *I* factor silencing have so far been detected only in the presence of full length *I* factors. As discussed above and in section 1.3.4 (Chapter 1), multiple copies of the *I* factor 5' UTR can mimic some of the effects of the inducer state normally associated with full length *I* factors. The aim of the work described in this chapter was to investigate whether the silencing effects mediated by multiple *I* factor 5' UTRs are also associated with parental effects similar to those reported by Jensen *et al*, 1995 and Jensen *et al*, 1999, but in the absence of full length *I* factors. The results are then discussed in the light of the gene silencing models outlined above.

In order to detect and quantify directionality in the silencing effects mediated by multiple copies of the *I* factor 5' UTR, test crosses were set up in both directions (mating schemes A and B, fig 4.2.1-1), between flies from strain 137*, that carry the chloramphenicol acetyl transferase gene driven by the *I* factor 5' UTR (186::CAT**, fig 4.2.1-2 A) and flies from strain 380***, that carry 10 copies of the *I* factor 5' UTR per haploid genome (186x2**** and 186x3, fig 4.2.1-2 B). To distinguish directionality due to maternal inheritance of *I* factor 5' UTRs from other parental effects associated with the 380 strain, control crosses were set up in both directions (mating schemes C and D, fig 4.2.1-1) between flies from strain 137 and flies from strain Ja. The Ja strain is genetically identical to the 380 strain except for the absence of the 186x2 and 186x3 constructs.

---

* This strain is a transgenic derivative of the eye colour mutant reactive strain w^k, and was constructed by Carol McLean for studies that led to the identification of the I factor 5' UTR as the transcriptional control region. In the paper by McLean, et al, 1993, this strain is called 186.137.

** This construct is a derivative of p186W8 (McLean *et al.*, 1993).

*** This strain is a derivative of the reactive *y* and *w* double mutant strain Ja and was contructed by Marie-Christine Chaboissier for studies on the titration model of I factor control. The name of this strain given in Chaboissier, *et al*, 1998, is C17P22C30P17H.

**** The names of the constructs abbreviated as 186x2 and 186x3 are C42PIY and pW83PI respectively (Chaboissier, *et al.*, 1998).

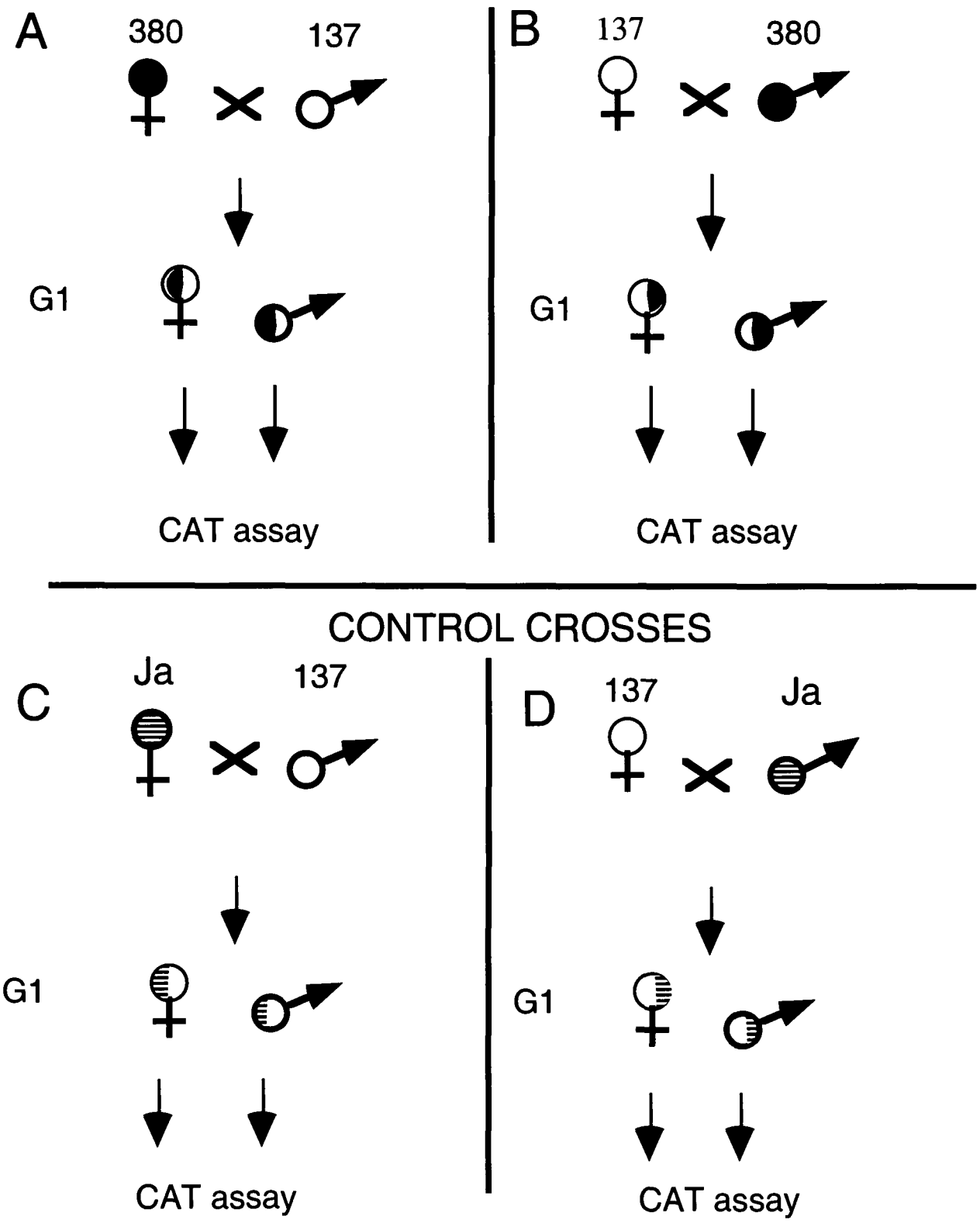


Fig 4.2.1-1. Mating scheme for detection of parental effects mediated by the I factor 5' UTR. In A, females carrying multiple copies of the I factor 5' UTR (constructs 186x2 and 186x3) were crossed with males carrying the 186::CAT construct, driven by one copy of the I factor 5' UTR. The reciprocal cross is shown in B. The control crosses are shown in C and D. The Ja strain is genetically identical to strain 380 but lacks the 186x2 and 186x3 constructs. The concentration of CATprotein was measured in the G1 progeny of each cross. In females, the CAT concentration was measured separately in ovaries and somatic tissues. The notation for transgenic homo- and heterozygous flies are as in figure 4.1.1-1.

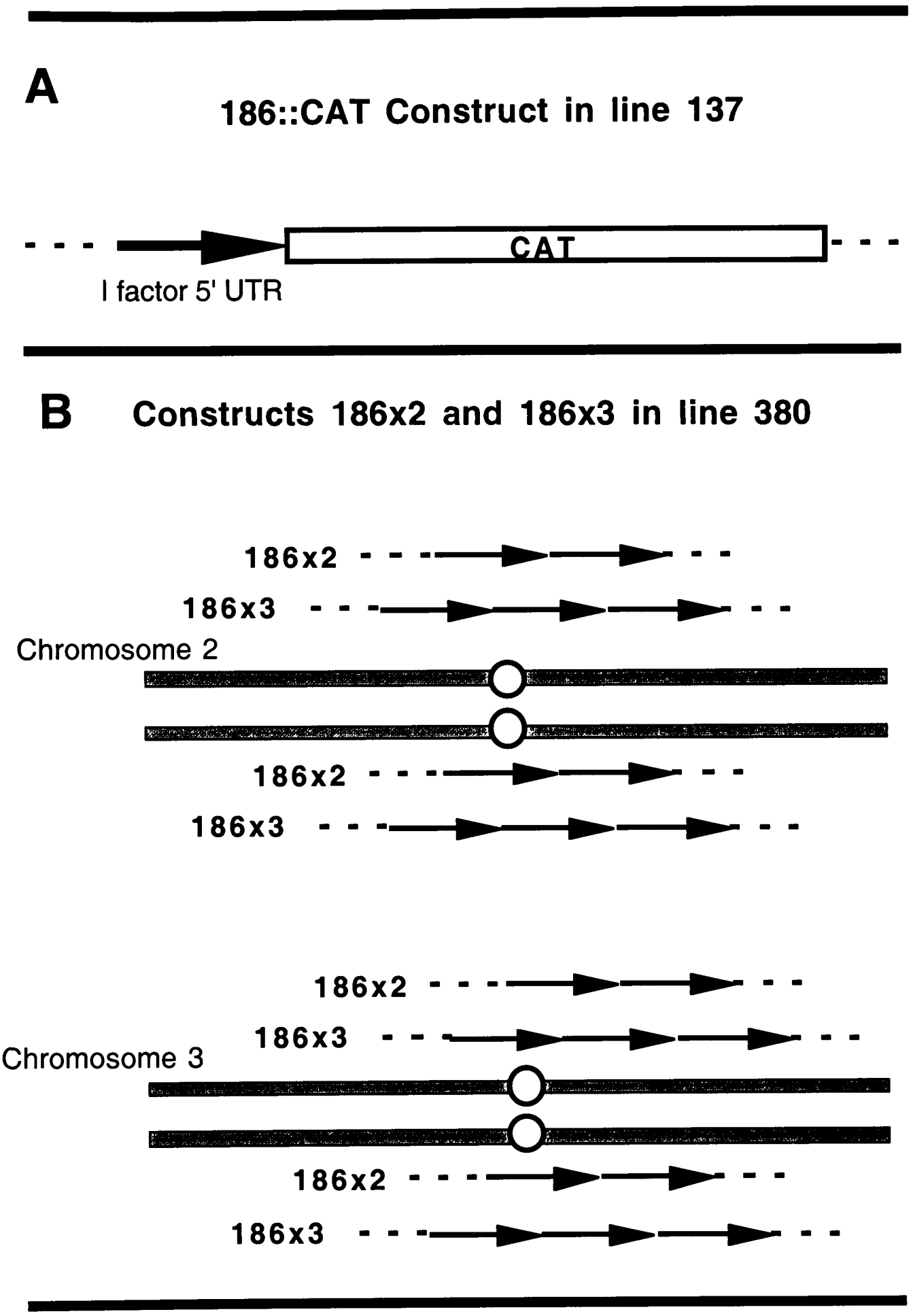


Figure 4.2.1-2: Diagrams of the constructs in lines 137 and 380 used for the crosses to detect parental effects of silencing mediated by the I factor 5' UTR. In A the 186::CAT transgene in line 137 is shown. The I factor 5'UTR is representd by a thick arrow poining towards the CAT open reading frame. In B, the genotype of line 380 with respect to the 186x2 and 186x3 construct is shown. The I factor 5' UTRs are shwon as arrows in head to tail arrangement. They are shwon juxtapposed to the the chromosomes to which they were mapped genetically by Chaboissier, et al, 1998. Dotted lines indicate non specified construct sequences.

The test crosses was repeated four times whilst control crosses were repeated three times under identical conditions. All stocks, prior to use in the crosses, had been maintained for >50 generations under conditions that favour high reactivity levels i.e. constant temperature (25°C) and short generations (Chapter 1, section 1.1.4). For maximum productivity, flies were reared on so called French media (Chapter 2, section 2.1.3) with additional baker's yeast. All flies were maitained at 25°C except for the brief periods of manipulations that took place at room temperature. Approximately 50 to 80 virgin females (10 per vial) and 25 to 40 males (5 per vial) were used in each of the four replicate matings, which corresponds to a total of about 320 virgin females and 160 males for each test cross and 240 virgin females and 120 males for each control cross. The parental (G_0) generation were allowed to lay eggs for up to 5 days and then removed. This ensured that the G_0 flies were usually no older than 5 days by the time they laid their last eggs in the G1 collection vials. Immediately after eclosion, the G1 progeny of each mating were transferred to fresh media and aged for 3 to 5 days prior to harvesting for extract preparation. In the case of the female G1 progeny, ovaries and carcasses* were separated manually under a dissecting microscope and seperate extracts prepared from the two body components. Male G1 progeny were used whole for extract preparation. All extracts, prior to determination of CAT concentration, had been stored for at least a week at -70°C after snap-freezing in liquid nitrogen.

CAT concentrations for each of the three sample categories were determined using a CAT specific enzyme-linked immunosorbant assay (ELISA) (Chapter2, section 2.2.4). The final CAT concentration in each extract is expressed as picograms (pg) of CAT per μg of total protein. Total protein concentrations were determined for a small aliquote of each extract by the Bradford method (Chapter2, section 2.2.3). Extract aliquotes for CAT-ELISA were then adjusted to contain between 50 and 100 μg total protein: the exact protein concentration was incorporated in the final calculation. The mean values of CAT concentration and standard errors (STE) in each sample category for each of the four matings in a cross are shown in table 4.2.1-I. Comparison of the mean CAT concentrations revelas that significant differences are seen: (i) between the values obtained for the ovarian fractions in the

---

* The term carcass is used in this context to indicate the totality of female tissues excluding the ovaries.

Table 4.2.1-I Mean CAT protein concentration in different tissues from G1 progeny of reciprocal crosses.

Cross	Ovaries	Female carcass	Male
380-137	0.083+/-0.01[ab]	0.382 +/-0.12	0.365+/-0.08
137-380	0.408 +/-0.06[ac]	0.216+/-0.05	0.257+/-0.06
Ja-137	1.47 +/-0.24[b]	0.215+/-0.1	0.390+/-0.12
137-Ja	1.65 +/-0.30[c]	0.304+/-0.13	0.355+/-0.10

Values are the mean of four measurements (three for control crosses) of CAT protein concentration of 100 flies (expressed in pg/mg of total protein) +/- s.e.m. p values from unpaired two-tailed t-tests: a) Maternal vs. paternal effect of transgenic 5' UTRs p=0.013. b)Effect of maternally inherited transgenic 5' UTRs versus control p=0.029. c) Effect of paternally inherited transgenic 5' UTRs versus control p=0.049. Comparisons within the female carcass and male sample category gave p>>0.050.

test versus the control crosses (380x137vs Jax137; p=0.029 and 137x380vs137xJa; p=0.049) and (ii) between the values obtained for the ovarian fractions of the reciprocal test crosses (380x137 vs 137x380; p=0.013). Thus, there is a significant drop in CAT levels detected in the ovaries of the progeny when at least one of the two parents carries multiple copies of the *I* factor 5' UTR: the effect, however, is much more accentuated when the multiple *I* factor 5'UTRs are inherited from the mother. In contrast, CAT concentrations measured in females carcasses and in males do not differ significantly between crosses (p>>0.050).

The mean CAT concentrations shown in table 4.2.1-I were then plotted as shown in figure 4.2.1-3. The silencing effect of multiple copies of the I factor 5' UTR on the 186::CAT transgene in the ovaries can be seen clearly in the histogram (black colums). Little variation is detected in female carcasses (stippled) and male (grey) tissues. The implications of these results are discussed further below.

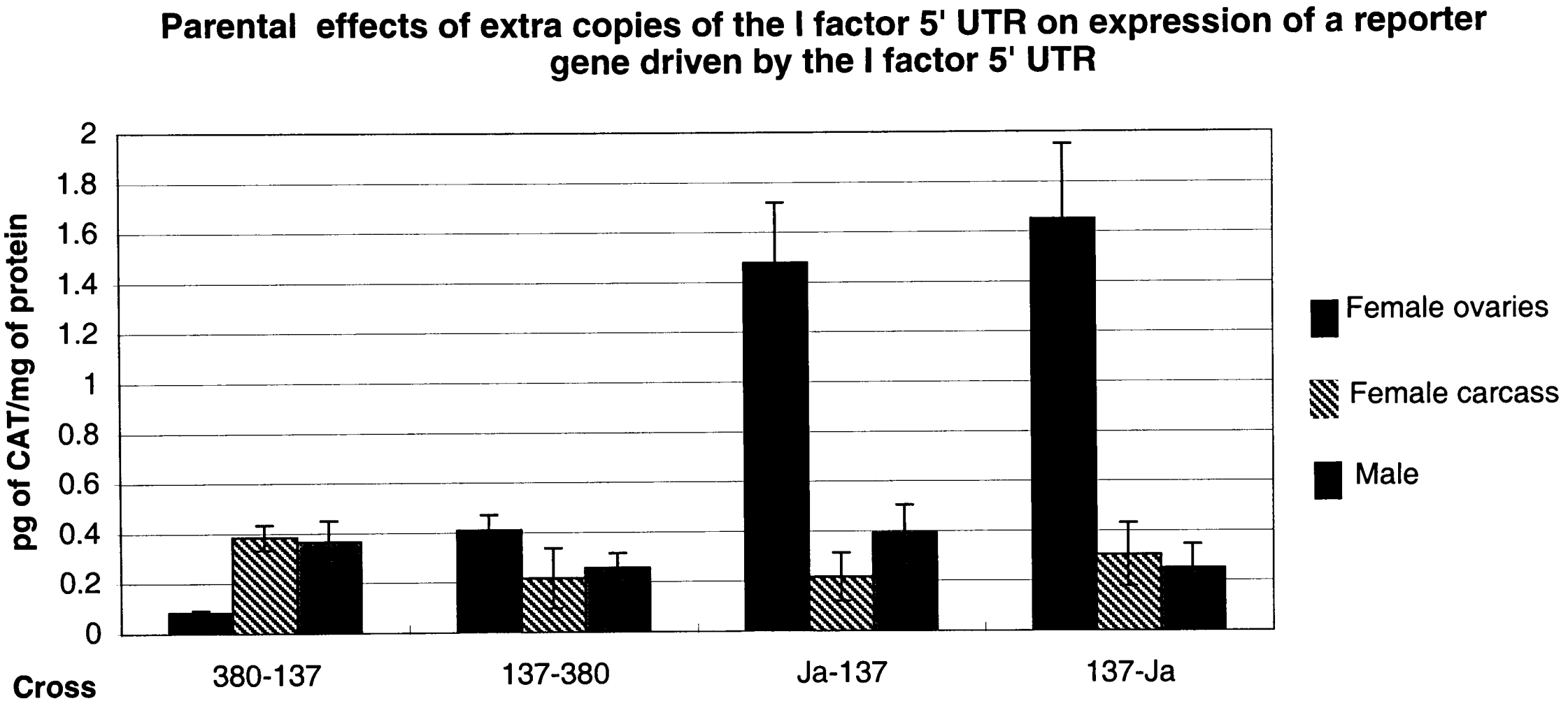


Fig 4.2.1-3 The transgenic line 380 which contains 10 copies of the I factor 5' UTR on a Ja background was crossed with line 137 which contains a chloramphenicol acetyl transferase reporter gene (CAT) driven by the I factor 5' UTR. Strain 137 was also crossed with strain Ja as a control. In female ovaries the presence of extra 5' UTRs significantly reduced expression. This reduction was also significantly greater when the 5'UTRs were inherited maternally. There were no significant differences in males or in female carcasses (female with dissected ovaries). In all crosses the maternal strain is named first. Mean CAT concentrations (+/- s.e.m.) were calculated from the values obtained from extracts prepared from 200 flies (100 females and 100 males) in each of the four matings in a cross.

4.3. Discussion

The work described in this chapter was aimed at investigating parental effects associated with silencing mediated by the *I* factor 5' UTR. The data obtained indicate that the sequences 1 to 186 of the *I* factor, representing the entire 5' UTR, mediate silencing of homologous sequences in *trans,* preferentially through the maternal germline. Preferential maternal transmission of silencing effects had been previously described for an *I* factor transgene harbouring a large internal deletion of ORF2 sequences and for increasing copies a 986 bp long fragment derived from ORF2 (Jensen, *et al.*, 1999). In both cases the silencing effects were measured as an ability of these constructs to attenuate the effects of *I-R* hybrid dysgenesis on embryonic lethality. Decreased *I* factor activity, however, could not be measured directly and had to be inferred from the fall in embryonic lethality associated with maternal inheritance of either of the two *I* factor containing constructs (Jensen, *et al.*, 1999). The work described here permitted direct quantification of maternal and paternal silencing effects of *I* factor sequences exerted on identical sequences derived from the *I* factor 5' UTR and linked to the chloramphenicol acetyl transferase reporter gene (186:CAT construct). Moreover, the approach chosen also permitted to measure reporter gene activity in the female soma and in males where the effects of hybrid dysgenesis are not detectable (Chapter 1, section 1.1.3).

Flies carrying 10 copies per haploid genome of the *I* factor 5' UTR (380 strain) were crossed in both directions with flies carrying the 186::CAT construct (137 strain). As a control identical crosses were set up between flies from the 137 strain and flies from the Ja strain, which is genetically identical to 380 but lacks the multiple 5'UTRs. The concentration of CAT enzyme was then measured in the G1 progeny of all the crosses. In agreement with previous findings (Chaboissier, *et al.*, 1998), the data obtained in this study show that multiple copies of the *I* factor 5' UTR significantly down regulate expression of the 186::CAT construct in the ovaries of G1 females. Moreover, ovary specific suppression of the 186::CAT reporter was five times stronger when the multiple copies of the *I* factor 5' UTR were inherited maternally. This five fold difference was calculated as the ratio of the mean CAT concentration values showed in table 4.2.1-I for the ovary measurements of 137/380 G1 females, versus 380/137 G1 females. This value is in very good agreement with the previously reported difference between the frequency of chromosomal contamination by *I* factor sequences in the germline of non-dysgenic (RSF) and dysgenic (SF) females (Picard, 1976). G1 females that received the multiple copies of the *I* factor 5'UTR from their mothers (380x137 cross) are, in fact, equivalent to

RSF females whilst those that received the multiple 5' UTR paternally (137x380 cross) are equivalent to SF females. CAT concentrations in the ovaries of control G1 females were almost identical independently of whether the 186::CAT construct had been inherited maternally or paternally. Thus, the single copy of the *I* factor 5' UTR linked to the reporter gene (186::CAT) does not show preferential maternal silencing in the ovaries. This may indicate that both the silencing and its preferential maternal transmission require multiple copies of *I* factor derived sequences. At low dosage neither of these effects are observed. Mean CAT concentration in the female soma and in males showed no dependence on parental transmission. This is in agreement with the apparent "neutrality" of these tissues to I factor activity.

The *I* factor sequences involved in the silencing interaction may be divided into a 'silencing' category and a 'silencing-sensitive' category. The present study has shown that the *I* factor 5' UTR can represent both the silencing and silencing-sensitive category. Previous work has revealed a requirement for transcription of silencing *I* factor-derived constructs, in the presence of transcribed silencing-sensitive active *I* factors . This led to the hypothesis that *I* factor silencing might be similar to dosage-sensitive posttranscriptional silencing (PTGS) processes described in plants (section 4.1.2). So far, however, it is unclear whether the transcription requirement for *I* factor silencing is indicative of a direct silencing role of *I* factor transcripts, as predicted for PTGS is plants, or whether transcription may be required to maintain a chromatin state that is essential for dosage sensitivity and subsequent silencing. Both the silencing 186x2 and 186x3 constructs, and the silencing-sensitive 186::CAT reporter used in the present study are believed to yield transcripts containing *I* factor sequences. Using constructs similar to those employed in the present work, it would be interesting to determine whether the silencing interactions require *I* factor containing transcripts to be produced also from the silencing-sensitive sequences. Experiments are currently underway to test the silencing-sensitivity of the 41-186 region from the *I* factor 5' UTR placed upstream of a hsp70 promoter (*I*. Clark, E. Hartswood and D. Finnegan pers. comm.). This heterologous enhancer-promoter system had been previously shown to drive strong ovarian expression of a Beta-galactosidase reporter gene (McLean, *et al.*, 1993). Since the 41-186 region of the *I* factor 5'UTR lacks the basic Inr and DPE promoter elements (Chapter 1, section 1.3.3), it is not expected to be included in transgene derived transcripts. Participation of this untranscribed *I* factor sequence in silencing processes will, therefore, have implications for models of *I* factor silencing.

The main features of *I* factor copy number control uncovered so far can be summarised as follows : (i) tissue specificity (Finnegan, 1989a; Bucheton, 1990); (ii) dosage sensitivity (Chaboissier, *et al.*, 1998; Jensen, *et al.*, 1999); (iii) preferred maternal inheritance of silencing effects (Jensen, *et al.*, 1995; Jensen, *et al.*, 1999) [This work]; (iv) persistance of silencing in the absence of *I* factor sequences for at least one generation (Jensen, *et al.*, 1995; Jensen, *et al.*, 1999). There are clearly two processes involved: one is copy number detection and the other is coordinated silencing of all the copies. Whilst both processes have to satisfy the criterion of tissue specificity the second must also survive female meiosis even in the absence of the main target sequences. The titration model (Finnegan, 1989a; Udomkit, *et al.*, 1996; Chaboissier, *et al.*, 1998) (Chapter 1, section 1.3.4) offered an elegant solution to all the problems described above: the titreteable factor would be ovary specific and titration itself would provide both copy number detection and silencing mechanisms. Moreover, persistance of the silencing effect in the absence of *I* factor sequences could be explained because of depletion of the titrateable factor from the maternal ovaries that may require one or more generations to be replenished. However, the experimental evidence has shown that *I* factor derived sequences other than the 5' UTR can also mediate this silencing effect suggesting that alternative mechanisms must be operating.

When *I* factor sequences first enter the reactive female germline, the progressive inactivation that correlates with increasng copy number, is reminescent of dosage-sensitive silencing processes such as cosuppression in plants and *Drosophila* (section 4.1.1). The main difference is the fact that *I* factor downregulation requires a few generations to become established. Moreover, all cosuppression phenomena characterised so far are somatic events, whereas *I* factor transposition and its downregulation only occur in the female germiline. The generational requirement of *I* factor silencing could be reflecting an inefficiency of host regulatory processes or perhaps the ability of the *I* factor to evade them for the first few generations. Plant retroviruses for example, have developed counterdefensive strategies to overcome host encoded PTGS processes (Kasschau and Carrington, 1998). The viral counterdefense, takes up the form of a protein that stimulates viral replication and diffusion through the plant tissues and that, like the host PTGS, acts postranscriptionally. In the case of the *I* factor, however, the time delay is also observed when the silencing *I* factor-derived sequences are not translated, suggesting perhaps that delayed downregulation is more likely to reflect the timing of host responses. Alternatively, the time delay may simply reflect a threshold effect of *I* factor copy number: below a certain dosage, silencing mechanisms are not triggered.

The generational gap would then represent the time required by the *I* factor to increase in copy number by means of transposition. This in agreement with the present data and with data from other work (McLean, *et al.*, 1993; Udomkit, *et al.*, 1996; Chaboissier, *et al.*, 1998) (jensen99). However, this hypothesis contradicts the results reported by Jensen *et al.*, 1995 where a single copy of a defective *I* factor construct had detectable effects on the severity of *I*-R hybrid dysgenesis.

Once *I* factor activity in the female germline has been suppressed, the silenced state can be transmitted maternally to homologous, unlinked *I* factor sequences. The transmission of the *I* factor silenced state to other homologous sequences is reminescent of paramutation in plants and related *trans*-sensing phenomena in *Drosophila* (sections 4.1.1 and 4.1.2). In this analogy, silencing *I* factor sequence are equivalent to paramutagenic alleles, while silencing-sensitive *I* factor sequences are equivalent to paramutable alleles. In plants paramutable alleles can spontaneously convert to the paramutagenic state. In the case of the *I* factor, a few generation in the female germline are sufficient for this conversion. In the opposite direction, paramutable alleles of the *I* factor can be regenerated from paramutagenic forms by one passage through the male germline. The paramutation analogy is strengthened further by the recent finding that paramutation-like phenomena in the pathogenic fungus *Phytophthora infestans* can persist even in the absence of the silencing transgene (van West *et al.*, 1999), as reported in the case of the *I* factor . In *Drosophila*, allelic interactions strongly reminescent of paramutation are believd to involve transmission of dominant silencing chromatin states between the interacting alleles,(section 4.1.1).

Preferential transmission of *I* factor silencing through the female germline is reminiscent of imprinting phenomena, yet a number of important differences suggests that the explanation for the parental effects associated with *I* factor silencing lies elsewhere. Firstly, one of the critical aspects of imprinting mechanisms is that silenced and active alleles do not interact even though they share the same nucleus (section 4.1.3), in contrast to what is observed with *I* factor silencing. Secondly, imprinting is not associated with maternal effects and the parental imprints should be detectable in progeny of both sexes. In contrast, *I* factor silencing is transmitted from mother to daughters, but not from mothers to sons, and its persistance in the female germline even in the absence of *I* factor sequences, implies a strong maternal effect.

If the resemblance of *I* factor silencing to known dosage-sensitive gene silencing processes underlies a genuine mechanistic similarity, *I* factor silencing and its

inheritance could represent the establishment and transmission of silencing chromatin states simlar to those proposed to explain trans-sensing phenomena in *Drosophila* (Henikoff, 1995; Henikoff, 1997; Henikoff and Comai, 1998). and paramutation in plants (Hollick, *et al.*, 1997). The parental effects of *I* factor silencing could then be explained in terms of preservation/loss of the silenced state through the female and male germline respectively. The known differences between female and male gametogenesis in *Drosophila* lend support to this possibility. It is known, for example, that whilst the basic chromatin organisation is preserved during oogenesis, *Drosophila* spermatogenesis is associated with dramatic changes in chromatin composition where nucleosomes are lost and replaced by protamines (Hennig, 1996; Lewin, 1998). This suggests that whilst the female germline offers a favourable environment for the preservation of epigentic imprints, the male germline is the place where they are most likely to be lost. Thus maternal inheritance of *I* factor silencing may result from the loss of the silencing imprint in the male germline* . Since the *I* factor does not transpose in male tissues, its active state is not detectable until it re-enters a reactive female germline.

One final consideration pertains to the possible nature of the silencing epigenetic imprint controlling *I* factor expression in the female germline. Some of the *Drosophila* dosage- and pairing-sensitive phenomena discussed above and in sections 4.1.1 and 4.1.2 invoke a role for PcG proteins. Additional experimental evidence has shown that PcG proteins can impart female specific meiotic inheritance of silenced states (Cavalli and Paro, 1998). Several aspects of PcG mediated silencing are strongly reminescent of the properties of *I* factor silencing suggesting a possible role for PcG proteins in *I* factor control (Chapter 5).

---

* Under special circumstances, the silencing imprint on the I factor can be lost also in the female germline. This can be seen when RSF females are crossed with reactive males: their F2 progeny will suffer hybrid dysgenesis observed as F3 embryonic lethality (Chapter 1, section 1.1.2).

Chapter 5

A possible role of Polycomb group proteins in regulation of I factor transposition.

5. A possible role of Polycomb group proteins in regulation of I factor transposition.

5.1. Introduction

In inducer females, the silent state of the *I* factor is retained throughout oogenesis and is transmitted faithfully to the female progeny even if no *I* factors were contributed by the father. Moreover, the silencing imprint associated with transgenes containing *I* factor sequences can be transmitted in *trans* either to other *I* factor containing transgenes or to active *I* factors. The efficiency of this silencing effect is partly dependent on the dosage of silencing transgenes and is enhanced when the latter are inherited maternally (Chapter 4). Similar phenomena of gene silencing which are dependent on the dosage of the homologous interacting sequences have been described in *Drosophila* and other experimental systems. Those gene silencing processes in *Drosophila* that most closely resemble *I* factor silencing, namely cosuppression and *trans*-inactivation of certain transgenes, appear to be mediated by protein factors from the Polycomb group (PcG) (Chapter 4). This and other evidence discussed below seem to make a strong case for a possible involvement of PcG proteins in *I* factor control in the female germline. The work described in this chapter represents a first attempt aimed at investigating the possibility of direct interaction between the *I* factor and POLYCOMB (Pc), a member of the *Drosophila* PcG.

5.1.1. Polycomb and other PcG proteins

Polycomb is one of 12 PcG genes characterised so far and these are listed in table 5.1.1.-I. The first member of the PcG to be identified genetically was called *extra sex combs* (*esc*), a name that reflects the common phenotype associated with PcG mutations, and that is used to identify new PcG members. The PcG phenotype is, in fact, more easily scored in males where the characteristic tufts of sex combs normally found only on the first pair of legs, develop also on the second and sometimes third pair of legs. PcG mutations, therefore, result in homeotic transformations whereby posterior segments acquire characteristic of more anterior ones. These phenotypes are a result of derepression of homeotic genes in

inappropriate parasegments (the 'out-of-phase' precursors of adult segments (Lawrence, 1993)).

Table 5.1.1-I : PcG genes in *D. melanogaster

Gene name	symbol
Polycomb	*Pc*
Polycomblike	*Pcl*
polyhomeotic	*ph*
Posterior sex combs	*Psc*
Additional sex combs	*Asx*
Enhancer of zeste	*E(z)*
Sex combs on midleg	*Scm*
Sex combs extra	*Sce*
extra sex combs	*esc*
super sex combs	*sxc*
pleiohomeotic	*pho*
multi sex combs	*mxc*

* **Adapted from Kennison, J.A., 1995**

Homeotic genes specify segment identity along the anteroposterior axis and are organised into two main gene clusters, the Bithorax complex (BX-C) and the Antennapedia complex (ANT-C) (Lawrence, 1993). In general, homeotic genes have complex regulatory regions consisting of multiple recognition sites for regulatory proteins that often extend several kb from the transcription units. The ON/OFF state of homeotic genes in individual parasegments is initially determined by a cascade of transcriptional activators and repressors encoded by the segmentation genes that come into action at the time of blastoderm formation (Lawrence, 1993; Kennison, 1997). Given the relatively rapid turnover of segmentation gene products, preservation of the cell specific pattern of active and inactive homeotic genes requires the products of the Polycomb group genes and of the functionally antagonistic genes from the Trithorax group (TrxG) (Lawrence, 1993; Hagstrom and Schedl, 1997; Kennison, 1997; Breiling and Paro, 1997).

Functional PcG and TrxG gene products ensure that the transcriptional states of homeotic genes are maintained through the many mitotic divisions that lead to formation of adult tissues. Although the products of PcG and TrxG genes are uniformly distributed throughout the developing embryo, their specificity of action appears to be determined by the transcriptional state of their target homeotic genes in individual cells (Breiling and Paro, 1997). PcG proteins recognise and interact with

homeotic genes in those cells where the latter had been silenced by segmentation repressors. Similarly, TrxG proteins recognise and interact with transcribed homeotic genes, which were originally switched on by segmentation activators.

The dominant, homozygous lethal *polycomb* (*Pc*) mutation was identified by P. Lewis in 1947 and the gene responsible was cloned more than 40 years later (Paro and Hogness, 1991). Three distinct transcripts are produced by the *Pc* gene: a 2.0 kb transcript that is most abundant in the early embryo, a 2.5 kb transcript that overlaps the temporal domain of the 2.0 transcript but is also found in later embryonic, larval and pupal stages and finally a 1.0 kb transcript that is most abundant in late third instar larvae (Zink and Paro, 1989). The polycomb protein is 390 aminoacids long with a relative molecular mass of 44K (Paro and Hogness, 1991). A schematic representation of this protein is shown in figure 5.1.1-1.

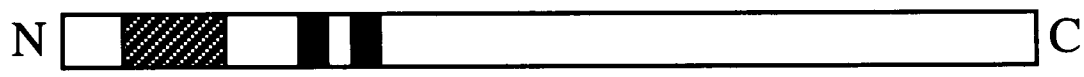


Fig: 5.1.1-1 Schemtic representation of the POLYCOMB protein (adapted from Breiling & Paro, 1997). The striped box represents the conserved chromodomain whilst the two smaller black boxes represent two consecutive histidine-rich domains of unknown function.

The most salient feature of this protein is the presence of a so called chromodomain (Paro and Hogness, 1991). This 48 aminoacids long motif close to the amino terminus is 65% identical to a conserved motif found in the *Drosophila* heterochromatin protein 1 (HP1). HP1 is the product of the *Su(var)205* gene, which as the name implies, was originally characterised as a modifier of position effect variegation (PEV). The *Drosophila* HP1 chromodomain has been used to identify related motifs in proteins from other organisms. HP1-like chromodomains have been found in other *Drosophila* species, in mice and humans (Singh *et al.*, 1991). In all cases analysed so far, these putative HP1 homologues have been found to be nuclear proteins that interact with chromosomes.

The existence of a chromodomain in the Pc protein represents the only structural feature relevant to the as yet hypothetical function of Pc as a chromatin binding protein. Point mutations within the Pc chromodomain, abolish the typical Pc binding pattern on polytene chromosomes, as detected by immunostaining (Eissenberg *et al.*, 1995). The HP1 and Pc proteins appear to have completely non-overlapping binding domains on polytene chromosomes: whilst HP1 is found preferentially associated with regions of constitutive heterochromatin, Pc bands can be seen extensively on euchromatic regions of the chromosome arms (Eissenberg, *et al.*, 1995). Replacing the HP1 chromodomain with the corresponding Pc motif causes the HP1 protein to acquire a Pc-like binding pattern on polytene chromosomes. Thus, the experimental evidence suggests that the Pc chromodomain might be a determinant of Pc binding specificity.

Within the PcG group, the chromodomain is unique to Pc. The other members of the PcG group each have one or more distinctive aminoacid motifs, often present in multiple copies (Breiling and Paro, 1997). The possible role of these motifs in specifying the function of PcG proteins remains to be determined. Current hypotheses suggest that at least some of these structural motifs might represent interfaces for protein-protein interaction (Eissenberg, *et al.*, 1995; Breiling and Paro, 1997). Genetic and biochemical evidence, in fact, implicates multiple PcG proteins in the control of individual homeotic genes (Pirrotta, 1997; Breiling and Paro, 1997). Combining heterozygous mutation in different PcG genes generally results in a more severe phenotype than those associated with the individual mutations. Moreover, immunostaining of polytene chromosomes with antibodies raised to different PcG proteins often gives overlapping banding patterns. Finally in immunoprecipitation experiments on *Drosophila* nuclear extracts, antibodies against individual PcG proteins associate specifically with large (~2.5 MDa) multiprotein complexes.

Putative multiprotein complexes containing products of various PcG genes are believed to assemble in correspondence of specific DNA sequences called Polycomb response elements (PREs) (Muller and Bienz, 1991; Simon *et al.*, 1993). PREs are between 100 and a few 100 bp long and show little or no sequence homology with each other. Multiple PREs exist upstream of individual homeotic genes within a cluster, each apparently specifying the transcriptional state of their cognate homeotic gene in adjacent parasegments. TrxG proteins have been shown to interact with DNA targets that are juxtaposed to or even overlap PREs (Strutt *et al.*, 1997; Pirrotta, 1998).

The mechanism by which PcG protein complexes are recruited at PREs is not known. Current models envisage recruitment of PcG proteins at PREs by protein factors that bind DNA directly but also interact with one or more members of the PcG (Eissenberg, *et al.*, 1995; Breiling and Paro, 1997; Pirrotta, 1998). Support for this hypothesis comes from the discovery at that least one PcG protein, pleiohomeotic, binds PRE DNA directly (Brown *et al.*, 1998). Adding to the complexity, however, is the fact that the composition of PcG protein complexes seems to vary between different PREs. One possibility is that the recruiting factor(s) determine the exact composition of the PcG complex at a given PRE. Alternatively the first phases of recruitment involve identical PcG components but other properties of the PRE or of the flanking sequences, eventually determine what PcG members take part in a given complex.

How binding of PcG proteins at PREs preserves the silent state of a target gene is still unclear. The so called spreading or mass action model, predicts that PcG complexes at PREs act as foci for the assembly of long range chromatin structures that lock the silent state in place (Pirrotta, 1997; Pirrotta, 1998). Experiments designed to investigate changes in chromatin structure, by determining chromatin accessibility using enzymatic probes, have yielded contrasting results. In some cases, reduced accessibility to RNA polymerase II was reported, whilst smaller polymerases such as the T7 enzyme, could efficiently transcribe genes silenced by PcG (McCall and Bender, 1996). In another study, the restriction patterns of PcG silenced loci were found to be indistinguishable from those of non-silenced control, suggesting that the DNA fibre was equally accessible to endonucleases in both cases (Schlossherr *et al.*, 1994). Moreover, cytological studies have found no evidence of relocation of PcG-silenced genes to particular sub nuclear compartments (Hagstrom and Schedl, 1997).

An alternative model of PcG-mediated silencing, invokes direct interaction between PcG protein complexes on different PREs through a looping out of the intervening sequences (Pirrotta, 1997; Pirrotta, 1998). Interactions between distant PREs might create a structure that ensures the preservations of an existing transcriptional state. This hop-and-skip model also envisages longer range trans interactions between PREs at different genomic locations. Some experimental evidence exists for *trans*-interactions mediated by PcG proteins (chapter 4 sections 4.1.1 and 4.1.3). Moreover *trans*-interactions between PRE-bound PcG proteins can help explain phenomena of transgene homing (Fauvarque and Dura, 1993; Kassis, 1994). Transgenes carrying PREs often insert in the vicinity of genomic PREs. One possible explanation for this

homing phenomenon is that when PRE-containing transgenes are introduced into the developing germline, PcG protein complexes form on the PRE in the transgene. *Trans*-interactions between the PcG complexes on the transgene and similar PcG complexes at their genomic locations homes the transgene close to a genomic PRE, increasing the probability of transgene insertion in its immediate vicinity.

The mechanism by which PcG complexes ensure that the silent state of a gene is maintained through cell division is still largely a matter of speculation. Cytological immunostaining studies have shown that, at the onset of mitosis, at least some PcG proteins, appear to dissociate from the chromosome and move to the cytoplasm (Hagstrom and Schedl, 1997). Thus, in order to explain mitotic survival of PcG imprints, some authors have hypothesised that at least one or more of the proteins in the PcG complex must remain associated with PREs even during passage of the replication fork (Hagstrom and Schedl, 1997). The marked PREs will then act as foci for the reassembly of the PcG complex after S phase. The hop-and-skip model of PcG silencing, offers an alternative explanation for the survival PcG silencing complexes through cell division (Pirrotta, 1997; Pirrotta, 1998). The complex looped structures, held at the base by PRE-bound PcG proteins may be replicated *in situ* loop by loop. As soon as one loop is replicated it immediately folds back in to the original conformation, as the replication fork passes through. In this view, undisturbed parts of the looped structure serve as templates for correct reassembly of individual loop components.

5.1.3. Maternal inheritance of PcG and TrxG imprints

The patterns of homeotic gene expression established in the somatic lineage during *Drosophila* development are believed to be absent or erased in germline cells. The available experimental evidence, in fact, suggests that the zygotic genome starts these patterns *de novo*. The germ cell precursors are the first to form when the *Drosophila* embryo is still a syncytium (Lawrence, 1993). This event precedes the establishment of homeotic expression patterns and therefore the coming into play of PcG and TrxG gene products. Nevertheless, all the identified PcG genes have strong maternal effects implying that considerable amounts of each gene product are deposited in the maturing oocyte by the mother (Breiling and Paro, 1997). This means that PcG and TrxG products are already present at the time of germ cell segregation, although their functions will not be required till the successive blastoderm stage.

Despite the apparent lack of a role for most PcG and TrxG proteins in the germline, these epigenetic factors have been found to mediate maternal inheritance of transcriptional states in experimental systems. The most detailed studies have focused on the antipodal actions of PcG and TrxG proteins at Fab-7 a PRE-containing sequence that is part of the complex regulatory region of the *Abdominal-B* gene, one of three homeotic genes in the BX-C (Busturia and Bienz, 1993). Biochemical studies of the Fab-7 fragment have shown that this 3.6 kb sequence contains overlapping binding sites for both PcG and TrxG proteins (Orlando and Paro, 1993; Strutt, *et al.*, 1997). Furthermore, *Fab-7*-bound PcG proteins have been found to compete efficiently against the powerful yeast GAL4 activator (Zink and Paro, 1995). The upstream activator sequence (UAS), which represents the *in vivo* target of the GAL4 protein, is no longer accessible for binding by this protein when linked transgenes also carry the Fab-7 fragment.

As part of a transgene that contains a UAS-linked *lac-Z* reporter and a GAL4 independent mini*white* gene, the Fab-7 fragment has been found to impose maternally heritable repressed and derepressed states on both transcription units (Cavalli and Paro, 1998). Transgene repression, detected as variegated or generally weak Beta-galactosidase staining and eye pigmentation, is PcG dependent and can be converted to a more strongly repressed (hyper repressed) state by mutations in TrxG genes or heat treatment of embryos. Hyper repression mediated by Fab-7-bound PcG is detectable also on the *scalloped* locus that lies close to the transgene insertion site, and results in deformed wing blades. Derepression of the UAS-*lac-Z* transcription unit is dependent on the presence of GAL4 activator synthesised from an unlinked GAL4 transgene driven by the hsp70 promoter. hsp-70-GAL4 transcription is induced by heat-shock pulses applied to embryonic and larval stages. The presence of the Fab-7 fragment ensures that Gal-4 dependent activation of the UAS-*lac-Z* gene results also in derepression of the mini white gene. Derepression, detected as uniform lac-Z staining and eye pigmentation, is believed to be mediated by TrxG products bounds to the Fab-7 element. The extent of derepression is decreased by the presence of a second Fab-7 fragment within the construct.

Transgene hyper repression and derepression are detected in both males and females developed from heat treated embryos or larvae, indicating that there is little or no difference in somatic inheritance of PcG/TrxG-mediated transgene imprints between the two sexes. In contrast, meiotic inheritance of the two transcriptional states occurs exclusively from mothers to daughters but not from mothers to sons. Moreover, transmission of silenced or active states has been shown to be independent of the

presence of the GAL4 transgene or of residual GAL4 activator. Males carrying hyper repressed or derepressed transgenes do not transmit either state to their progeny.

A possible explanation for the events taking place at the Fab-7 containing construct, suggest direct competition between PcG and TrxG factors (Cavalli and Paro, 1998). This idea is supported by the fact that both hyper repression and derepression are completely reversible. In absence of heat treatment, the interplay between PcG and TrxG factors results in variegated transgene expression. Mild heat treatment of transgenic embryos, by transferring them from 18°C to 28 °C for a brief period, apparently shifts the balance in favour of PcG and the transgenes become hyper repressed. The biochemical reasons for this temperature dependent effect are as yet unexplained (Cavalli and Paro, 1998). Transgene derepression by the GAL4 activator, is believed to result in displacement of PcG proteins, although the exact mechanics of this event are not clear given that the domains of PcG and GAL4 binding are completely non overlapping (Zink and Paro, 1995). Derepression affects also the miniwhite gene whose activation is not under GAL4 control. The active chromatin state initiated by GAL4 binding, shifts the balance in favour of TrxG factors which will then ensure its perpetuation through mitosis and meiosis. These evens are similar to the establishment and maintenance of homeotic gene expression patterns.

The stochastic component that seems to determine at least in part the dynamics of the PcG/TrxG balance, is reflected in the frequency with which the resulting transcriptional states are inherited. In the absence of heat treatment, 22% of the flies have a pale yellow eye phenotype associated with hyper repression, whilst only 8% display variegation with some red ommatidia, which indicates clonal derepression. In flies that were treated at 28°C as embryos, hyper repression reaches a frequency of 100 %, but upon female meiosis, only 15 % of the daughters (in addition to the baseline 22%) will inherit a hyper repressed miniwhite transgene and show pale yellow eyes. Derepression, on the other hand, reaches a frequency >70% in flies where the GAL4 activator was induced, and 30 % of the female progeny of derepressed mothers will inherit an active miniwhite gene giving uniform red eyes.

Recent work in *Caenhorabditis elegans* has uncovered specific roles for PcG homologues in female germline development and chromatin silencing. Normal germline development in *C. elegans* depends mainly on four genes, named *mes-2*, *mes-3* , *mes-4* and *mes-6*, where mes stands for maternal-effect sterile (Capowski *et al.*, 1991). As their name implies the common mutant phenotype associated with

mutations in any of the mes genes, is sterility of the F1 hermaphrodite progeny of a mutant mother. This is the reason why mes phenotypes are also referred to as grandchildless. The sterility of mes-mutant hermaphrodites appears to be caused by degeneration of germline cells during larval development and current models suggest that the underlying molecular defect might be global de silencing of normally suppressed genes (Korf *et al.*, 1998).

Molecular analysis of the MES proteins has revealed that *mes-2* and *mes-6* encode the *C. elegans* homologues of the *Drosophila* PcG proteins, Enhancer of zeste E(z) (Holdeman *et al.*, 1998) and extra sex combs (esc) (Korf, *et al.*, 1998) respectively. Reinforcing the structural homology is the fact that MES-2 and MES-6 both localise to nuclei and one requires the presence of the other for proper function. This is in keeping with the evidence from *Drosophila* where, E(z) and esc proteins have been shown to interact directly with each other (Jones *et al.*, 1998). This homology is of particular interest given also that amongst the *Drosophila* PcG proteins E(z) is the only one so far for which there exists evidence for a role in germline development. Certain temperature sensitive alleles of *E(z)* result in sterility of mutant flies at restrictive temperatures (Phillips and Shearn, 1990). Surprisingly, *mes-2* and *mes-6* appear to be the only PcG homologues in the 90% complete sequence of the *C. elegans* genome (Korf, *et al.*, 1998), excluding, however, cec-1, a *C. elegans* soma specific chromodomain protein (Agostoni *et al.*, 1996). This has of course important implication in terms of evolution of PcG function and hypotheses have been put forward suggesting that the germline restricted function of *C. elegans* mes genes might represent the archetypal function of PcG/MES proteins (Korf, *et al.*, 1998). The somatic specific roles for PcG in *Drosophila* and for vertebrate homologues might represent roles acquired relatively recently in evolutionary time.

In addition to their essential role in germline development, *mes* genes have also been implicated in phenomena of transgene array silencing in *C. elegans* (Kelly and Fire, 1998). Repetitive transgene arrays introduced in *C. elegans* are silenced specifically in the germline and the silencing is transmitted preferentially from the mothers. Mutations in any of the four *mes* genes increase the frequency of transgene reactivation to a greater extent in the germline of hermaphrodites than in the male germline. The involvement of MES proteins in transgene silencing is in agreement with their putative role as germline specific silencers, and also suggests their participation in defense mechanisms involving the silencing of repetitive sequences such as retroviruses and transposable elements. Unlike cosuppression, however, transgene array silencing in *C. elegans* is not a strictly dosage sensitive process

because transgenes in non repetitive arrays are not silenced. Rather it is more reminiscent of transgene array silencing in *Drosophila* and RIGs in plants (section 4.1.1 Chapter 4). However, apart from its germline specificity and its requirement from PcG homologues, transgene array silencing in *C. elegans* offers another striking similarity with *I* factor silencing: transgene silencing is progressive and requires several generations to become established (Kelly *et al.*, 1997).

In summary, PcG- and TrxG- mediated control of transgenes carrying PREs shares a number of features with aspects of *I* factor control, for example the apparently repressive effects of temperature changes and the preferential inheritance through the maternal germline. To strengthen the case for PcG involvement in *I* factor control is the finding that phenomena of progressive gene silencing and maternal inheritance in a completely different experimental system, also depend at least in part on PcG homologues. PcG (and possibly TrxG) involvement in *I* factor control in the female germline of *Drosophila* will be further discussed in the light of the results presented below.

5.2. Results

5.2.1. A putative pleiohomeotic binding site in the I factor 5' UTR

Pleiohomeotic (PHO), is the only protein from the PcG group, so far, to show direct DNA binding abilities in vitro. Its DNA-binding properties were first uncovered in studies aimed at identifying the protein functions mediating pairing-sensitive silencing of transgenes carrying a 176 bp fragment from the control region of the *Drosophila* engrailed gene (Brown, *et al.*, 1998). Deletion of a 17 bp sequence from within the 176 bp fragment, abolished its trans-silencing abilities. The protein that specifically recognised the 17 bp sequences was identified and found to be the product of the pleiohomeotic gene. Comparison of the PHO aminoacid sequence with the data base revealed its homology to the mammalian transcription factor Yin Yang-1 (YY1) (Brown, *et al.*, 1998).

YY1 is a ubiquitous and multifunctional transcription factor that can act as either an activator or a repressor in a promoter-specific manner (Shi *et al.*, 1997). YY1 has been implicated in regulation of viral and cellular promoters and is believed to play

an important role in the regulation of promoters that are located downstream of the transcription start site. Amongst the internal promoters bound by YY1 is that of the human LINE element *L1Hs* (Hariharan *et al.*, 1991; Flanagan *et al.*, 1992). The *L1Hs* promoter is part of the 900 bp long 5'UTR of these elements. The first 100 or so bp of the *L1Hs* 5'UTR appear to be crucial for *L1Hs* expression since deletion of this sequence results in a 300-fold reduction in transcription in culture cells. YY1 appears to interact specifically with the first 40 nucleotides of the *L1Hs* 5'UTR. Many of the promoters recognised by YY1 contain the conserved consensus 5'-CCATNTT-3', which, where present, marks the binding site for YY1. The CCATNTT consensus has been found also in the first 40 nucleotides of the L1 5'UTR at position 21-13 on the bottom strand.

The YY1 protein has distinct functional domains that can account for its multifunctional behaviour. The N terminus of this protein contains two acidic regions that are associated with transactivator function (Bushmeyer *et al.*, 1995). This is followed by a spacer region with modulatory function and by a four zinc finger domain that is essential for both protein-DNA and protein-protein interactions. The YY1 zinc finger domain appears to mediate most of the transcriptional repressor activity of this protein (Galvin and Shi, 1997). The similarity between PHO and YY1 is restricted to the 118 aminoacid long region that encodes the four zinc finger domain, suggesting that perhaps the function of PHO may be limited to repression, in keeping with the role attributed to other members of the PcG group (Brown, *et al.*, 1998). The presence within the PHO protein of a conserved zinc finger domain that can potentially mediate specific DNA and protein interactions, coupled with the experimentally proven DNA-binding ability of PHO, have led to the suggestion that this protein may act as a recruiting factor for PcG proteins at PREs. It was therefore of particular interest in the context of the present work to determine the presence of a PHO binding consensus, even more so in the light of the fact that the consensus has been found in the 5' UTR of *L1Hs*.

A homology search of the *I* factor 5' UTR was therefore carried out using the CCATNTT sequence. Only one imperfect match was found on the entire 186 bp UTR at nucleotide positions 6-12 of the 5' UTR on the upper strand, giving the sequence **CCAcTTC**. The small case letters indicate departures from the main core consensus CCAT as defined by Brown, and co-workers (pers. comm). When the search was extended to the entire *I* factor sequence, two additional putative PHO binding sites were found, both within ORF2: one, at position 1589-1595, has sequence **CCATCTC** and one with sequence **CCATTTC** at position 2140-2146.

The implications for the presence and location of these putative PHO binding sites are discussed below.

5.2.2. Is Pc protein recruited to the site of multiple I factor 5' UTR insertions ?

Pal-Bhadra *et al.*, 1997, found that Pc is specifically recruited to the site of suppressed transgenes (high dosage) as determined by polytene immunostaining. If, however, the same transgene insertions are not suppressed (low dosage), no Pc binding is detected (Chapter 4). The next important question, therefore, was to determine whether Pc is recruited at the site of "silencing" 5' UTRs of the *I* factor carried by the 380 strain (Chapter 4). As shown in figure 4.2.1-2, Chapter 4, the four insertions of constructs 186x2 and 186x3, had been genetically assigned to chromosome 2 and 3 (Chaboissier *et al.*, 1998). For precise mapping they needed to be assigned to the arms of polytene chromosomes, where the banding patterns permit precise localisation.

Mapping of 186x2 and 186x3 constructs on polytene chromosomes

Chromosome preparations from salivary glands of 380 third instar larvae, were hybridised with a mixture of digoxygenin labelled probes (Chapter 2, section 2.2.6) derived from the *yellow* and *white* genes that mark the 186x2 and 186x3 constructs respectively. Detection of hybridisation bands required absorption of anti-digoxygenin antibodies linked to alkaline phosphatase followed by addition of substrate (Chapter 2, section 2.1.1). Hybridisation was expected on six distinct sites: two on the X chromosomes where the endogenous yellow and white gene reside, two on chromosome 2 and two on chromosome 3. As shown in **plate I, panel D** only five bands could be clearly localised. The two close to the tip of the X chromosome (yellow arrow heads) identify the polytene regions 1B and 3 C of this chromosome where the endogenous *yellow* and *white* genes are located respectively. These sites are clearly visible also in **panel E** and in **panel F** although they are not marked in the latter because of crowding of the chromosome arms. Two bands representing insertions of the 186x2 and 186x3 transgenes are visible on the distal end of the left arm of chromosome 2 (2L, red arrow heads) and these are indicated in panels **D**, **E**, **F**. A close up of the 2L arm is shown in **plate II-panel A** and higher magnification shots from different preparations are shown in **panels B** and **C** of the same plate. **Panel C** in **plate II** also shows the exact location of the insertions at polytene bands

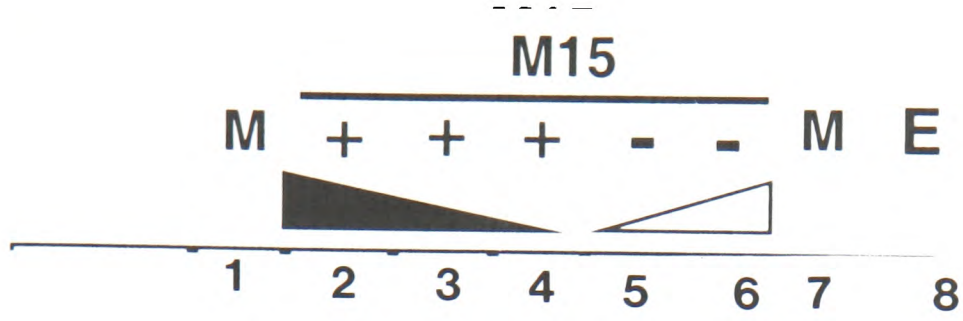
25F/26A and 27C/D. Since the transgene insertions were hybridised to equimolar mixtures of the yellow and white probes, it is not possible to deduce the order of the 186x2 and 186x3 insertions along the 2L arm. The 25F/26A and 27C/D insertions are visible also in **plate VIII, top panel** (red arrow heads) where the tip of chromosome 2 is to the right and therefore the transgene insertions follow in numerical order from right to left.

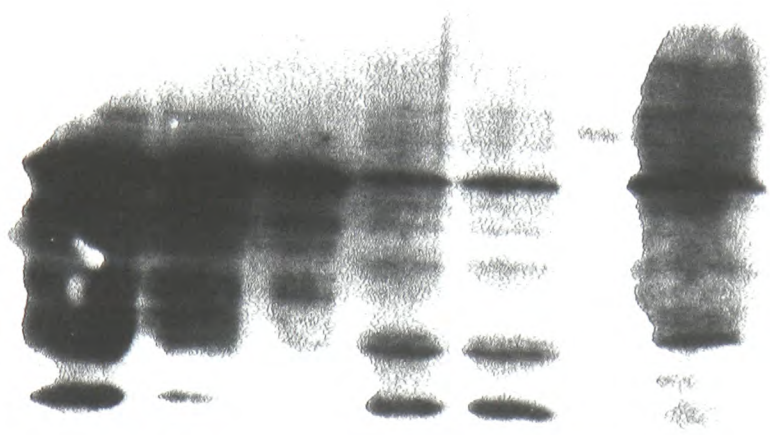
The only other visible hybridisation band identifies transgene insertions in the proximal end of the right arm of chromosome 3 shown in **plate *I*, panel D** and **plate II, panel B** (green arrow heads). Higher magnification images from two different preparations also shown in **plates VIII and IX, top panels** (red arrow heads). The exact location of the 3R insertion is 84B/C. The 25F/26/A, 27C/D and 84B/C insertions are the only ones (together with the 1B and 3C bands) that were consistently visible in all preparations. Additional hybridisation bands were on occasion seen on the distal end of 3R and even on the proximal end of the X chromosome. But these represented isolated sightings that were not reproducible and were not pursued further. The insertions that gave reliable signals map to the chromosomes to which they were assigned through crosses with strains carrying marked balancer chromosomes (Chaboissier, *et al.*, 1998) thus confirming their genetic location. The simplest explanation as to why one of the expected insertions on chromosome 3 is not visible is that the only detectable hybridisation band on 3R actually identifies two closely linked insertions.

Immunostaining of polytene chromosomes with anti-Pc antibodies

Preliminary attempts to determine whether Pc protein co-localises to the sites of transgene insertions were made by polytene chromosome immunostaining using a non-fluorescent detection method (Chapter 2, section 2.2.6). The affinity purified anti-Pc antibodies (Chapter 2, section 2.2.6) to be used were first tested by Western blotting against bacterial extract from induced and uninduced cells from the M15 bacterial strain (Chapter 2, section 2.1.2). This strain carries a plasmid encoding the portion of Pc protein against which the antibodies were raised. As a positive control, extracts were prepared also from 18-22 hour old embryos (Chapter 2, section 2.2.8), where high levels of Pc protein are detected (Zink and Paro, 1989). The results are shown in figure 5.2.2-1. Lanes 2 to 4 have decreasing amounts of crude extracts from induced M15 cells, whilst lanes 5 to 6 have increasing amounts of uninduced M15 cells. The molecular weight markers in lanes 1 and 7 were loaded as reference but they cannot help identify Pc protein since the latter migrates abnormally on SDS-

Figure 5.2.2-1: Western blot analysis using affinity purified Anti-Pc antibodies. Lanes M 1 and 7: prestained molecular weight marker; lanes 2 to 4, crude extracts from IPTG-induced M15 cells, expressing recombinant LacZ-Pc fusion, decreasing amounts of extract have been loaded in lanes 2 to 4; lanes 5 to 6: leaky expression of LacZ-Pc fusion from uninduced M15 cells; lane E 8 : crude extracts from 18-22 hour-old embryos at the time of maximum Pc expression. Pc indicates the main Polycomb isoform whilst Pci indicates a minor isoform not detected in recombinant bacterial cells. The 44 K Pc protein migrates with an apparent size of approximately 66 K on SDS-PAGE.





◄ Pc

◄ Pci

PAGE (Zink and Paro, 1989). The main band in embryo extracts (labelled Pc in lane 8), identifies the Pc product, seen as comigrating bands of decreasing intensity in extracts from induced M15 cells (lanes 2 to 4). Fainter comigrating bands are detected also in extracts from uninduced M15 cells, perhaps due to leaky expression of the lacZ-Pc fusion transgene. The minor Pc band visible in embryo extracts (labelled Pci in lane 8) has no counterparts in bacterial extracts and most probably represents one of the common Pc isoforms (Zink and Paro, 1989).

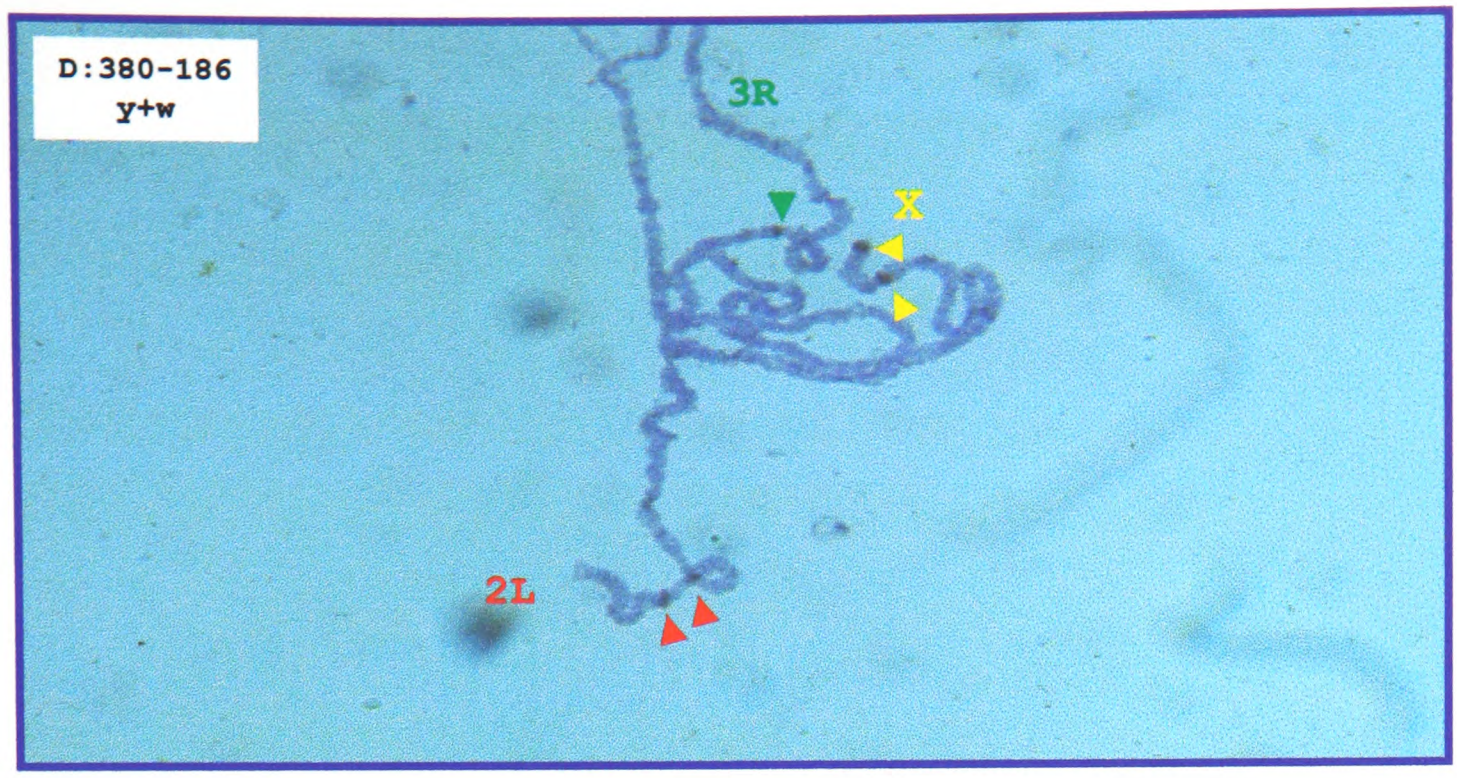
Since the chemical treatment that renders the chromosomes suitable for DNA *in situ* hybridisation also removes most of the main chromosome antigens, and vice versa, the procedures to detect protein antigens render the chromosomes poor substrates to hybridisation of DNA probes, probe and antibody generated signals cannot be detected on the same preparations. Thus, anti-Pc antibodies had to be absorbed to fresh chromosome preparations from salivary glands of 380 larvae. As a control, it became necessary to map polycomb sites to the regions of interest also on chromosomes from the genetically equivalent Ja strain, which, however, lacks transgene insertions. Prior to antibody absorption, the chromosome squashes were treated with formaldehyde, which crosslinks protein antigens to the surface of chromosomes so that they are not lost during the numerous manipulations (chapter 2. section2.2.6). Secondary antibodies linked to Horse Radish Peroxidase (HRP) were then absorbed and addition of substrate (Chapter 2, section 2.2.6) results in appearance of red/brown bands at the sites of hybridisation.

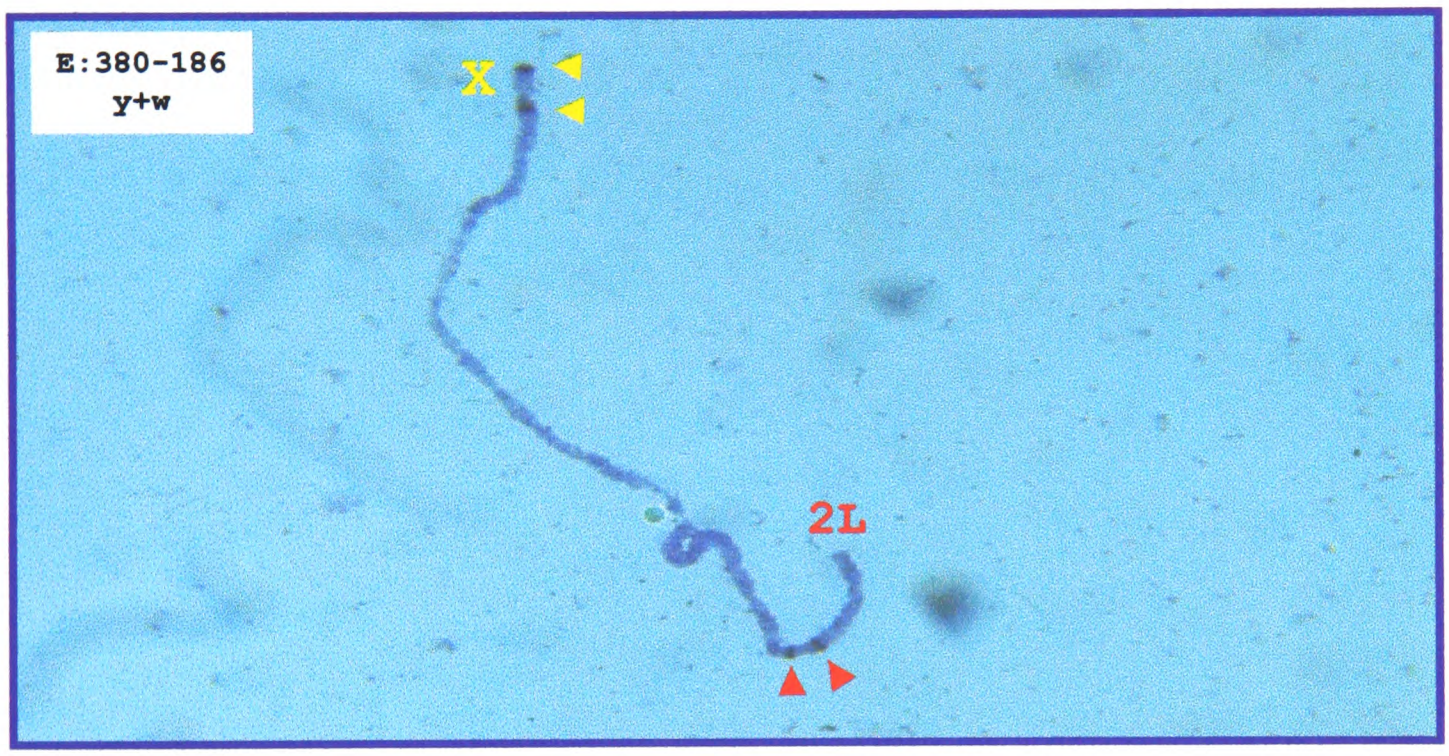
The Pc binding patterns on polytene chromosomes from the 380 strain, detected by HRP-linked secondary antibodies are shown in **plates III** and **IV panels A** and **D** (labelled 380-186 anti-Pc). The equivalent binding patterns detected on polytene chromosomes from the Ja strain are shown in **plate III, panels B** and **C** and in **plate IV, panel E**. Where appropriate, chromosome arms are identified. The resolution of bands stained with the products of HRP was certain localisation of fainter Pc bands in the regions of interest on 2L (bracketed in red in preparations from both 380 and Ja, where resolution did not permit unambiguous identification of bands) and 3R. The fact that Pc maps to over 100 sites on polytene chromosomes (Zink and Paro, 1989) did not render the task any simpler. Preliminary evidence suggested, however, that strong Pc bands were present on Ja chromosomes at position 25F/26A (**plate III, panel C**) and 84B/C (**plate III, panel B** indicated by a green arrow head). This indicated that at least two of the transgenes had inserted in close proximity of pre-existing Pc binding sites on polytene chromosomes. Resolution of Pc bands could be improved by increasing the contrast between the colour of the Pc bands against the

chromosomal background. This could be achieved using immunofluorescence. However, the information gained from careful examination of Pc patterns visualised with HRP could be used to ease identification of the 2L and 3 R chromosome arms. Proceeding from the distal end and moving down the two chromosome arms, strong Pc binding was detected at the telomeres of both 2L and 3R and numerous "landmark" Pc bands were detected.

Chromosome squashes from both the 380 and Ja strains were prepared for antibody absorption following a procedure very similar to that used for HRP detection, the main difference being that the secondary antibodies were linked to biotin rather than to HRP (Chapter 2, section 2.2.6). Detection involved binding of avidin conjugated to the fluorescent dye Texas red (Chapter 2, section 2.2.6) to the antibody-biotin complexes. The chromosomes were then visualised by DAPI staining. The 25F/26A and 27C/D bands were more easily visualised because the proximal region in which they are located is more frequently in an outstretched configuration that permits easy reading of polytene bands. Pc bands in regions 25F/26A and 27C/D were compared between 380 (**plate V, top panel, plate VI, panels A** and **B** and **plate VII top panel**) and Ja (**bottom panels** in **plates V, VI** and **VII**) chromosomes. Inspection of the fluorescent stained chromosomes revealed Pc bands in correspondence of the 25F/26A and 27C/D regions of 2L, thus suggesting that both of the insertions on 2L have inserted in close proximity of or overlapping pre-existing site of Pc binding sites in this region. Co-localisation with pre-existing Pc bands was confirmed also for the 84B/C transgene insertion by direct comparison between the hybridisation band detected with the yellow/white probes on 3R of 380 chromosomes (**plates VIII** and **IX, top panels**) and the corresponding region on Ja chromosomes stained for fluorescent detection of Pc bands (**plates VIII** and **IX bottom panels**). Within the limits of resolution, a site of Pc binding is detected in the 84B/C region of Ja chromosomes which corresponds to the location of the transgene on 380 chromosomes. The co-localisation of 186x2 and 186x3 constructs with pre-existing Pc binding sites on polytene chromosomes, is in a sense unfortunate because it does not give a clear cut answer to the question of whether multiple copies of the *I* factor 5' UTR recruit PcG complexes at their sites of insertion. However, given the unusual ability of sequences with a PRE function to interact with each other in trans, the present findings could still provide evidence of PcG involvement in *I* factor control, a discussed below.

Plate I, panels D, E and **F**: DNA in situs on polytene chromosomes from third instar larvae of the 380 strain. The probe was an equimolar mixture of digoxygenin (DIG) labelled fragments the from coding region of the yellow and white genes, which detects the homologous sequences linked to the 186x2 and 186x3 constructs as well as the endogenous sequences. Detection was by absorption of Alkaline phosphates linked anti-DIG antibodies. 3R: right arm of chromosome 3 , the arrow head points to hybridisation signal (green arrow head); X chromosome and the position of the endogenous yellow and white genes are indicated (yellow arrow heads); 2L: the left arm of chromosome 2 , red arrow heads indicate the position of two hybridisation signals on this chromosome arm.


D:380-186
y+w
3R
X
2L


E:380-186
y+w
X
2L

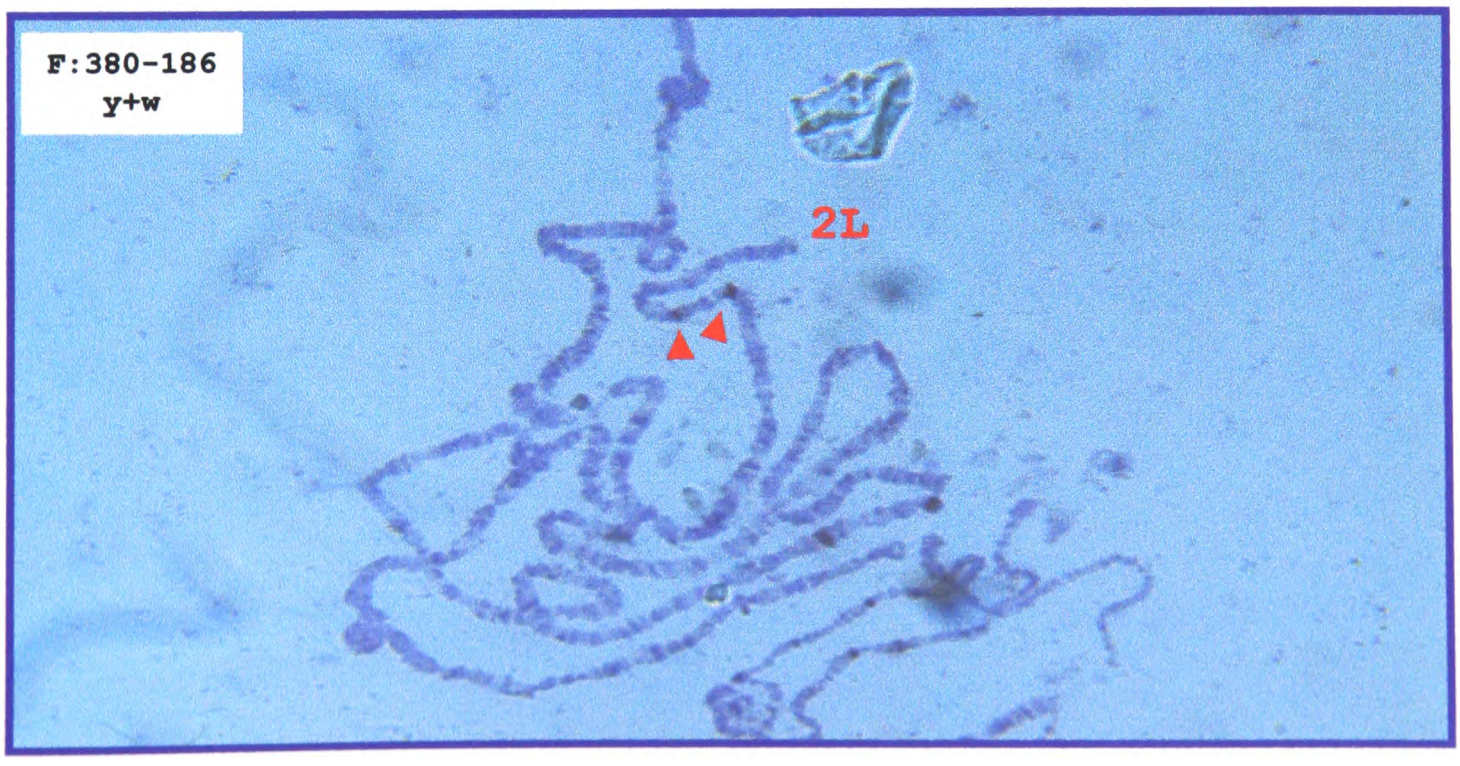
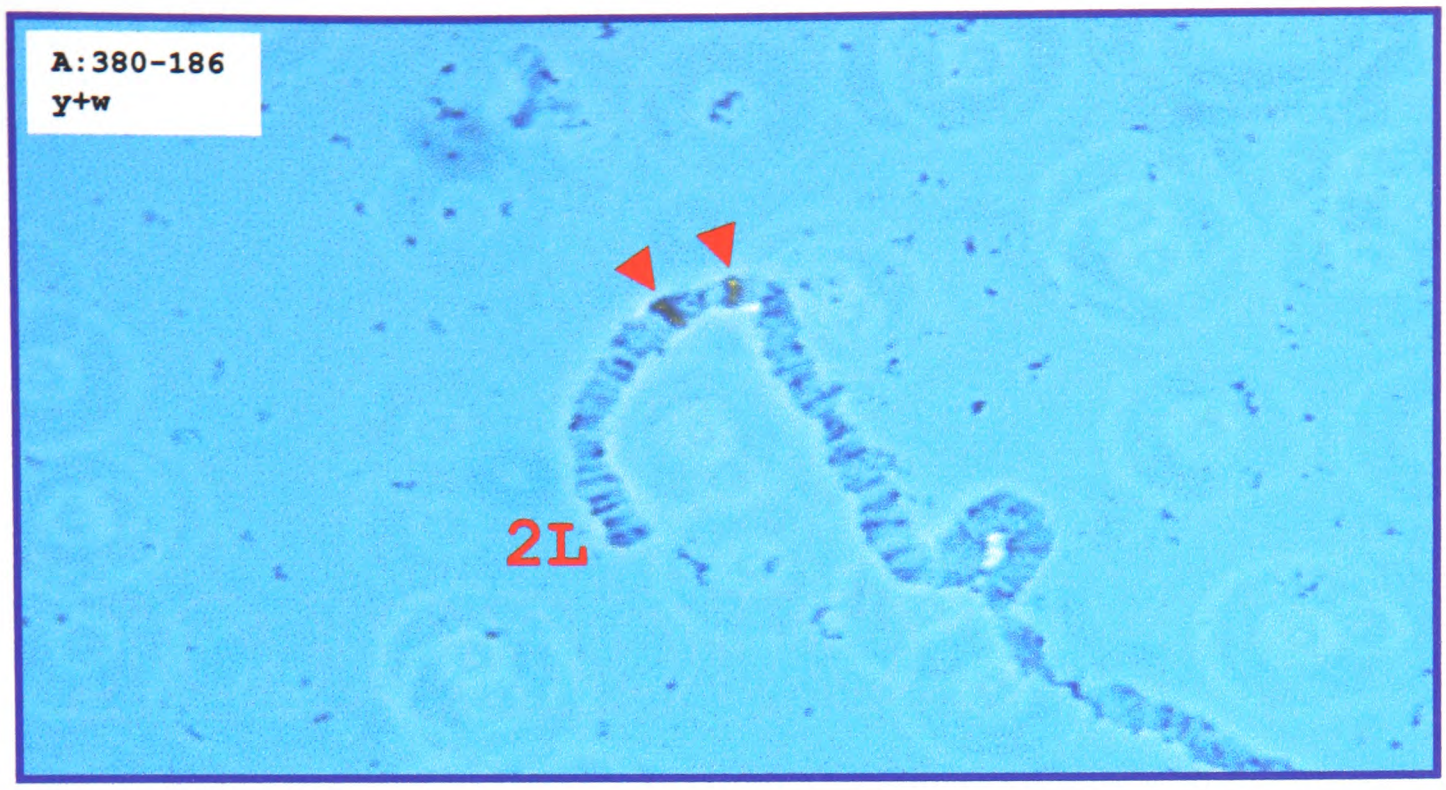
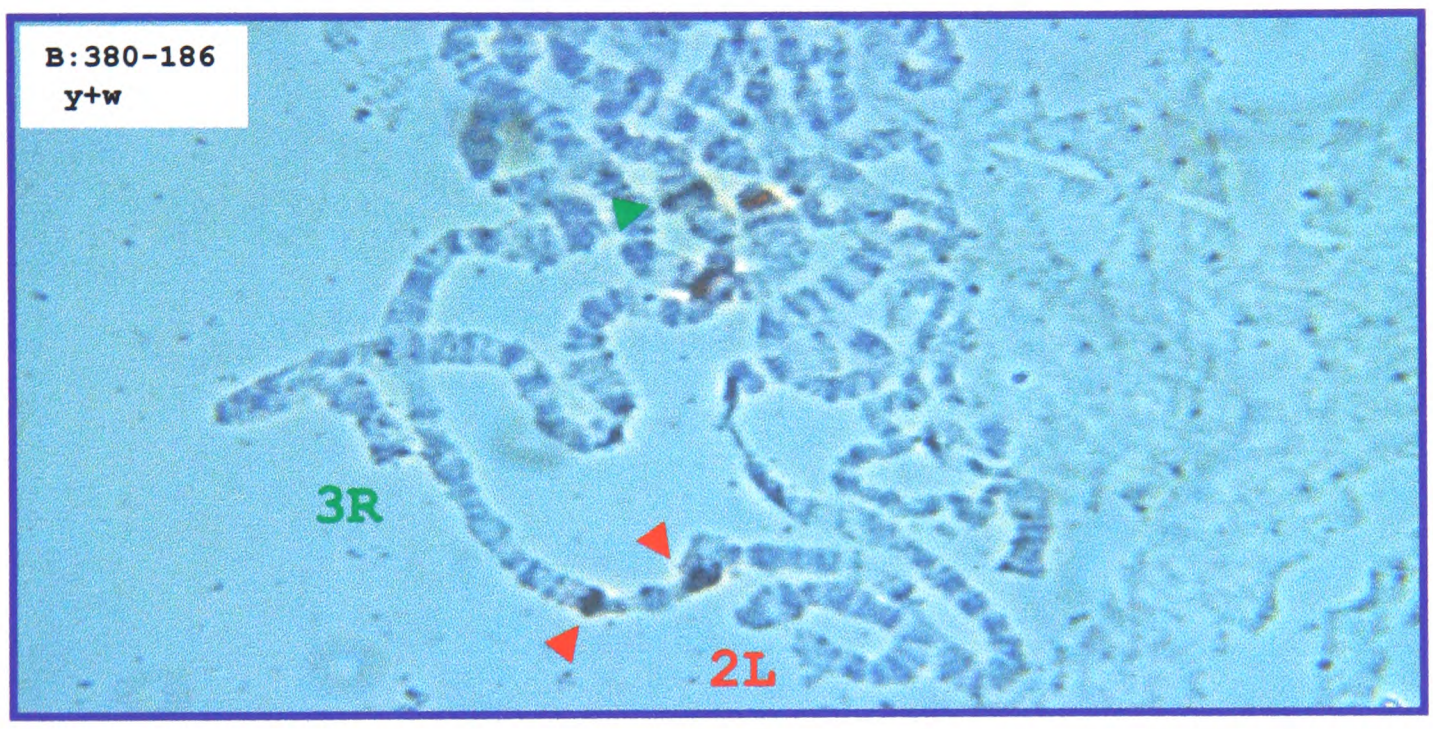

F:380-186
y+w
2L

Plate II, panels **A**: high magnification image of 2L shown in plate I, **panels B** and **C**: high magnification of 2L and 3 R (**B**) and 2L only (**C**). The exacts location of the 2L insertions is indicated (red arrow head panel **C**): from distal to proximal they are 25F/26A and 27C/D.


A:380-186
y+w
2L


B:380-186
y+w
3R
2L

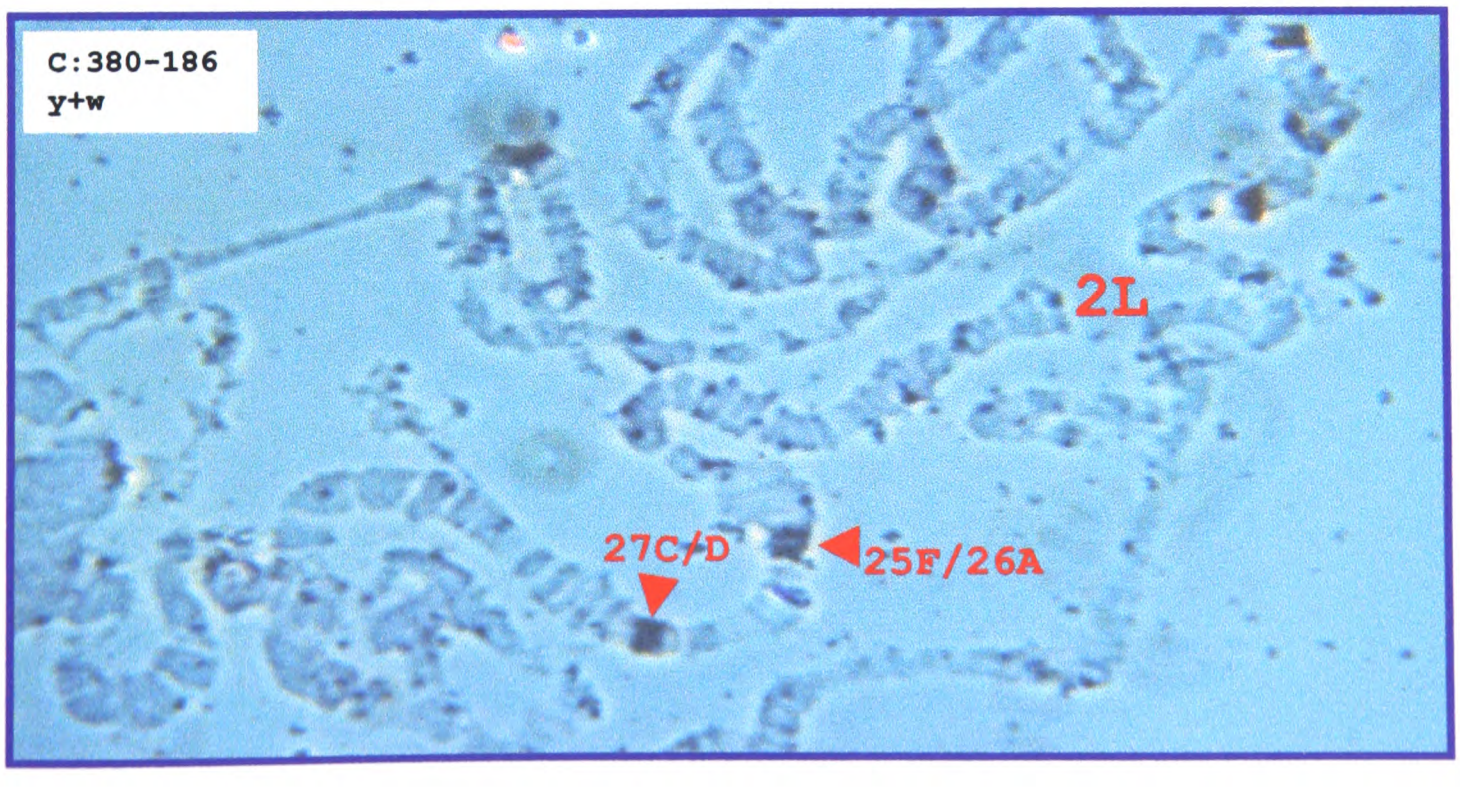
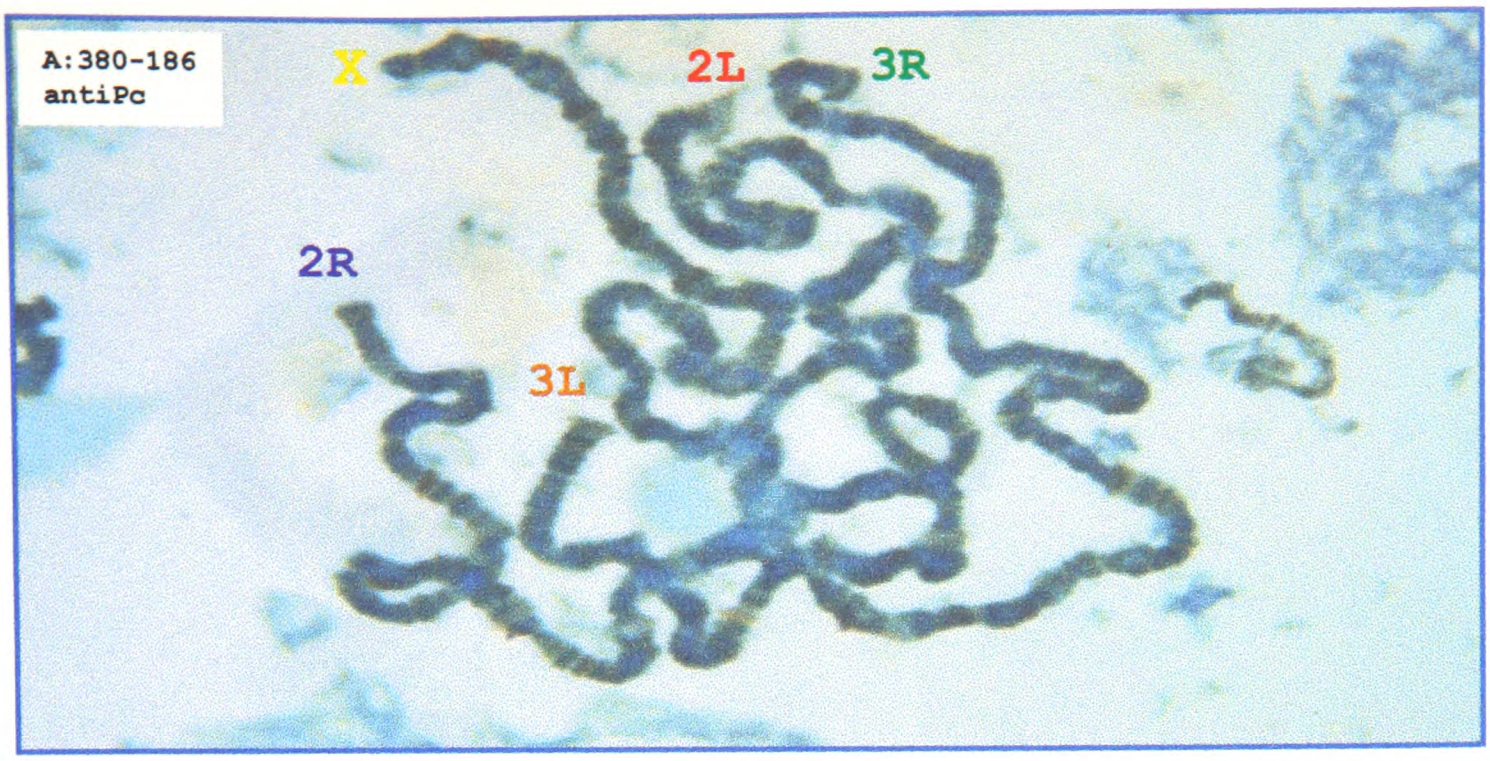
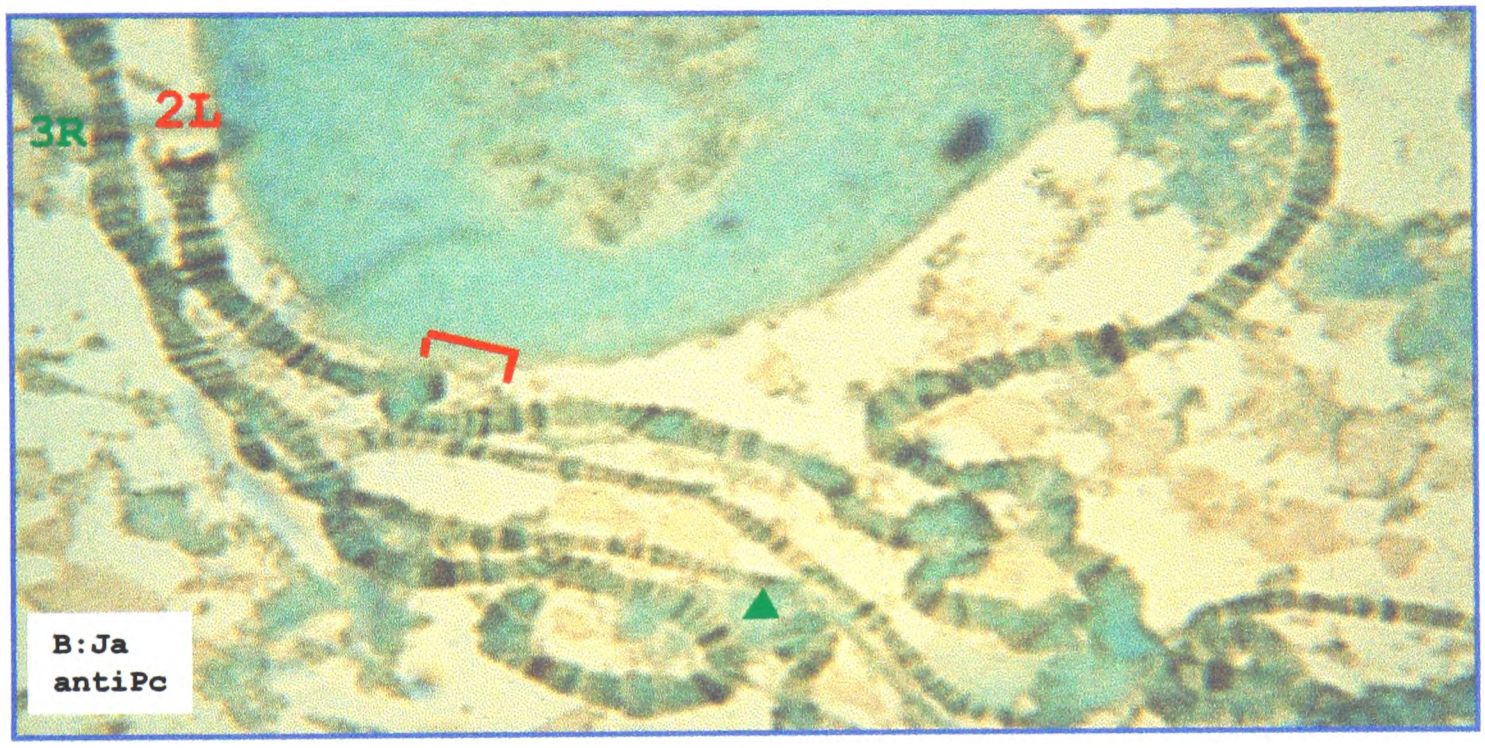

C:380-186
y+w
2L
27C/D
25F/26A

Plate III: Immunostaining of polytene chromosomes with affinity purified anti-Pc antibodies and detection by absorption of Horse-Radish-Peroxidase conjugated secondary antibodies. Numerous Pc-binding sites are detectable. The identity of chromosome arms is indicated. **Panel A**: chromosome from larvae of the 380 strain: 2R: right arm of chromosome 2; 3L left arm of chromosome 3; X chromosome; 2L: left arm of chromosome 2 3 R: right arm of chromosome 3. **Panels B** and **C**:: chromosome preparations from the Ja strain. The proximal end of 3R and the distal end of 2L are clearly visible in **panel B**. The region of transgene insertion is indicated by a green arrow head (3 R) and a red square bracket (2L). **Panel C**: Chromosome 2L and the region of transgene insertion are indicated. The X chromosome is also identified.


A:380-186
antiPc
X
2L
3R
2R
3L


3R
2L
B:Ja
antiPc

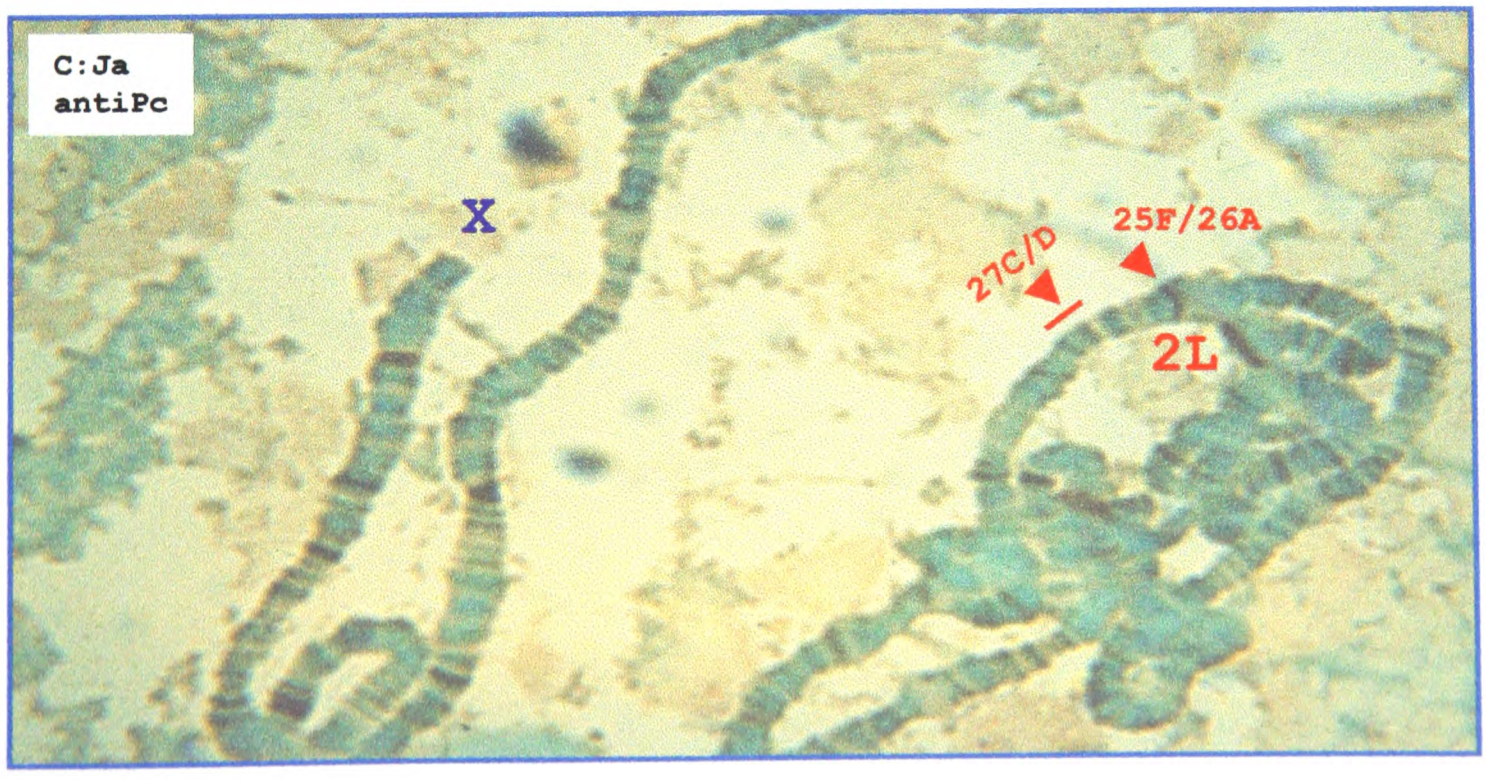
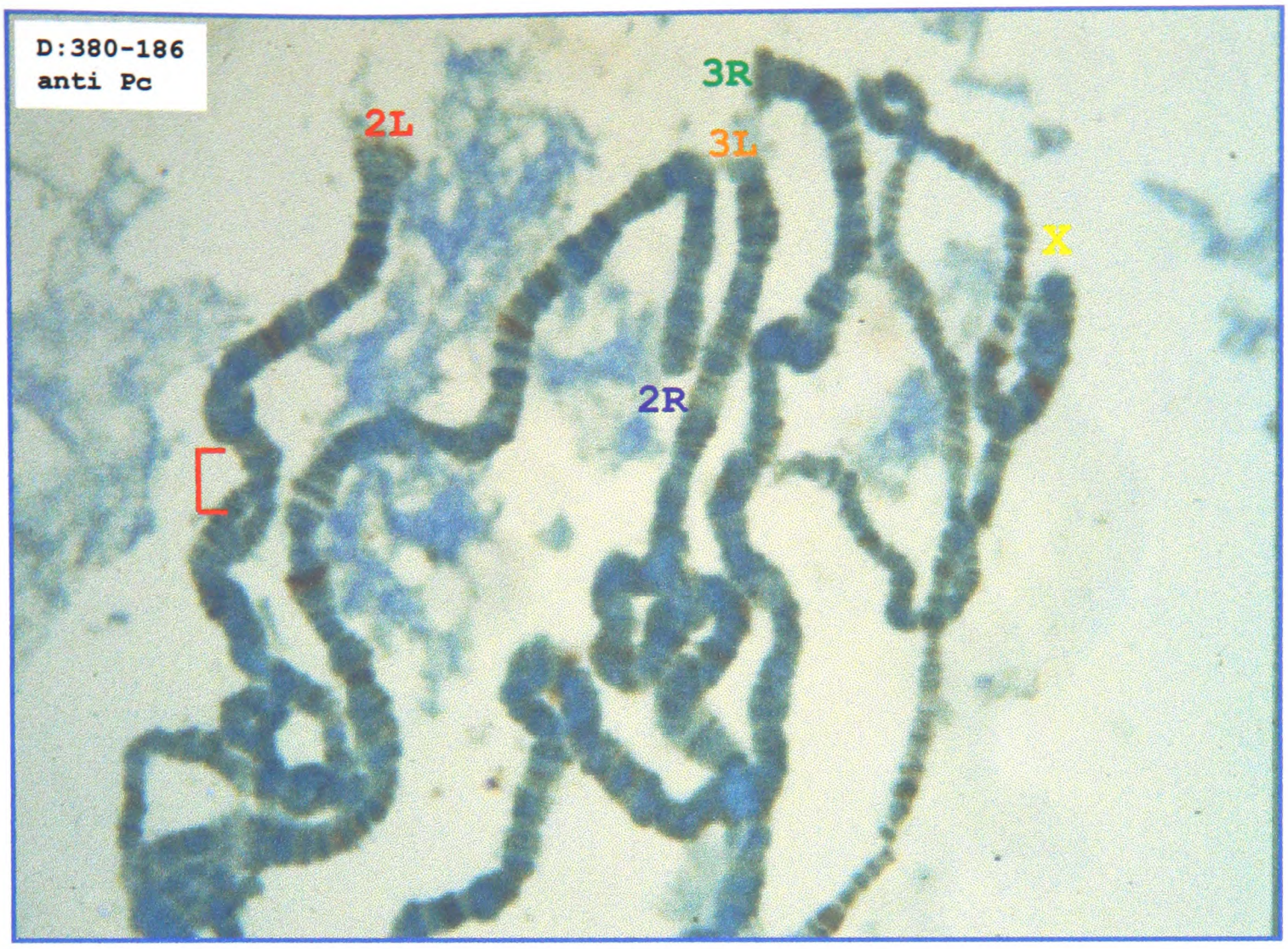

C:Ja
antiPc
X
25F/26A
27C/D
2L

Plate IV: Immunodetection of Pc binding site on polytene chromosomes. **Panel D**: 380 chromosomes. **Panel E**: Ja chromosomes. Notations are as in plate III.


D:380-186
anti Pc
2L
3R
3L
X
2R

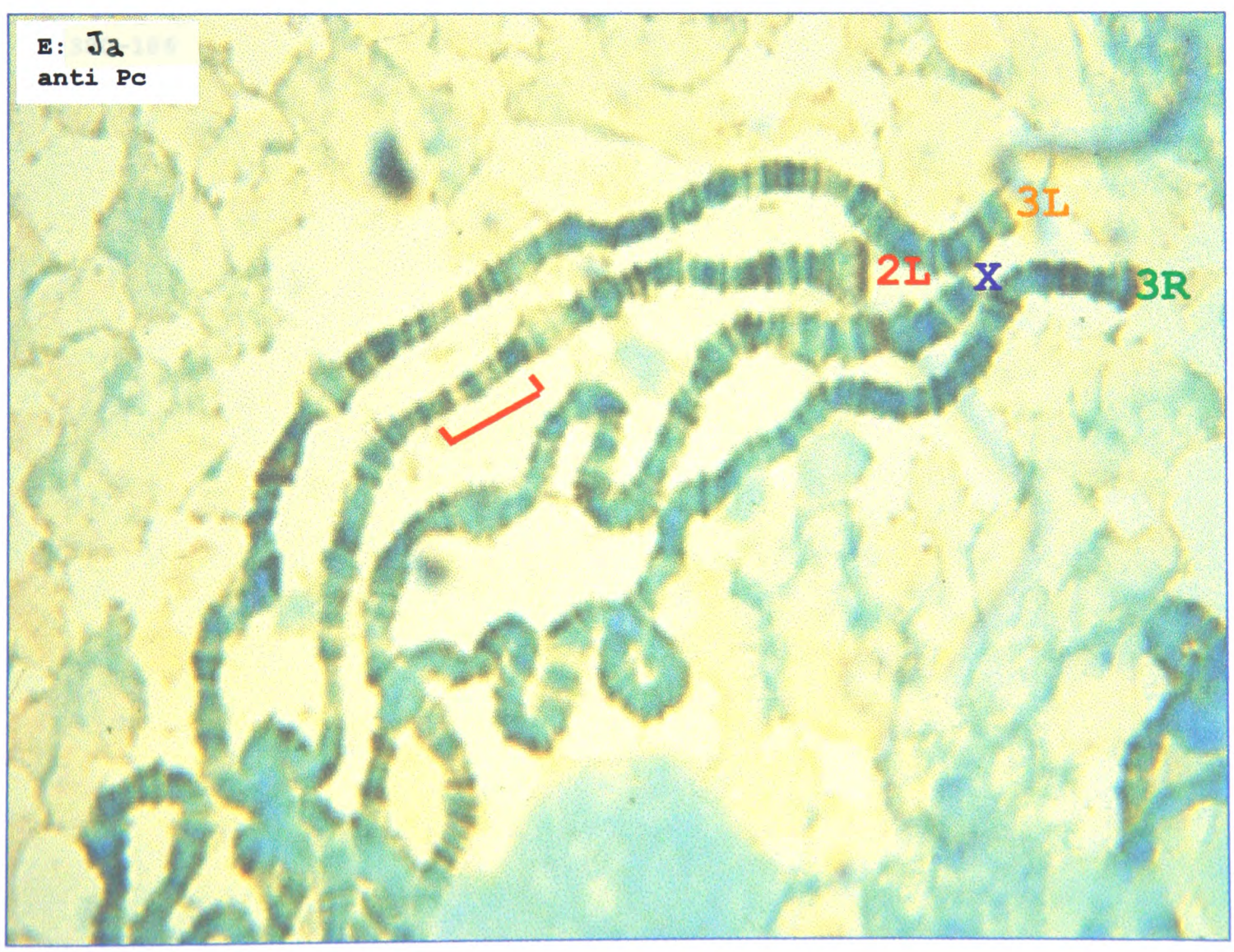
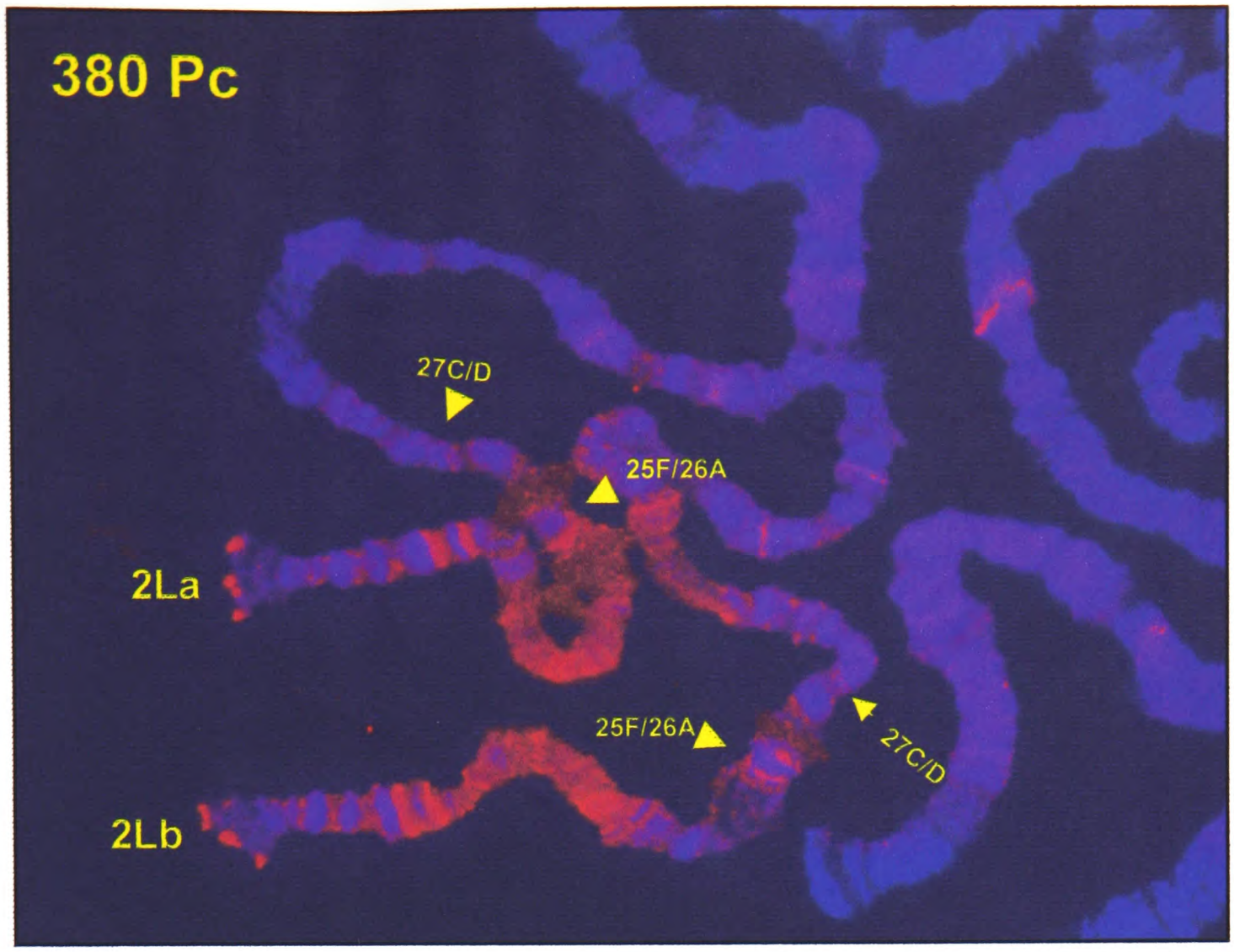

E: Ja
anti Pc
3L
2L
X
3R

Plate V: Fluorescent immunostaining of Polytene chromosomes from the 380 (top) and Ja (bottom) stains, for detection of Pc binding sites. The fluorescent red dye is linked to an avidin carrier that interacts with the biotinylated secondary antibodies. Primary antibodies are affinity purified anti-Pc antibodies. Chromosomes are stained with DAPI. In the top panel the site of transgene insertion are indicated for the unwound distal end of 2L in a 380 background. In the bottom panel, the same region in 2 L is identified and sites of transgene insertion are indicated. Sites of Polycomb binding exist in the Ja strain at positions 25F/26A and 27 C/D.


380 Pc
27C/D
25F/26A
2La
25F/26A
27C/D
2Lb

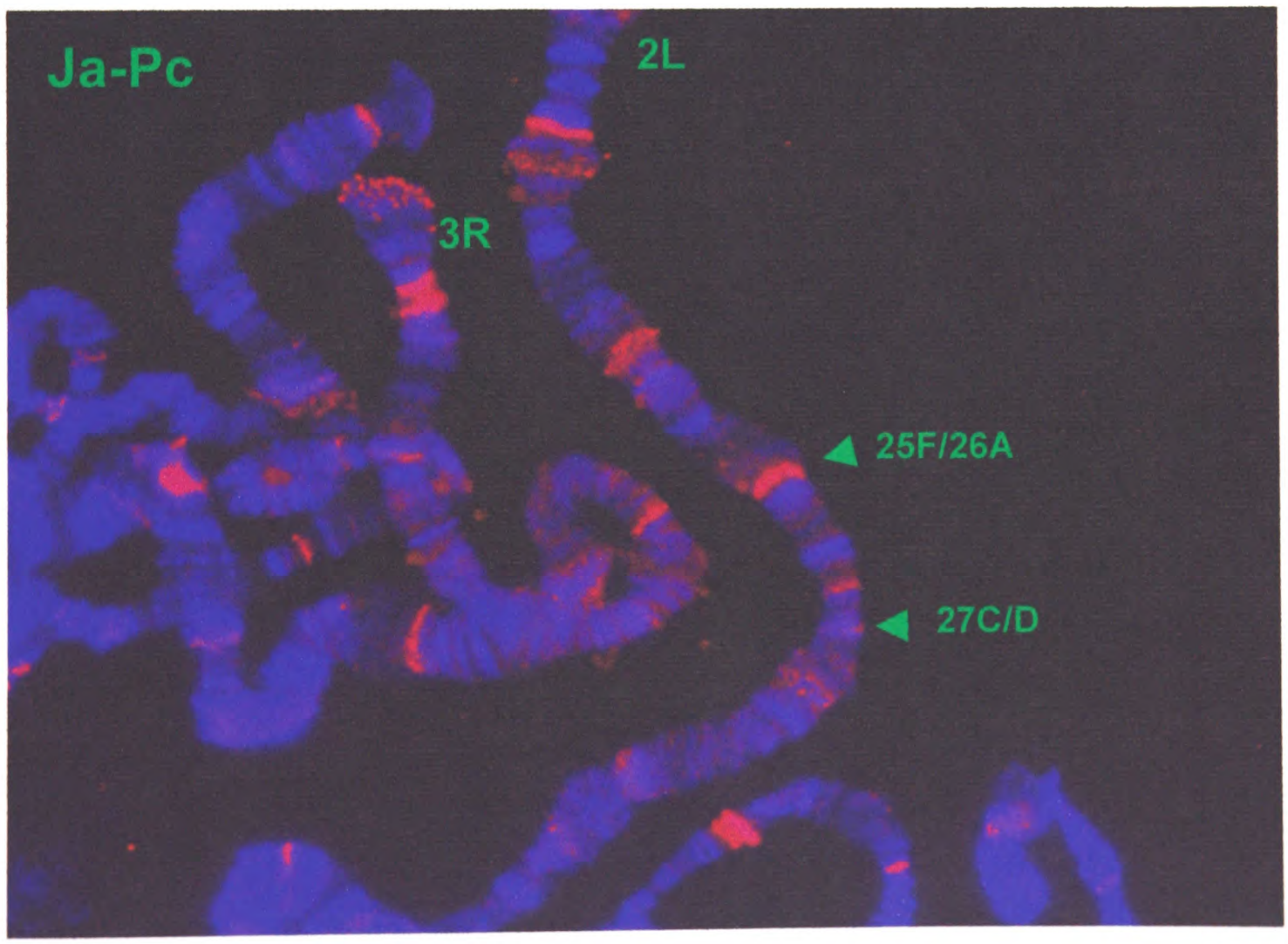
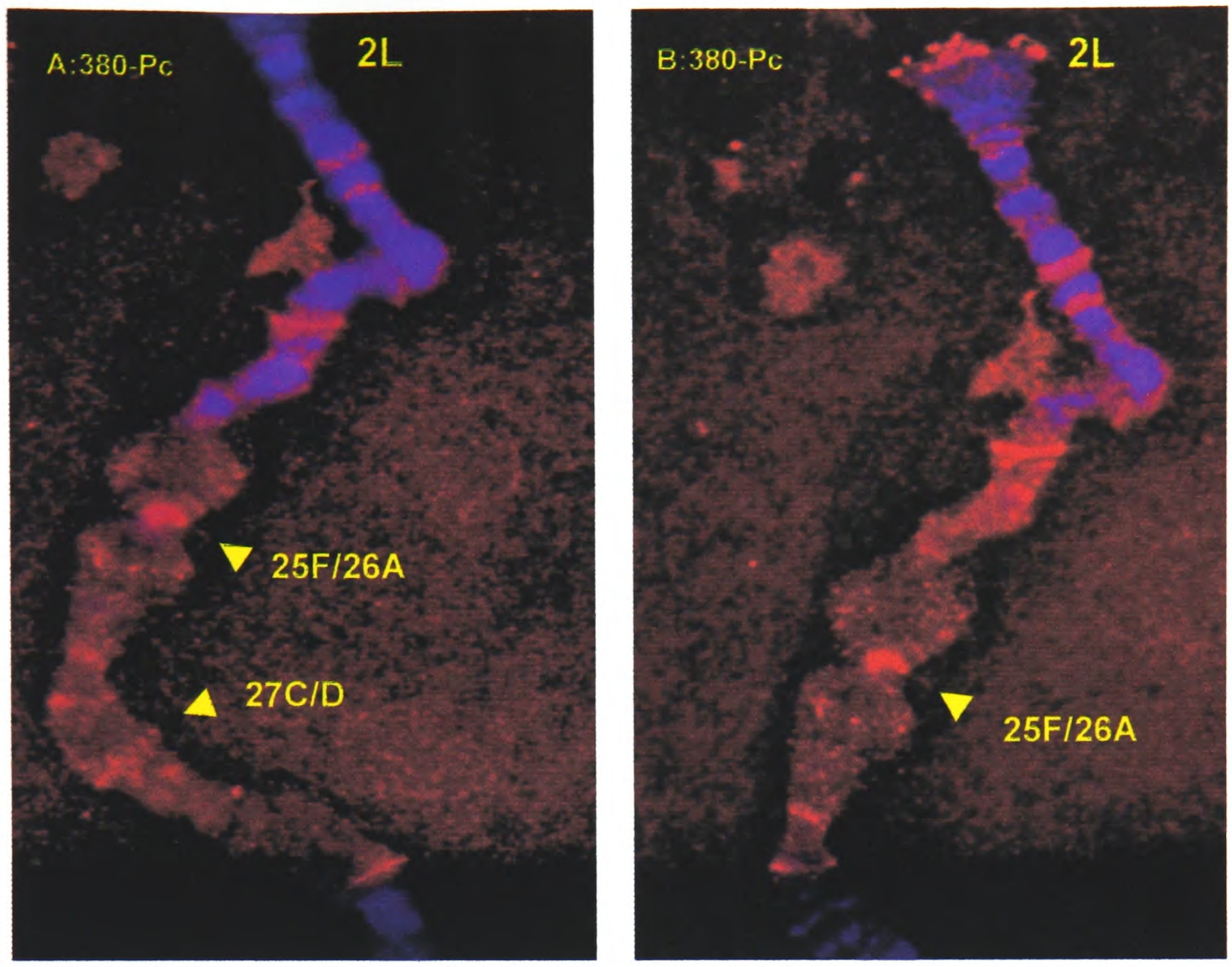

Ja-Pc
2L
3R
25F/26A
27C/D

Plate VI: Fluorescence immunostaining of polytene chromosomes from the 380 (top panels A and B) and Ja (bottom panel) strains, using affinity purified anti-Pc antibodies. High resolution images of 2L. The sites of transgene insertion on this chromosome arm are indicated. The strong Pc staining characteristic of the telomere region of 2L is visible (see also plates below).


A:380-Pc
2L
25F/26A
27C/D
B:380-Pc
2L
25F/26A

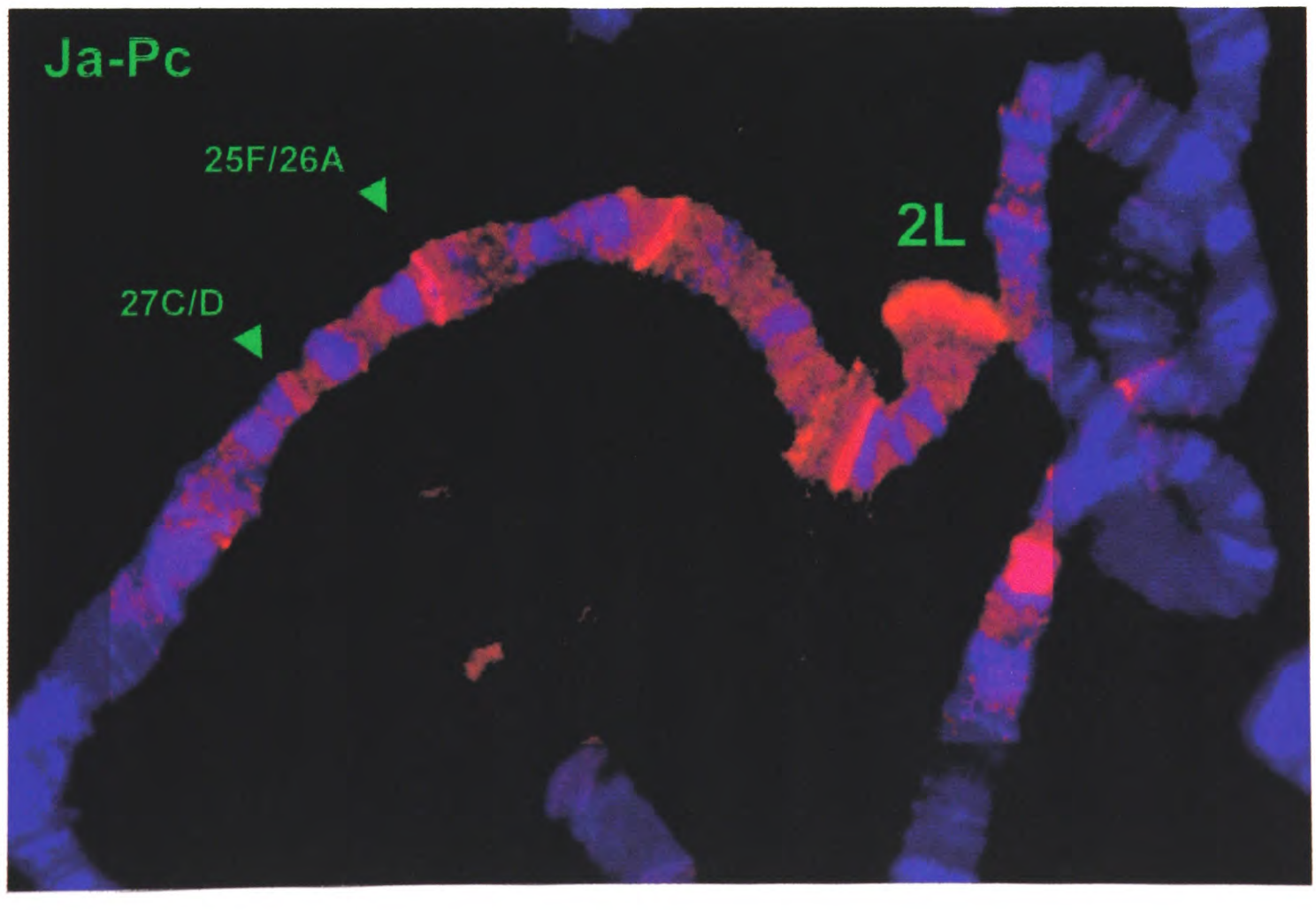
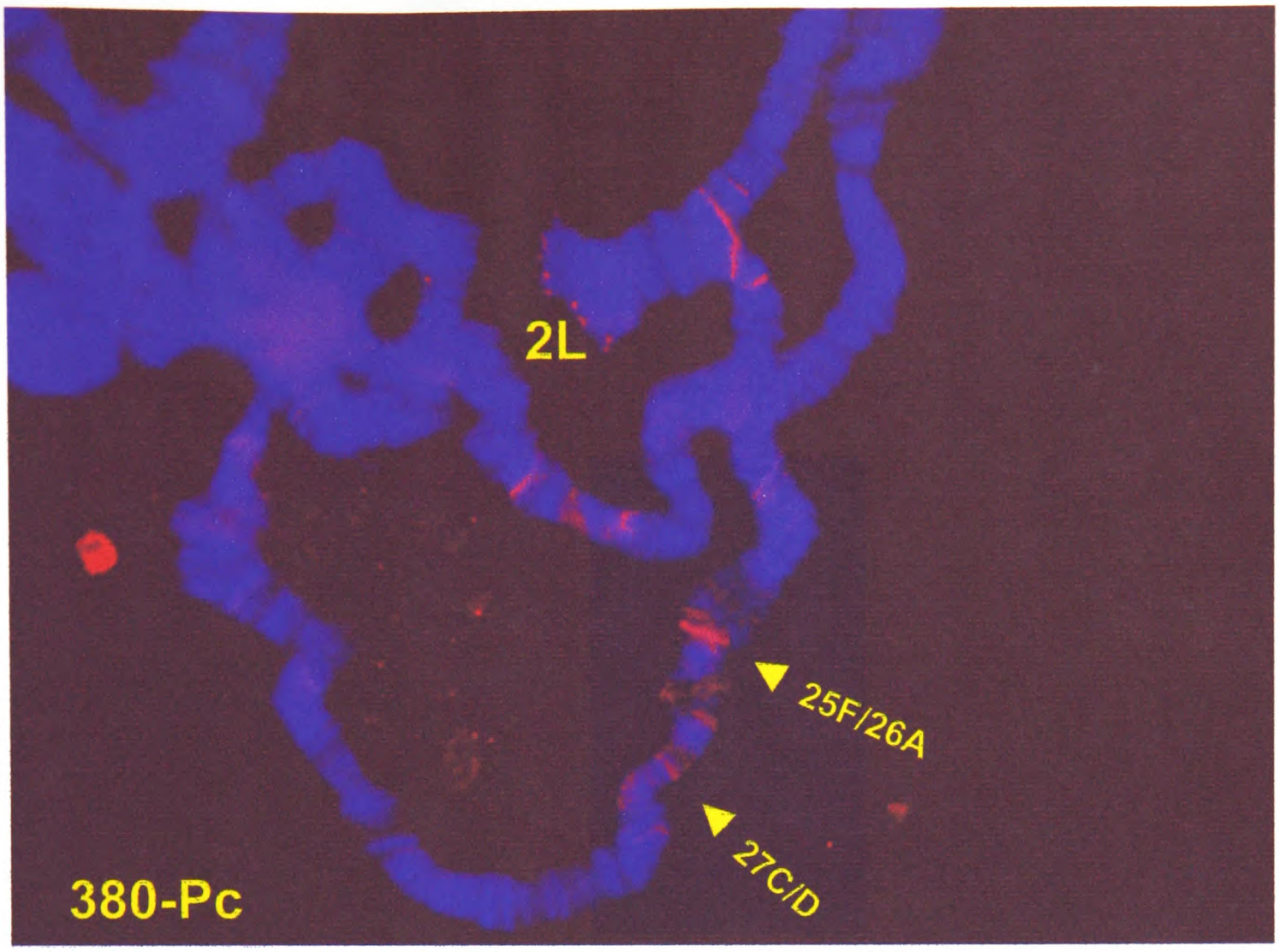

Ja-Pc
25F/26A
27C/D
2L

Plate VII: Fluorescence immunostaining of polytene chromosomes from the 380 (top panel) and Ja (bottom panel) strains, using affinity purified anti-Pc antibodies. High resolution images 2L showing the presence of Pc bands at the sites of transgene insertion in both strains.


2L
25F/26A
27C/D
380-Pc

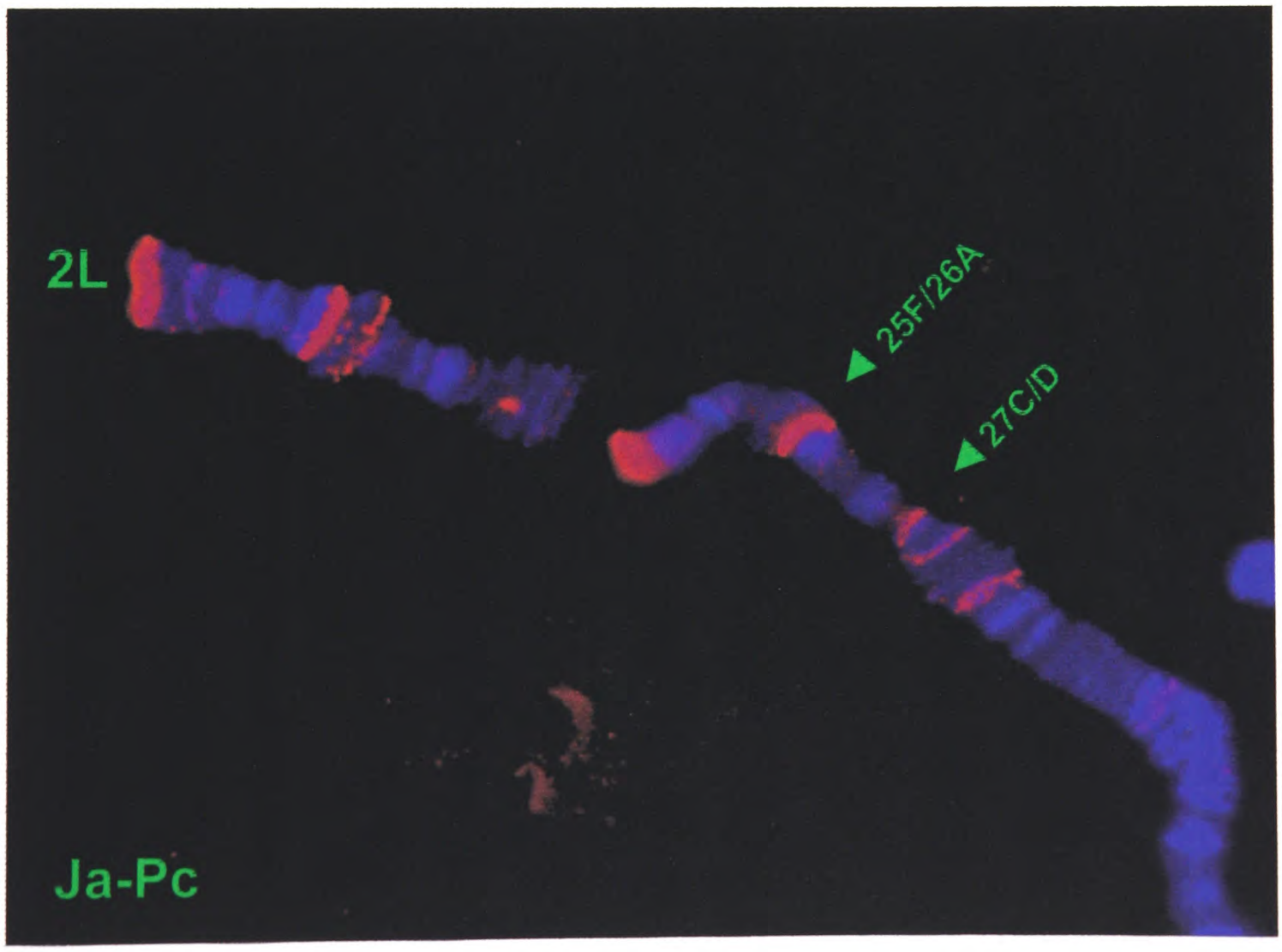
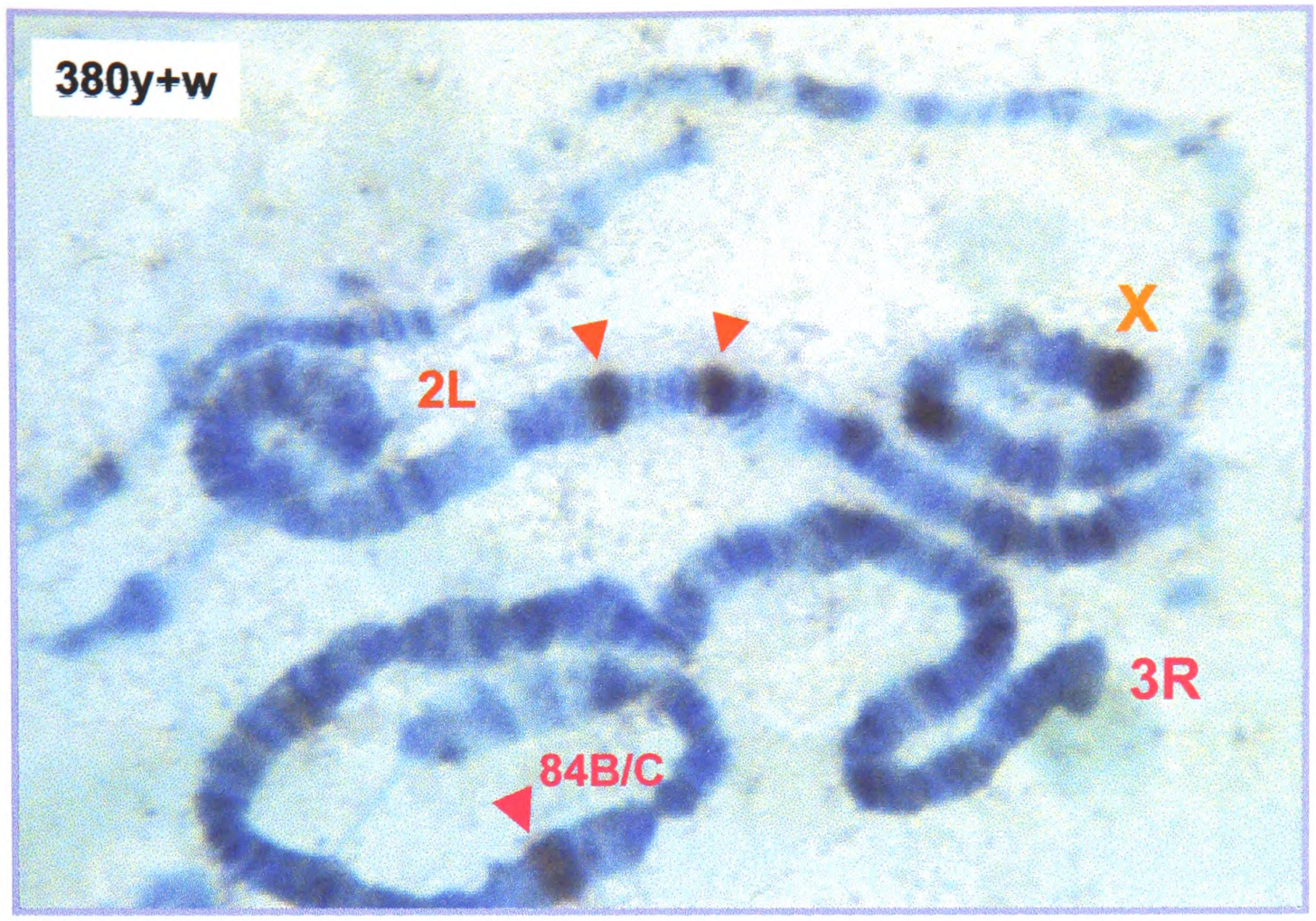

2L
25F/26A
27C/D
Ja-Pc

Plate VIII: Direct comparison of the site of transgene insertion detected by DNA in situ on 380 chromosomes and the corresponding region on Ja chromosomes stained with anti-Pc antibodies and fluorescent secondary antibody conjugates. Special attention is given to the 84B/C region on 3R, which coincides with a strong hybridisation signal on 380 chromosomes (top panel). A site of Pc binding is detectable on the same region on chromosomes from the Ja strain (bottom panel). Chromosome notations are as in previous plates.


380y+w
2L
X
3R
84B/C

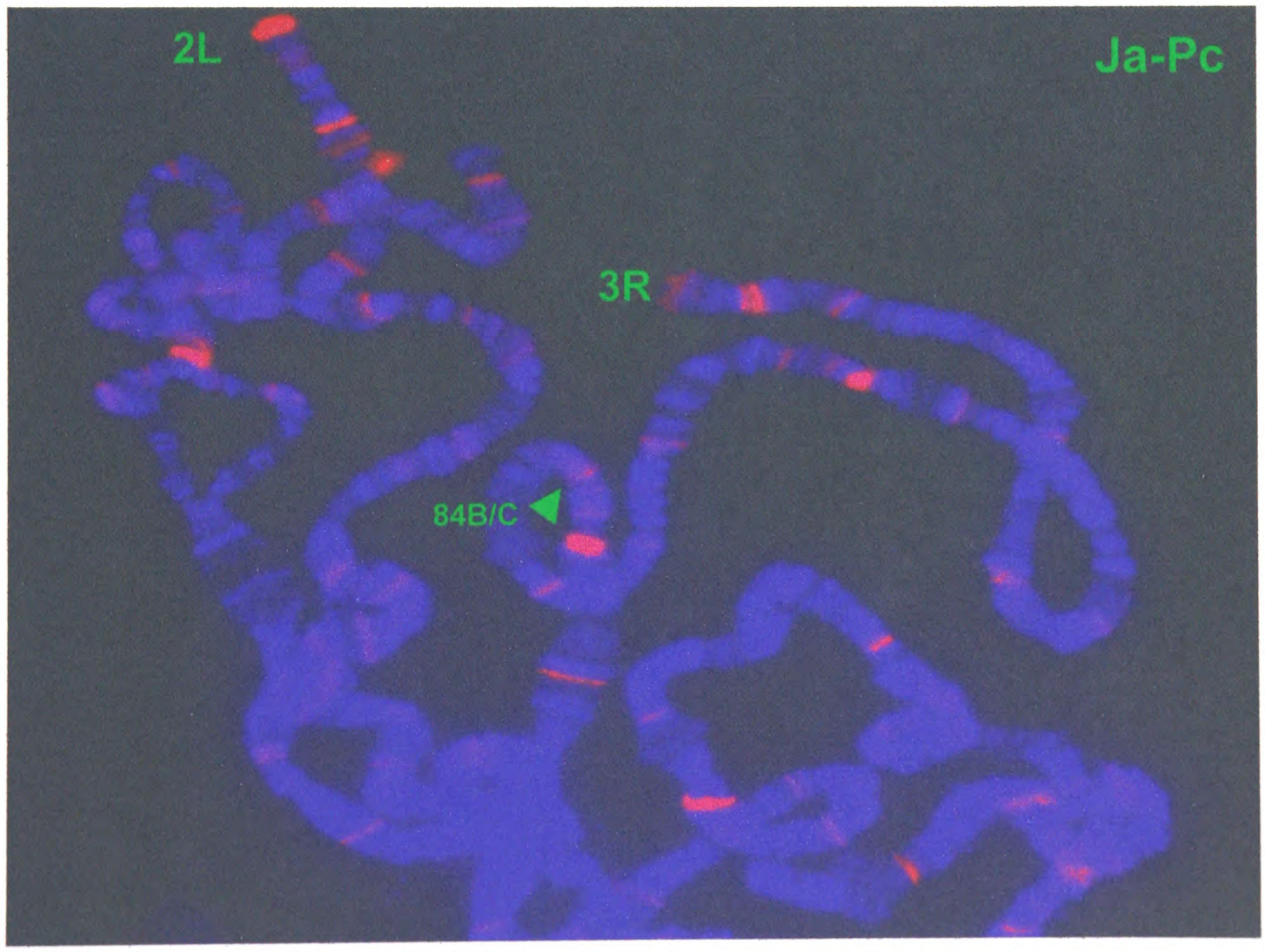
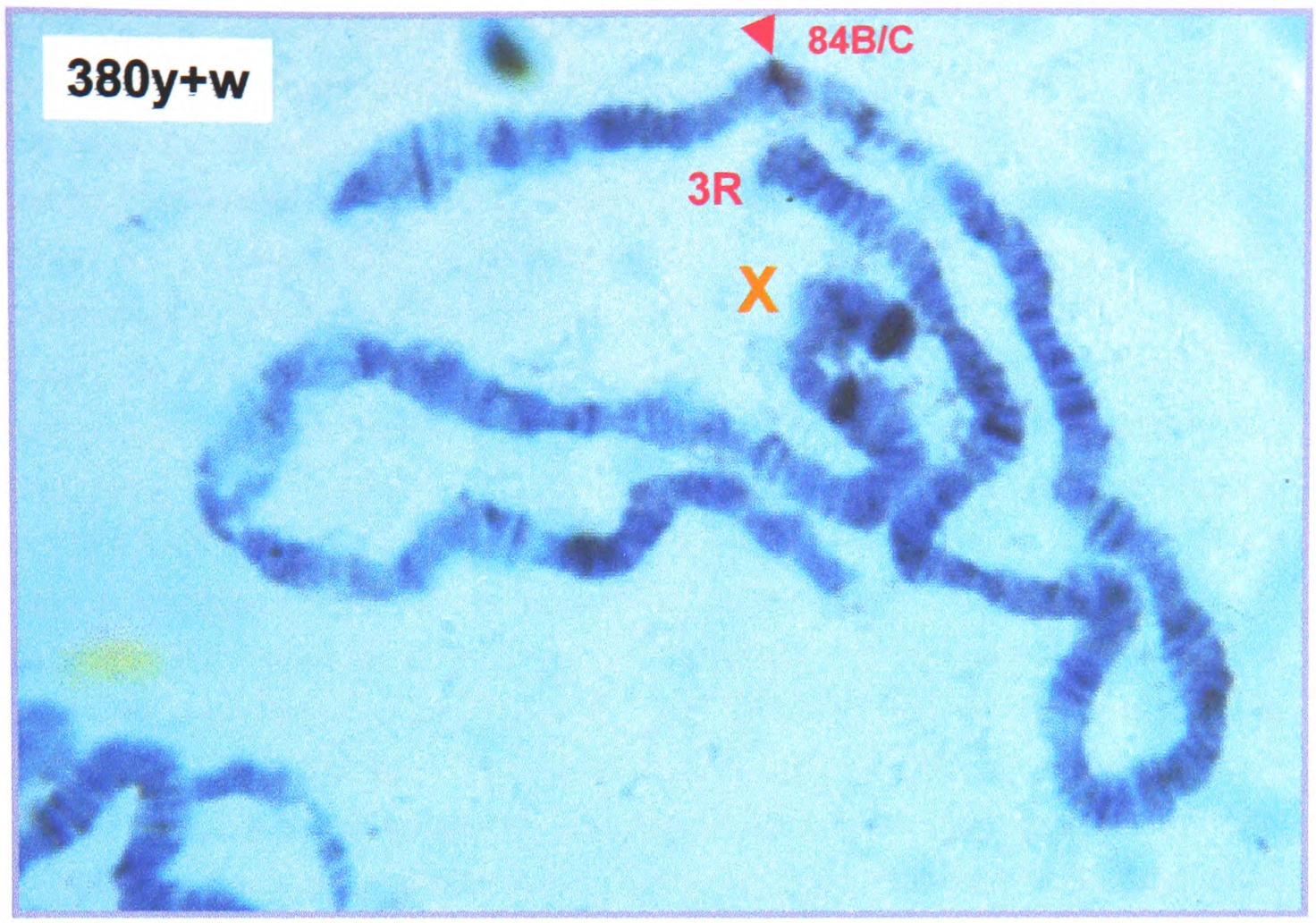
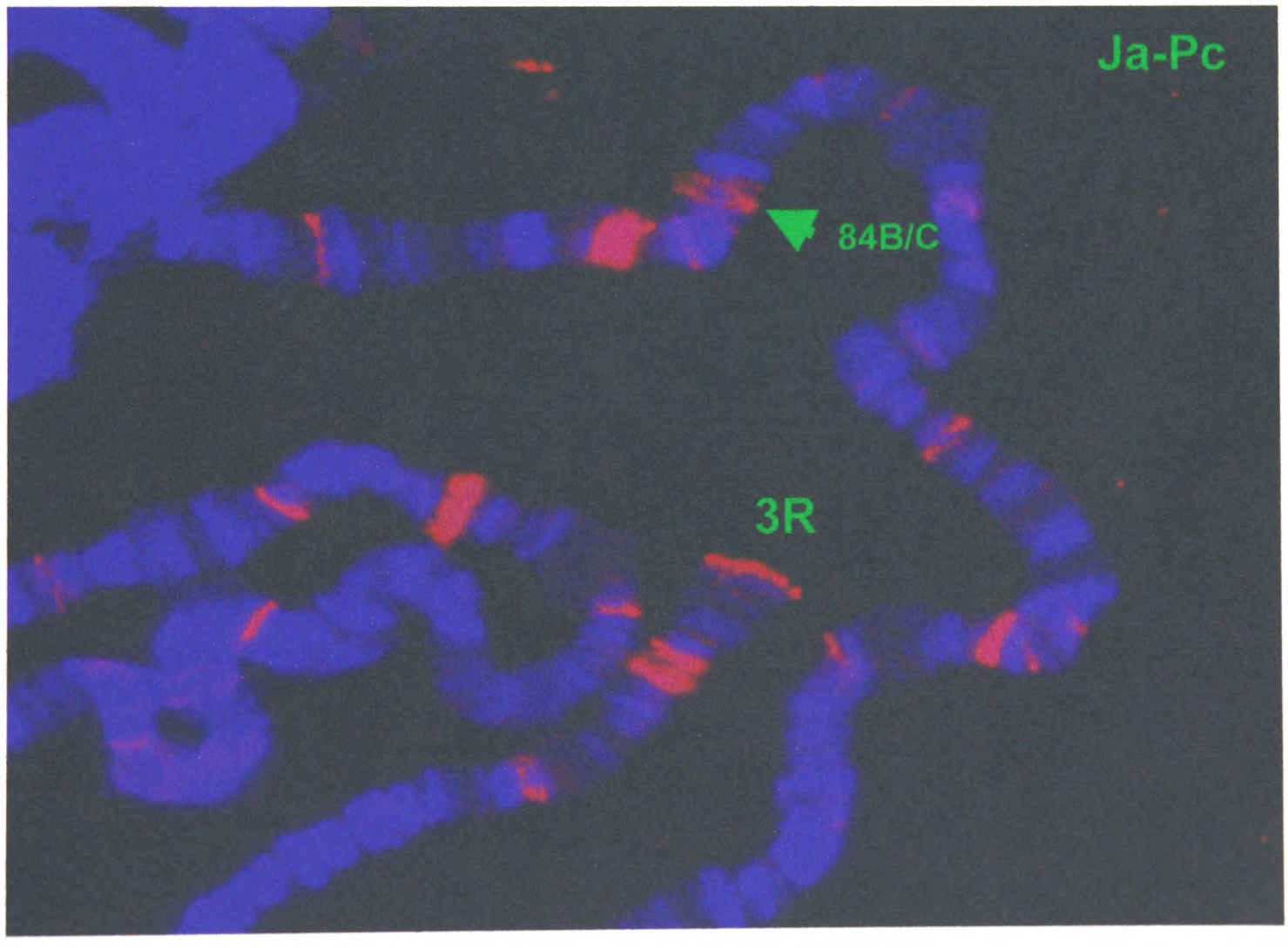

2L
Ja-Pc
3R
84B/C

Plate IX: High magnification images of the proximal end of 3R showing signal from DNA probe hybridisation at 84B/C on 380 chromosomes (top panel). The same region is shown in strain Ja where Pc bands are detected by immunofluorescence. A strong Pc hybridisation band is detected at 84B/C. Note also the characteristic "dotted" Pc signal detected on the telomeric region of 3R.


380y+w
84B/C
3R
X


Ja-Pc
84B/C
3R

5.3. Discussion

In the present chapter the circumstantial evidence for the involvement of epigenetic factors from the *Drosophila* PcG in *I* factor control was brought to the test. On the one hand, the existence of a binding site for a PcG homologue in the 5' UTR of the human L1 element (section 5.2.1), prompted a search for a similar site within the *I* factor 5' UTR. Since PHO is believed to act as link between DNA and the other members of the PcG group (Brown, *et al.*, 1998), the presence of a PHO binding site within the *I* factor would further strengthen the case in favour of a role of PcG proteins in *I* factor control. An imperfect match to the CCATNTT consensus recognised by the pleiohomeotic (pho) /Yin Yang 1 (YY1) proteins, was found at position 6-12 of the *I* factor 5' UTR. The *I* factor putative site, CCAcTTC, has a base substitution within what is believed to be the critical core consensus CCAT. Extending the search for putative pho/YY1 binding sites revealed the existence of two additional matches at the 5' end of ORF2. The sequence of these sites, starting with the one closest to the 5' end, is CCATCTC and CCATTTC. Although at this stage it is impossible to go beyond speculation, it is interesting to note that all three putative binding sites are within those *I* factor sequences that have so far been tested and found to have silencing properties (Chapter 4). However, only biochemical studies coupled to *in vivo* studies testing binding and the effects of mutations on *I* factor regulation can clarify the role, if any, of these sequences. If this consensus truly identifies PREs, then the *I* factor has potentially multiple PREs within its sequence, in keeping with current models of PRE function that envisage co-operative interactions between PcG proteins bound to many weak proto-PRE sites (Pirrotta, 1998).

On the other hand, direct interaction between the *I* factor and Polycomb protein could be tested because of the availability of strain 380 that carries multiple *I* factor 5' UTR which have been shown to mediate silencing of homologous sequences preferentially through the maternal germline (Chapter 4). Independent studies had shown Pc recruitment to the sites of suppressed transgenes present in multiple unlinked copies (Pal-Bhadra, *et al.*, 1997). One possible explanation for the results of Pal-Bhadra, *et al*, 1997, is that the *Adh* sequences within the *w-Adh* transgene carry a weak PRE that is normally not recognised by PcG proteins, as shown by the fact that there is no detectable PcG binding at the *Adh* locus in a non-transgenic background (Pirrotta, 1997). Increasing the copy number of the *Adh* gene and of the associated proto PRE would render the *Adh* transgene a substrate for PcG mediated suppression. *I* factor silencing, associated with progressive increase in copy number

might have similar mechanics. Furthermore, PREs have also been shown to mediate suppression of nearby genes in a temperature dependent manner. Exposing embryos carrying the PRE containing transgene to higher temperatures, results in increased levels of PcG mediated suppression. Similarly, if SF females (or their mothers) are exposed to a temperature change, usually a temperature increase, there is a decline in F2 embryonic lethality associated with *I* factor transposition, suggesting that perhaps similar suppressing effect operate on the *I* factor in response to temperature changes. The PRE mediate silencing is inherited exclusively from mother to daughter in a fashion reminiscent of *I* factor mediated silencing. In *C. elegans*, maternal inheritance of transgene silencing is controlled at least in part by two genes that are homologous to the *Drosophila* PcG members, *Enhancer of zeste* and *Extra sex combs.* Transgene silencing in *C. elegans*, as is the case for the *I* factor, is a progressive event that requires a number of generations to be established (section 5.1.3).

In order to investigate recruitment of PcG proteins at the site of insertion of transgene constructs carrying multiple copies of the *I* factor 5' UTR, the transgene insertions were mapped to polytene chromosomes from salivary glands of larvae from the 380 strain. Only three of the expected four insertions carried by this strain were detected and were mapped to positions 25F/26A, 27 C/D on 2L and 84B/C on 3 R. Since two of the four transgene insertions had been previously mapped genetically to chromosome 3 (Chaboissier, *et al.*, 1998), one possibility is that the single band at 84B/C represents two closely linked insertions that cannot be resolved on polytene chromosomes. Mapping of Pc binding sites to the same chromosomal regions on chromosomes from transgenic and nontransgenic strains, was first attempted using a non fluorescent dye.

Preliminary data suggested that insertions at positions 25F/26A and 84 B/C, co-localise with pre-existing Pc binding sites. However, given the large number of Pc bands on polytene chromosomes, and the relatively poor resolution of Pc bands obtained using this method, detection efficiency was increased by using immunofluorescence. Comparison of the corresponding chromosomal regions in transgenic and non-transgenic strains, revealed that, within the limits of resolution, all three insertions reside in close proximity of pre-existing Pc bands. The colocalisation of the constructs carrying multiple copies of the *I* factor 5' UTR with chromosomal sites of Pc binding found also in non transgenic strains, might be entirely fortuitous. This possibility is not at all unlikely because of the high density of Pc sites on polytene chromosomes. Pal-Bhadra, *et al*, 1997, also reported that a

number of their constructs had inserted at pre-existing Pc sites. Of course the colocalisation precludes any possibility of determining whether Pc was recruited *de novo* at the sites of transgene insertion.

The alternative possibility is that colocalisation represents the reciprocal event to Pc recruitment. In other, words, instead of PcG proteins being recruited at the site of transgene insertion, it is the transgenes that were recruited at sites of PcG binding at the time they were introduced in the developing germline. This "homing" phenomenon has already been reported for PRE containing transgenes (section 5.1.1), and has been explained by invoking direct recruitment of PcG proteins on the transgene at the time of germline injection. The transgene-bound PcG proteins will then direct the transgene to genomic PREs thus increasing the chance of transgene integration in their proximity. Homing could explain colocalisation of Pc and the 186x2 and 186x3 constructs used in this study as well as account for similar colocalisation events reported by Pal-Bhadra et al, 1997, given that their constructs were shown to be substrates of PcG binding. Naturally the homing process is not expected to be 100% efficient but it would interesting to test whether, increasing the number of copies of putative PRE sequences *in cis*, increases the frequency with which the transgene carrying them inserted close to other PRE sequences. This hypothesis could be tested experimentally by performing large scale transgenesis using constructs like 186x2 and 186x3 and comparing the frequency of colocalisation with respect to derivative constructs that lack the *I* factor 5' UTRs.

One legitimate objection to the approach used in this study, is the fact that all known silencing effects mediated by *I* factor sequences take place in the germline, whilst salivary glands are somatic tissue. It could be argued, however, that the main role of PcG proteins is maintenance of patterns of homeotic gene expression in the somatic cell lineage. In designing this experiment, the hope was that Pc recruitment as detected on salivary glands chromosomes would reflect events in the germline. Despite the fact that nurse cells in *Drosophila* ovaries are polyploid, the multiple chromosome copies in these cells are not organised as in polytene chromosomes and therefore do not give the regular banding patterns. However, there are known mutations in the *ovarian tumor* (*otu*) gene which cause a number of germline abnormalities, including the appearance of pseudo nurse cells with true polytene chromosomes (King *et al.*, 1981). Thus future studies aimed at detecting PcG recruitment on polytene chromosomes at the site of transgene insertions carrying I factor sequences, should be conducted in an *otu* mutant background.

Thus the work described here, has identified putative pleiohomeotic binding sites within the I factor 5' UTR and the 5' end of ORF2. Biochemical studies are required to test whether these sites are in fact recognised by pleiohomeotic protein. Moreover, recruitment of PcG proteins on trasgenes carrying multiple "silencing" copies of the *I* factor 5' UTR was investigated here. The data obtained do not provide unmabigous evidence of PcG involvement. Yet they do suggest that interactions between PcG proteins and *I* factor sequences might occur at the time of transgenesis and result in colocalisation of the transgenes at known sites of polycomb binding.

6. Concluding remarks and future directions

The work described here was aimed at investigating a number of properties associated with the I factor 5' UTR, a 186 bp sequence that drives transcription of this retroelement. Physical studies of the I factor 5' UTR have uncovered a region of great flexibility located almost exactly in the centre of this 186 bp molecule. The biological relevance of this sequence-dependent flexibility could be investigated further *in vitro* by looking at the effects of protein binding on structural parameters of the I factor as detected by changes in electrophoretic mobility. Multiple copies of the I factor 5' UTR have been shown to mediate inheritance of silenced states preferentially through the maternal germline. The reversible and heritable nature of I factor silencing suggests a role for epigenetic mediators from the *Drosophila* Polycomb group. A putative binding site for the PcG protein pleiohomeotic has been found by sequence analysis in the 5' UTR of the I factor. Two other putative sites have been found in the 5' end of ORF2. *In vitro* studies will be required to determine whether the I factor CCACTTC consensus is bound by pleiohomeotic protein. In addition, preliminary evidence suggests that PcG proteins might recognise and interact with transgenes that carry multiple copies of the I factor 5' UTR and mediate their preferential insertion in proximity of pre-existing sites of PcG binding, as detected on polytene chromosomes. Further investigation will be required to confirm this observation. However, genetic approaches, looking at the effects on I factor control of mutations in Polycomb group genes, will certainly provide a more efficient means to identify mediators of I factor silencing. Other aspects of I factor control that require elucidation relate to the mechanism and timing of I factor silencing. The evidence suggests that distinct mechanisms operate in the female germline and in the soma. Whilst somatic control of I factor transposition is most certainly post transcriptional and dosage insensitive, germline silencing is dosage sensitive and seems to have both transcriptional and posttranscriptional components. Future investigations will be aimed at unravelling the specific steps involved in I factor control and at investigating the role of I factor encoded functions in transposition and its regulation.

7.References

7.References

Abad, P., Vaury, C., Pelisson, A., Chaboissier, M.-C., Busseau, I., and Bucheton, A. (1989). A long interspersed repetitive element, the I factor of Drosophila tessieri, is able to transpose in different Drosophila species. Proc. Natl. Acad. Sci. USA *86*, **8887-8891**.

Agostoni, E., Albertson, D., Wittman, C., Hill, F., Tobler, H., and Muller, F. (1996). cec-1, a soma-specific chromobox gene in C. elegans. Dev. Biol. *178*, **316-326**.

Alexeev, D.G., Lipanov, A.A., and Skuratovskii, I.Y. (1987). Structure of polyd(A)-polyd(T) is a B type double helix with a distinctively narrow minor groove. Nature *325*, **821-823**.

Amrein, H., and Axel, R. (1997). Genes expressed in neurons of adult male Drosophila. Cell *88*, **459-469**.

Anderson, J.N. (1986). Detection, sequence patterns and function of unusual DNA structures. Nucl. Acids. Res. *14*, **8513-8533**.

Arkhipova, I.R. (1995). Promoter elements in Drosophila melanogaster revealed by sequence analysis. Genetics *139*, **1359-1369**.

Arnott, S., Chandrasekaran, R., Hall, I.H., and Puigjaner, L.C. (1983). Heteronomous DNA. Nucl. Acids. Res. *11*, **4141-4155**.

Arnott, S., Chandraserkaran, R., Hall, I.H., Puigjaner, L.C., Walker, J.K., and Wang, M. (1982). DNA secondary structures: helices, wrinkles, and junctions. ColdSrping Harbour Symp. Quant. Biol. *XLVII*, **53-65**.

Assaad, F.F., Tucker, K.L., and Signer, E.R. (1993). Epigenetic repeat-induced gene silencing (RIGS) in Arabidopsis. Plant Mol. Biol. *22*, **1067-1085**.

Bechert, T., Heck, S., Fleig, U., Diekmann, S., and Hegemann, J.H. (1999). All 16 centromere DNAs from Saccharomyces cerevisiae show DNA curvature. Nucl. Acids. Res. *27*, **1444-1449**.

Bender, W., Akam, M., Korch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B., and Hogness, D.S. (1983). Molecular genetics of the bithorax complex of Drosophila melanogaster. Science *221*, **23-29**.

Biemont, C. (1986). Polymorphism of the mdg-1 and I mobile elements in Drosophila melanogaster. Chromosoma *93*, **393-397**.

Biemont, C. (1992). Population genetics of transposable DNA elements. Genetica *86*, **67-84**.

Biessman, H., Mason, J.M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K.L., and Pardue, M.L. (1990). Addition of telomere-associated HeT-A DNA sequences "heals" broken chromosome ends in Drosophila. Cell *61*, **663-673**.

Bingham, P.M. (1997). Cosuppression comes to the animals. Cell *90*, **385-387**.

Birchler, J.A., Pal-Bhadra, M., and Bhadra, U. (1999). Less from more: cosuppression of transposable elements. Nat. Genet. *21*, **148-149**.

Blackman, R.K., and Gelbart, W.M. (1989). The transposable element hobo of Drosophila melanogaster. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington DC: Am. Soc. Microbiol.), pp. **523-529**.

Blesa, D., and Martinez-Sebastian, M.J. (1997). bilbo, a non-LTR retrotransposon of Drosophila subobscura: a clue to the evolution of LINE-like elements in Drosophila. Mol. Biol. Evol. *14*, **1145-1153**.

Bogenhagen, D.F., and Brown, D.D. (1981). Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell *24*, **261-270**.

Bolshoy, A., McNamara, P., Harrington, R.E., and Trifonov, E.N. (1991). Curved DNA without A A: experimental estimates of all 16 DNA wedge angles. Proc. Natl. Acad. Sci. USA *88*, **2312-2316**.

Bossi, L., and Smith, D.M. (1984). Conformational change in the DNA associated with an unusual promoter mutation in a tRNA operon of Salmonella. Cell *39*, **643-652**.

Bouhidel, K., Terzian, C., and Pinon, H. (1994). The full-length transcript of the I factor, a LINE element of Drosophila melanogaster, is a potential bicistronic RNA messenger. Nucl. Acids Res. *22*, **2370-2374**.

Bracco, L., Kotlarz, D., Kolb, A., Diekman, S., and Buc, H. (1989). Synthetic curved DNA sequences can act as transcriptional activators in Escherichia coli. EMBO J. *8*, **4289-4296**.

Bregliano, J.-C., and Kidwell, M.G. (1983). Hybrid Dysgenesis determinants. In Mobile Genetic Elements, Shapiro, ed.: Academic Press Inc.), pp. **363-410**.

Bregliano, J.-C., Laurencon, A., and Degroote, F. (1995). Evidence for an inducible repair-recombination system in the female germ line of Drosophila melanogaster.I.Induction by inhibitors of nucleotide synthesis and by gamma rays. Genetics *141*, **571-578**.

Bregliano, J.C., Picard, G., Bucheton, A., Pelisson, A., Lavige, J.M., and L'Heritier, P. (1980). Hybrid dysgenesis in Drosophila melanogaster. Science *207*, **606-611**.

Breiling, A., and Paro, R. (1997). Mechanisms of heritable gene silencing during developement of Drosophila. In Nuclear organisation, chromatin structure, and gene expression., R. van Driel and A. P. Otte, eds. (Oxford- New York: Oxford University Press), pp. **236-249**.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.-L., and Kassis, J.A. (1998). The Drosophila polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol. Cell *1*, **1057-1064**.

Brukner, I., Dlakic, M., Savic, A., Susic, S., Pongor, S., and Suck., D. (1993). Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. Nucl. Acids Res. *21*, **1025-1029**.

Brukner, I., Sanchez, R., Suck, D., and Pongor, S. (1995). Sequence-dependent bending propensity of DNA as reveled by DNase I: parameters for trinucleotides. EMBO J. *14*, **1812-1818**.

Bucheton, A. (1979). Non mendelian female sterility in Drosophila melanogaster: influence of aging and thermic treatment. III. Cumulative effects induced by these factors. Genetics *93*, **131-142**.

Bucheton, A. (1990). I transposable elements and I-R hybrid dysgenesis in Drosophila. Trends in Genetics *6*, **16-21**.

Bucheton, A., Lavige, J.-M., Picard, G., and L'Heritier, P. (1976). Non Mendelian female sterility in Drosophila melanogaster: quantitative variations in the efficiency on inducer and reactive strains. Heredity *36*, **305-314**.

Bucheton, A., Paro, R., Sang, H.M., Pelisson, A., and Finnegan, D.J. (1984). The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell *38*, **153-163**.

Bucheton, A., Simonelig, M., Vaury, C., and Crozatier, M. (1986). Sequences similar to the I transposable element involved in I-R hybrid dysgenesis in D. melanogaster occur in other Drosophila species. Nature *322*, **650-652**.

Bucheton, A., Vaury, C., Chaboissier, M.-C., Abad, P., Pelisson, A., and Simonelig, M. (1992). I elements and the Drosophila genome. Genetica *86*, **175-190**.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & Dev. *10*, **711-724**.

Burkhoff, A.M., and Tullius, T.D. (1987). The unusual conformation adopted by adenine tracts in kinetoplast DNA. Cell *48*, **935-943**.

Bushmeyer, S., Park, K., and Atchison, M.L. (1995). Characterisation of functional domains within the multifunctional trasncription factor YY1. J. Biol. Chem. *270*, **30213-30220**.

Busseau, I., Chaboissier, M.-C., Pelisson, A., and Bucheton, A. (1994). I factors in Drosophila melanogaster: transposition under control. Genetica *93*, **101-116**.

Busseau, I., Pelisson, A., and Bucheton, A. (1989). I elements of Drosophila melanogaster generate specific chromosomal rearrangements during transposition. Mol. Gen. Genet. *218*, **222-228**.

Busseau, I., Pelisson, A., and Bucheton, A. (1989a). Characterisation of 5' truncated copies of the I factor in Drosophila melanogaster. Nucl. Acids Res. *17*, **6939-6945**.

Busturia, A., and Bienz, M. (1993). Silencers in Abdominal-A, a homeotic Drosophila gene. EMBO J. *12*, **1415-1425**.

Calladine, C.R., Drew, H.R., and McCall, M.J. (1988). The intrinsic curvature of DNA in solution. J Mol. Biol. *201*, **127-137**.

Capowski, E.E., Martin, P., Garvin, D., and Strome, S. (1991). Identification of grandchildless loci whose producys are required for normal germ-line development in the nematode Caenorhabditis elegans. Genetics *129*, **1061-1072**.

Carmena, M., and Gonzalez, C. (1995). Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in Drosophila melanogaster. Chromosoma *103*, **676-684**.

Cavalli, G., and Paro, R. (1998). The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell *93*, **505-518**.

Chaboissier, M.-C., Bucheton, A., and Finnegan, D.J. (1998). Copy number control of a transposable element, the I factor, a LINE-like element in Drosophila. Proc. Natl. Acad. Sci. USA *95*, **11781-11785**.

Chaboissier, M.-C., Bucheton, A., and Finnegan, D.J. (1999). In Press.

Chaboissier, M.-C., Busseau, I., Prosser, J., Finnegan, D.J., and Bucheton, A. (1990). Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in Drosophila melanogaster. EMBO J. *9*, **3557-3563**.

Chaboissier, M.-C., Lemeunier, F., and Bucheton, A. (1995). IR hybrid dysgenesis increases the frequency of recombination in Drosophila melanogaster. Genet. Res. Camb. *65*, **167-174**.

Cogoni, C., and Macino, G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, **166-169**.

Crawford, D.L., Segal, J.A., and Barnett, J.L. (1999). Evolutionary analysis of TATA-less proximal promoter function. Mol. Biol. Evol. *16*, **194-207**.

Crothers, D.M., Haran, T.E., and Nadeau, J.G. (1990). Intrinsically bent DNA. J. Biol. Chem. *265*, **7093-7096**.

Crouse, H.V. (1960). The controlling element in sex chromosome behaviour in Sciara. Genetics *45*, **1429-1443**.

Crozatier, M., Vaury, C., Busseau, I., Pelisson, A., and Bucheton, A. (1988). Structure and genomic organisation of I elements involved in I-R hybrid dysgenesis in Drosophila melanogaster. Nuc. Acids. Res. *16*, **9199-9213**.

Csink, A.K., Linsk, R., and Birchler, J.A. (1994). The Lighten up (Lip) gene of Drosophila melanogaster: a modifier of retroelement expression, position effect variegation and white locus insertion alleles. Genetics *138*, **153-163**.

Csink, A.K., Linsk, R., and Birchler, J.A. (1994a). Mosaic suppressor, a gene in Drosophila that modifies retrotransposon expression and interacts with zeste. Genetics *136*, **573-583**.

Daniels, D.L., De Wet, J.R., and Blattner, F.R. (1980). New map of bacteriophage lambda. J. Virol. *33*, **390-400**.

Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A. (1990). Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics *124*, **339-355**.

Danilevskaya, O.N., Arkhipova, I.R., Traverse, K.K., and Pardue, M.L. (1997). Promoting in tandem: the promoter for telomer transposons HeT-A and implications for the evolution of retroviral LTRs. Cell *88*, **647-655**.

Davis, N.A., Majie, S.S., and Kahn, J.D. (1999). TATA box DNA deformation with and without the TATA box-binding protein. J. Mol. Biol. *291*, **249-265**.

Dawson, A., Hartswood, E., Paterson, T., and Finnegan, D.J. (1997). A LINE-like transposable element in Drosophila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. EMBO J. *16*, **4448-4455**.

de La Roche Saint Andre, C., and Bregliano, J.-C. (1998). Evidence for a multistep control in transposition of I factor in Drosophila melanogaster. Genetics *148*, **1875-1884**.

Depicker, A., and Van Montagu, M. (1997). Posttranscriptional gene silencing in plants. Curr. Opin. Cell Biol. *9*, **373-382**.

Di Franco, C., Galuppi, D., and Junakovic, N. (1992). Genomic Distribution of transposable elements among individuals of an inbred Drosophila line. Genetics *86*, **1-11**.

Di Gabriele, A.D., Sanderson, M.R., and Steitz, T.A. (1989). Crystal lattice packing is important in determining the bend of DNA dodecamer containing an adenine tract. Proc. Natl. Acad. Sci. USA *86*, **1816-1820**.

Di Nocera, P.P., and Casari, G. (1987). Related polypeptides are encoded by Drosophila F elements, I factors, and mammalian L1 sequences. Proc. Natl.Acad. Sci. USA *84*, **5843-5847**.

Di Nocera, P.P., Contursi, C., and Minchiotti, G. (1994). LINE-related elements in Drosophila melanogaster. Genetica *94*, **173-180**.

Di Nocera, P.P., Digau, M.E., and Dawid, I.B. (1983). A family of ologadenylated transposable sequences in Drosohpila melanogaster. J. Mol. Biol. *168*, **715-728**.

Di Nocera, P.P., Graziani, F., and Lavarogna, G. (1986). Genomic and structural organisation of Drosophila melanogaster G elements. Nucl. Acids Res. *14*, **675-691**.

Dickerson, R.E., Drew, H.R., Conner, B.N., Kopka, M.L., and Pjura, P.E. (1982). Helix geometry and hydration in A.DNA, B-DNA, and Z-DNA. Cold Spring Harbour Symp. Quant. Biol. *XLVII (1)*, **13-24**.

Diekmann, S. (1989). Definitons and nomenclature of nucleic acid structure parameters. EMBO J. *8*, **1-4**.

Diekmann, S. (1992). Analysing DNA curvature in polyacrylamide gels. Methods in Enzymology *212*, **30-46**.

Diekmann, S., and Lilley, D.M.J. (1987). The anomalous migration of a stable cruciform: temperature and salt dependence, and some comparisons with curved DNA. Nucl. Acids Res. *15*, **5765-**.

Diekmann, S., and Wang, J.C. (1985). On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mibilities. J. Mol. Biol. *186*, **1-11**.

Dimitri, P., Arca', B., Berghella, L., and Mei., E. (1997). High genetic instability of heterochromatin after transposition of the LINE-like I factor of Drosophila melanogaster. Proc. Natl.Acad. Sci. USA *94*, **8052-8057**.

Dimitri, P., and Junakovic, N. (1999). Revising the selfish DNA hypothesis. Trends Genet. *15*, **123-124**.

Dorer, D.R. (1997). Do transegene arrays form heterochromatin in vertebrates? Transgenic Res. *6*, **6-10**.

Dorer, D.R., and Henikoff, S. (1994). Expansion of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. Cell *77*, **1-20**.

Drew, H.R., and Travers, A.A. (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. *186*, **773-790**.

Driver, A., Lacey, S.F., Cullingford, T.E., Mitchelson, A., and O'Hare, K. (1989). Structural analysis of Doc transposable elements associated with mutations at the white and suppressor of forked loci of Drosophila melanogaster. Mol. Gen. Genet. *220*, **49-52**.

Eden, S., and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. Curr. Opin Genet. Dev. *4*, **255-259**.

Efstratiadis, A. (1994). Parental imprinting of autosomal mammalian genes. Curr. Opin. Genet. Dev. *4*, **265-280**.

Eickbush, T.H. (1992). Transposing without ends: The non-LTR retrotransposable elements. The New Biologist *4*, **430-440**.

Eickbush, T.H. (1997). Telomerase and retrotransposons: which came first? Science *277*, **911-912**.

Eissenberg, J.C., Elgin, S.C.R., and Paro, R. (1995). Epigentic regulation in Drosophila: a conspiracy of silence. In Chromatin structure and gene expression., S. C. R. Elgin, ed. (Oxford: IRL Press), pp. **147-171**.

Engels, W.R. (1979a). Extrachromosomal control of mutability. Proc. Natl. Acad. Sci. USA *76*, **4011-4015**.

Engels, W.R. (1979b). Hybrid Dysgenesis in Drosophila melanogaster:Rules of inheritance of female sterility. Genetical Research *33*, **219-236**.

Engels, W.R. (1980). Hybrid dysgenesis in Drosopila and the stocastic loss hypothesis. Cold Spring Harbor Symposia on Quantitative Biology *45*, **561-565**.

Engels, W.R. (1989). P elements in Drosophila. In Mobile DNA, D. E. Berg, Howe, M.M., ed. (Washington D.C.: American Society for Microbiology), pp. **437-484**.

Fairall, L., Martin, S., and Rhodes, D. (1989). The DNA binding site of the Xenopus transcription factor IIIA has a non-B-form structure. EMBO J. *8*, **1809-1817**.

Falb, D., and Maniatis, T. (1992). Drosophila transcriptional repressor protein that binds specifically to negative control elements in the fat body enhancer. Mol. Cell. Biol. *12*, **4093-4103**.

Fauvarque, M.-O., and Dura, J.M. (1993). ?? Genes Dev. *7*, **1508-1520**.

Fawcett, D.H., Lister, C.K., Kellet, E., and Finnegan, D.J. (1986). Transposable elements controlling I-R hybrid dysgenesis in D.melanogaster are similar to mammalian LINEs. Cell *47*, **1007-1015**.

Feng, Q., Moran, J.V., Kazazian Jr, H.H., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, **905-916**.

Feng, Q., Schumann, G., and Boeke, J.D. (1998). Retrotransposon R1Bm endonuclease cleaves the target sequence. Proc. Natl. Acad. Sci. USA *95*, **2083-2088**.

Finnegan, D.J. (1989a). The I factor and I-R hybrid dysgenesis in Drosophila melanogaster. In Mobile DNA, D. E. Berg, Howe,M.M., ed. (Washington DC: American Society for Microbiology), pp. **503-518**.

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

Finnegan, D.J. (1990). Transposable elements and DNA transposition in Eukaryotes. Curr. Opin. Cell. Biol. *2*, **471-477**.

Finnegan, D.J. (1993). New primers for an old enzyme. Curr. Biol. *3*, **468-470**.

Finnegan, D.J. (1997). Transposable elements: How non-LTR retrotransposons do it. Curr. Biol. *7*, **R245-R248**.

Fire, A., Xu, S., Montogmery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caernorhabditis elegans. Nature *391*, **806-811**.

Fitzgerald, D.J., Dryden, G.L., Bronson, E.C., Williams, J.S., and Anderson, J.N. (1994). Conserved patterns of bending in satellite and nucleosome positioning DNA. J. Biol. Chem. *19*, **21303-21314**.

Flanagan, J.R., Becker, K.G., Ennist, D.L., GLeason, S.L., Driggers, P.H., Levi, B.-Z., Appela, E., and Pzato, K. (1992). Cloning of a negative transcription factor that binds to the upstream conserved region of Maloney murine leukemia virus. Mol. Cell. Biol. *12*, **38-44**.

Flavell, R.B. (1994). Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc. Natl. Acad. Sci. USA *91*, **3490-3496**.

Gabor Miklos, G.L., and Cotsell, J.N. (1990). Chromosome structure at interfaces between major chromatin types: alpha- and beta-heterochromatin. BioEssays *12*, **1-6**.

Gabrielian, A., Simoncsits, A., and Pongor, S. (1996). Distribution of bending propensity in DNA sequences. FEBS Lett. *393*, **124-130**.

Gabrielian, A., Vlahovicek, K., and Pongor, S. (1997). Distribution of sequence-dependent curvature in genomic DNA sequences. FEBS Lett. *406*, **69-74**.

Galas, D.J., Eggert, M., and Waterman, M.S. (1985). Rigorous pattern recognition methods for DNA sequences:analysis of promoter sequences from Escherichia coli. J. Mol. Biol. *186*, **117-128**.

Galvin, K.M., and Shi, Y. (1997). Multiple mechanisms of transcriptional repression by YY1. Mol. Cell. Biol. *17*, **3723-3732**.

Garrick, D., Fiering, S., Martin, D.I.K., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nat. Genet. *18*, **56-59**.

Gartenberg, M.R., and Crothers, D.M. (1991). Synthetic DNA bending sequences increase the rate of in vitro transcription initiation at the Escherichia coli lac promoter. J. Mol. Biol. *219*, **217-230**.

George, J.A., and Eickbush, T.H. (1999). Conserved features at the 5' end of Drosophila R2 retrotransposable elements: implications for transcription and translation. Insect Mol. Biol. *8*, **3-10**.

Gill, G. (1994). Taking the initiative. Curr. Biol. *4*, **374-376**.

Gindhart, J.G., and Kaufman, T.C. (1995). Identification of Polycomb and trithorax group responsive elements in the regulatory region of the Drosophila homeotic gene sex combds reduced. Genetics *139*, **797-814**.

Golic, K.G., Golic, M.M., and Pimpinelli, S. (1998). Imprinted control of gene activity in Drosophila. Curr. Biol. *8*, **1273-1276**.

Goodsell, D.S., and Dickerson, R.E. (1994). Bending and curvature calculations in B-DNA. Nucl. Acids. Res. *22*, **5497-5503**.

Goodsell, D.S., Kaczor-Grzeskowiak, M., and Dickerson, R.E. (1994). The crystal structure of CCATTAATGG: implications for bending of B-DNA at T-A steps. J. Mol. Biol. *239*, **79-96**.

Grant, S.R. (1999). Dissecting mechanisms of posttranscriptional gene silencing: divide and conquer. Cell *96*, **303-306**.

Green, M.M. (1973). Some observations and comments on mutable and mutator genes in Drosophila. Genetics (suppl.) *73*, **187-194**.

Green, M.M. (1977). Genetic instability in Drosophila melanogaster: De novo induction of putative insertion mutations. Proc. Natl. Acad. Sci. *74*, **3490**.

Hagerman, P.J. (1984). Further evidence of static curvature of a subfragment{..}. Proc. Natl. Acad. Sci. USA *81*, **4632-4636**.

Hagerman, P.J. (1985). Sequence dependence of the curvature of DNA: a test of the phasing hypothesis. Biochemistry *24*, **7033-7037**.

Hagerman, P.J. (1990). Sequence-directed curvature of DNA. Annu. Rev. Biochem. *59*, **755-781**.

Hagstrom, K., and Schedl, P. (1997). Remembrance of things past: maintaining gene expression patterns with altered chromatin. Curr. Opin. Genet. Dev. *7*, **814-821**.

Harada, K., Yukuhiro, K., and Mukay, T. (1990). Transposition rates of movable genetic elements in Drosophila melanogaster. Proc. Natl. Acad. Aci. USA *87*, **3248-3252**.

Hariharan, N.D., Kelley, E., and Perry, R.P. (1991). Delta, a transcription factor that binds to downstream elements in several polymerase II promoters is a functionally versatile zinc finger protein. Proc. Natl. Acad. Sci USA *88*, **9799-9803**.

Haron, T.E., Kahn, J.D., and Crothers, D.M. (1994). Sequence elements responsible for DNA curvature. J. Mol. Biol. *244*, **135-143**.

Hazelrigg, T., and Petersen, S. (1992). An unusual genomic position effect on Drosophila white gene expression: pairing dependence, interactions with zeste and molecular analysis of revertants. Genetics *130*, **125-138**.

Henikoff, S. (1995). Gene silencing in Drosophila. Curr. Top. Microbiol. Immunol. *197*, **193-208**.

Henikoff, S. (1997). Nuclear organisation and gene expression: homologous pairing and long range interactions. Curr. Opin. Cell Biol. *9*, **388-395**.

Henikoff, S., and Comai, L. (1998). Trans-sensing effects: the ups and downs of being together. Cell *93*, **329-332**.

Hennig, W. (1996). Spermatogenesis in Drosophila. Int. J. Dev. Biol. *40*, **167-176**.

Hiraizumi, Y. (1971). Spontaneous recombination in Drosophila melanogaster males. Proc. Natl. Acad. Sci. *68*, **268-270**.

Hiraoka, Y., Denburgh, A.F., Parmelee, S.J., Rykowski, M.C., Agard, D.A., and Sedat, J.W. (1993). The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis. J. Cell Biol. *120*, **591-600**.

Hoffmann, A., Oelgeschlager, T., and Roeder, R.G. (1997). Considerations of transcriptional control mechanisms: Do TFIID-core promoter elements recapitulate nucleosome-like functions? Proc. Natl. Acad. Sci. USA *94*, **8928-8935**.

Holdeman, R., Nehrt, S., and Strome, S. (1998). MES-2, a maternal protein essential for viability of the germline in C. elegans, is homologous to a Drosophila polycomb group protein. Development *125*, **2457-2467**.

Hollick, J.B., Dorweiler, J.E., and Chandler, V.L. (1997). Paramutation and related allelic interactions. Trends Genet. *13*, **302-308**.

Houck, M.A., Clark, J.B., Peterson, K.R., and Kidwell, M.G. (1991). Possible horizontal transfer of Drosophila genes by the mite Proctolaelaps regalis. Science *253*, **1125-1129**.

Huey, R.B., Wakefield, T., Crill, W.D., and Gilchrist, G.W. (1995). Within- and between-generation effects of temperature on early fecundity of Drosophila melanogaster. Heredity *74*, **216-223**.

Hunter, C.A. (1993). Sequence-dependent DNA structure: the role of base stacking interactions. J. Mol. Biol. *230*, **1025-1054**.

Hutchison III, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R., and Edgell, M.H. (1989). LINEs and related retroposons: long interspersed repeated sequences in the Eukaryotic genome. In Mobile DNA, D. E. Berg, Howe,M.M., ed. (Washington DC: Am.Soc.for Microbiol.), pp. **503-617**.

Imasheva, A.G., Loeschcke, V., Zhivotovsky, L.A., and Lazebny, O.E. (1998). Stress temperature and quantitative variation in Drosophila melanogaster. Heredity *81*, **246-253**.

Jack, J.W., and Judd, B.H. (1979). Allelic pairing and gene regulation: a model for the zeste-white interaction. Proc. Natl. Acad. Sci. USA *76*, **1368-1372**.

Jackson, M., Heller, D., and Leinwand, L. (1985). Transcriptional measurements of mouse repeated DNA sequences. Nucl. Acids Res. *13*, **3389-3403**.

Jakubczak, J.L., Burke, W.D., and Eickbush, T.H. (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA *88*, **3295-3299**.

James, A.C., Azevedo, R.B.R., and Partridge, L. (1997). Genetic and environmental responses of Drosophila melanogaster from a latitudinal cline. Genetics *146*, **881-890**.

Jensen, S., Cavarec, L., Gassama, M.-P., and Heidmann, T. (1995). Defective I elements introduced into Drosophila as transgenes can regulate reactivity and prevent I-R hybrid dysgenesis. Mol. Gen. Genet. *248*, **381-390**.

Jensen, S., Caverec, L., Dhellin, O., and Heidmann, T. (1994a). Retrotransposition of a marked Drosophila LINE-like I element
in cells in culture. Nuc. Acids. Res. *22*, **1484-1488**.

Jensen, S., Gassama, M.-P., and Heidmann, T. (1994). Retrotransposition of the Drosophila LINE I element can induce deletion in the target DNA: A simple model also accounting for the varaibility of the normally observed target site duplications. Biochem. Byophys. Res. Comm. *202*, **111-119**.

Jensen, S., Gassama, M.-P., and Heidmann, T. (1999). Taming of transposable elements by homology-dependent gene silencing. Nat. Genet. *21*, **209-212**.

Jensen, S., and Heidmann, T. (1991). An indicator gene for detection of germline retrotransposition in transgenic Drosophila demonstrates RNA-mediated transposition of the LINE I element. EMBO J. *10*, **1927-1937**.

Jones, C.A., Ng, J., Peterson, A.J., Morgan, K., Simon, J., and Jones, R.S. (1998). The Drosophila esc and E(z) are direct partners in Polycomb-group-mediated repression. Molec. Cell. Biol *18*, **2825-2834**.

Karpen, G.H. (1994). Position-effect variegation and the new biology of heterochromatin. Curr. Opin. Genet. Dev. *4*, **281-291**.

Kasschau, K.D., and Carrington, J.C. (1998). A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. Cell *95*, **461-470**.

Kassis, J.A. (1994). Unusual properties of regulatory DNA from the Drosophila engrailed gene: Three " pairing sensitive" sites within a 1.6 Kb region. Genetics *136*, **1025-1038**.

Kawamoto, T., Makino, K., Orita, S., Nakata, A., and Kakunaga, T. (1989). DNA bending and binding factors of the human beta-actin promoter. Nucl. Acids Res. *17*, **523-537**.

Kelly, W.G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development *125*, **2451-2456**.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed C. elegans gene. Genetics *146*, **227-238**.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of ds RNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell *95*, **461-470**.

Kennison, J.A. (1997). The Polycomb and Trithorax group proteins of Drosophila: trans-regulators of homeotic gene function. Annu. Rev. Genet. *29*, **289-303**.

Kidwell, M.G. (1979). Hybrid dysgenesis in Drosophila melanogaster: the relationship between the P-M and I-R interaction systems. Genetical Research *33*, **205-217**.

Kidwell, M.G. (1983). Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA *80*, **1655-1659**.

Kidwell, M.G., and Kidwell, J.F. (1976). Selection for male recombination in Drosophila melanogaster. Genetics *84*, **333-351**.

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. (1977). Hybrid dysgenesis in Drosophile melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics *86*, **813-833**.

Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene *85*, **15-23**.

King, R.C. (1970). Ovarian development in Drosophila melanogaster., 1st Edition (New York & London: Academic Press).

King, R.C., Riley, S.F., Cassidy, J.D., White, P.E., and Paik, Y.K. (1981). Giant polytene chromosomes from ovaries of a Drosophila mutant. Science *212*, **441-443**.

Koo, H.-S., and Crothers, D.M. (1987). Chemical departures of DNA bending at Adenine-Thymine tracts. Biochem. *26*, **3745-3748**.

Koo, H.-S., Wu, H.-M., and Crothers, D.M. (1986). DNA bending at Adenine-Thymine tracts. Nature *320*, **501-506**.

Korf, I., FAn, Y., and Strome, S. (1998). The Polycomb group in Caenorhabditis elegans and maternal control of germline development. Development *125*, **2469-2478**.

Lachaume, P., Bouhidel, K., Mesure, M., and Pinon, H. (1992). Spatial and temporal expression of the I factor during oogenesis in Drosophila melanogaster. Developement *115*, **729-735**.

Lachaume, P., and Pinon, H. (1993). Germline expression of I factor, a functional LINE from fruit fly, is positively regulated by reactivity, a peculiar cellular state. Molecular and General Genetics *240*, **277-285**.

Lajoinie, O., Drake, M.E., Dastugue, B., and Vaury, C. (1995). Aberrant pre-mRNA maturation is caused by LINE insertions into introns of the white gene of Drosophila melanogaster. Nuc. Acids Res. *23*, **4015-4022**.

Langley, C., Brookfield, J.F.Y., and Kaplan, N. (1983). Transposable elements in Mendelian populations.I. Theor. Genetics *104*, **457-471**.

Laurencon, A., and Bregliano, J.-C. (1995). Evidence for an inducible repair-recombination system in the female germ-line of Drosophila melanogaster. II. Differential sensitivity to gamma rays. Genetics *141*, **579-585**.

Laurencon, A., Gay, F., Ducau, J., and Bregliano, J.-C. (1997). Evidence for an inducible repair-recombination system in the female germ-line of Drosophila melanogaster.III. Correlation between reactivity levels, cross over frequency and repair eficiency. Genetics *146*, **1333-1344**.

Lavige, J.-M. (1986). I-R system of hybrid dysgenesis in Drosophila melanogaster: further data on the arrest of development of the embryos from SF females. Biology of the Cell *56*, **207-216**.

Lawrence, P.A. (1993). The making of a fly: The genetics of animal design. (Oxford: Blackwell Scientific Publications).

Leigh-Brown, A.J., and Moss, J.E. (1987). Transposition of the I element and copia in a natural population of Drosophila melanogaster. Genet. Res. Camb. *49*, **121-128**.

Lerman, L.S., and Frisch, H.L. (1982). Why does the elctrophoretic mobility of DNA in a gel varies with the length of the molecule. Biopolymers *21*, **995-997**.

Levene, S.D., and Zimm, B.H. (1989). Understanding the anomalous electrophoresis of bent DNA molecules: a reptation model. Science *245*, **396-399**.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F. (1993). Transposons in place of telomeric repeats at a Drosophila telomere. Cell *75*, **1083-1093**.

Lewin, B. (1998). The mystique of epigenetics. Cell *93*, **301-303**.

Linial, M., and Shlomai, J. (1987a). The sequence-directed bent structure in kinetoplast DNA is recognised by an enzyme from Crithidia fasciculata. J. Biol. Chem. *262*, **15194-15201**.

Linial, M., and Shlomai, J. (1987b). Sequence-directed bent DNA helix is the specific binding site for Crithidia fasciculata nicking enzyme. Proc. Natl. Acad. Sci. USA *84*, **8205-8209**.

Linial, M., and Shlomai, J. (1988). Bent DNA structures associated with several origins of replication are recognised by a unique enzyme from trypanosomatids. Nucl. Acids. Res. *16*.

Lipanov, A., Kopka, M.L., Kaczor-Grzeskowiak, M., Quintana, J., and Dickerson, R.E. (1993). Structure of the B-DNA decamer CCAACITTGG in two different space groups: conformational flexibility of B-DNA. Biochem. *32*, **1373-1389**.

Lloyd, V.K., Sinclair, D.A., and Grigliatti, T.A. (1999). Genomic imprinting and position-effect variegation in Drosophila melangoster. Genetics *151*, **1503-1516**.

Luan, D.D., and Eickbush, T.H. (1995). RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. *15*, **3882-3891**.

Luan, D.D., and Eickbush, T.H. (1996). Downstream 28S gene sequences on the RNA template affect the choice of primer and the accuracy of initiation by the R2 reverse transcriptase. Mol. Cell. Biol. *16*, **4726-4734**.

Luan, D.D., Korman, M.H., Jakubzack, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, **595-605**.

Lumpkin, O.J., and Zimm, B.H. (1982). Mobility of DNA in gel electrophoresis. Biopolymers *21*, **2315-2316**.

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

Marini, J.C., Effron, P.N., Goodman, T.C., Singleton, C.K., Wells, R.D., Wartell, R.M., and Englund, P.T. (1984). Physical characterisation of a kinetoplast DNA fragment with unusual properties. J. Biol. Chem. *259*, **8974-8979**.

Marini, J.C., and Englund, P.T. (1983). Correction. Proc. Natl. Acad. Sci. USA *80*, **7678**.

Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982a). Bent helical structure in kinetoplast DNA. Proc. Natl. Acad. Sci. USA *79*, **7664-7668**.

Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982b). A bent helix in kinetoplast DNA. Cold Spring Harbor Symp. Quant. Biol. *XLVII*, **279-283**.

Marinkovic, D., and Bajraktari, I. (1988). Parental age dependent changes as a source of genetic variation in Drosophila melanogaster. Genetics *77*, **113-121**.

Martienssen, R.A., and Richards, E.J. (1995). DNA methylation in eukaryotes. Curr. Opin. Genet. Dev. *5*, **234-242**.

Martin, F., Maranon, C., Olivares, M., Alonso, C., and Lopez, M.C. (1995). Characterisation of a non-long terminal repeat retrotransposon cDNA (LITc) from Trypanosoma cruzi: homology of the first ORF with the Ape family of DNA repair enzymes. J. Mol. Biol. *247*, **49-59**.

Matzke, M.A., and Matzke, A.J.M. (1995). Homology-dependent gene silencing in transgenic plants: what does it realy tell us ? Trends Genet. *11*, **1-3**.

McCall, K., and Bender, W. (1996). Probes for chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. EMBO J *15*, **569-580**.

McLean, C., Bucheton, A., and Finnegan, D.J. (1993). The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of Drosophila melanogaster, contains an internal promoter and sequences that regulate expression. Molecular and Cellular Biology *13*, **1042-1050**.

Meller, V.H., Wu, K.H., Roman, G., Kuroda, M.I., and Davis, R.L. (1997). roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosagfe compensation system. Cell *88*, **445-457**.

Mette, M.F., van der Winden, J., Matzke, M.A., and Matzke, A.J.M. (1999). Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters. EMBO J. *18*, **241-248**.

Meyer, P., Heidmann, I., and Niedendhof, I. (1993). Differences in DNA methylation are associated with a paramutation phenomenon in transgenic petunia. Plant J. *4*, **89-100**.

Minchiotti, G., Contursi, C., and Di Nocera, P.P. (1997). Multiple downstream promoter modules regulate the transcription of the Drosophila melanogaster I, Doc and F elements. J. Mol. Biol. *267*, **37-46**.

Mizrohki, L.J., Obolenkova, L.A., Priimagi, A.F., Illyin, Y.V., Gerasimova, T.I., and Georgiev, G.P. (1985). The nature of unstable insertion mutations and reversions in the locus cut of Drosophila melanogaster: molecular mechanisms of transposition memory. EMBO J. *4*, **3781-3787**.

Mizrokhi, L.J., Georgieva, S.G., and Ilyin, Y.V. (1988). jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the Internal promoter by RNA polymerase II. Cell *54*, **685-691**.

Mizuno, T. (1987). Random cloning of bent DNA segments from Escherichia coli chromosome and primary charaterisation of their structure. Nucl. Acids Res. *15*, **6827-6841**.

Mizuno, T., and Itoh, K. (1988). Random cloning of bent DNA segments from Saccharomyces cerevisiae and primary charcterisation of their structure. Mol. Gen. Genet. *214*, **249-256**.

Moran, J.V., DeBernardinis, R.J., and Kazazian Jr, H.H. (1999). Exon shuffling by L1 retrotransposons. Science *283*, **1530-1534**.

Mount, S.M., Green, M.M., and Rubin, G.M. (1988). Partial revertants of the transposable element-assciated suppressible allele white-apricot in Drosophila melanogaster: structures and responsiveness to genetic modifiers. Genetics *118*, **221-234**.

Muller, J., and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryos. EMBO J. *10*, **3147-3155**.

Munteanu, M.G., Vlahovicek, K., Parthasarathy, S., Simon, 1., and Pongor, S. (1998). Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. Trends Biochem. Sci. *23*, **341-347**.

Murhpy, M.R., Fawlkes, D.M., and Fitzgerald-Hayes, M. (1991). Analysis of centromeric function in Saccharomyces cerevisiae using synthetis centromere mutants. Chromosoma *101*, **189-197**.

Murphy, N.B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M., and Pays, E. (1987). Trypanosome brucei repeated element with unusual structure and transcriptional properties. J. Mol. Biol. *195*, **855-871**.

Nelson, H.C.M., Finch, J.T., Luisi, B.F., and Klug, A. (1987). The structure of an oligod(A)-oligod(T) tract and its biological implication. Nature *330*, **221-226**.

Ng, H.-H., and Bird, A. (1999). DNA methylation and chromatin modification. Curr. Opin. Genet. Dev. *9*, **158-163**.

Novina, C.D., and Roy, A.L. (1996). Core promoters and transcriptional control. Trends Genet. *12*, **351-355**.

Ohyama, T., Tsujibayashi, H., Tagashira, H., Inano, K., ueda, T., Hirota, Y., and Hashimoto, K. (1998). Suppression of electrophoretic anomaly of bent DNA segements by the structural property that causes rapid migration. Nucl. Acids Res. *26*, **4811-4817**.

Orlando, V., and Paro, R. (1993). Mapping Polycomb-repressed domains using in vivo formaldehyde cross-linked chromatin. Cell *75*, **1187-1198**.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (1997). Cosuppression in Drosophila: Gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. Cell *90*, **479-490**.

Panning, B., and Jaenisch, R. (1998). RNA and the epigenetic regulation of X chromosome inactivation. Cell *93*, **305-308**.

Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F., and Traverse, K.L. (1996). Drosophila telomeres: new views on chromosome evlution. Trends Genet. *12*, **48-52**.

Pardue, M.L., Danilevskaya, O.N., Traverse, K.L., and Lowenhaupt, K. (1997). Evolutionary links between telomerases and transposable elements. Genetica *100*, **73-84**.

Paro, R., and Hogness, D. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc. Natl. Acad. Sci. USA *88*, **263-267**.

Pasero, P., Sjakste, N., Blettry, C., Got, C., and Marilley, M. (1993). Long-range organisation and sequence-directed curvature of Xenopus laevis satellite 1 DNA. Nucl. Acids Res. *21*, **4703-4710**.

Peifer, M., and Bender, W. (1986). The anterobithorax and bithorax mutations of the bithorax complex. EMBO J. *5*, **2293-2303**.

Pelisson, A. (1981). The I-R system of hybrid dysgenesis in Drosophila melanogaster: are I factor insertions responsible for the mutator effect of the I-R interaction? Molecular and General Genetics *183*, **123-129**.

Pelisson, A., and Bregliano, J.-C. (1987). Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in Drosophila melanogaster. Molecular and General Genetics *207*, **306-313**.

Phillips, M.D., and Shearn, A. (1990). Mutations in polycombeotic, a Drosophila polycomb group gene, cause a wide range of maternal and zygotic phenotypes. Genetics *125*, **91-101**.

Picard, G. (1976). Non mendelian female sterility in Drosophila melanogaster: hereditary transmission of I factor. Genetics *83*, **107-123**.

Picard, G. (1978b). Non mendelian female sterility in Drosophila melanogaster: sterility in stocks derived from the genotypically inducer or reactive offspring of SF and RSF females. Bilogie Cellulaire *31*, **245-254**.

Picard, G. (1978c). Non medelian female sterility in Drosophila melanogaster: sterility in the daughter progeny of SF anf RSF females. Biologie Cellulaire *31*, **235-244**.

Picard, G., Bregliano, J.-C., Bucheton, A., Lavige, J.-M., and Kidwell, M.G. (1978). Non-mendelian female sterility and hybrid dysgenesis in Drosophila melanogaster. Genetical Research *32*, **275-287**.

Picard, G., Bucheton, A., Lavige, J.-M., Fleuriet, A., Bregliano, J.-C., and L'Heritier, P. (1972). Further data on non mendelian female sterility in D. melanogaster. Dros. Inf.Ser. *49*, **45**.

Picard, G., and L'Heritier, P. (1971). A maternally inherited factor inducing sterility in Drosophila melanogaster. Dros.Inf.Serv. *46*, **54**.

Picard, G., Lavige, J.-M., Bucheton, A., and Bregliano, J.-C. (1977). Non Mendelian female sterility in Drosophila melanogaster: physiological pattern of embryo lethality. Biologie Cellulaire *29*, **89-98**.

Picard, G., and Pelisson, A. (1979). Non mendelian female sterility in Drosophila melanogaster: characterisation of the non-inducer chromosomes of inducer strains. Genetics *91*, **473-489**.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese., C., and Gatti, M. (1995). Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. Proc. Natl. Acad. Sci. USA *92*, **3804-3808**.

Pirrotta, V. (1990). Transvection and long-disatance gene regulation. Bioessays *12*, **409-414**.

Pirrotta, V. (1997). Chromatin-silencing mechanisms in Drosophila maintain patterns of gene expression. Trends Genet. *13*, **314-318**.

Pirrotta, V. (1998). Polycombing the genome: PcG, TrxG, and chromatin silencing. Cell *93*, **333-336**.

Plaskon, R.R., and Wartell, R.M. (1987). Sequence distribution associated with DNA curvature are found upstream of strong E.coli promoters. Nucl. Acids Res. *15*, **785-796**.

Poncin, M., Hartmann, B., and Lavery, R. (1992). Conformation sub-states in B-DNA. J. Mol. Biol. *226*, **775-794**.

Pritchard, D.K., and Schubiger, G. (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleoplasmic ratio. Genes & Development *10*, **1131-1142**.

Pritchard, M.A., Dura, J.-M., Pelisson, A., Bucheton, A., and Finnegan, D.J. (1988). A cloned I factor is fully functional in Drosophila melanogaster. Mol. Gen. Genet.

Proust, J., and Prudhommeau, C. (1982). Hybrid dysgenesis in Drosophils melanogaster. I. Further evidence for and characterisation of the mutator effect of the inducer-reactive interaction. Mutation Research *95*, **225-235**.

Proust, J., Prudhommeau, C., Ladeveze, V., Gotteland, M., and Fontyne-Branchard, M.C. (1992). I-R hybrid dysgenesis in Drosophila melanogaster. Use of in situ hybridisation to show the association of I factor DNA with induced sex-linked recessive lethals. Mut. Res. *268*, **265-285**.

Prudhommeau, C., and Proust, J. (1990). I-R hybrid dysgenesis in Drosophila melanogaster: nature and site specificity of induced recessive lethals. Mut. Res. *230*, **135-157**.

Radic, M.Z., Lundgren, K., and Hamkalo, B.A. (1987). Curvature of mouse DNA and condensation of heterochromatin. Cell *50*, **1101-1108**.

Rhodes, D. (1979). Nucleosome cores reconstituted from poly(dA-dT) and the octamer of histones. Nucl. Acids Res. *6*, **1805-1816**.

Robertson, H.M., and Lampe, D.J. (1995). Distribution of transposable elements in arthropods. Annu. Rev. Entomol. *40*, **333-357**.

Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends Biochem. Sci. *21*, **327-335**.

Rogina, B., and Helfand, S.L. (1995). Regulation of gene expression is linkd to life span in adult Drosophila. Genetics *141*, **1043-1048**.

Ronsseray, S., Lemaitre, B., and Coen, D. (1993). Maternl al inheritance of P cytotype in Drosophila melanogaster: a " pre-P cytotype" is strictly extra-chromosomally transmitted. Molecular and General Genetics *241*, **115-123**.

Rubin, G.M., Kidwell, M.G., and Bingham, P.M. (1982). The Molecular Basis of P-M Hybrid Dysgenesis: The Nature of induced mutations. Cell *29*, **987-994**.

Ryan, K.A., Shapiro, T.A., Rauch, C.A., and Englund, P.T. (1988). Replication of kinetoplast DNA in trypanosomes. Annu. Rev. Microbiol. *42*, **339-358**.

Ryder, K., Silver, S., De Lucia, A.L., Fanning, E., and Tegtmeyer, P. (1986). An altered DNA conformation in origin region I is a determinanat for the binding of SV40 large T antigen". Cell *44*, **719-725**.

Sabl, J.F., and Henikoff, S. (1996). Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. Genetics *142*, **447-458**.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning. A laboratory manual., 2nd Ed Edition, Volume 1, C. Nolan, ed. (NY: Cold Spring Harbor Laboratory Press.).

Sang, H.H., Pelisson, A., Bucheton, A., and Finnegan, D.J. (1984). Molecular lesions associated with white gene mutations induced by I-R hybrid dysgenesis in Drosophila melanogaster. EMBO J. *3*, **3079-3085**.

Schlossherr, J., Eggert, H., Paro, R., Cremer, S., and Jack, R.S. (1994). Gene inactivation in Drosophila mediated by the Polycomb gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fiber. Mol. Gen. Genet. *243*, **453-462**.

Schwarz-Sommer, Z., Leclercq, L., Gobel, E., and Saedler, H. (1987). Cin4, an insert altering the structure of the A1 gene in Zea mays, exhibits properties of non viral retrotransposons. EMBO, J. *6*, **3873-3880**.

Seleme, M.d.C., Busseau, I., Malinsky, S., Bucheton, A., and Teninges, D. (1999). High-frequency retrotransposition of a marked I factor in Drosophila melanogaster correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. Genetics *151*, **761-771**.

Selker, E.U. (1997). Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? Trends Genet. *13*, **296-301**.

Selsing, E., Wells, R.D., Alden, C.J., and Arnott, S. (1979). Bent DNA: visualisation of a base-paired and stacked A-B conformational junction. J. Biol. Chem. *254*, **5417-5422**.

Sezutsu, H., Nitasaka, E., and Yamazaki, T. (1995). Evolution of the LINE-like I element in the Drosophila melanogaster species group. Mol. Gen. Genet. *249*, **168-178**.

Shapiro, T.A. (1993). Kinetoplast DNA maxicircles: networks within networks. Proc. Natl. Acad. Sci. USA *90*, **7809-7813**.

Shapiro, T.A., and Englund, P.T. (1995). The structure and replication of kinetoplast DNA. Annu. Rev. Microbiol. *49*, **117-143**.

Shi, Y., Lee, J.-S., and Galvin, K.M. (1997). Everything you have ever wanted to know about Yin Yang 1... Biochim. Biophys. Acta *1332*, **F49-F66**.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Conner, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev. Biol. *158*, **131-144**.

Simonelig, M., Bazin, C., Pelisson, A., and Bucheton, A. (1988). Transposable and nontransposable elements similar to the I factor involved in inducer-reactive (IR) hybrid dysgenesis in Drosophila melanogaster coexist in various Drosophila species. Proc. Natl. Acad. Sci. USA *85*, **1141-1145**.

Simpson, L. (1979). Isolation of maxicircle component of kinetoplast DNA fom hemoflagellate protozoa. Proc. Natl. Acad. Sci. USA *76*, **1585-1588**.

Simpson, R.T., and Kunzler, P. (1979). Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. Nucl. Acids Res. *6*, **1387-1415**.

Singer, M.F. (1992). SINEs and LINEs: highly specialised short and long interspersed sequences in mammalian genomes. Cell *28*, **433-434**.

Singer, M.F., and Skowronski, J. (1985). Making sense out of LINEs: long interspersed repeat sequences in Mammalian genomes. TIBs *10*, **119-122**.

Singh, P.B., Miller, J.R., Pearce, J., Kothary, R., Burton, R.D., and Paro, R. (1991). A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. Nucl. Acids. Res. *19*, **789-794**.

Smale, S.T., and Baltimore, D. (1989). The "initiator" as a transcriptional control element. Cell *57*, **103-113**.

Smale, T.S. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim. Biophys. Acta *1351*, **73-88**.

Snyder, M., Buchman, A.R., and Davis, R.W. (1986). Bent DNA at a yeast autonomously replicating sequence. Nature *324*, **87-89**.

Spana, C., and Corces, V.G. (1990). DNA bending is a determinant of binding specificity for a Drosophila zinc finger protein. Genes Dev. *4*, **1505-1515**.

Spradling, A.C., and Rubin, G.M. (1982). Transposition of cloned p elements into Drosophila germ-line chromosomes. Science *218*, **341-347**.

Stacey, S.N., Lansman, R.A., Brock, H.W., and Grigliatti, T.A. (1986). Distribution and conservation of mobile elements in the genus Drosophila. Mol. Biol. Evol. *3*, **522-534**.

Steinemann, M., Steinemann, S., and Lottspeich, F. (1993). How Y chromosomes become genetically inert. Proc. Natl. Acad. Sci. USA *90*, **5737-5741**.

Strutt, H., Cavalli, G., and Paro, R. (1997). Co-localisation of polycomb protein and GAGA factor on regulatory elements repsonsible for the maintenance of homeotic gene expression. EMBO J. *16*, **3621-3632**.

Sved, J. (1976). Hybrid Dysgenesis in Drosphila melanogaster: a possible explanation in terms of spatial organisation of chromosomes. Aust. J. Biol. Sci. *29*, **375-388**.

Tartof, K.D., and Henikoff, S. (1991). Trans-sensing effects from Drosophila to humans. Cell *65*, **201-203**.

Tatout, C., Docquier, M., Lachaume, P., Mesure, M., Lecher, P., and Pinon, H. (1994). Germline expression of a functional LINE from Drosophila melanogaster: fine characterization allows for potential investigations of trans-regulators. Int. J. Dev. Biol. *38*, **27-33**.

Travers, A.A. (1989). DNA conformation and protein bending. Annu. Rev. Biochem. *58*, **427-452**.

Trifonov, E.N., and Sussman, J.L. (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. Proc. Natl. Acad. Sci. USA *77*, **3816-3820**.

Udomkit, A., Forbes, S., Delgleish, G., and Finnegan, D.J. (1995). BS a novel LINE-like element in Drosophila melanogaster. Nucl. Acids Res. *23*, **1354-1358**.

Udomkit, A., Forbes, S., McLean, C., Arkhipova, I., and Finnegan, D.J. (1996). Control of expression of the I factor, a LINE-like transposable element in Drosophila melanogaster. EMBO, J. *15*, **3174-3181**.

Ulyanov, N.B., and James, T.L. (1995). Statistical analysis of DNA duplex structure features. Met. Enzymol. *261*, **90-115**.

Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., and Razin, A. (1982). The absence of detectable methylated bases in Drosophila melanogaster. FEBS lett. *146*, **148-152**.

van West, P., Kamoun, S., van't Klooster, J.W., and Govers, F. (1999). Internuclear gene silencing in Phytophthora infestans. Mol. Cell *3*, **339-348**.

Vaucheret, H., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Mourrain, P., Palauqui, J.-C., and Vernhettes, S. (1998). Transgene induced gene silencing in plants. Plant J. *16*, **651-659**.

Vaury, C., Abad, P., Pelisson, A., Lenoir, A., and Bucheton, A. (1990). Molecular characteristics of the heterochromatic I elements from a reactive strain of Drosophila melanogaster. J. Mol. Evol. *31*, **424-431**.

Vaury, C., Bucheton, A., and Pelisson, A. (1989). The Beta heterochromatic sequences flanking the I elements are themselves defective transosable elements. Chromosoma *98*, **215-224**.

Vaury, C., Drake, M.E., Lajoinie, O., and Dastugue, B. Unpublished. .

Vaury, C., Pelisson, A., Adad, P., and Bucheton, A. (1993). Properties of transgenic strains of Drosophila melanogaster containing I transposable elements from Drosophila tessieri. Genet. Res. Camb. *61*, **81-90**.

Verrijzer, C.P., and Tjian, R. (1996). TAFs mediate transcriptional activation and promoter selectivity,. Trends Biochem. Sci. *21*, **338-342**.

Vieira, C., Lepetit, D., Dumont, S., and Biemont, C. (1999). Wake up of transposable elements following Drosophila simulans world wide colonisation. In press.

Wallarth, L.L. (1998). Unfolding the mysteries of heterochromatin. Curr. Opin. Genet. Dev. *6*, **147-153**.

Weis, L., and Reinberg, D. (1992). Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J. *6*, **3300-3309**.

Willard, H.R., and Saltz, H.K. (1997). Remodelling chromatin with RNA. Nature *386*, **228-229**.

Williams, J.S., Eckdahl, T.T., and Anderson, J.N. (1988). Bent DNA as a replication enhancer in Saccharomyces cerevisiae. Mol. Cell. Biol. *8*, **2763-2769**.

Wu, H.-M., and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending. Nature *308*, **509-513**.

Xiong, Y., and Eickbush, T.H. (1988). Functional expression of a sequence specific endonuclease encoded by the retrotransposon R2Bm. Cell *55*, **235-246**.

Yean, D., and Gralla, J. (1997). Transcription reinitiation rate: a special role for the TATA box. Mol. Cell Biol. *17*, **3809-3819**.

Ytterborn, K.H. (1985). Hybrid dysgenesis: evolution of different I-R properties in related strains of Drosophila melanogaster. Hereditas *102*, **165-170**.

Zahn, K., and Blattner, F.R. (1985). Sequence-induced DNA curvature at the bacteriophage lambda origin of replication. Nature *317*, **451-453**.

Zahn, K., and Blattner, F.R. (1987). Direct evidence for DNA bending at the lambda replication origin. Science *236*, **416-422**.

Zawel, L., and Reinberg, D. (1993). Initiation of transcription by RNA polymerase II: a multi-step process. Prog. Nucl. Acid Res. & Mol. Biol. *44*, **67-108**.

Zhurkin, V.B., Lysov, Y.P., and Ivanov, V.I. (1979). Anisotropic flexinility of DNA and nucleosomal structure. Nucl. Acids. Res. *6*, **1081-1096**.

Zink, B., and Paro, R. (1989). In vivo binding pattern of a trans-regulator of homeotic genes in Drosophila melanogaster. Nature *337*, **468-471**.

Zink, B., and Paro, R. (1995). Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. EMBO J. *14*, **5660-5671**.

Zwaan, B.J., Bijlsma, R., and Hoekstra, R.F. (1992). On the developmental theory of ageing.II.The effect of developmental temperature on longevity in realation to adult body size in Drosophila melanogaster. Heredity *68*, **123-130**.