Transcriptional control of the mitotic regulator *string,* in *Drosophila*

A thesis submitted for the degree of Doctor of Philosophy

by

Briony Patterson, B.Sc. (Hons)

Departments of Biochemistry and Genetics
University of Adelaide
Adelaide, S.A., 5005
Australia

December 1996

TABLE OF CONTENTS

ABSTRACT

Before morphogenesis can begin, positional information must be in place to direct the complex array of cellular processes. In a *Drosophila* embryo the transcription factors encoded by the patterning genes are expected to provide this information. However, the genes regulated in this manner, to bring about morphogenesis, have so far been elusive. One potential candidate is *string*, a homologue of the mitotic initiator *cdc25* from *Schizosaccharomyces pombe*. Early in embryogenesis a complex spatio-temporal pattern of *string* transcription partitions the embryo into mitotic domains which delineate larval organ primordia. The complexity and timing of *string* transcription, at such an early stage, suggests that this morphogenetic event is regulated by the patterning genes.

If *string* is integrating patterning gene information it is anticipated that the promoter may be comprised of many position specific elements (PSE's), each defining individual mitotic domains. To test this, promoter fragments from the *string* gene were used in *lacZ* reporter constructs to look for these PSE's. Initial constructs failed to identify any such elements. However, the addition of a proximal *stg* promoter fragment, to these constructs, allowed the identification of some *stg* regulatory regions that activated transcription in specific domain-like patterns. In particular, PSE's for cycle 14 domains 1, 2, and 21, were identified as well as distinct PSE's for different subsets of domains N and M.

The PSE's that activated transcription in domains 1 and 2 were defined to a region of 1.4kb by further construct generation and a detailed analysis of the regulation of mitotic domain 2 was undertaken. The early patterning genes *buttonhead (btd)* and *snail (sna)* were found to be required for the transcriptional regulation of domain 2, and their products were shown bind specifically to the defined 1.4kb region.

This work has demonstrated that *string* is a downstream target of the patterning genes, making a direct connection between patterning information and morphogenesis, which suggests that mitotic timing forms an independent and important part of morphogenesis.

STATEMENT

This work contains no material which has been accepted for the award of any other degree or diploma in any university or tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent for my thesis, when deposited in the University library, being made available for loan and photocopying.

Briony Patterson, December 1996

ACKNOWLEDGMENTS

Firstly, I want to thank Rob Saint for giving me the opportunity to do this work. His supervision has always been supportive while giving me the freedom to make my own decisions, good or bad, and I feel that I have learnt much from this.

To the members of the Saint lab who were there at the beginning (all those years ago): Pete Wigley, Richard D'Andrea, Rick Tearle, Paul Moretti, Louise O'Keefe, Bill Kalionis and Gary Hime. I am very grateful to them all for their help and encouragement.

Thanks also to all of the present members of the Saint Lab for creating a supportive and interactive work environment. In particular to: Helena for her advice and genuine interest in my work, Steve and Dan for help on the computers, Julie for injecting some of my constructs, Michelle for her competent cells, Lynn for being a transcription soul mate, Julianne for her cakes, Michael for the late night chats, and Louise for the coffee breaks and all of those words of encouragement.

Also to Paul Moir, for fixing anything that was broken, and the central services unit staff for providing plenty of buffers, media and clean glassware which were much appreciated.

I would also like to acknowledge Bruce Edgar and Dara Lehman for helpful discussions and during my short time in Seattle, from which I benefited a great deal. I am very grateful to have had this opportunity.

One sure way to slow down your Ph.D is to have a baby! I have absolutely no regrets though. Having Alice around has given us so much enjoyment and she helps me to keep science and life in perspective.

Lastly, to Stanley for always being there when I need him, for whatever reason. Talking science, finishing my minipreps, cooking dinner, etc. etc. Thanks Stanno.

CHAPTER 1: INTRODUCTION

In *Drosophila*, the pattern of cellular proliferation required to form a larva from a fertilised egg follows an invariant pathway. Initially, a succession of rapid divisions gives rise to an undifferentiated single cell epithelium, the cellular blastoderm. Proliferation after this time follows a very distinct and controlled pattern that is coupled to other morphogenetic events. Studying how this patterned proliferation is regulated at such an early stage in development may give us some insights into the role of proliferation during development

1-1 PROLIFERATION IN THE *Drosophila* EMBRYO

1-1.1 Syncytial divisions

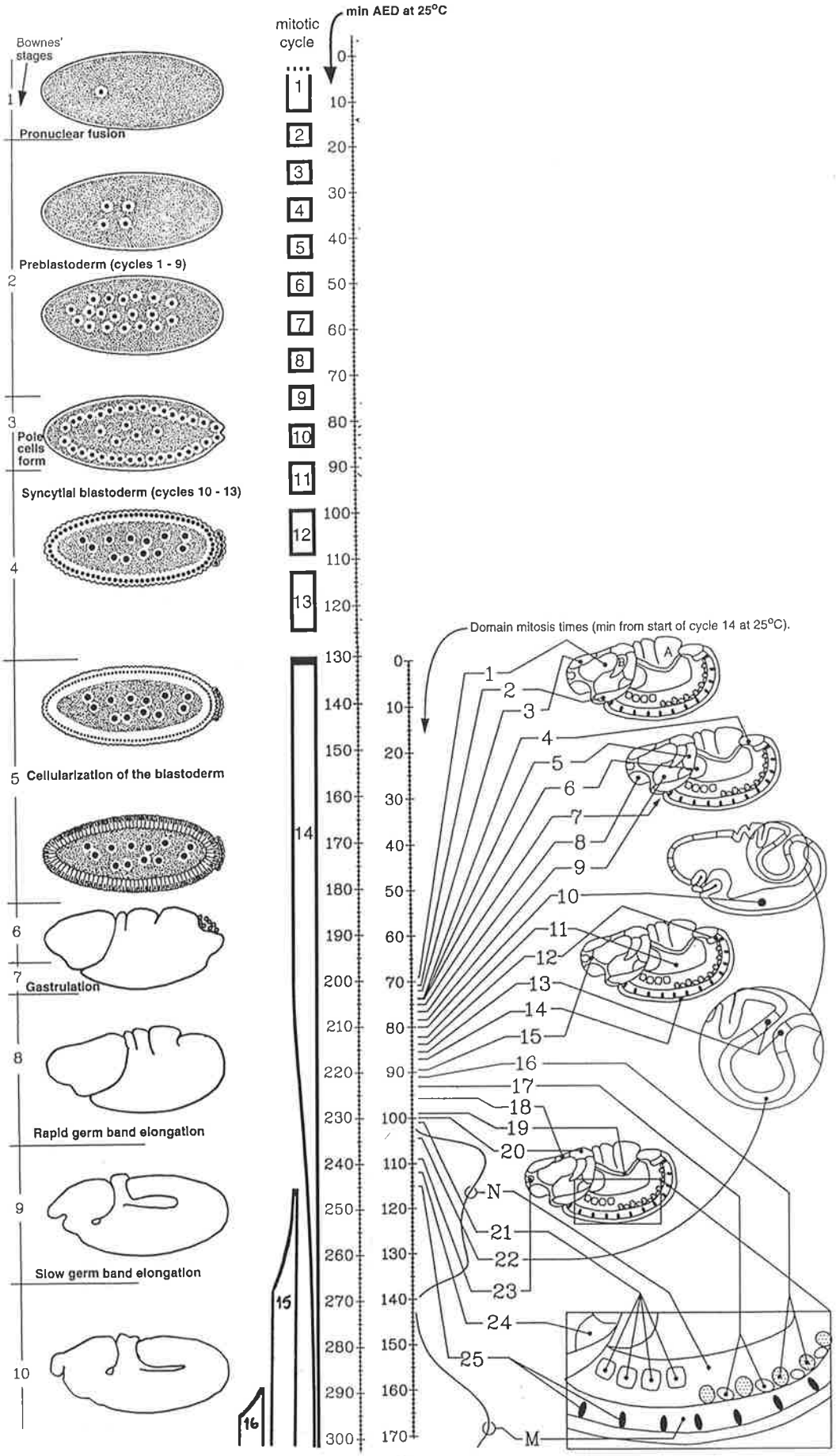
In the newly fertilised *Drosophila* embryo the male and female pronuclei replicate individually and then, using a common mitotic spindle, undergo the first mitosis. No cytokinesis is associated with this pronuclear division, leaving the two resulting zygotic nuclei to share the same cytoplasm (B. P. Sonnenblick 1950). These nuclei proceed rapidly and synchronously through a further 12 rounds of mitosis (making a total of 13 cycles of division) before cell membranes are finally formed (F. R. Turner and A. P. Mahowald 1976). During these early cycles, as the number of nuclei increases they describe a hollow ellipsoid shape below the cortex of the embryo. The nuclei then migrate through the cortex to the periphery, forming a syncytial blastoderm by the completion of cycle 10 (V. E. Foe and B. M. Alberts 1983). From then on, the rate of mitosis slows slightly for each successive round of division, increasing from 8 minutes at cycle 10, to 17 minutes for cycle 13 (V. E. Foe 1989; V. E. Foe and B. M. Alberts 1983). The rigid synchronicity of mitosis is relaxed slightly during cycles 10 through to 13 such that mitosis occurs in metachronous waves which move rapidly across the blastoderm from each pole towards the equator (V. E. Foe and B. M. Alberts 1983). There is a significant pause in proliferation following cycle 13 and it is during this period that membranes extend from the periphery to surround each of the blastoderm nuclei and generate the cellular blastoderm (summarised in figure 1-1).

1-1.2 Cellular blastoderm, pole cells and yolk nuclei

Of the nuclei generated by 13 rounds of division, approximately 6000 contribute to the cellular blastoderm (F. R. Turner and A. P. Mahowald 1976) forming the somatic component of the embryo. Of the remaining nuclei, a small number migrate to the reach the egg membranes in the posterior region one cycle ahead of most of the nuclei (in cycle 9) and then cellularise (in cycle 10), these are the germ cell progenitors or 'pole cells' (F. R. Turner and A. P. Mahowald 1976). Once they have cellularised, proliferation in the pole cells follows a schedule independent of the somatic nuclei (G. M. Technau and J. A. Campos-Ortega 1986). A proportion of the blastoderm nuclei never migrate to the periphery, or, begin to migrate but then fall back to the interior of the embryo (V. E. Foe and B. M. Alberts 1983). These are the yolk nuclei which will later become polyploid (A. V. Smith and T. L. Orr-Weaver 1991).

Figure 1-1

A time line showing the proliferative events that occur in the first 300 minutes of embryogenesis. The embryo pictures on the left identify the morphogenetic events that are occurring as the embryo progresses through the mitotic cycles. On the right is a detailed schedule of the mitotic domains of cycle 14. The tapering line in mitotic cycle 14 represents the decreasing number of cells that still remain in cycle 14. This is gradually replaced by a tapering line that represents cells that have entered mitosis of cycle 15 and likewise for cycle 16. Reproduced (in a modified form) from Foe *et al.*,(1993).


Bownes' stages
mitotic cycle
min AED at 25°C
Pronuclear fusion
Preblastoderm (cycles 1 - 9)
Pole cells form
Syncytial blastoderm (cycles 10 - 13)
Cellularization of the blastoderm
Gastrulation
Rapid germ band elongation
Slow germ band elongation
Domain mitosis times (min from start of cycle 14 at 25°C).
A
N
M

1-1.3 Cycle 14 The first post-blastoderm division

For the first 55 minutes of cycle 14, cellularisation is in progress and although the cells enter cycle 14, and replicate their DNA, mitosis of cycle 14 does not immediately follow (B. A. Edgar and P. H. O'Farrell 1990; B. A. Edgar and G. Schubiger 1986; S. L. McKnight and O. L. Miller Jr. 1977). This is the first time a gap phase (G2) is introduced into the cell cycle, unlike the previous cycles where nuclei are either in mitosis or DNA replication (B. A. Edgar and G. Schubiger 1986). The length of this G2 phase varies across the embryo (B. A. Edgar and P. H. O'Farrell 1990) and as a result, mitosis of cycle 14 is asynchronous. However, within this asynchrony a complex pattern arises where regions of cells divide in synchrony (V. Hartenstein and J. A. Campos-Ortega 1985). These regions, termed mitotic domains, have been mapped in detail and found to be invariant from one embryo to another, in both pattern and relative timing of mitosis (V. E. Foe 1989).

1-1.4 Mitotic domains

Foe (1989) identified 25 mitotic domains during cycle 14 which cover the embryo in a bilaterally symmetric pattern (Figure 1-2). These domains enter mitosis over a period of 45 minutes with domain 1 ($\partial_{14}1$) being the first (70 minutes into cycle 14) and domain 25 ($\partial_{14}25$), the last (115 minutes into cycle 14).

A domain may consist of a single region of cells if it spans the lateral plane of symmetry of the embryo. Domains in this class are; $\partial_{14}3$, 8, 15, 18, 20 and 23 in the head region of the embryo and $\partial_{14}10$ which forms a broad band along the length of the ventral side of the embryo (see figure 1-2). Alternatively, a domain may consist of two identical regions of cells on the left and right sides of the embryo such as domains; $\partial_{14}1$, 2, 9, and 24 in the head and $\partial_{14}4$, 5, 6, 7, 12, 13, 14, 19 and 22 in the trunk. Some domains are comprised of a pair of regions that are identical on the left and right sides that are metamerically reiterated along the body of the embryo, as in $\partial_{14}11$, 16, 17, 21 and 25.

While the majority of the cellular blastoderm goes to form these 25 mitotic domains, there are two regions of the embryo which Foe termed 'non-dividing' domains A and B (see figure 1-2, panels A and C). No mitosis was observed in these regions during the period when the post-blastoderm divisions were under way. This suggests that domains A and B arrest their proliferation in G2 of cycle 14. However, later on the nuclei of domain A have been observed to enter a modified version of the cell cycle, termed endoreplication, where replication proceeds in the absence of mitosis (A. V. Smith and T. L. Orr-Weaver 1991).

The complexity of the cycle 14 mitotic domains in the region of the ventral neurectoderm lead Foe (1989) to group most of these cells into two domains where mitosis was late and asynchronous, domains N and M (∂_{14}N and ∂_{14}M). In these two domains the length of the cycle 14 G2 varies widely. The cells of ∂_{14}N begin to divide as $\partial_{14}21$ enters mitosis (103 minutes into cycle 14) and continue to divide for nearly 40 minutes, well after $\partial_{14}25$ has completed its mitosis. ∂_{14}M begins mitosis after the completion of ∂_{14}N and extends from 140 to 190 minutes into cycle 14.

A detailed study of the pattern of mitosis in the ventral neurectoderm (VN) has revealed many more metamerically reiterated domains within Foe's ∂_{14}N and M (V. Hartenstein *et al.*, 1994) (figure 1-3). Incorporation of a nucleotide base analogue (BrdU), allows domains of cells to be identified, not just for the brief period when synchronous mitosis can be visualised but for the longer period from the initiation of DNA replication to the following round of mitosis. Three longitudinal strips were identified

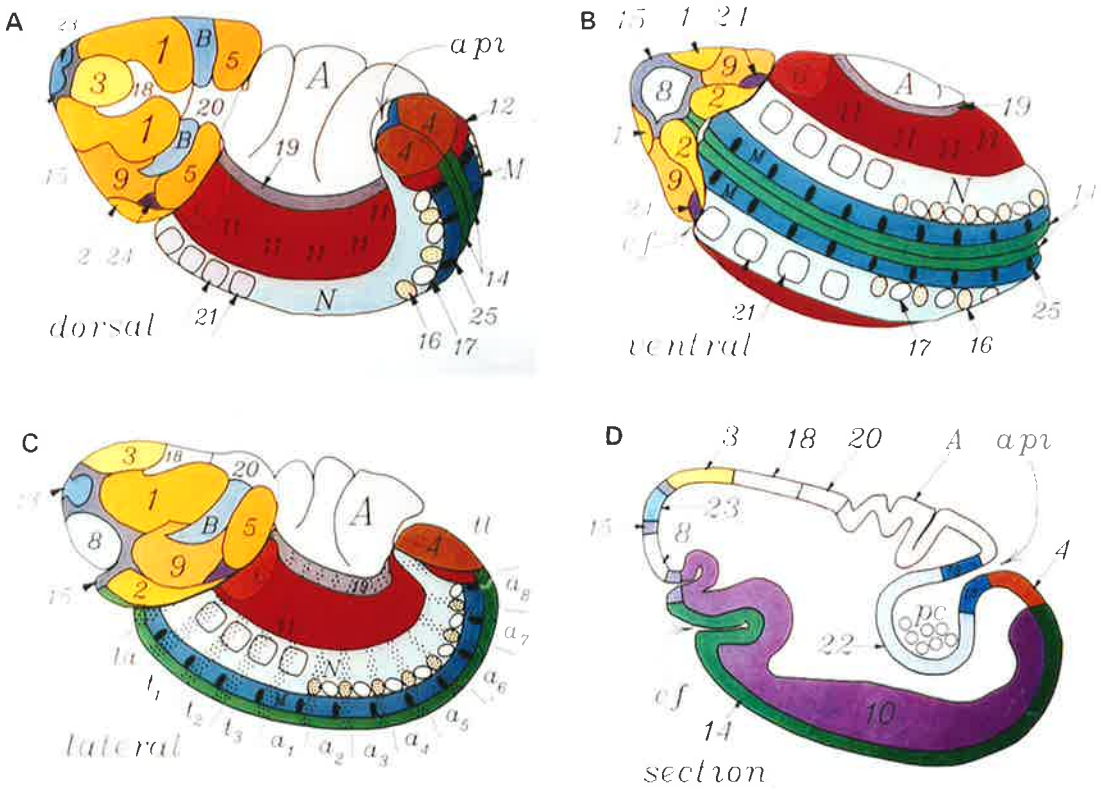


Figure 1-2

Atlas of the mitotic domains of cycle 14, projected onto an early germband extending embryo. Dorsal (**A**), ventral (**B**), lateral (**C**) and midsagittal-section (**D**) views are shown with the anterior of the embryo to the left and the dorsal surface upwards except in the ventral view where the embryo has been rotated such that the dorsal surface faces into the page. Domains 1 to 25 and domains N and M are identified as well as the non-dividing domains A and B. In the lateral view, segments are marked; labial (la), thoracic (t1 to t3), and abdominal (a1 to a8) and stippling shows the location of Engrailed protein. Other structures marked are; the lumen of the amnioproctodeal invagination (api), the pole cells (pc), the cephalic furrow (cf), and the tail region (tl). Reproduced from Foe (1989).

within the VN, the lateral, intermediate and medial zones (∂_{14}N corresponds to the lateral and intermediate VN, and ∂_{14}M to the medial VN, see figure 1-3). Along each of these three zones there are metamerically repeated domains (with some slight differences in the pattern between the thoracic and abdominal segments), the lateral VN (lVN) has two domains per parasegment, lVN1 and 2 while the intermediate VN (iVN) and the medial VN (mVN) have at least four. Some of the earlier of these domains were identified by Foe from within ∂_{14}N and M; ∂_{14}16 and 17 correspond to iVN1 and 3 respectively*, domain ∂_{14}21 corresponds to lVN1, and ∂_{14}25 is mVN4. The timing of replication in the remaining domains has been described by Hartenstein *et al.,* (1994), and falls within the limits that Foe determined for ∂_{14}N and M. Including the data of this study, there are now at least 31 mitotic domains at cycle 14 that enter mitosis over a period of 120 minutes, by which time many of the earlier mitotic domains have already entered mitosis of cycle 15, and some even that of cycle 16 (V. E. Foe *et al.,* 1993).

*Although Hartenstein *et al.,* (1994) report ∂_{14}16 and 17 corresponding to iVN3 and 1 respectively, it seems more likely that this would be the other way around. Foe mapped ∂_{14}16 within the stripe of engrailed expression which marks the anterior third of the parasegment. The nomenclature used by Hartenstein and co-workers, numbers the domains sequentially from the anterior of each parasegment, therefore iVN1 must be within the engrailed stripe and not iVN3 (see figure 1-3).

1-1.5 Cycles 15 and 16

Cycles 15 and 16 represent the last two divisions for the vast majority of the cells of the developing embryo. A detailed study of these cycles has not yet been achieved because of technical problems associated with the increased complexity of the embryo at this stage of development. However, while domains have been found with the same boundaries in cycle 15 as in cycle 14 (V. E. Foe *et al.,* 1993), the boundaries of many of the cycle 15 domains appear to be further subdivided (V. E. Foe *et al.,* 1993; V. Hartenstein *et al.,* 1994). As in cycle 14, DNA replication in these cycles follows immediately after the previous mitosis (and requires the same time duration) implying that further changes in the mitotic domain pattern are again determined by varying the length of G2 (B. A. Edgar and P. H. O'Farrell 1990) .

1-1.6 Proliferation in the neural tissues

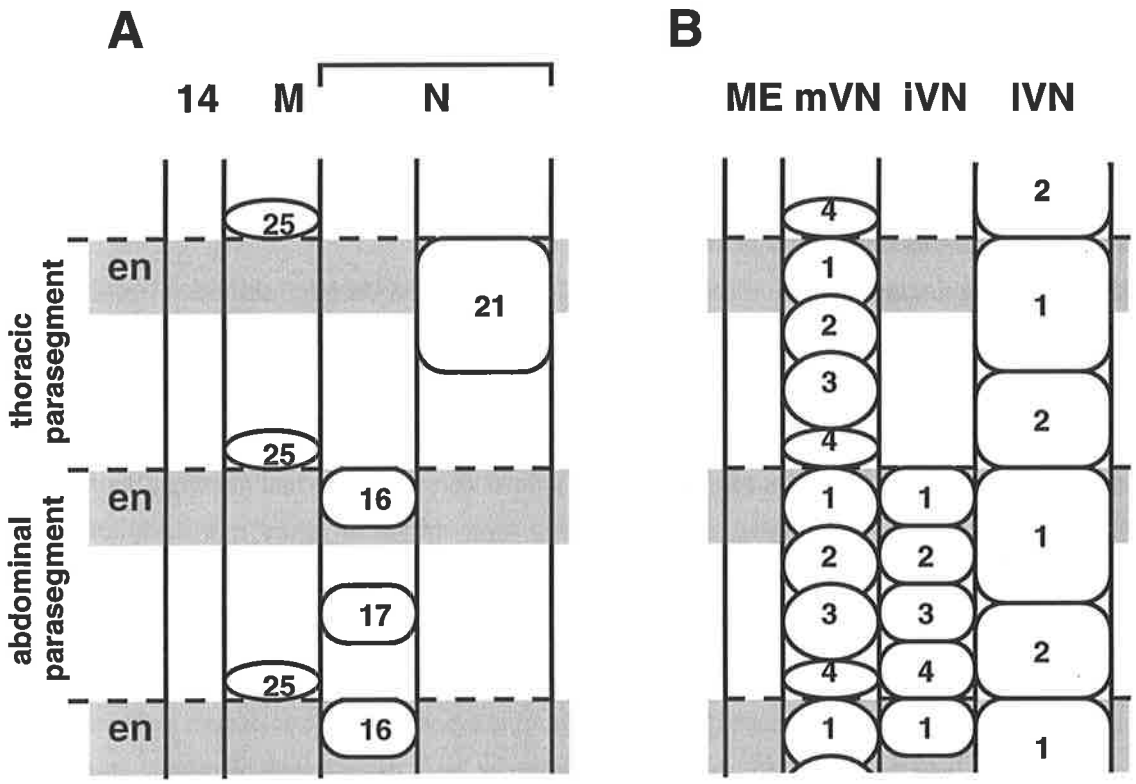
Most cells of the embryo cease to proliferate following their sixteenth mitosis and enter their first G1 phase (B. A. Edgar and P. H. O'Farrell 1990). However, the tissues that form the nervous system of the larva and the adult fly continue to proliferate in a precise spatio-temporal pattern for significantly longer (reviewed in (V. E. Foe *et al.,* 1993)). In some cases, up to 9 more divisions have been identified after cycle 16 (reviewed in (B. A. Edgar 1995)) and for at least the first neuroblast division it appears that S phase follows immediately after mitosis, suggesting that the timing of this division is still regulated by the length of the G2 phase (K. Weigmann and C. F. Lehner 1995).

1-1.7 Endoreplication

Later on in embryogenesis and larval life, the G1 arrested cells that form the larval tissues enter an endoreplication cycle to become polyploid. No cell division occurs in these endoreplication cycles as they consist only of a period of DNA replication followed by a gap phase (A. V. Smith and T. L. Orr-

Figure 1-3

A mitotic domain map of the ventral neurogenic region showing how the domains assigned by Foe (1989) (**A**) align with those of Hartenstein *et al.*, (1994)(**B**). The two panels (A and B) represent a thoracic and an abdominal parasegments from the right hand side of the ventral neurectoderm with stripes of engrailed expression identifying the anterior third of each parasegment. Domain 14 marks the ventral midline of the embryo which corresponds to the mesectoderm (ME). Domain M corresponds to the medial ventral neurectoderm (mVN), with domain 25 and mVN4 being the same cells. Domain N is divided into domains of the intermediate and lateral VN (iVN and lVN) from within which, Foe's domain's 16,17 and 21 can be identified. The numbering of new domains follows their position within the parasegment (from anterior to posterior) rather than their order of entry into mitosis. Although Foe observed domain 21 only in the thoracic segments and domains 16 and 17 only in the abdominal segments, Hartenstein *ct al.*, (1994) suggest that 21 may also be present in abdominal segments. The pattern of mitosis in the iVN of the thoracic parasegment is not described in sufficient detail in Hartenstein *et al.*, (1994) to include this information.


A
14
M
N
thoracic parasegment
abdominal parasegment
en
en
en
25
25
25
21
16
17
16
B
ME
mVN
iVN
lVN
4
1
2
3
4
1
2
3
4
1
1
2
3
4
1
2
1
2
1
2
1

Weaver 1991). The G1 arrested cells that go to form the imaginal tissues of the adult do not become polyploid, rather they remain in G1 until specific times in larval development when they re-initiate proliferation to form the adult structures (reviewed in (B. A. Edgar 1995)).

1-1.8 Mitotic domains delineate larval organ primordia

The earliest visible morphogenetic event in the cellular blastoderm is the commencement of gastrulation, 55 minutes into cycle 14. The cephalic fold and the ventral furrow invaginate and then, at about the same time as the first domains are entering mitosis, germ band extension begins. The size and shape of particular domains and the timing of their mitosis is tightly linked to the events of gastrulation suggesting that they serve a function in the development of the embryo. Indeed, mitotic domains exhibit specialised cell biological behaviours some of which can be identified even before they undergo their synchronous mitosis (V. E. Foe 1989). Dramatic changes in cell shape or the orientation of mitotic spindles can distinguish a domain and reveal underlying differences in cell commitment even at this early stage in development.

As the embryo develops beyond mitosis of cycle 14, differences between domains become more apparent. However, cell movements make it difficult to relate the developing tissues to the mitotic domains of the cellular blastoderm. Foe (1989) made use of 'fate maps' compiled from a large number of tissue ablation and transplantation studies, to trace mitotic domains to their location in the larva. From this she was able to determine that the cells of mitotic domains do in fact identify the embryonic primordia of particular larval organs, confirming that they reflect an early manifestation of cell commitment (V. E. Foe 1989).

Although it is commonly accepted that organogenesis requires controlled cellular proliferation. It is striking that the exact timing of mitosis in primordial larval tissues should be so intricately regulated at a stage when it is hardly possible to distinguish them, morphologically. Why this occurs is the subject of much speculation (see section 1-5). An understanding of how the pattern of mitotic domains is determined may uncover some clues towards answering this question.

1-2 *string* IS RESPONSIBLE FOR THE MITOTIC DOMAIN PATTERN

1-2.1 The *string* phenotype

In a near saturation mutagenesis screen of the *Drosophila* genome for recessive lethal mutations that alter the pattern of the larval cuticle, eight alleles of a locus called *string (stg)* were uncovered (G. Jürgens *et al.*, 1984). In these mutants, the number of denticle rows per segment was strongly reduced. *stg* mutant embryos were subsequently found to have far fewer cells than normal embryos (B. A. Edgar and P. H. O'Farrell 1989), suggesting a defect in cell proliferation. A more detailed analysis revealed that while progression through the first thirteen division cycles was unaffected, any further rounds of mitosis failed to occur (B. A. Edgar and P. H. O'Farrell 1989). The DNA content of the nuclei where proliferation had arrested also revealed that replication had occurred after mitosis 13 and therefore the mutant embryos were arrested in G2 of cycle 14.

Although no further divisions were detected after cycle 13, other developmental processes such as gastrulation and differentiation continue virtually unaffected (B. A. Edgar and P. H. O'Farrell 1989; A.

P. Gould *et al.,* 1990; P. Hartenstein and J. W. Posakony 1990), suggesting that the *stg* mutation is having an effect specifically on the progression of the cell cycle.

It is likely that the arrest in *stg* mutants, at G2 phase of cycle 14 represents the earliest point at which zygotic transcription is required for cell proliferation in the embryo. It has been shown that effectively no zygotic transcription occurs in the embryo before cycle 10 (B. A. Edgar and G. Schubiger 1986) and, using α-amanitin as a transcription inhibitor, that zygotic transcription is not required for cell cycle progression until cycle 14 (B. A. Edgar *et al.,* 1986; B. A. Edgar and P. H. O'Farrell 1989). It may be expected, then, that the phenotype of a zygotically expressed gene affecting cell cycle progression would not be apparent until at least cycle 14, with the requirements for earlier cycles being met by maternal products. Arrest at G2 of cycle 14 makes *stg* the earliest known zygotic mutant that affects cell cycle progression. The fact that *stg* is the earliest, and only identified, zygotic mutant to block cell cycle progression at this stage also suggests that zygotic expression of *stg* is the limiting factor for entry into mitosis of cycle 14.

1-2.2 Cloning the *string* gene

The availability of P element insertion mutants that were unable to complement *stg* mutants obtained in the ethyl methane sulfonate (EMS) screen of Jürgens *et al.*,(1984) facilitated the cloning of the *stg* cDNA. The P element insertion l(3)neo62 (at 99A) was found to be unable to complement the EMS *stg* mutants and further, transheterozygotes between l(3)neo62 and one of the EMS mutants (stg^{7L105}) showed weak *stg* cuticle phenotypes (B. A. Edgar and P. H. O'Farrell 1989). Plasmid rescue was utilised to obtain genomic DNA surrounding the P element insertion site, which was then hybridised to developmental Northern blots to identify transcripts that were expressed during periods of cell proliferation. A 2.8kb transcript from the region closest to the P element insertion site was found to be expressed at high levels during the early cycles (1-13), then absent in embryos that had completed cycle 13 but not yet entered mitosis of cycle 14. It was observed again at high levels in embryos undergoing cycles 14 and 15 but was barely detectable later in embryogenesis, when the majority of proliferation has ceased. This evidence strongly suggests that the 2.8kb cDNA corresponds to the *stg* locus and is supported by the fact that embryos homozygous for a null allele of *stg* fail to express this transcript, as determined by whole mount *in situ* hybridization (B. A. Edgar and P. H. O'Farrell 1989).

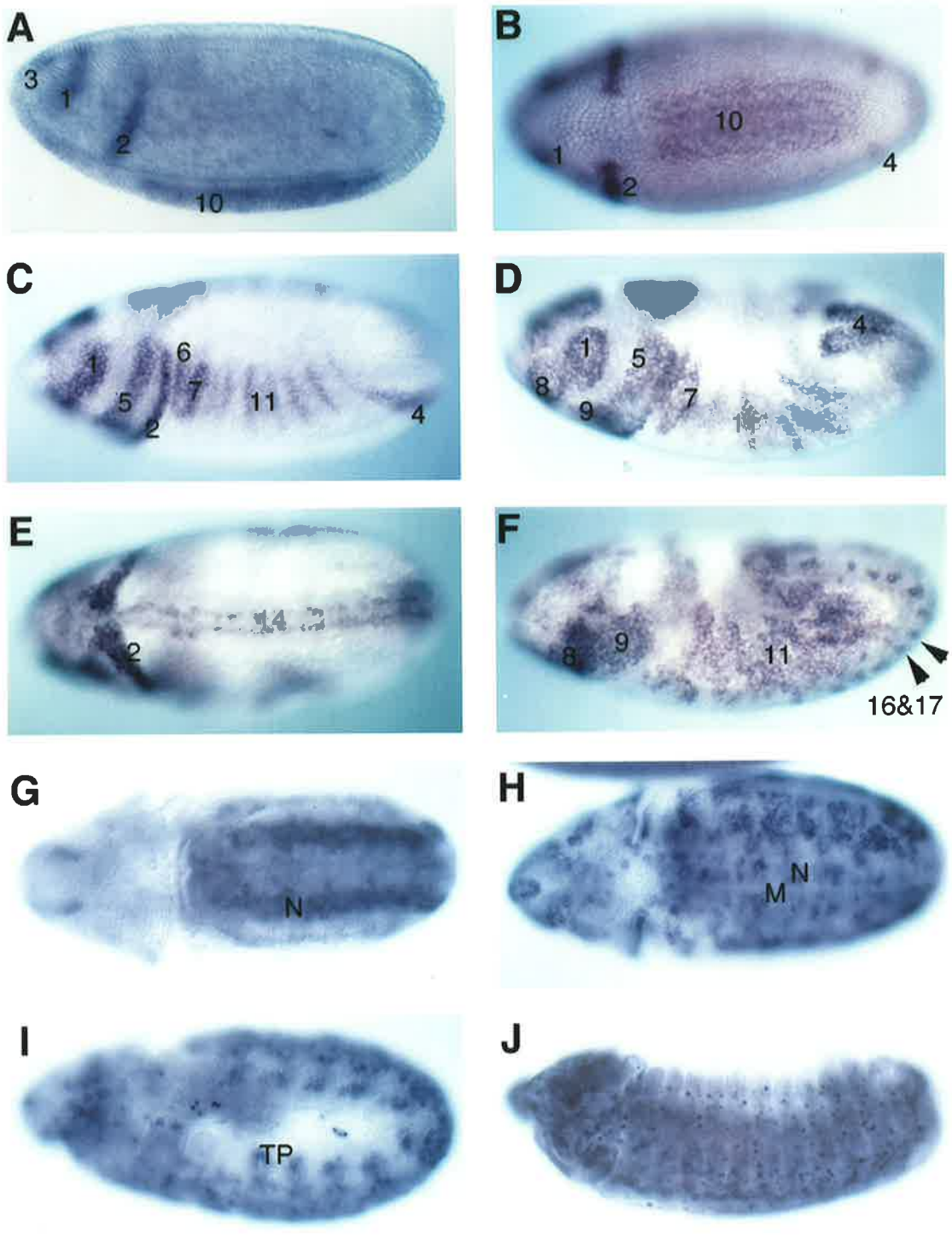
During the early division cycles (1-13) a 3.0kb transcript was also detected, by Northern analysis, suggesting that there may be a maternal specific component to this gene (B. A. Edgar and P. H. O'Farrell 1989).

1-2.3 *string* is a homologue of the mitotic inducer *Cdc25*

Database comparisons using sequence from the *stg* cDNA clone revealed significant sequence similarity to the mitotic inducer *cdc25* from *Schizosaccharomyces pombe*. The homology between *stg* and *cdc25* is sufficient for *stg* to replace *cdc25* in *S.pombe*, suggesting that these proteins perform similar functions in widely diverged organisms (B. A. Edgar and P. H. O'Farrell 1989). Proteins containing structural homology to *cdc25* have also been identified in *S.cerevisiae* (MIH1)(S. T. Bissen 1995), *Dictyostelium* (J. B. A. Millar *et al.,* 1991), *Aspergillus* (M. J. O'Connell *et al.,* 1992) *Arabidopsis* (J. B. A. Millar *et al.,* 1991), *Drosophila* (*twine,* see section 1-2.4(i))(L. Alphey *et al.,* 1992; C. Courtot

Figure 1-4

Expression of the *stg* mRNA as detected by *in situ* hybridization, showing the cycle 14 mitotic domains and some of the later patterns. All views are lateral unless described otherwise. **A**, following cellularisation (stage 6), expression can be seen in ∂_{14}1, 2, 3 (early), and 10. **B**, at a very similar stage, a ventral view, identifies ∂_{14}1,2,4 and 10. **C**, as gastrulation proceeds (stage 7), expression is still present in ∂_{14}1,2, and 4 but now it is also visible in ∂_{14}5,6,7, and 11. **D**, as rapid germband extension commences (stage 8), the same domains described in C are still present but in this view, ∂_{14}8 and 9 can also seen. **E**, a ventral view at a similar stage to that of D, from this view ∂_{14}2 and 14 can be seen. **F**, as rapid germband elongation progresses, expression in the head can be seen in ∂_{14}8 and 9 and in the trunk, ∂_{14}11,16 and 17 are visible. **G**, by the end of rapid germband elongation (stage 8), expression in many of the early domains has disappeared and in this dorsal view ∂_{14} N can be detected along the germband, the expression domain in the head is unidentified although it could ∂_{14}15, or an early cycle 15 domain. **H**, a ventral view during gnathal and clypeolabral lobe formation (stage10), along the ventral midline, a few cells from ∂_{14}N and more in ∂_{14}M are expressing *stg*, in the more lateral regions cycle 15 domains can be detected, these may be the cells of ∂_{15}11. In the head, cycle 15 expression domains are also visible. **I**, later in gnathal and clypeolabral lobe formation (stage 10), expression can be seen in the tracheal placodes (TP) as they are invaginating and also, in the ventral neurogenic region and the brain where expression continues well after most other regions of the embryo have ceased to divide. **J**, divisions continue in the developing nervous system throughout embryonic development, as shown here by the significant amount of *stg* expression still occurring during germband shortening in the tissues of the central and peripheral nervous system (stage 14).


A
3
1
2
10
B
10
1
2
4
C
1
6
5
7
11
2
4
D
1
5
4
8
9
7
E
14
2
F
8
9
11
16&17
G
N
H
M
N
I
TP
J

et al., 1992; J. Jimenez *et al.,* 1990), leech (A. Kakizuka *et al.,* 1992)*Xenopus* (J. B. A. Millar *et al.,* 1991), mouse (*cdc25A, cdc25B, and cdc25C*) (A. Kakizuka *et al.,* 1992; J. B. A. Millar *et al.,* 1991; D. Wickramasinghe *et al.,* 1995; S. Wu and D. J. Wolgemuth 1995), and human (*cdc25A, cdc25B, cdc25C*) (J. B. A. Millar *et al.,* 1991; P. Russell *et al.,* 1989). Several of these have also been shown to functionally complement *cdc25* in *S.pombe,* further confirming it's conserved function as a mitotic inducer (K. Sadhu *et al.,* 1990).

Importantly, mutations in the *cdc25* gene of *S.pombe* result in cells arresting in G2 of the cell cycle, (P. Nurse *et al.,* 1976) further indicating that *stg* and *cdc25* are functionally homologous.

1-2.4 *string* expression precedes mitosis throughout embryogenesis

Whole mount *in situ* hybridisation to visualise *stg* expression at different developmental stages revealed a complex pattern that suggested that *stg* controls mitosis in the *Drosophila* embryo. The *stg* message is expressed throughout the embryo during the first thirteen division cycles and then disappears during cellularisation (B. A. Edgar and P. H. O'Farrell 1989; F. A. Myers *et al.,* 1995). *stg* reappears, almost immediately, this time in a pattern identical to that of the mitotic domains but preceding it by 20 to 25 minutes (B. A. Edgar and P. H. O'Farrell 1989) (figure 1-4).

(i) Maternal *string* transcription

stg expression prior to cellularisation is provided maternally as it is present even in unfertilised eggs (B. A. Edgar and P. H. O'Farrell 1989). The dramatic disappearance of this pool of *stg* message at cellularisation suggests that the stability of the transcript is being specifically altered. Aneuploidy studies covering the X and autosomes 2 and 3 have failed to uncover a zygotic gene that is responsible for the degradation of the maternal *stg* message (F. A. Myers *et al.,* 1995). However, by blocking all zygotic transcription in the early embryo, with α-amanitin, Edgar and Datar (1996) found that embryos failed to arrest at G2 of cycle 14 (B. A. Edgar and S. A. Datar 1996). Instead, they entered an extra syncytial division, suggesting that one or more zygotic gene products are required for the degradation of the maternal *stg* message.

The specific degradation of the *stg* message at the time when the early divisions cease and it's role as a mitotic activator in the cell cycle suggest that *stg* is required for the progression of the early division cycles. Interestingly though, embryos produced from germline clones devoid of wildtype *stg* message appear completely normal (B. A. Edgar and S. A. Datar 1996). However, another functional *cdc25* homologue, *twine*, also exists in *Drosophila*. Although *twine* is required for meiosis in the male and female germline, maternal transcripts are also present in the embryo until cellularisation, when they are abruptly degraded (L. Alphey *et al.,* 1992). *twine* is also dispensable during these early divisions, however, the removal of both *stg* and *twine* maternal products has a severe effect on egg production. This suggests that these two *cdc25* homologues perform an overlapping function during oogenesis and the early embryonic cell cycles (B. A. Edgar and S. A. Datar 1996).

(ii) Zygotic *string* transcription

The spatio-temporal pattern of zygotic *stg* expression precedes and exactly defines each mitotic domain, further implicating it in the control of the mitotic domain pattern. *stg* expression is therefore absent in only the two non-dividing domains (A and B). The patterns are so identical that minor variations in the shape of a domain of *stg* expression, that can be observed between embryos, are also reflected in the shape of the mitotic domain (B. A. Edgar and P. H. O'Farrell 1989). Any

differences that do occur between the shape of a domain of *stg* expression and the corresponding mitotic domain result only from changes in cell shape and position. For example, *stg* expression in $\partial_{14}10$ defines a broad band along the ventral surface of the embryo, but by the time these cells enter mitosis, gastrulation has commenced and they have been internalised by formation of the ventral furrow.

In most domains, the time between the initiation of *stg* expression and mitosis is constant, making the order of appearance of the domains of *stg* expression the same as for the mitotic domains. So as soon as cellularisation is complete, or even a little before, *stg* is expressed in $\partial_{14}1$ and then $\partial_{14}2$, and about 20 minutes later $\partial_{14}1$ enters mitosis, closely followed by $\partial_{14}2$. One exception to this is $\partial_{14}10$. Here *stg* expression initiates at about the same time as that in $\partial_{14}1$ and 2, however, mitosis will occur in $\partial_{14}1$ through to 9 before $\partial_{14}10$ finally enters mitosis. Accumulation of the *stg* message in $\partial_{14}10$ is noticeably slower than in most other domains, suggesting that *stg* transcript must reach a particular threshold before mitosis is initiated. This threshold effect can also be observed within the cells of a domain as *stg* mRNA does not necessarily accumulate evenly and minor fluctuations in abundance are reflected in the timing of mitosis in individual cells (B. A. Edgar and P. H. O'Farrell 1989).

In most cycle 14 domains, once mitosis is complete, *stg* transcripts rapidly disappear. For cycles 15 and 16, *stg* mRNA again accumulates in a pattern that predicts the position and timing of the mitotic domains. The tissues of the peripheral and central nervous system continue to express *stg* transcripts well after cycle 16, in accordance with their extended proliferation program. This pattern is also dynamic, and it has been shown that the first neuroblast division is anticipated by the expression of *stg* (K. Weigmann and C. F. Lehner 1995). This suggests that *stg* is still acting as a rate limiting factor for entry into mitosis, however, the possible addition of a G1 phase in some of these cell cycles may mean that additional factors are also involved in regulating proliferation (B. A. Edgar *et al.*, 1994).

In summary, the striking similarity between the pattern of zygotic *stg* transcription and the subsequent domains of mitosis indicates that *stg* is a key regulator of the mitotic domain pattern.

1-2.5 Ectopic *stg* expression can induce mitosis

It has been shown that uniform *stg* expression, driven by a heat shock inducible promoter, during G2 of cycle 14, advances the timing of mitosis across the embryo (B. A. Edgar and P. H. O'Farrell 1990). Mitosis occurs in a some-what random pattern, with longer pulses of *stg* expression giving a more synchronous result. This suggests that *stg* is the rate limiting factor for progression into mitosis of cycle 14. Interestingly, in these embryos where mitosis of cycle 14 is induced synchronously, the following rounds of mitosis resume their normal patterns of division, suggesting that the timing of each round of mitosis and *stg* transcription is controlled independently.

Ectopic expression of *stg* in cells that have just passed through mitosis revealed that *stg* is not sufficient to immediately induce another round of mitosis and it was shown that nuclei began to incorporate BrdU within 5 minutes of the induced mitosis (B. A. Edgar and P. H. O'Farrell 1990). This suggests that all of the components, necessary for DNA replication are constitutively present and that there may be a replication feedback mechanism that prevents cells from entering mitosis while they are replicating their DNA. Indeed, 45 minutes after mitosis, when DNA replication is complete, cells could again be driven into mitosis (B. A. Edgar and P. H. O'Farrell 1990).

Mitosis in cycles 15 and 16 can also be brought forward by the ectopic expression of *stg*. During these cycles there is no window when all of the cells of the embryo are in G2, however, regions of the embryo that are In G2 when *stg* Is ectopically expressed will be driven into mitosis (B. A. Edgar and P. H. O'Farrell 1990). So, *stg* is sufficient to induce mitosis and the subsequent round of DNA replication in cycles 14, 15 and 16. However, in wild-type embryos, the majority of the cells of the embryo do not replicate their DNA after mitosis of cycle 16, rather they enter their first G1, from where they can no longer be driven into mitosis by expressing *stg*.

1-2.6 Summary

The above evidence provides a strong argument for control of the mitotic domain pattern by regulated *stg* expression. The mutant phenotype of *stg*, arresting cell division, but not differentiation, at G2 of cycle 14 is consistent with *stg* being a mitotic inducer that plays a specific role in cell cycle control (see 1-3.4). That *stg* expression anticipates the pattern of mitosis so precisely suggests that it is regulating the timing of mitosis. Further, the absence of any other zygotic mutants that block cell cycle progression at the same point or earlier suggests that there is no other single zygotic gene, acting earlier than *stg*, that could be fulfilling this role. Finally, that ectopic *stg* expression can drive cells from G2 into mitosis indicates that *stg* is the only rate limiting component for entry into mitosis. Therefore *stg* is acting as a rate limiting component at this stage of development, regulating the timing of mitosis and subsequent cell cycle progression via its complex pattern of expression.

1-3 CELL CYCLE CONTROL THE TRANSITION FROM G2 TO MITOSIS

1-3.1 Introduction

Studies of the molecular basis of the transition from G2 into mitosis, in Schizosaccharomyces *pombe* and *Xenopus laevis,* have identified key molecules that respond to external signals to regulate progression. Homologues of these key molecules have since been identified in all species from *Drosophila* to humans demonstrating the highly conserved nature of cell cycle control.

1-3.2 Cyclins and cyclin dependent kinases

Among the most conserved of the cell cycle regulators are the cyclins and the cyclin dependent kinases (Cdks), which together, control the progression of the cell cycle. At specific stages of the cell cycle, different cyclin molecules are expressed and are necessary for the activity of the cdks with which they associate. Active cdk-cyclin complexes then phosphorylate substrates that perform specific functions within the cell cycle.

Entry into mitosis is known to be regulated by activation of the mitosis promoting factor (MPF), which is composed of a protein kinase subunit, Cdc2, and a positive regulatory subunit, usually Cyclin B (G. Draetta *et al.,* 1989; W. G. Dunphy *et al.,* 1988; J. Gautier *et al.,* 1988) (figure 1-5). Cdc2 kinase is present throughout the cell cycle, so, it is the accumulation of Cyclin B (and Cyclin A) during G2 that controls the level of MPF in a cell (A. W. Murray and M. W. Kirschner 1989; M. J. Solomon *et al.,* 1990). The accumulation of Cyclin B is sufficient to trigger mitosis in some organisms such as *Xenopus,* at some developmental stages (A. W. Murray and M. W. Kirschner 1989; M. J. Solomon *et al.,* 1990), however, it has been shown in *S.pombe* that the over-expression of Cyclin B does not

advance the timing of mitosis (I. Hagan *et al.,* 1988). This suggests that additional factors are expressed in G2, apart from cyclin B, which contribute to mitotic initiation.

Activation of MPF also requires that the Cdc2 kinase be phosphorylated on threonine 161. This is performed by the Cdk-activating kinase (CAK), the catalytic subunit of which is itself a Cdk (Cdk7) (D. Fesquet *et al.,* 1993; R. Y. C. Poon *et al.,* 1993; M. J. Solomon *et al.,* 1993)(figure 1-5). CAK activity requires Cdk7 to associate with Cyclin H. Unlike Cdc2-Cyclin B, however, it appears that CAK activity is not specific to G2 (R. P. Fisher and D. O. Morgan 1994; M. Matsuoka *et al.,* 1994), suggesting that the timing of this phosphorylation of MPF does not play a critical role in the induction of mitosis.

1-3.3 Negative regulators of MPF

Apart from the positive regulators of MPF, molecules that affect the timing of mitosis by inactivating MPF have been identified. This negative regulation appears to be critical to ensure that the cell is ready to enter mitosis and, further, removal of this inhibition can then produce the sudden burst of MPF activity that is characteristic of mitotic induction.

The main inhibitory effect on Cdc2-Cyclin B is the phosphorylation of the Cdc2 on tyrosine residue 15 and, to a lesser extent (in metazoans), threonine residue 14 (using the numbering for *S.pombe*)(figure 1-5). Prior to mitosis, Cdc2 has been found to be highly phosphorylated on these residues which overlap the ATP binding domain thus inactivating MPF (G. Draetta *et al.,* 1988; M. J. Solomon *et al.,* 1990). The dephosphorylation of these residues then results in the activation of MPF (K. L. Gould and P. Nurse 1989; W. Krek and E. A. Nigg 1991; C. Norbury *et al.,* 1991) (and reviewed in (P. Nurse 1990)). The importance of this inhibitory phosphorylation has been demonstrated by specifically mutating these residues, to prevent phosphorylation. In yeast and mammalian tissue culture cells this has been found to advance the timing of mitosis. Even in cells that have been arrested during DNA replication, entry into mitosis can occur if these residues are not phosphorylated (T. Enoch *et al.,* 1991; C. Smythe and J. W. Newport 1992). This inhibitory signal is therefore critical to prevent mitosis from initiating before the completion of S phase.

The Wee1 tyrosine/serine kinase is primarily responsible for the phosphorylation of Cdc2 on tyrosine 15 (C. Featherstone and P. Russell 1991; K. Lundgren *et al.,* 1991; P. Russell and P. Nurse 1987b)(figure 1-5), and a membrane-associated, Wee1 related, kinase Myt1 is thought to phosphorylate threonine 14 (S. Atherton-Fessler *et al.,* 1994; S. Kornbluth *et al.,* 1994; P. A. Mueller *et al.,* 1995).

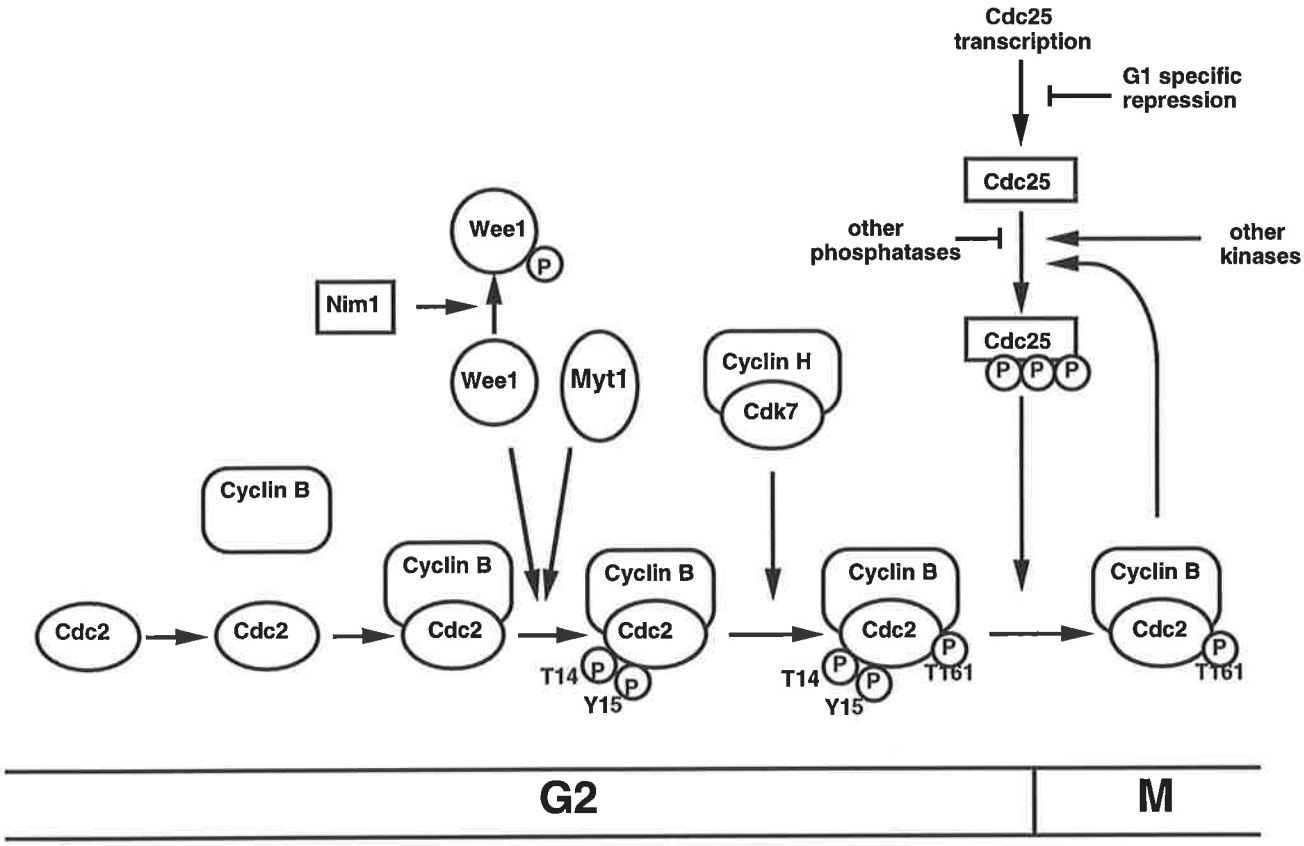
1-3.4 Removal of inhibitory phosphorylation induces mitosis

In *S.pombe*, Wee1 itself has been shown to be phosphorylated by another protein kinase, Nim1. This phosphorylation of Wee1 prevents it from phosphorylating Cdc2 and thus permitting mitotic activation (T. R. Coleman *et al.,* 1993; L. L. Parker *et al.,* 1993; P. Russell and P. Nurse 1987a; L. Wu and P. Russell 1993)(figure 1-5). It is thought that Nim1 may respond to nutrition, establishing a link between cell physiology and the cell cycle (H. Feilotter *et al.,* 1991), however, this role may be specific to *S.pombe* as Nim1 homologues have not yet been identified in higher eukaryotes.

Although Nim1 inactivates Wee1, it is not responsible for the dephosphorylation of cdc2-cyclin B that will trigger mitosis. The identification of the *cdc25* gene from *S. pombe* as a positive regulator of mitosis provided a likely candidate for this role. Mutations in the *cdc25* cause cells to arrest in G2 and

Figure 1-5

A schematic diagram summarising the molecular events involved in the progression towards mitosis. Cdc2 is present throughout the cell cycle and as G2 phase progresses Cyclin B accumulates and associates with Cdc2, forming MPF (or pre MPF). Phosphorylation events then control the activation state of MPF. Wee1 and Myt1 inhibit the activation by phosphorylating tyrosine 15 and threonine 14 of Cdc2, Wee1 can also be negatively regulated by Nim1 which phosphorylates Wee1 to prevent it from inactivating Cdc2. Cdk7-Cyclin H must phosphorylate threonine 161 for MPF activation. Finally, Cdc25 must dephosphorylate threonine 14 and tyrosine 15 to activate MPF and allow entry into mitosis. Cdc25 is present only just prior to entry into mitosis, as *cdc25* transcription is repressed outside of G2. Once translated the Cdc25 protein must still be phosphorylated on several sites to be active, this is regulated by several phosphatases and kinases. One of the activating kinases is active MPF, which then generates a positive feedback loop which can result in a sudden increase in active Cdc25 and MPF.


Cdc25
transcription
G1 specific
repression
Cdc25
other
phosphatases
other
kinases
Cdc25
P
P
P
Wee1
P
Nim1
Wee1
Myt1
Cyclin H
Cdk7
Cyclin B
Cyclin B
Cyclin B
Cyclin B
Cyclin B
Cdc2
Cdc2
Cdc2
Cdc2
Cdc2
Cdc2
T14
P
P
Y15
T14
P
P
Y15
P
T161
P
T161
G2
M

never enter mitosis (P. Fantes 1979; P. Nurse *et al.,* 1976). Further, Russell and Nurse (1986) found that *cdc25* acts as a rate limiting inducer of mitosis, as over expression caused mitosis to initiate at a reduced cell size (P. Russell and P. Nurse 1986). However, it was not until homology was identified between Cdc25 and a serine/tyrosine protein phosphatase (PTPase) from vaccinia virus, VH1, that it was suggested that it may directly dephosphorylate the Cdc2 kinase (S. Moreno and P. Nurse 1991).

In vitro experiments have found that bacterially expressed Cdc25 (from several different species including *Drosophila*) can dephosphorylate bacterially expressed or highly purified Cdc2-CyclinB complexes on the tyrosine and possibly the threonine residue (J. Gautier *et al.,* 1991; U. Strausfeld *et al.,* 1991). Furthermore, a mutant form of Cdc25, where one conserved residue within the homologous catalytic site of the VH1 PTPase is altered, can no longer dephosphorylate these residues. *In vivo*, this mutation has also been shown to abolish the activity of Cdc25 (W. G. Dunphy and A. Kumagai 1991; J. Gautier *et al.,* 1991; U. Strausfeld *et al.,* 1991). These results identify Cdc25 as the phosphatase that is directly responsible for MPF activation (figure 1-5).

Although the phosphorylation of Cdc2-Cyclin B, particularly on tyrosine 15, is critical for mitotic control in organisms such as *S. pombe* and *D. melanogaster*, it is of little importance in at least one organism. The replacement of cdc2 with a mutant form that cannot be phosphorylated on the equivalent tyrosine residue in the budding yeast *S. cerevisiae*, showed no measurable effect on mitosis (A. Amon *et al.,* 1992; P. K. Sorger and A. W. Murray 1992). This suggests that other controls may be involved in the activation of MPF in some organisms.

1-3.5 Cell cycle regulation of Cdc25

If Cdc25 is a rate limiting step in the transition from G2 to mitosis, then it should be active only in late G2 of the cell cycle. Cdc25 has been shown to be regulated transcriptionally as well as by post-translational modifications. In *Xenopus*, Cdc25 is present throughout the cell cycle but is activated by extensive phosphorylation only upon entry into mitosis (T. Izumi *et al.,* 1992; A. Kumagai and W. G. Dunphy 1992). This phosphorylation of Cdc25 appears be negatively regulated by several phosphatases that sense the physiology of the cell and inhibit phosphorylation if the cell is not ready for mitosis (reviewed in (R. W. King *et al.,* 1994)). One factor that phosphorylates Cdc25 is Cdc2-Cyclin B (T. Izumi and J. L. Maller 1993), revealing a positive feedback loop that presumably acts to augment Cdc25 activation and induce a sudden burst of MPF activation (figure 1-5).

In *S.pombe* the levels of both *cdc25* mRNA and Cdc25 protein have been found to increase as cells proceed through interphase, reaching a peak at the beginning of mitosis (S. Moreno *et al.,* 1990), this suggests that the specific transcription of *cdc25* during G2 may regulate Cdc25 accumulation. Regulation of human *cdc25C* also appears to be transcriptionally controlled, with transcriptional activation being repressed specifically in G1 (F. C. Lucibello *et al.,* 1995) (figure 1-5).

In the leech embryo, different cell lineages appear to regulate cdc25 by using different mechanisms. In the micromeres, *cdc25* is specifically expressed in G2 of the cell cycle indicating that it is being transcriptionally regulated, while in the blast cells its expression appears to be constitutive, suggesting that its activity is regulated post translationally, possibly by phosphorylation (S. T. Bissen 1995).

1-3.6 *string* is transcriptionally regulated in *Drosophila*

In *Drosophila*, the tyrosine phosphorylated form of Cdc2 has been found to accumulate during G2 of cycle 14 and then disappear following the zygotic expression of *Stg* (B. A. Edgar *et al.*, 1994), indicating that the phosphorylation of cdc2 is critical for mitotic control.

During cycle 14, *stg* transcription, the accumulation of the Stg protein, and the resulting pattern of mitotic domains in the embryo all follow a very similar spatio-temporal pattern (B. A. Edgar and P. H. O'Farrell 1989; B. A. Edgar *et al.*, 1994). This indicates that the timing of mitosis is regulated by expression of the *stg* transcript and that post-translational events do not play a major role in the timing of Stg activation. However, the accumulation of *stg* transcripts could be regulated either by RNA stability or by *de novo* transcription. Whole mount *in situ* hybridisations revealed that in cells where *stg* was accumulating, two intense dots could be detected in the nucleus, representing nascent transcripts from each copy of the *stg* gene. The absence of these nuclear dots in cells where *stg* transcripts could not be detected implies that *stg* transcription reflects *stg* expression and therefore the complex pattern of *stg* expression is determined transcriptionally (P. H. O'Farrell *et al.*, 1989).

1-4 THE REGULATION OF ZYGOTIC *string* TRANSCRIPTION

1-4.1 Patterning in the *Drosophila* embryo

The newly cellularised embryo shows no outward sign of the complex developmental events that are about to follow. Already though, positional information exists in the *Drosophila* embryo. Even prior to fertilisation, maternal transcripts and their products are specifically localised within the oocyte, giving the embryo its polarity and mediating the spatial specific expression of the pattern-formation genes. The complex pattern of *stg* transcription that begins once cellularisation is complete probably relies on this underlying patterning information to determine the boundaries of its domains.

(i) Dorsal-ventral patterning of the embryo

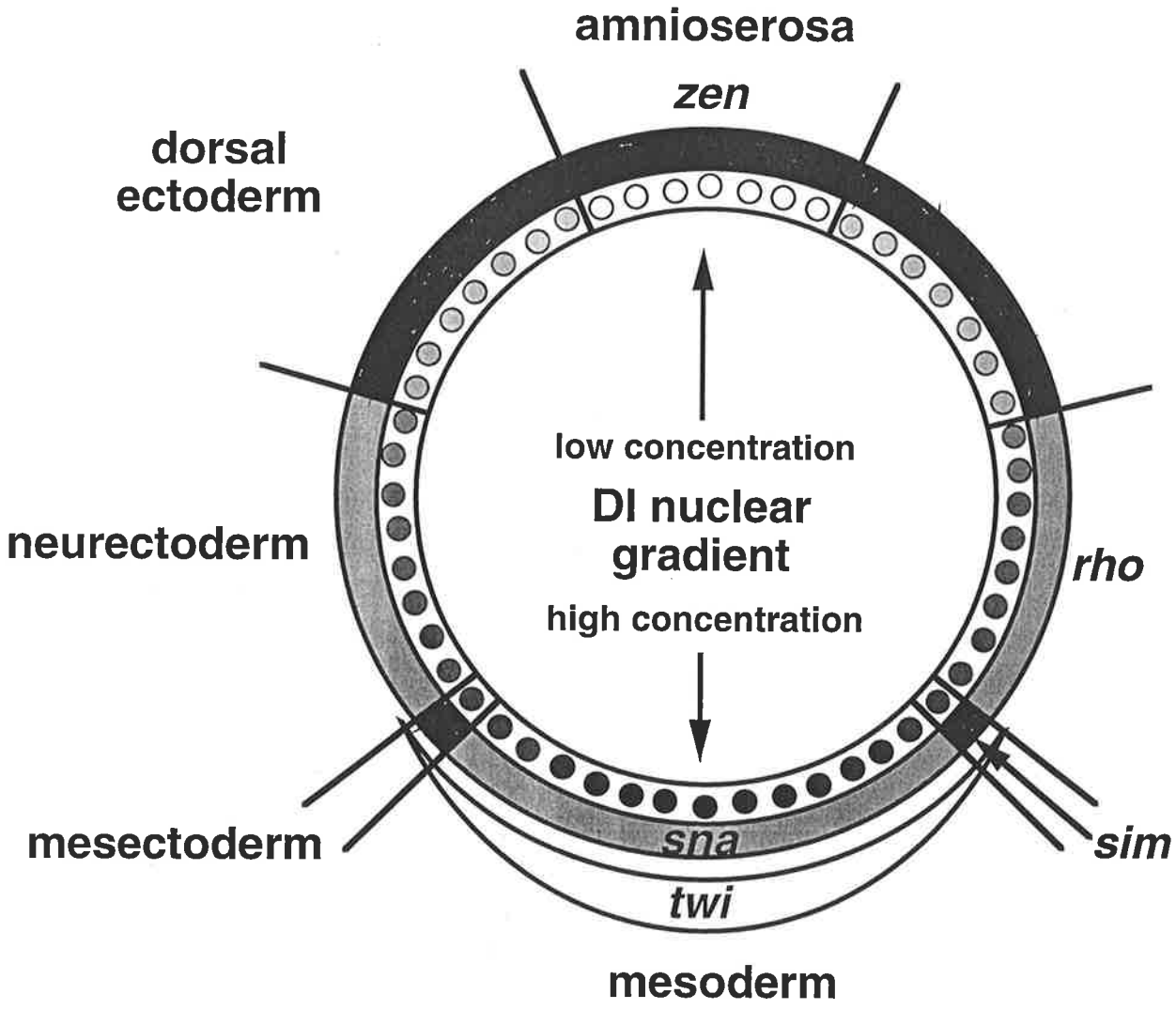
Several primary embryonic tissue types require dorsal-ventral polarity to determine their position and relative size. The mesoderm derives from the ventral-most cells of the embryo and is flanked by a thin strip of mesectodermal tissue. The neurectoderm and the dorsal ectoderm occupy the lateral regions and the amnioserosa derives from the most dorsal cells of the embryo (figure 1-6).

During oogenesis and the early cleavage divisions a series of maternal effect genes are involved in setting up polarity in the dorsoventral system. Upon formation of the syncytial blastoderm, one of these genes, *dorsal (dl)*, relays this information via a gradient of nuclear localisation of its gene product. In ventral regions of the embryo, Dl is selectively transported into the nucleus. More laterally, a smaller proportion of Dl is localised to the nucleus, while on the dorsal surface, Dl protein remains completely in the cytoplasm (figure 1-6) (S. Roth *et al.*, 1989; C. A. Rushlow *et al.*, 1989; R. Steward 1989). *dl* encodes a transcription factor with homology to the avian oncogene *rel* (T. D. Gilmore and H. M. Temin 1986) and the mammalian B cell transcriptional activator NF-kB (P. A. Baeuerle and D. Baltimore 1988). Nuclear localised Dl can thus interact with the zygotic genes that define the tissues along the dorsal-ventral axis (figure 1-6).

In the ventral-most region of the embryo, two genes; the helix-loop-helix transcription factor *twist (twi)*(B. Thisse *et al.*, 1988) and the zinc finger transcription factor *snail (sna)*(J. L. Boulay *et al.*, 1987), are required for the formation of the mesoderm. Both *twi* and *sna* are first expressed during cycle 12 as a narrow stripe along the ventral surface of the embryo. Expression gradually expands in both, and

Figure 1-6

A cross section through a *Drosophila* embryo showing the different tissues that form along the dorsal-ventral axis and some of the genes that are specifically expressed in these tissues. The inner circle represents the blastoderm embryo and shows the nuclear gradient of Dorsal protein. The outer circle schematises the patterns of gene expression; *zen* in the dorsal ectoderm and amnioserosa, *rho* in the neurectoderm (and possibly the mesectoderm), *sim* in the mesectoderm, and *sna* in the mesoderm. The crescent shape underneath the circles represents *twi* expression which is strong in the mesoderm but weaker at the edges and in the mesectoderm.


amnioserosa
zen
dorsal ectoderm
low concentration
Dl nuclear gradient
high concentration
neurectoderm
rho
mesectoderm
sna
sim
twi
mesoderm

by cellularisation *twi* expression forms a band a few cells wider than the presumptive mesoderm and *sna* expression develops sharp boundaries that precisely define the presumptive mesoderm (M. Leptin 1991). *twi* and *sna* are activated only by relatively high concentrations of Dl due to the presence of binding sites in their promoters that respond only to high concentrations of Dl (J. Jiang and M. Levine 1993; C. Thisse *et al.*, 1991). The presence of Twi as well as Dl binding sites in the *sna* promoter appears to be important in refining the sharp boundaries of *sna* expression as the gradient of Dl alone is not sufficient to generate this sharp on-off switch (Y. T. Ip *et al.*, 1992b).

In the ventral neurectoderm, *dl* is involved in activating genes such as *rhomboid (rho)*(also known as *veinlet (ve),* figure 1-6). The *rho* promoter contains Dl binding sites that are closely linked to basic helix-loop-helix (bHLH) DNA binding domains. It appears that co-operative binding between Dl and a ubiquitously expressed bHLH protein is required to achieve activation in these lateral regions where Dl alone is insufficient. To prevent *rho* expression in the ventral region, Sna binding sites are present in the *rho* promoter which overlap the bHLH binding sites and effect the repression of *rho* transcription wherever Sna is expressed (Y. T. Ip *et al.*, 1992a; J. Jiang and M. Levine 1993).

Another gene, *single-minded (sim),* is expressed specifically in the mesectoderm, its expression forming two single cell lines that border the mesoderm (figure 1-6). Both *twi* and *dl* have been implicated in the activation of *sim* expression(C. Rushlow and K. Arora 1990) and, like *rho*, *sim* utilises Sna to repress its transcription in the more ventral tissues (Y. Kasai *et al.*, 1992).

Finally, Dl is also able to regulate gene transcription of dorsally expressed genes such as *zerknült (zen).* Dl binds to the *zen* promoter and represses its transcription in ventral and lateral regions (J. Jiang *et al.*, 1991). These very same Dl sites that cause repression in *zen* are also capable of activating *twi* which suggests that Dl is acting in combination with other transcription factors to repress *zen* (J. Jiang *et al.*, 1992; D. Pan and A. J. Courey 1992).

The products of all of the above mentioned genes (apart from *rho,* which encodes a putative membrane receptor (E. Bier *et al.*, 1990)) are themselves transcription factors. Sna and Twi have already been shown to be involved in the transcriptional regulation (see above), *sim* is a member of the bHLH transcription factor class (J. R. Nambu *et al.*, 1991) and *zen* is a homeodomain containing transcription factor (T. Hoey and M. Levine 1988). Their spatial specific expression in the cellularised embryo provides tissue specific information for the developmental events that follow.

(ii) Anterior-posterior patterning in the trunk of the embryo

In the trunk region of the *Drosophila* embryo, a series of three gnathal, three thoracic and eight abdominal segments must be generated (figure 1-7). Proper formation of structures along this anterior-posterior axis (excluding the first two gnathal segments) are mediated by the maternal factors *bicoid (bcd)* and *hunchback (hb). bcd* encodes a homeodomain transcription factor (W. Driever and C. Nüsslein-Volhard 1988) and *hb*, a zinc finger type transcription factor (D. Tautz *et al.*, 1987).

During oogenesis *bcd* messenger RNA localises to the anterior pole of the oocyte. Following fertilisation, translation of this message during the syncytial divisions, produces a gradient of Bcd protein activity as it diffuses towards the posterior of the embryo (W. Driever and C. Nüsslein-Volhard 1988). In contrast, the maternal *hb* mRNA is initially present throughout the embryo but translation of this message is blocked in the posterior half of the embryo by the *nanos* gene product such that Hb is only found in the anterior half of the embryo (E. R. Gavis and R. Lehmann 1992; D. Tautz and C. Pfeifle 1989) (see figure 1-7).

The gap genes, which are also transcription factors, respond to different levels of Bcd and Hb activity and are transcriptionally activated in broad stripes along the syncytial blastoderm (figure 1-7). *hb* itself is zygotically activated by the binding of Bcd and maternal Hb to its promoter (W. Driever *et al.,* 1989; M. Simpson-Brose *et al.,* 1994), to produce a broad band of expression in the anterior of the embryo. The gap gene *Krüppel (Kr)* responds to Bcd protein in the absence of Hb and is therefore activated just posterior to the Hb expressing region. The *Kr* promoter was found to contain high affinity Bcd binding sites which can activate *Kr* even at extremely low Bcd concentrations. (M. Hoch *et al.,* 1991). Other members of the gap genes; *Knirps (Kni), giant (Gt)* and *tailless (tll)* are also regulated by analogous mechanisms to define broad regions along the embryo during the syncytial stage (reviewed in (M. Pankratz and H. Jäckle 1993) (see figure 1-7).

Although transcription of the gap genes generates discrete bands of expression, their transcription factor products appear to diffuse in the syncytium to form localised but overlapping gradients. These act in combinations to activate transcription of the pair-rule genes in seven evenly spaced stripes, defining the segmental pattern of the embryo.

The pair-rule gene *hairy (h)* requires greater than 14kb of upstream promoter DNA to effect its correct pattern of seven stripes (G. Riddihough and D. Ish-Horowicz 1991). Its pattern appears discretely over time with stripe 1 being the first to appear, followed by stripes 3, 2, 4, 7, 5 and lastly 6 (figure 1-7). Transcriptional activation in the separate *h* stripes can be achieved by placing different promoter fragments, from within the 14kb, upstream of a basal promoter/*lacZ* reporter gene construct and transforming *Drosophila.* This shows that there are discrete regulatory units for specific stripes (M. J. Pankratz *et al.,* 1990). However, it appears that not all of the stripes can be separated without affecting their normal expression pattern, suggesting that regulatory sequences for different stripes are in some cases very close to, or even overlapping each other (G. Riddihough and D. Ish-Horowicz 1991).

Another pair-rule gene, *even skipped (eve),* is also expressed in seven stripes along the embryo, however, its stripes are staggered relative to those of *h* and so it requires different combinations of gap genes for its regulation (figure 1-7).Study of the *eve* stripe 2 regulatory region has identified binding sites for Bcd, Hb, Gt and Kr within a 700bp fragment. In regions of high Gt or Kr concentration, activation by Bcd and Hb is repressed so that *eve* is only expressed in the narrow region between Gt and Kr (S. Small *et al.,* 1991). The gradients of Kr and Kni are critical regulators of the expression of *h* in stripe 6. Kr binds strongly to the stripe 6 regulatory region so that even low levels of Kr can repress *h*, Kni binds weakly to the same region of DNA and activates *h* but only where the Kni concentration is high (M. J. Pankratz *et al.,* 1990) (see figure 1-7).

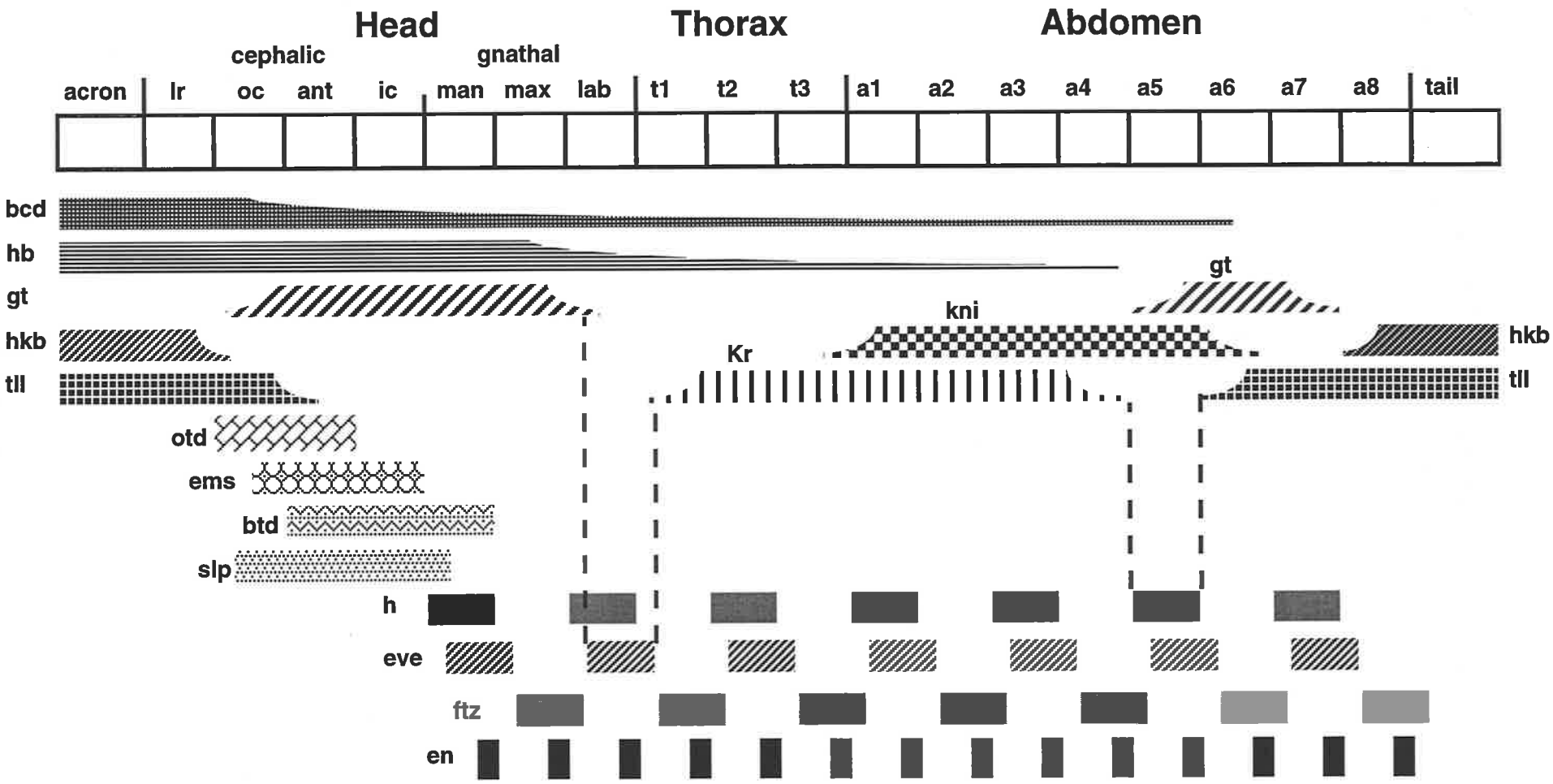
The protein products of the pair-rule genes, which are again transcription factors, define discrete bands along the embryo even though they are expressed prior to cellularisation (B. A. Edgar *et al.,* 1987).

At the same time as the pair rule genes are being activated, homeotic gene expression is initiated in an overlapping pattern such that parasegments of the developing embryo express particular sets of homeotic genes. These homeotic genes are regulated by the maternal and gap genes and all encode homeodomain type transcription factors (reviewed in (P. W. Ingham 1988)).

Following cellularisation, segment polarity genes, including *engrailed (en)* and *wingless (wg),* are expressed in fourteen narrow stripes that define different regions within each segment. The En

Figure 1-7

A summary diagram showing some of the patterning genes that are expressed along the anterior-posterior axis. The boxes at the top represent the segments of the embryo. The non segmented region at the anterior terminus is the acron, the head segments are: labral (lr), ocular (oc), antennal (ant), intercalary (ic), mandibular (man), maxilary (max), and labial (lab). There are three thoracic (t1 to t3) and then 8 abdominal (a1 to a8) segments and the most posterior non-segmented region is the tail. The maternally expressed genes *bcd* and *hb* form gradients that together define the boundaries of expression for many of the zygotic gap genes such as: *hunchback (hb), giant,(gt) Krüppel (Kr)*, and *knirps (kni)*. The terminal maternal coordinate system (not shown) is responsible for the expression of the *huckebein* and *tailless* gap genes. Both of these systems are required for the expression of the head specific gap genes: *orthodenticle (otd), empty spiracles (ems), buttonhead (btd),* and *sloppy-paired (slp)*. The primary pair-rule genes such as: *hairy (h)* and *even skipped (eve)* respond to the overlapping gradients produced by the gap genes to generate stripes that identify the segments, with each being out of register with the other. Stripe 2 of *eve* requires activation by *bcd* and *hb*, anterior repression by *gt* and posterior repression by *Kr*. Stripe 6 of *h* requires activation by *kni*, anterior repression by *Kr* and posterior repression by *tll* and the absence of *kni*. The pair rule gene *fushi tarazu (ftz)* is activated directly by the primary pair rule genes. The segment polarity genes such as *engrailed (en)* are then expressed in particular segmental regions. The homeotic genes (not shown) generate the different segmental identities and are activated directly by the gap genes. The boundaries of expression shown are approximations only.


Head
Thorax
Abdomen
cephalic
gnathal
acron
lr
oc
ant
ic
man
max
lab
t1
t2
t3
a1
a2
a3
a4
a5
a6
a7
a8
tail
bcd
hb
gt
hkb
tll
otd
ems
btd
slp
h
eve
ftz
en
kni
Kr

protein encodes a homeodomain DNA binding protein and is expressed in stripes one cell wide which demarcate the anterior limit of each parasegment (reviewed in (P. W. Ingham 1988)).

By the time the cellular blastoderm is formed, many genes that encode transcription factors have spatially restricted patterns of expression in the trunk region of the anterior to posterior axis. These genes then direct target genes that bring about development.

(iii) Patterning in the head of the embryo

The *Drosophila* head is composed of seven distinct segments, four cephalic (labral, ocular, antennal, and intercalary) and three gnathal segments (mandibular, maxillary, and labial, see figure 1-7). In *bcd* mutants there is a massive deletion of these segments, however the anterior gap gene *hb* has little effect on the formation of the cephalic or the most anterior of the gnathal segments (mandibular) (W. Driever *et al.*, 1989). This suggests that other Bcd activated genes, expressed anterior to *hb*, are involved in the formation of these head structures. Five genes that appear to perform such functions are; *crocodile (croc)* (U. Häcker *et al.*, 1995), *orthodenticle (otd), empty spiracles (ems), buttonhead (btd)* (S. M. Cohen and G. Jürgens 1990), and *sloppy paired (slp)* (U. Grossniklaus *et al.*, 1994) (see figure 1-7).

These genes all encode transcription factors: *croc* and *slp* products belong to a family of transcription factors that contain a fork head DNA binding domain (U. Grossniklaus *et al.*, 1992; U. Häcker *et al.*, 1995), *otd* and *ems* are homeobox containing genes (D. Dalton *et al.*, 1989; R. Finkelstein *et al.*, 1990; U. Waldorf and W. J. Gehring 1992) , and *btd* encodes a zinc finger transcription factor that is related to human Sp1 (E. Wimmer *et al.*, 1993). These genes appear to act in a manner analogous to the gap genes of the trunk region, as mutations in any of these, result in the deletion of segmental regions corresponding to the region of gene expression (E. R. Gavis and R. Lehmann 1992). The patterns of expression of these genes, however, do not form adjacent stripes, rather they are broad and overlapping, with different anterior and posterior boundaries such that each segment or segmental region expresses different combinations of gene products (see figure 1-7 and (U. Grossniklaus *et al.*, 1994)). This information has been proposed to be sufficient to determine segmental identity and polarity, given that pair-rule and homeotic genes do not play a major role in the establishment of the five most anterior head segments (reviewed in (S. M. Cohen and G. Jürgens 1990; G. Jürgens and V. Hartenstein 1993)).

The specific expression of each of the head specific gap genes depends on Bcd activity (D. Dalton *et al.*, 1989; Q. Gao *et al.*, 1996; U. Grossniklaus *et al.*, 1994; U. Häcker *et al.*, 1995; E. Wimmer *et al.*, 1995) However, in the most terminal regions of the embryo the gap genes *huckebein (hkb)* and *tailless (tll)*, which are activated by the terminal maternal coordinate system, are also required (Q. Gao *et al.*, 1996; U. Grossniklaus *et al.*, 1994; U. Häcker *et al.*, 1995; E. Wimmer *et al.*, 1995) (see figure 1-7). Further to this, correct expression also requires information from the dorsoventral coordinate system (S. M. Cohen and G. Jürgens 1990; Q. Gao *et al.*, 1996; U. Grossniklaus *et al.*, 1994; E. Wimmer *et al.*, 1995).

1-4.2 The patterning genes provide the spatial information for *string* transcription

The combination of the patterning information of the anterior-posterior and dorsal-ventral axes subdivides the cellular blastoderm into many distinct regions. This spatial specific expression provides

the embryo with a network of transcription factors which direct morphogenesis. Mutations in the patterning genes have been observed to alter the pattern of *stg* transcription, showing that they play a critical role in determining the pattern of mitosis in the embryo (B. A. Edgar *et al.*, 1994).

The expression patterns of some patterning genes correlate well with specific mitotic domains and mutations in these genes can alter a specific mitotic domain. For example, in *sna* mutants, $\partial_{14}10$ *stg* expression and mitosis are absent (K. Arora and C. Nüsslein-Volhard 1992; B. A. Edgar *et al.*, 1994). In these embryos the region of $\partial_{14}10$ (which defines the presumptive mesoderm) divides as if it were part of $\partial_{14}14$ (the presumptive mesectoderm). Presumably this is a result of the mis-expression of mesectodermal specific genes that are normally repressed by *sna* (see 1-4.1(i)). In *twi* mutant embryos $\partial_{14}10$ is also disrupted but this time the mitotic domain is not completely mis-specified and the tissue corresponding to $\partial_{14}10$ does not express *stg* or divide (K. Arora and C. Nüsslein-Volhard 1992; B. A. Edgar *et al.*, 1994). In another example, *btd* mutant embryos fail to express *stg* in $\partial_{14}2$, which correlates with the region of *btd* expression (B. A. Edgar *et al.*, 1994).

Many of the early patterning genes, such as *bcd* and *Kr*, have been shown to cause gross alterations in the pattern of *stg* expression (and the mitotic domains) (B. A. Edgar *et al.*, 1994; V. E. Foe and G. M. Odell 1989). Some of the effects of these mutations are likely to be indirect given that the expression of the patterning genes that these genes regulate is also perturbed and that their regions of expression are too large to define any mitotic domains. However it is possible that the gap genes work in combination to control *stg* expression, as they do to direct pair-rule gene expression (see 1-4.1(ii)). For example, $\partial_{14}11$ may be regulated in this manner, as it has a pair-rule type periodicity to its expression that is altered by gap gene mutations but not by pair-rule mutations (B. A. Edgar *et al.*, 1994). The expression of the pair-rule and segment polarity genes appears not to be required to define any of the early mitotic domains, although they are required for the later domains of cycle 14 and those of cycle 15 do require. This seems to reflect the more detailed nature of the domains às development of the embryo proceeds.

1-4.3 *string* transcription is complex

Analysis of *stg* transcription has revealed a very complex promoter region. Genomic fragments that cover the *stg* transcript and large regions of the surrounding DNA have been transformed into *Drosophila* to try and rescue the mutant *stg* phenotype. However, fragments that extend either 30kb upstream of the transcript or 20 kb downstream have rescued only some domains of *stg* expression (B. A. Edgar *et al.*, 1994)(Edgar & Lehman, personal communication). Complementary to this, a small deletion within the promoter of the *stg* transcript removes *stg* expression in only a couple of domains, whereas a larger undefined deletion removes *stg* expression in nearly all domains (B. A. Edgar *et al.*, 1994). The combined information from these experiments reveals that the *stg* promoter is large and that its regulation involves position specific elements within the promoter since different genomic fragments are responsible for different patterns of *stg* transcription.

1-4.4 Does *string* integrate the patterning information?

Section 1-4.3 (above) indicates that *stg* is integrating spatial information to achieve spatial-specific *stg* transcription. Although the patterning genes are required for this to occur, it is not yet clear whether they provide this information directly or whether intermediate genes are involved.

Intermediate genes could, in theory, be regulated by the patterning genes, with each mitotic domain being specified by a single intermediate or master gene. However, the lack of early mutants that alter only single mitotic domains, or even a few mitotic domains suggests that these genes do not exist, although, if each domain had its own master gene, such mutants may have been difficult to identify.

In some cases potential master genes have been identified, whose expression coincides with a particular domain. However, none so far have been found to effect *stg* expression in the domain in which they are expressed. For example, the patterning gene *sim*, which precisely defines $\partial_{14}14$ (the mesectoderm), does not affect *stg* transcription in $\partial_{14}14$ (B. A. Edgar *et al.*, 1994), and *collier (col)* whose expression co-coincides with $\partial_{14}2$ also does not appear to affect *stg* transcription in $\partial_{14}2$ (M. Crozatier *et al.*, 1996). However, it is possible that there are other master genes that do affect *stg* transcription.

It seems likely that different domains have arisen in different ways, some using master genes, others using the patterning genes directly. If the timing of mitosis is critical then master genes may provide a means of delaying the timing of *stg* transcription in a particular domain.

1-4.5 Developmental control of the cell cycle

The patterning genes, or master regulator genes under the control of the patterning genes, play an essential role in the development of the *Drosophila* embryo. That these genes interact with the promoter of a conserved component of the cell cycle to effect the timing of mitosis provides an example of the cell cycle responding to developmental rather than cell cycle signals.

Another *Drosophila* cell cycle gene, *cyclin E*, also appears to be developmentally regulated during embryogenesis (J. A. Knoblich *et al.*, 1994; H. E. Richardson *et al.*, 1993). *cyclin E* is required for the transition from G1 into S phase, and its down regulation in the ectodermal cells following cycle 16 results in their G1 arrest, preventing further division. In the developing neural tissues however, *cyclin E* expression continues during this extended period of proliferation. The fact that *cyclin E* expression is virtually unaffected in embryos where the cell cycle has been blocked (as it is in *stg* mutants) suggests that its expression, at this stage, is using developmental rather than the usual cell cycle regulation (J. A. Knoblich *et al.*, 1994).

Given that regulated cell proliferation is crucial for development it seems logical that developmental cues should directly affect cell cycle components, especially those that are rate limiting for entry into S phase or mitosis.

1-5 THE SIGNIFICANCE OF MITOTIC DOMAINS

1-5.1 The roles of mitosis in morphogenesis

The most obvious role for mitosis during embryogenesis is its proliferative function of producing more cells. The control of this process, in addition to selective cell death, is essential for tissues to develop a particular morphology.

Different morphologies can be achieved by altering the polarity of cell division to result in the formation of distinct groups of cells. For example, the cells of mitotic domains 8 and 9 divide along axes that are perpendicular to the embryonic surface such that half of the resulting cells remain on the embryonic surface and the other half form an internal layer (V. E. Foe 1989). Many of the other domains divide along axes parallel to the embryonic surface which presumably increases the area of

the epithelial layer. Although there are distinct cell shape changes associated with the process of invagination, that occur as part of gastrulation (M. Costa *et al.,* 1994; M. Leptin and B. Grunwald 1990), it seems likely that changing cell densities across the embryo, as different regions divide may also play some role in these folding events. The ventral furrow, for example, invaginates just prior to mitosis in this region ($\partial_{14}10$) (V. E. Foe 1989) and it seems feasible that mitosis at this stage may help to push the newly invaginated cells further into the interior of the embryo. Asymmetric cell division, where the resulting daughter cells are of different sizes will also affect the morphology of a proliferating tissue, particularly when it is associated with polarised cell division. Given that the different morphological consequences that can result from mitosis appear to play a significant role in the development of tissues, it seems likely that the timing and relative order of these mitotic events may also be important. Each tissue must follow its own morphological pathway such that one does not interfere with or affect another.

A developmental event that is linked to mitosis is delamination. One example of this occurs in the ventral neurectoderm (VN) where individual cells delaminate from the epithelium and move into the interior of the embryo to form the neural precursor cells (neuroblasts) of the larval central nervous system (V. Hartenstein and J. A. Campos-Ortega 1984). Hartenstein *et al.*, (1994) noted that the delamination of these neuroblasts occurs just prior to division, while the cells surrounding a particular neuroblast divide immediately prior to the delamination event. It was observed that the cytoskeletal changes that occur during delamination and mitosis are very similar and "rounding up" of cells occurs before each of these events. However mutually excusive differences between delaminating and dividing cells then become apparent. Hartenstein *et al.*, (1994) speculate that the rounded up state reflects a "mitosis/delamination ready state" and that possibly the "molecular machinery" underlying mitosis is also employed for delamination. Given then that mitosis and delamination appear to be linked it seems likely that this may place some constraints upon the timing of mitosis particularly in the ventral neurogenic region.

That the mitotic domains appear to be coordinated with the events of gastrulation has lead to the suggestion that this may be necessary since they require different cytoskeletal organisations (V. E. Foe *et al.,* 1993; V. Hartenstein *et al.,* 1994). In which case it may be most important to ensure that mitosis does not happen when a particular morphogenetic event is occurring, such as delamination. However it may also be beneficial to undergo mitosis soon before or after such an event given that there are some similarities in the early organisation of the cytoskeleton for both mitosis and delamination.

1-5.2 The roles of mitosis in differentiation

Asymmetric cell divisions can affect development by producing daughter cells with different cell fates. In the polarised cell divisions of mitotic domain 9, two gene products that are involved in determining cell fate, *numb* and *prospero,* have been shown to segregate specifically into the basal daughter cell (R. Kraut *et al.,* 1996). Kraut *et al.*, (1996) showed that to do this they require the Inscuteable protein which becomes apically localised prior to mitosis and is responsible for the perpendicular axis of division in these cells. This separation of differentiative signals into one of the two layers of cells that result from this perpendicular mitosis allows them to assume a different fate from their sibling cells that are left on the surface of the embryo.

Similarly, cell fate in the developing nervous system has been shown to require cell cycle progression (K. Weigmann and C. F. Lehner 1995). The fate of particular cells in the CNS requires the expression of *eve*, however, in a *stg* mutant no neuroblast divisions occur and *eve* expression is absent. Interestingly, by providing *stg* ectopically in a *stg* mutant background, at the time of normal neuroblast divisions, mitosis occurs and *eve* is also expressed. The fact that progression through S phase is also required for *eve* expression but not cytokinesis shows that it is cell cycle progression rather than just *stg* expression that allows *eve* to be expressed. However, it appears that the timing of this *stg* regulated cell division is also critical for *eve* expression.

Another role for mitosis in differentiation is associated with the length of time between subsequent mitotic cycles. Shermoen and O'Farrell (1991) showed that the transcription of genes is aborted between mitotic cycles and that the length of the transcription unit and of the cell cycle determines whether a transcript will be completed between subsequent rounds of mitosis (P. H. O'Farrell 1992; A. W. Shermoen and P. H. O'Farrell 1991). The *kni* cognate gene, *knirps-related (knrl)*, differs from *kni* mainly in the length of its primary transcript, which results the presence of a 19.1kb rather than 0.9kb intron. This prevents *knrl* from functioning during abdominal segmentation (M. Rothe *et al.,* 1992). So, the variation in mitotic cycle length following cycle 13 may allow different genes to be expressed only in domains where the cell cycle length is sufficient for the completion of transcription.

1-5.3 The requirement for mitotic domains

The above sections provide some evidence for the requirement for mitotic domains. However, there is also evidence to suggest that the mitotic domains are to some degree dispensable. Edgar and O'Farrell (1990) produced a synchronous cycle 14 mitosis in wildtype embryos by heat shock induction of *stg* and found that 40% of embryos still hatched and 62% of these survived to adulthood. Even in the embryos that did not hatch there was no observable effect on the pattern of the cuticle, suggesting that to a large degree development was unaffected. However, repeated heatshock inductions did have more severe effects with less embryos surviving (B. A. Edgar and P. H. O'Farrell 1990). These results suggest that, either the effects of the mitotic domains on development are subtle or that compensatory mechanisms exist. This is supported by the phenotype of *stg* mutants which show that development proceeds to a remarkable degree even in the absence of cell division (B. A. Edgar and P. H. O'Farrell 1989; A. P. Gould *et al.,* 1990; P. Hartenstein and J. W. Posakony 1990), suggesting that mitosis and morphogenesis are independent processes that occur in parallel.

The presence of genes that are expressed in a domain specific manner such as *sim* and *col* that are not involved in the regulation of *stg* (see 1-4.4) further indicates that differentiation and mitosis may be occurring in parallel. Some level of interplay between these processes must occur though and one example has been observed in mitotic domain 14 (the mesectoderm). Here *sim* has been found not to be required for *stg* expression in $\partial_{14}14$ (B. A. Edgar *et al.,* 1994) however, in *sim* mutants $\partial_{14}14$ does not enter mitosis (J. R. Nambu *et al.,* 1991). This suggests that if the differentiation of the mesectoderm is arrested cell proliferation is also affected, this could be controlled by preventing the phosphorylation of the Stg protein, for example. The tight coupling of the mitotic domain pattern and the morphogenetic movements in the embryo also indicates that each is dependent on the other. Between individual embryos there can be significant variation in developmental timing, although this

always affects both morphogenesis and mitosis such that their relationship to each other is strictly maintained (V. E. Foe and G. M. Odell 1989).

That the mitotic domains have only subtle effects on development does not reduce their significance in terms of evolution. Any alteration that increases the survival of the species will be maintained into the next generation even if it is subtle. It may not be that mitosis has to occur at a specific time in each domain, however, if there are some specific times when it would be detrimental to the survival of the embryo then these must be avoided. It is possible that the complexity of *stg* transcription has arisen to serve this function.

Mitotic domains have been identified in other species suggesting that the evolution of developmental regulation of the cell cycle is not unique to *Drosophila*. Embryos of the *Calliphora vomitoria* blowfly have a very similar pattern of mitotic domains to that of *Drosophila* even though they two diverged about 60 million years ago (V. E. Foe and G. M. Odell 1989). In zebrafish there are 3 mitotic domains that arise at the midblastula transition however, two of these form extra embryonic tissues leaving only one to generate all of the embryonic lineages. Within this embryonic domain there are no further subdivisions which suggests that cell cycle length is not involved in determining cell fate as in *Drosophila*. However, the formation of the 3 mitotic domains occurs just prior to the first morphogenetic movement in the embryo (epiboly) further suggesting a link between mitosis and morphogenesis (D. A. Kane *et al.*, 1992). In leech embryos, particular cell lineages have distinctly longer G2 phases relative to the rest of the embryo, and it has been suggested that this co-ordinates their divisions with other aspects of embryogenesis (S. T. Bissen 1995; S. T. Bissen and D. A. Weisblat 1989).

To what extent these different mitotic cycles are regulated by *cdc25* homologues is unclear. A *cdc25* homologue appears to be responsible for some of the cell-specific timings of cell divisions in leech (S. T. Bissen 1995)(see 1-3.5). In mouse the expression patterns of *cdc25* homologues have revealed complex patterns suggesting that they may be developmentally regulated (A. Kakizuka *et al.*, 1992; D. Wickramasinghe *et al.*, 1995; S. Wu and D. J. Wolgemuth 1995). At least one of these homologues, *cdc25B*, has also been found to be specifically expressed in G2 in tissue culture cells (A. Kakizuka *et al.*, 1992), as has the human *cdc25C* (F. C. Lucibello *et al.*, 1995). Whether any of these homologues are regulated developmentally rather than just by cell cycle factors remains to be seen.

The most likely role for the mitotic domains in development appears to be to assist particular morphogenetic events in subtle ways. In *Drosophila*, and possibly other insects, this seems to have been utilized extensively. Perhaps a higher degree of coordination is required here because of the more rapid nature of embryonic development.

1-6 THIS STUDY

This thesis describes a study of the transcriptional regulation of *stg*. The possibility that *stg* may be integrating information from the underlying network of patterning genes suggested that its regulatory region may contain many separable enhancer regions. To test this hypothesis, promoter analysis of *stg* was initiated. This involved taking *stg* promoter fragments and inserting them upstream of a *hsp70/lacZ* reporter gene construct. These constructs were then transformed into *Drosophila*, using

P element mediated transformation, and the resulting embryos were tested for patterns of *lacZ* expression.

Chapter 3 details the initial constructs generated which contained fragments spanning the 30kb upstream and 20 kb downstream of the *stg* transcript. No enhancer regions were identified in this search that reflected real patterns of *stg* transcription.

In chapter 4, proximal fragments from the *stg* promoter were included with various upstream fragments to test the hypothesis that basal *stg* sequences are required to achieve correct transcriptional activation. The inclusion of a 500bp proximal fragment allowed the identification of enhancers for several cycle 14 domains, within the first 17kb upstream of the *stg* transcript. However, the location of enhancer sequences for many more domains remain unidentified. For some domains it is known that the enhancers lie within the regions already tested, although for others it appears that they are located further upstream.

In chapter 5, an attempt to determine whether the patterning genes are directly regulating *stg* transcription was undertaken by performing a detailed analysis of the enhancer region for one of the identified domains, $\partial_{14}2$. This revealed that the two early patterning genes, *btd* and *sna*, are likely to be directly responsible for the transcriptional activation of *stg* in this domain.

Chapter 6 discusses the implications of the results obtained in the previous chapters and suggests further experiments.

CHAPTER 2: MATERIALS AND METHODS

2-1 MATERIALS

2-1.1 Chemical reagents

All reagents were of analytical grade, or the highest grade obtainable.

2-1.2 Enzymes

Restriction endonucleases:	Boehringer Mannheim, New England Biolabs, and Pharmacia
Alkaline calf intestinal phosphatase and proteinase K:	Boehringer Mannheim
T4 DNA ligase and T4 DNA polymerase:	Promega
Klenow and T4 Polynucleotide Kinase	Bresatec
DNase, RNase and lysozyme	Sigma

2-1.3 Radio-labelled compounds

α-^{32}P-dATP (3000Ci/mmole):	Bresatec
α-^{35}S-dATP (1500Ci/mmole)	
γ-^{32}P-dATP (4000Ci/mmole)	
α-^{32}P-dGTP (3000Ci/mmole)	

2-1.4 *E. coli* strains

DH5α: F'/*endA1 hsdR17* ($r_K^- m_K^+$) *supE44 thi- 1 recA1 gyrA* (Nal^r) *relA1 Δ(lacIZYA-argF)U169 deoR* (ø80 *dlacΔ(lacZ)M15*) (S. F. Ausubel *et al.*, 1994).

XL1blue: F' ::Tn*10 pro* A^+B^+ $lacI^q$ *Δ(lacZ)M15/recA1 endA1 gryA96* (Nal^r) *thi hsdR17* ($r_K^- m_K^+$)*supE44 relA1 lac* (W. O. Bullock and e. al. 1987).

JM110: F' *traD36* $lacI^q$ *Δ(lacZ)M15 proA*$^+$*B*$^+$*/rpsL* (Str^r) *thr leu thi lacY galK galT ara fhuA dam dcm supE44 Δ(lac-proAB)* (C. Yanisch-Perron *et al.*, 1985).

2-1.5 *Drosophila* strains

Unless otherwise indicated, strains are as described (D. L. Lindsley and G. G. Zimm 1992) and obtained from the Indiana Stock Centre, Bloomington, IA.

Canton-S

w^{1118}

twi^1

sna^{18}

btd^{XG}

ems^{7D}

sim/lacZ (J. R. Nambu *et al.*, 1990)

w;Adh/CyO;Δ2-3Sb/TM6b

2-1.6 Media and buffers

(i) Media

All buffers and media were prepared with distilled and deionised water and sterilised by autoclaving, except heat labile reagents, which were filter sterilised.

All bacterial strains were propagated in LB-broth or on LB-agar plates.

L-broth:	1% (w/v) amine A, 0.5% yeast extract, 1% NaCl, pH7.0.
SOC:	2% tryptone, 0.5% yeast extract, 100mM NaCl, 25mM KCl, 100mM $MgCl_2$,100mM $MgSO_4$, 0.2% glucose.
Plates:	liquid broth with 1.5% bacto-agar.

Where required for selection ampicillin was added to a final concentration of 100 µg/ml.

Drosophila culture media:	10% treacle, 20% yeast, 1% agar, 10% polenta, 2.5% tegosept ,1.5% propionic acid

(ii) Buffers

Commonly used buffers were:

Protein gel running buffer :	1.5% tris base, 7.2% glycine, 0.5% SDS.
Protein 'sample' buffer:	62.5 mM Tris-HCl pH 6.8, 10% glycerol, 2% SDS, 5% 2-ß-mercaptoethanol, 0.00125% bromophenol blue.
PBS:	7.5 mM Na_2HPO_4, 2.5 mM NaH2PO4, 145 mM NaCl.
PBST:	PBS + 0.1% Tween 20
TBE:	50 mM Tris-borate pH 8.3, 1 mM EDTA.
TE:	10 mM Tris-HCl pH 7.4, 0.1 mM EDTA.
TAE:	40 mM Tris-acetate pH 8.2, 1 mM EDTA.
10 x agarose gel load buffer:	80% glycerol, 50 mM EDTA, 0.1% bromophenol blue
Sequencing gel load buffer:	98% deionised formamide, 10 mM EDTA pH 8.0, 0.025% xylene cyanol, 0.025% bromophenol blue.

2-1.7 Libraries

Drosophila melanogaster cosmid library (J.W. Tamkun)

2-1.8 Plasmids

(i) Vectors

pBluescript KS+ and SK+ (Stratagene)

pGEX-1 and 2T (D. B. Smith and K. S. Johnson 1988)

pUCBM20 (Boehringer Mannheim)

HZ50PL (Y. Hiromi and W. J. Gehring 1987)

(ii) Recombinant plasmids not described in text

p*Caspew* (obtained from R.Saint)

btd (kuntsplice) (E. Wimmer *et al.*, 1993)

GST-Btd (E. Wimmer *et al.*, 1993)

pNB40 *sna* (cDNA) (M. Leptin 1991)

pNB40 *twi* (cDNA) (M. Leptin 1991)

BK-95 (ems) (B. Kalionis and P. H. O'Farrell 1993)

pNB40 *col* (cDNA)(M. Crozatier *et al.*, 1996)

pπ25.7*wc*(Δ2-3) (G. Rubin, UC Berkeley, CA)

2-1.9 Oligonucleotides

(i) Sequencing Primers

Reverse:	5'-d(AACAGCTATGACCATG)-3'
T3:	5'-d(ATTAACCCTCACTAAAGGGA)-3'
T7:	5'-d(TAATACGACTCACTATAGGG)-3'
SK	5'-d(CGCTCTAGAACTAGTGGATC)3'
KS	5'-d(TCGAGGTCGACGGTATC)3'
stg#1	5'-d(TTTAACCATAATTTGG)3'
stg#2	5'-d(CAGCGCCTGCCGTTTGG)3'

(ii) other primers

in vitro mutagenesis primers:

178/179:	5'-d(GGACTGCTGACTTCAGCTGATGCTCGG)3'
1167/1168:	5'-d(CGCACTAAGTGCACGGGTGGCGGCC)3'
1184/1185:	5'-d(GTGGCGGCCGTGCACGTGGCTGCATTG)3'

EMSA double stranded oligonucleotides:

wt *stg:*	5'-d(GTGGGCGGGTGGCGGCCGTGGGCGTGGCTGCAT)3'
	3' (CACCCGCCCACCGCCGGCACCCGCACCGACGTA)5'
mut *stg:*	5'-d(GTGGACGGGTGACGGCCGTGGACGTGGCTGCAT)3'
	3' (CACCTGCCCACTGCCGGCACCTGCACCGACGTA)5'

consensus Sp1 binding site oligo (Promega):

5'-d(ATTCGATCGGGGCGGGGCGAGC)3'

3' (TAAGCTAGCCCCGCCCCGCTCG)5'

2-1.10 Molecular weight markers

(i) DNA

λ DNA digested with *Bst*EII and *Sal*I produces fragments of sizes (in kb): 14.14, 7.24, 4.82, 4.32, 3.68, 3.13, 2.74, 2.32, 1.93, 1.37, 1.26, 0.70, 0.45, 0.22 and 0.11

(ii) Protein

Prestained high molecular weight markers (GIBCO BRL).

2-2 METHODS

Miscellaneous, well established molecular biological techniques were carried out according to the protocols published previously (S. F. Ausubel *et al.*, 1994; J. Sambrook Fritsch, E.F., & Maniatis, T. 1989).

2-2.1 Restriction analysis of DNA

DNA was digested with restriction endonucleases under conditions recommended by the suppliers. 1/10th volume of agarose gel load buffer was added and samples were run on a 1.0 or 2.0% agarose horizontal minigel (Hoefer HE 33) in TAE buffer at 5-10 V/cm. DNA was visualised by staining the gel with 10 µg/ml ethidium bromide and viewing under UV light.

2-2.2 DNA fragment purification

DNA was isolated from agarose gel slices by application of the frozen gel slice to a syringe plugged with glass wool. The liquid from the gel was squeezed out, phenol/chloroform extracted and then ethanol precipitated prior to resuspension in a suitable amount of water.

2-2.3 Creation of recombinant plasmids

Plasmid vector DNA was prepared by digestion with the appropriate restriction endonuclease in the presence of 1 U CIP to remove 5' terminal phosphates. Linear vector molecules were then phenol/chloroform extracted and purified on Sepharose CL-6B mini-columns (see section 2.2.12). Ligations of 100 ng total DNA were performed with insert:vector of 3:1 in 10-20µl 30mM Tris-HCl pH 7.8, 10mM $MgCl_2$, 10mM DTT, 0.5mM ATP and 1 U T4 DNA ligase at 18^{o}C for 4-16 h.

2-2.4 Transformation of recombinant molecules

(i) Heat shock method

A 50 ml mid-log phase culture of DH5α, XL1blue, or JM110 was harvested, resuspended in 20ml 50mM $CaCl_2$ and left on ice for 20 min. The cells were harvested and carefully resuspended in 2ml 50mM $CaCl_2$. 100 µl cell suspension was typically mixed with 5 µl ligation mix and left on ice for 30 min before heat shock at 37^{o}C for 3 min. The mixture was incubated at 37^{o}C for 30 min following addition of 0.5 ml L-broth, plated on L-broth plates with ampicillin, and grown at 37^{o}C for 16 h. For 'blue/white' selection of pBluescript or pUCBM20 recombinant clones 10 µl each of 20% IPTG and 10% BCIG was added to plating mixture.

(ii) Electroporation method

500mL of LB-broth was inoculated with 100th volume of a fresh overnight culture (usually of XLI-blue) and grown at 37^{o}C until the O.D.600 was ~0.7-8. The flask was chilled on ice and then the bacteria pelleted. After a wash with 500ml then 250ml of ice cold water and a wash with 10ml of ice cold 10% glycerol, the cell pellet was resuspended in 1.5ml of ice cold 10% glycerol, dispensed as 45µl aliquots, snap frozen and then stored at -80^{o}C. One aliquot of cells was thawed, mixed with 5µl of salt free DNA and then electroporated at 2.5kV (Bio-Rad). The cells were rescued in 1ml of SOC, incubated at 37^{o}C for 0.5hr and then plated.

2-2.5 Colony cracking for analysis of recombinant clones

Lids were removed from microfuge tubes and 15µl of cracking solution was dispensed into each one. A colony was picked up with a yellow tip, gently touched to a fresh LB plate (a masterplate) then transferred to a microfuge tube containing the solution. The tip was 'swirled' until the solution rose up into it by capillary action. The tubes were incubated at 65^{o}C with an additional 'swirl' for 15 minutes.

The samples were loaded into a non-submerged agarose gel and were 'run in' at 30V. After this time the gel was submerged with the running buffer and subjected to electrophoresis at 90V until the bromophenol blue dye reached the bottom of the gel Any clones that migrated slower in the gel than the control parental colony were selected for further analysis.

Cracking solution: 50mM NaOH, 0.5% SDS, 5mM EDTA, 10% Glycerol, 0.025%, bromophenol blue

2-2.6 Isolation of plasmid DNA

(i) 'Miniprep'

A single colony was used to inoculate 2ml of L-broth plus ampicillin and incubated for 5-16 h at 37^{o}C with shaking. DNA was isolated by the "boiled lysis" method (G. Murphy and T. Kavanagh 1988)to the stage of isopropanol precipitation, where the pellet (typical yield 10 µg) was resuspended in 20 µl of water.

(ii) Large scale preparation

A single colony was used to inoculate 25ml of L-broth plus ampicillin and incubated for 16 at 37^{o}C with shaking. DNA of high purity was isolated by using a 'Qiagen' midi kit (Qiagen Inc.).

2-2.7 Plating the cosmid library

Cosmids were plated onto LB plates overlayed with nitrocellulose filters, containing 50 µg/ml ampicillin and supplemented with 10% glycerol and then grown at 30^{o}C for 16 h. In the order of 25,000 recombinant cosmids were plated at a density of 3000 per 15cm plate, giving a >99% probability of isolating any single copy sequence. Replica filters were made and processed as described in (N. H. Brown and F. C. Kafatos 1988) and then hybridised in plastic petri dishes using procedures in section 2-2.9. The master plates were stored at -20^{o}C.

2-2.8 Radiolabelling of DNA fragments

DNA fragments were labelled by Klenow catalysed α-^{32}P-dATP incorporation in random oligonucleotide primed synthesis products (A. P. Feinberg and B. Vogelstein 1983) using a "Megaprime" kit (Amersham). Unincorporated nucleotides were removed by size exclusion spun column chromatography using Sephadex G-25, DNA grade (Pharmacia) as described elsewhere (J. Sambrook Fritsch, E.F., & Maniatis, T. 1989).

2-2.9 Hybridisation of radiolabelled probes to membrane immobilised nucleic acids

Filters were pre-hybridised with hybridisation mix (50% formamide, 5 x SSC, 0.5% blotto, 100 µg/ml sonicated and denatured salmon sperm DNA) at 42^{o}C for at least 2 h. Heat denatured and snap cooled radiolabelled probe was added to the membranes with fresh hybridisation mix and incubated at 42^{o} for 4-16 h.

Membranes were washed at high stringency with two 10 min washes each of 2 x SSC, 0.1% SDS at rt, then 0.1 x SSC, 0.1% SDS at 65^{o}C.

2-2.10 Autoradiography

Membranes or dried gels were exposed for variable periods to X-Omat AR X-ray film (Kodak) in an autoradiography cassette (Ilford) at rt, or at -80^{o}C in the presence of a calcium tungstate intensifying screen. Alternatively, exposure was to a pre-erased phosphorimager capture screen followed by laser scanning and image analysis (Fujix BAS1000 scanner and MacBas version 2 software).

2-2.11 Isolation of cosmid DNA

If the positive clone could not be clearly identified on the master plates a second round of screening was performed prior to the isolation of cosmid DNA. Putative positive clones were selected and were replated as in 2-2.7(scaled down), except that the plates were not overlayed with nitrocellulose and did not contain 10% glycerol.

Duplicate positive clones were identified and inoculated into 50mls of L-broth containing 50µg/ml ampicillin and grown at 37^{o}C for 16 hours. Cells were then pelleted at 4000K for 5 minutes, and the supernatant removed. The pellet was resuspended in 2mls of TES (25mM Tris HCl pH 8.0, 10mM EDTA pH 8.0, 15% sucrose) and then lysed by the addition of 4mls of 0.2M NaOH, 1%SDS (made up fresh), gently mixed and left at room temperature for 5 minutes. Chromosomal DNA, high molecular weight RNA and protein/membrane complexes were then precipitated by adding 3mls of ice cold KAcF (2.5M KAc, 4.5% Formic acid), mixing gently, and leaving on ice for 5 minutes. The precipitate was then pelleted at 15000K for 15 minutes at 4^{o}C and the supernatant was then transferred to a fresh tube. 5µl of RNase A (10mg/ml) and 2µl of proteinase K (20mg/ml) were then added and left at 37^{o}C for 30 minutes. Following this treatment the DNA was extracted twice with an equal volume of phenol/chloroform and then ethanol precipitated. The DNA pellet was typically resuspended in 200µl.

2-2.12 Nucleotide sequence analysis

(i) Sequencing template preparation

9 µg of plasmid DNA was RNase treated, alkali denatured and purified on a Sepharose CL-6B mini column as described (G. Murphy and T. Kavanagh 1988). 3 µg of this was annealed with 10 ng of primer at 37^{o}C for 1 h.

(ii) Sequencing reactions

DNA was sequenced by the dideoxy method (F. Sanger *et al.*, 1977) using α-^{35}S-dATP and a Sequenase Sequencing Kit (United States Biochemical).

(iii) Electrophoresis

Products of sequencing reactions were resolved on 0.4 mm 6% acrylamide (acrylamide:bis-acrylamide, 20:1), 8 M urea, TBE gels. Gels were dried on a vacuum gel drier at 80^{o}C for 30 min onto 3MM paper (Whatman), and autoradiographed for 16 h.

2-2.13 *In vitro* mutagenesis of plasmid clones

In vitro mutagenesis was performed according to manufacturers protocol (Stratagene), using XLI blue and the helper phage M13/K07 to generate the single stranded DNA. Putative mutant

clones were analysed by restriction digestion for the incorporation of the new restriction site and then sequenced to confirm the alterations.

2-2.14 Maintenance of *Drosophila* stocks

Stocks were routinely cultured at 25^{o}C in plastic vials or plastic bottles containing *Drosophila* culture medium. Stocks needed for collections of large numbers of eggs were maintained in a population cage.

2-2.15 P element mediated transformation of *Drosophila*

(i) Micro-injection of embryos

High purity DNA for injection was prepared using the method described in section 2-2.6(ii). The construct DNA (at a concentration of 700ng/µl) and the transposase activity plasmid, pπ25.7*wc*(Δ2-3) (at concentration of 300ng/µl) were combined in injection buffer (5mMKCl, 0.1mM PO_4 pH6.8). Manually dechorinated w^{1118} embryos, staged between 30 to 60 minutes AED at 18^{o}C, were aligned on a strip of non-toxic rubber cement (Earth), in a humidified room to prevent excessive desiccation, and then covered with a drop of light paraffin oil. The posterior end of each embryo was then micro-injected with the above DNA mix and embryos were left to hatch and crawl into the yeast paste that encircled them.

(ii) Screening for transformants

Adults that developed from the injected embryos were individually crossed to w^{1118} virgins or males and transformed lines were identified amongst the progeny by the w^{+} eye colour marker. The eye colours obtained varied from very pale yellow to strong orange but were consistent for any independent event, except that males often showed a stronger eye colour than females.

(iii) Creating stable lines of transformants

Independent transformants were crossed to the doubly balanced stock: *w;Adh/CyO;Δ2-3Sb/TM6b* and in the next generation, male transformant flies, carrying the *Cyo* and *TM6b* chromosomes were selected and crossed back to w^{1118} virgins. The progeny of this cross were scored to determine whether the P element insert was segregating from either the second chromosome (by the presence of no curly winged flies (*Cy*) with coloured eyes), or the third chromosome (by the presence of no flies with an increased number of bristles on the humeral plate (*Hu*) with coloured eyes). If the P element insertion was on the X chromosome then no male flies with coloured eyes would be detected amongst the progeny. Any lines that did not segregate with one of these three chromosomes were assigned to the fourth chromosome and discarded as long as at least three other insertion events that were not on the fourth chromosome were identified.

Once the chromosome of insertion was determined, stable lines were generated by homozygosing the P element insert, or if this was lethal, the insertion was maintained over a balancer chromosome such as *Cyo* or *TM6b*.

2-2.16 Harvesting and 'fixing' *Drosophila* embryos

Embryos were collected on grape juice agar plates smeared with yeast. They were then harvested and washed thoroughly in a sieve using copious amounts of 'embryo wash buffer'. The sieve was then transferred into a container with 50% commercially available bleach (2% sodium

hypochlorite) for 2 minutes to de-chorionate the embryos. The embryos were once again washed in the sieve thoroughly using 'embryo wash buffer'. They were then transferred to a glass scintillation vial containing a two-phase mix of 4ml of 4% formaldehyde in PBS (made fresh by boiling paraformaldehyde in PBS) and 4ml of heptane. The vial was then shaken on an orbiting platform such that the interface between the liquid phases was disrupted and the embryos were bathing in an emulsion, for between 15 and 30 minutes to 'fix' the embryos. The bottom phase (aqueous) was drawn off and replaced with 4ml of methanol and the vial was shaken vigorously for 1 minute to de-vitellinise the embryos. The de-vitellinised embryos sink from the interface and can be collected from the bottom phase (methanol). Embryos were rinsed several times in methanol at which point they were either stored at -20^{o}C in methanol or processed for whole mount *in situ* hybridisation or immuno-staining.

embryo wash buffer: 0.7% NaCl, 0.15% Triton X-100

2-2.17 Whole mount immuno-staining of *Drosophila* embryos

The methanol was removed from embryos in a microfuge tube and replaced with PBST. Several rinses were done using PBST followed by a single wash for 30 minutes. They were then 'blocked' in 1ml of PBST containing 5% Blotto (commercially available skim milk powder) for at least 1 hour. The blocking solution was removed and the primary antibody was added (rabbit anti ßgal antibody), diluted 1 in 250 in fresh blocking solution (usually 200µl). The embryos were routinely incubated with gentle agitation at 4^{o}C overnight. The next day, the antibody solution was removed and the embryos were washed extensively in PBST (several changes of buffer over a 2 hour time period). The embryos were then incubated with secondary antibody conjugated with horseradish peroxidase diluted in fresh blocking solution for at least 2 hours at room temperature with gentle agitation. Following a period of washing as for the primary antibody, the antibody localisation was detected colourimetrically. The embryos were incubated in a solution of 0.5mg/ml DAB, 0.045% H_2O_2, until the staining had developed (as assayed on a dissecting microscope), and then rinsed thoroughly using PBST prior to mounting in PBS/80% glycerol.

The rabbit ß-galactosidase antibody was raised at the University of Adelaide animal house by injecting rabbits with the ß-galactosidase purified protein (Boehringer Mannheim). After three boosts, the serum was collected and the IgG's were purified on a protein A column according to the manufacturers specifications (Biorad).

2-2.18 Whole mount *in situ* RNA hybridisations

(i) Preparation of digoxigenin incorporated antisense RNA probes

Clones for making antisense RNA probes were selected such that the insert of interest had a promoter for SP6, T7 or T3 RNA polymerase at the 3' end of the coding sequence (not necessarily full length). *In vitro* transcription was then used to incorporate DIG-11-dUTP into the RNA product. This was done using the DIG RNA labelling kit, obtained from Boehringer Mannheim, according to the manufacturers specifications.

Typically 0.5 or 1µl of the above probe was then added to 50µl of prehybe solution (see below), heated to 95^{o}C for 1 minute and the snap chilled in wet ice.

(ii) Hybridization and detection

This method is slightly modified (D. Tautz and C. Pfeifle 1989). All washes are 1ml volume for 5 minutes unless stated otherwise and were carried out under gentle agitation.

Fixed embryos (as described in 2-2.16), usually about 50µl in an microfuge tube, were rinsed several times in ethanol, then washed once in 50% ethanol, then washed three times in PBST before one 60min wash in PBST. The embryos were then treated with proteinase K (50µg/ml) in PBST for 5 to 7 minutes. This digestion was stopped using three washes of PBST containing 2mg/ml glycine, followed by two PBST washes. The embryos were then post fixed in 4% formaldehyde in PBST for 20 minutes and then washed five times in PBST.

500µl of prehybe solution (50% deionised formamide, 5XSSC, 50µg/ml heparin, 0.1% Tween 20, 100µg/ml sonicated boiled salmon sperm DNA) was then added and the embryos were incubated at 55°C for 60 minutes. 450µl of the prehybe solution was then removed and the probe was added, gently mixed, and then left overnight at 55°C.

The following day, the probe was drawn off and 1ml of prehybe solution was added and left at 55°C for 60 minutes. 500µl of this prehybe was then removed and 500µl of PBST (pre heated to 55°C) was added and left at 55°C for 30 minutes. This was followed by five 1ml washes at 55°C for 20 minutes each.

The embryos were then incubated in PBST containing a 1 in 2000 dilution of the anti-digoxigenin antibody conjugated to alkaline phosphatase (Boehringer Mannheim) for 60 minutes at room temperature. This was followed by four 20 minute washes.

Colour detection of the antibody was then performed by washing three times in dig staining buffer (100mM NaCl, 50mM $MgCl_2$, 100mM Tris pH 9.0, 0.1% Tween 20) and then adding 1 ml of staining solution (4.5µl of 100mg/ml NBT made up in 70% dimethylformamide (Boehringer Mannheim) and 3.5µl of 50mg/ml BCIP in dimethylformamide (4-toluidine salt, Boehringer Mannheim) in 1ml of staining buffer). This reaction was left for one to several hours in the dark at room temperature, and was monitored intermittently using a dissecting microscope. Once the staining reaction had proceeded far enough it was stopped by several washes in PBST containing 20mM EDTA. Embryos were then mounted in PBS containing 80% glycerol and 20mM EDTA.

2-2.19 Light microscopy and photography

Embryos were mounted under a coverslip supported by pieces of double sided tape and the edges were then sealed with commercially available nail varnish. They were then viewed on a Zeiss Axiophot light microscope using 10X eyepiece lenses and Plan-Neofluar 20X/0.5 or 40X/0.75 objectives with DIC optics. Photographs were taken with a Zeiss Microphot system and recorded on Ektachrome 160T reverse colour film (Kodak).

Slides were scanned with a Kodak RFS 2035 Film Scanner at >500 dpi. Adobe photoshop 3.0.4 was used for image preparation. Colour prints were obtained using a Kodak XLT7720 Digital Continuous Tone Printer.

2-2.20 Expression of bacterial fusion proteins

Clones in pGEX plasmids were transformed into bacterial strains DH5α or XLI-blue. A single colony was transferred into a flask containing LB-broth and 100µg/ml ampicillin and grown overnight at

37°C as a pre-culture. A 1:10 dilution of this culture was made in a flask containing LB-broth and 150-250μg/ml of ampicillin (the higher concentration of antibiotic serves to maintain the plasmid under the potentially stressful conditions of induction), incubated at 37°C until the $O.D._{600}$ reached 0.6-0.8. IPTG was added to a final concentration of 0.1-0.3mM. The culture was then incubated at 37°C for a further 3-6 hours to allow the accumulation of expressed protein. The bacteria were then pelleted at low speed and the medium discarded.

2-2.21 Harvesting soluble fusion protein

A 500ml culture of each of the GST clones was induced as described in section 2-2.20. The cells were then pelleted at 6000xg for 10 minutes, washed in 20ml of LB, repelleted at 6000xg for 10 minutes and suspended in 1.3ml of buffer A (40mM Tris-HCl pH 7.7, 25% [w/v] sucrose, 0.2mM EDTA, 1mM DTT, 1mM PMSF, 1mM sodium metabisulfite). 0.5ml of buffer A containing 1mg/ml lysozyme was then added and the mixture was incubated at 4°C for 24 hours. A two thirds volume of 10M urea was then added to give a final concentration of 4M urea, this mixture was also incubated at 4°C for 24 hours before being centrifuged at 63,000xg for 1 hour. The supernatant (containing the solubilised protein) was then dialysed against 200ml of buffer B (20mMTris-HCl pH 7.7, 50mM KCl, 10mM $MgCl_2$, 1mM EDTA, 10μM $ZnSO_4$, 20% [v/v] glycerol, 1mM DTT, 0.2mM PMSF, and 1mM sodium metabisulfite) containing 1M urea, at 4°C for 90 minutes. The mixture was then dialysed twice against 500ml of buffer B, for 2 hours and overnight. The resulting protein was then stored at -80°C. The amount of fusion protein present in the extract was then estimated by SDS PAGE (see below) and comparison to protein size standards of known concentration.

This method is largely based upon that of (J. T. Kadonaga *et al.,* 1987).

2-2.22 Protein gel electrophoresis

All SDS-PAGE of protein samples and Coomassie blue staining of gels was performed exactly as described elsewhere (E. Harlow and D. Lane 1988).

2-2.23 Electrophoretic mobility shift assays (EMSA)

(i) Band shifting assays

The 1.4kb BamHI to XhoI fragment was purified on an agarose gel as described in 2-2.2 and then digested with Sau3IIA. 1μg of this digest was then end filled using Klenow 5U, 1μl of $\alpha^{32}P$ labelled dATP and 1μl of each 2mM unlabelled dCTP, dGTP, and dTTP in klenow buffer (50mM Tris HCl pH 7.2, 10mM $MgSO_4$, 0.1mM DTT) in a total volume of 50μl for 19 minutes at rt. The unincorporated nucleotides were then removed as in section 2-2.8.

20 fmoles of the labelled DNA was then incubated in a reaction volume of 10μl, on ice for 20 minutes, in binding buffer (4% glycerol, 1mM $MgCl_2$, 0.5mM EDTA, 0.5mM DTT, 50mM NaCl, 10mM Tris-HCl, pH 7.4, 0.08mg/ml herring sperm DNA) containing between 0 and 4μl of protein extract. Unlabelled specific competitor DNA at 10X, 100X or 1000X the concentration of the labelled DNA was added for 20 minutes prior to addition of the labelled DNA. Following this, the reactions were loaded directly onto a 10% non-denaturing 30:1 polyacrylamide gel containing 10% glycerol and 0.5X TBE, without the addition of any load buffer. The protein/DNA complexes were resolved by electrophoresis

at 10V/cm in 0.5X TBE at 4°C. Gels were then dried on a vacuum gel drier at 80°C for 30 min onto 3MM paper (Whatman), and autoradiographed.

(ii) Double stranded oligonucleotide EMSA

The complementary single stranded oligonucleotides were annealed in the presence of annealing buffer (50mM Tris HCl pH 8.0, 40mM KCl, 6mM $MgCl_2$) by heating to 95°C for 2 minutes followed by a slow cool to room temperature. The DNA was then run on a 20% non-denaturing 30:1 polyacrylamide gel to resolve the double stranded product. This band was excised and the oligo eluted (J. Sambrook Fritsch, E.F., & Maniatis, T. 1989).

Approximately, 1.75 pmoles of this oligonucleotide was then kinased using 5U of T4 polynucleotide kinase and 1µl of $\gamma\,^{32}P$ labelled dATP in kinasing buffer (50mM Tris-HCl pH7.6, 10mM $MgCl_2$, 5mM DTT, 0.1mM spermidine, 0.1mM EDTA) and incubating at 37°C for 30 minutes. The unincorporated nucleotides were then removed as in section 2-2.8.

8.75 fmoles of double stranded labelled oligo was incubated in binding buffer in a total volume of 10µl on ice for 20 minutes, in the presence of none or 0.5µl of protein extract. Unlabelled specific competitor oligonucleotides were added at 10X, 100X or 1000X the concentration of the labelled DNA for 20 minutes prior to the addition of the labelled oligonucleotide. Following this, the reactions were loaded directly onto a 12% non-denaturing 30:1 polyacrylamide gel containing 10% glycerol and 0.5X TBE, without the addition of any load buffer. Electrophoresis was performed as described above.

2-2.24 Regulatory considerations

All manipulations involving recombinant DNA were carried out in accordance with the regulations and approval of the Genetic Manipulation Advisory Committee and the University Council of the University of Adelaide.

All manipulations involving animals were carried out in accordance with the regulations and approval of the Animal Ethics Committee and the University Council of the University of Adelaide.

2-3 ABBREVIATIONS

Abbreviations are as described in "Instructions to authors", *Biochem. J.* (1978) 169, 1-27.
In addition:

APS	ammonium persulphate
BCIG	5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bisacrylamide	N,N'-methylene-bisacrylamide
bp	base pair
BSA	bovine serum albumin
blotto	skim milk powder
CIP	alkaline calf intestinal phosphatase
DIC	differential interference contrast
dpi	dots per inch
DTT	dithiothreitol

EMSA	electrophoretic mobility shift assay
GST	glutathione-s-transferase
IPTG	isopropyl-ß-D-thiogalactopyranoside
kb	kilobase
kD	kilodalton
NBT	4-Nitro blue tetrazolium chloride
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PMSF	phenylmethanesulfonylfluoride
rpm	revolutions per minute
rt	room temperature
SDS	lauryl sulphate sodium salt (sodium dodecyl sulphate)
TEMED	N,N,N',N'-tetramethylethylenediamine
U	unit

CHAPTER 3: INITIAL PROMOTER ANALYSIS OF *STRING*

3-1 BACKGROUND

Promoter analysis of upstream *stg* sequences, utilising a *lacZ* reporter gene vector, was initiated prior to the work detailed in this thesis. My honours thesis describes the generation of five promoter constructs that cover the first 8.7kb upstream of the *stg* TATA box. These constructs used a Kpn I restriction enzyme site that was engineered into the TATA box using *in vitro* mutagenesis. The constructs all began from this Kpn I site and extended upstream for; 0.5, 1.2, 3.1, 6.2, and 8.7kb (see figure 3-1A). This strategy was undertaken in case particular basal sequences, located close to the TATA box, were required in combination with upstream enhancer elements to achieve normal *stg* transcription.

After P element mediated transformation of these constructs into *Drosophila*, embryos were tested for the expression of the ß-galactosidase reporter gene. This was done by assaying for activity of the protein using the colour substrate 5-Bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (BCIG).

All of the constructs expressed ß-galactosidase in a broad stripe on the ventral side of the embryo at cellularisation, which corresponds to the mesoderm and mitotic domain 10 (figure 3-1 B). This was the only cycle 14 mitotic domain that was identified and its presence in all of the constructs suggested that the enhancers responsible were located within the first 0.5kb upstream of the *stg* TATA box. Some other regions of ß-galactosidase expression were observed only in the 8.7kb construct, however the timing of their expression suggested that they may have been cycle 15 domains rather than those of cycle 14 (figure 3-1 C). All constructs also expressed ß-galactosidase in a reiterated pattern of lateral patches at the completion of rapid germband elongation (figure 3-1D), and during germband shortening in the central nervous system (CNS) and, at a low level, in the peripheral nervous system (PNS) (figure 3-1E). These common expression patterns also suggest that the enhancers for these regions lie in the first 0.5kb upstream of *stg* .

3-2 CHARACTERISATION OF THE DOMAIN 10 ENHANCER REGION

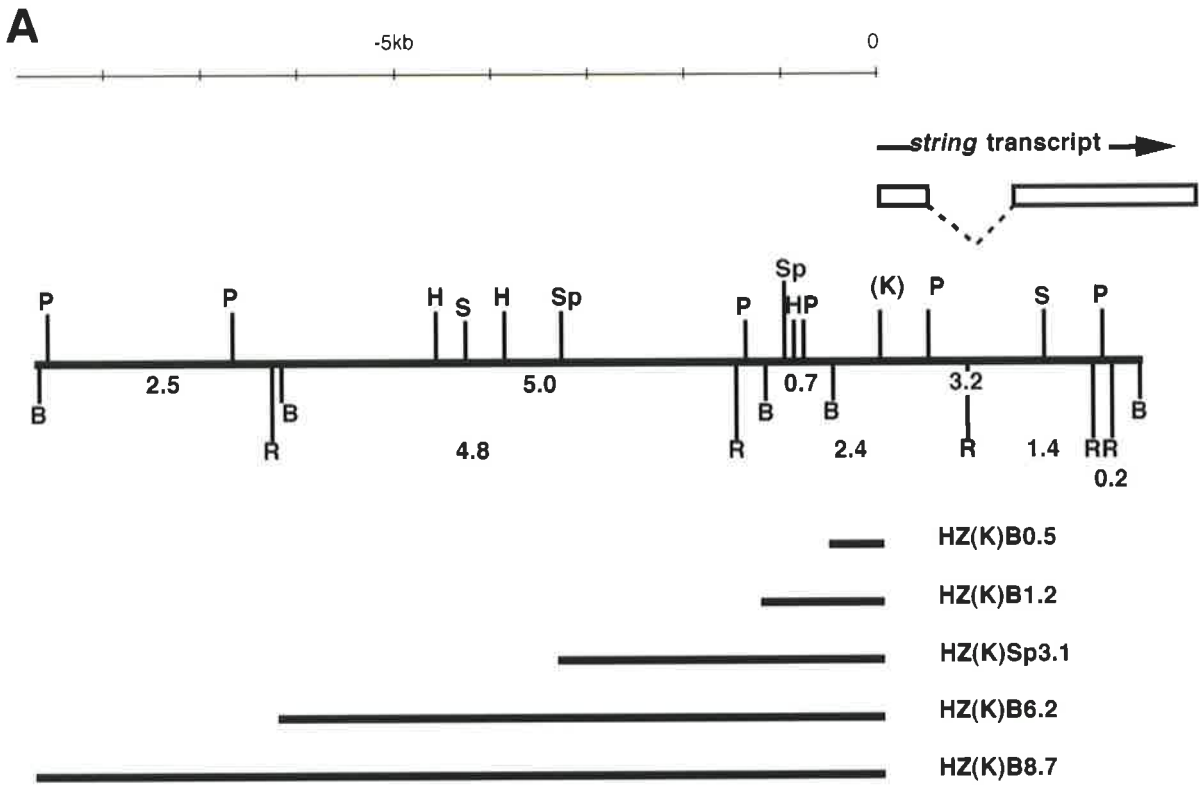
3-2.1 Background

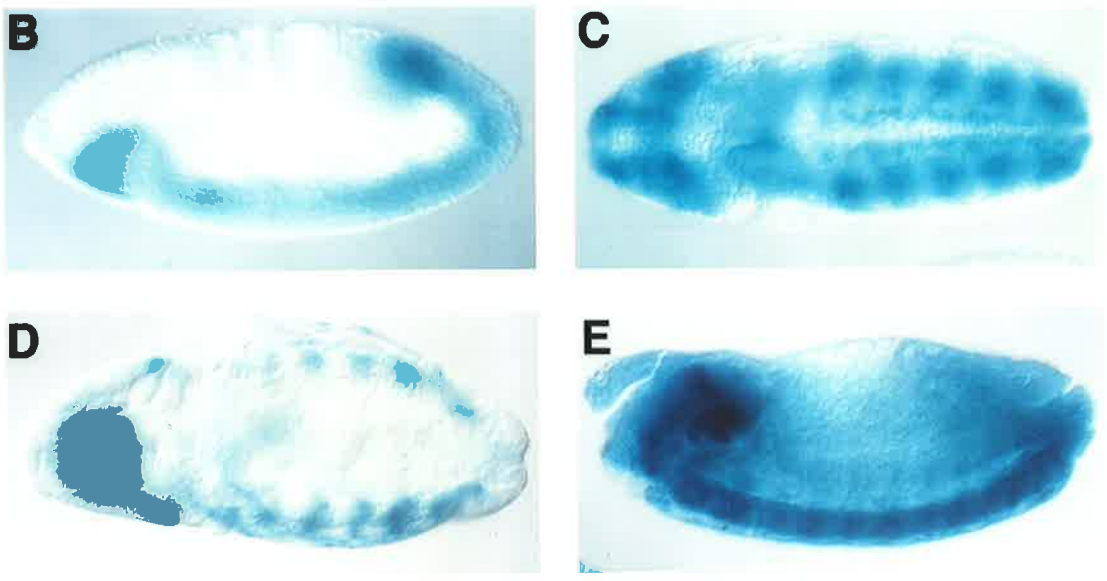
To further define the domain 10 enhancer activity within the 500bp region, deletion studies were initiated. However, the possibility arose that this mesodermal pattern of transcriptional activation of *lacZ* was artefactual. Doyle *et al.* (1989) discovered that particular fragments of the *zen* promoter, when fused to *lacZ* in a transformation vector derived from Carnegie 20 (G. M. Rubin and A. C. Spradling 1982), resulted in *lacZ* transcription in the mesoderm (H. J. Doyle *et al.*, 1989).As *zen* is normally only expressed in the cells of the dorsal surface, this result was unexpected. It was found that the *rosy (ry)* transcription unit which is present within the vector to act as the eye colour marker for selecting transformants was responsible for this artefactual expression. Placing the same promoter fragments, that gave the mesodermal expression patterns, into a vector containing the *white, (w)* transcription unit, rather than *ry*, was sufficient to remove this artefactual expression (H. J. Doyle *et al.*, 1989).

The fact that *ry* was found to be normally expressed in the mesoderm suggested that the transcription unit used in the Carnegie 20 vector may still contain a mesodermal enhancer element (H.

Figure 3-1

A genomic map showing the contiguous constructs made and the expression patterns obtained in embryos transgenic for these constructs when stained for ß-galactosidase activity. **A**, map showing the *stg* transcript and the first 8.7kb upstream of the transcription start site with constructs shown below. Restriction sites mapped are: EcoRI (R), BamHI (B), PstI (P), HindIII (H), SacI (S), SpeI (Sp) and KpnI (K). The KpnI site, is that created at the TATA box at -30bp by *in vitro* mutagenesis. The constructs exist in the vector HZ50PL, hence the letters HZ in the name of each. The restriction sites that delimit the *stg* sequences are also present in the name with each construct beginning from the introduced KpnI site (K) and extending upstream to genomic sites as indicated on the map. The numbers refer to the size of the construct. **B**, a lateral view of an embryo carrying the construct HZ(K)B0.5, during rapid germband extension (stage 8) expression can be seen in the region corresponding to the mesoderm. **C**, a dorsal view of an embryo carrying the construct HZ(K)B8.7, during slow germband elongation (stage 9), expression can be seen in patches along the germband and in the head, however, these expression patterns were thought to be too late to reflect cycle 14 domains. **D**, a ventral view of an embryo carrying the construct HZ(K)B0.5, towards the end of slow germband elongation (stage 9), lateral patches of expression can be seen as well as a strong patch of expression in the head. **E**, a lateral view of an embryo carrying the construct HZ(K)B0.5 following germband retraction and head involution (stage 14), expression can be seen in the ventral nerve cord and brain as well as fading expression in the PNS (out of the plane of focus).


A
-5kb
0
string transcript
P
P
H
S
H
Sp
P
Sp
HP
(K)
P
S
P
B
R
2.5
4.8
5.0
0.7
2.4
3.2
1.4
0.2
HZ(K)B0.5
HZ(K)B1.2
HZ(K)Sp3.1
HZ(K)B6.2
HZ(K)B8.7


B
C
D
E

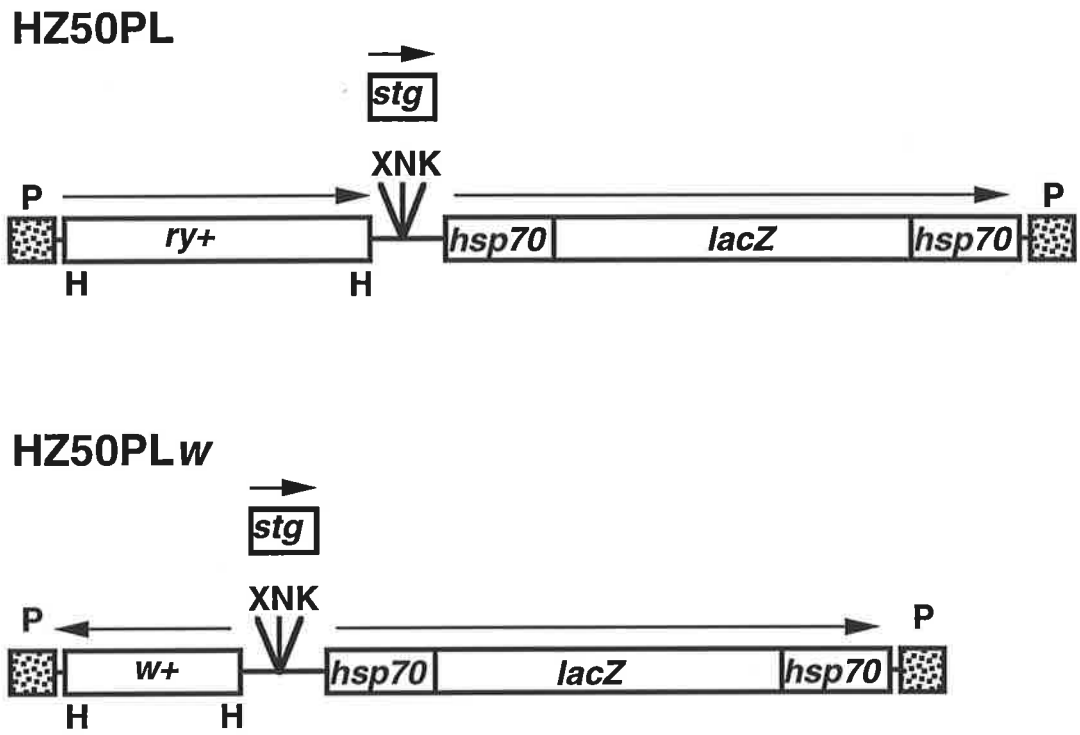


Figure 3-2

The structure of the vectors HZ50PL and HZ50PL*w*. Only the portion of the vector that integrates into the *Drosophila* genome is shown. Patterned boxes represent the P element inverted repeats that are present at each end of the vector. The *hsp70/lacZ* fusion gene is shown and the arrow above it shows the direction of transcription. The polylinker contains an XbaI (X), a NotI (N), and a KpnI site (K) into which the *stg* sequences are inserted. The arrow above the *stg* sequence represents the orientation of the fragment with the arrow pointing towards the *stg* transcript. The vector HZ50PL contains the *ry* transgene which is transcribed towards the polylinker and HZ50PL*w* contains the mini *w* transgene which is transcribed away from the polylinker.

J. Doyle *et al.,* 1989). Although not all fragments inserted in this transformation vector showed mesodermal expression, particular sequences from the *zen* promoter consistently showed this pattern. This suggested that the *ry* mesodermal enhancer element acts in combination with some sequences from the *zen* promoter which are repressed in the presence of other *zen* sequences.

The constructs described in section 3-1 were also generated in a vector called HZ50PL (Y. Hiromi and W. J. Gehring 1987), derived from Carnegie 20. Hiromi *et al.* (1987) inserted a *hsp70/lacZ* fusion gene into the existing polylinker of Carnegie 20 to make HZ50PL. This fusion gene consisted of a *hsp70* basal promoter and 3' sequences fused to the coding region of the E.coli *lacZ* gene. The basal promoter contained *hsp70* sequences including a TATA box, cap site, leader sequence and the first 7 amino acids of the coding region. This basal promoter is not heatshock responsive as it contains only 17bp of *hsp70* sequence 5' to the TATA box which does not include any of the major heatshock elements. At the 3' end of the *lacZ* the *hsp70* sequences encode the polyadenylation signal. Three unique restriction sites; Xba I, Not I, and Kpn were also inserted into the Carnegie 20 polylinker just 5' to the *hsp70/lacZ* fusion gene (see figure 3-2). The reported analysis of the expression pattern in embryos carrying this vector stated that no consistent pattern of expression was detected, with most lines showing no expression at all. However, Hiromi *et al.* (1987) noted in their materials and methods section that they observed some line dependent expression in the ventral midline and/or mesodermal anlage at the germband extension stage in the HZ50PL vector and some constructs generated in this vector.

3-2.2 Construction of a new transformation vector

For our analysis of the *stg* promoter it was quite conceivable that we would see expression in the mesoderm, as this defines a mitotic domain. However, the possibility that this may have been artefactual lead us to test the 0.5kb fragment in a transformation vector using the *white* gene as the selectable eye colour marker. The 7.2kb *ry* transcription unit which was originally cloned into Carnegie 2 as a Hind III fragment to make Carnegie 20, was removed from HZ50PL by performing a Hind III partial digest. In its place the 4.1kb mini *white* gene was isolated using Eco RI, from the plasmid p*Caspew*, the EcoRI sites were then end filled using Klenow and Hind III linkers were ligated to this fragment to allow it to be cloned into the same site as the *ry* gene is inserted in HZ50PL. This new vector was termed HZ50PL*w,* and apart from the fact that the w^+ transcription unit faces in the opposite direction to that of ry^+ in HZ50PL, no other features of the vector were altered (figure 3-2).

3-2.3 Testing for mesodermal expression in the new transformation vector

The 0.5kb *stg* fragment that gave mesodermal expression exists as a Kpn I to Bam HI fragment in pBluescript (pBST) as the clone p*stg*(K)B0.5, this fragment was subcloned into the new vector HZ50PL*w* in the same manner as it was originally inserted into HZ50PL. The insert was released using KpnI and Xba I which cuts in the polylinker outside of the Bam HI site, this fragment was then inserted into HZ50PL*w,* Kpn I to Xba I, to make the clone HZ*w*(K)B0.5.

Following transformation of HZ*w*(K)B0.5 into *Drosophila*, three independent lines were tested for ß-galactosidase expression. None of the lines showed any expression in the mesoderm, revealing that the mesodermal staining previously observed was artefactual (compare figures 3-3A&B).

Comparisons between HZ(K)B0.5 and HZ*w*(K)B0.5 were performed using RNA *in situ* hybridisations to detect the *lacZ* message (as detailed in section 4-2.3(i)). By detecting transcriptional activation, rather than protein expression, it became clear that the mesodermal transcription in HZ(K)B0.5 was occurring too early to represent *stg* transcription in domain 10 and that the anterior and posterior limits of the expression extended beyond those of domain 10. Further, the very strong expression observed was uncharacteristic of domain 10 which is noted for its slow and gradual accumulation of *stg* transcripts (compare figures 3-3A and 1-4A&B).

Some patterns were identified in the construct HZ*w*(K)B0.5 that were not previously or clearly detected using the ß-galactosidase activity assay. This resulted from the sensitivity of the RNA *in situ* hybridisations, which allowed the detection of patterns that were quite weak when the colour substrate reaction was allowed to proceed for longer periods of time. Some variation in the intensity of expression of these different patterns between independent lines was also observed. A description of the expression pattern therefore includes all of the patterns observed even though some of them were extremely faint in particular lines.

Prior to gastrulation, a background expression pattern of broad and diffuse stripes along the anterior to posterior axis of the embryo (figure 3-3B) was observed when the RNA detection was allowed to proceed for a long time. This pattern faded as germband extension began (figure 3-4A), and no further patterns were detected until the end of rapid germband elongation (stage 8). At this time, a series of lateral stripes became evident along the germband (figure 3-4B). These patches appear to be the cycle 15 equivalent of domain 11. In cycle 14, domain 11 exists as a series of 5 broad lateral stripes. In cycle 15 however, this domain apparently subdivides into 10 anterior mitotic domains (one for each of the segments T1 through to A8) and 10 posterior mitotic domains (V. E. Foe *et al.,* 1993). It is possible then, that the thin stripes of expression generated by this construct represent either the anterior or the posterior set of these mitotic domains.

Shortly after this, expression begins in the ventral neurogenic region in a complex pattern, that appears to represent cycle 14 divisions in a subset of domains N and M and also later cycles of domains M and N (figure 3-4C&D). Prior to germband shortening (stage 11), expression can still be seen in the ventral neurogenic region as well as in the developing brain lobes, reflecting the later divisions in this region (figure 3-4E). Following germband retraction (stage 13), expression is still apparent in the brain lobes and to a lesser degree in the ventral nerve cord and PNS (figure 3-4F).

One independent line of the HZ*w*(K)B0.5 construct also showed an additional expression pattern which is detailed in section 4-2.3(ii).

3-2.4 Comparison of *stg* and *zen* promoter sequences.

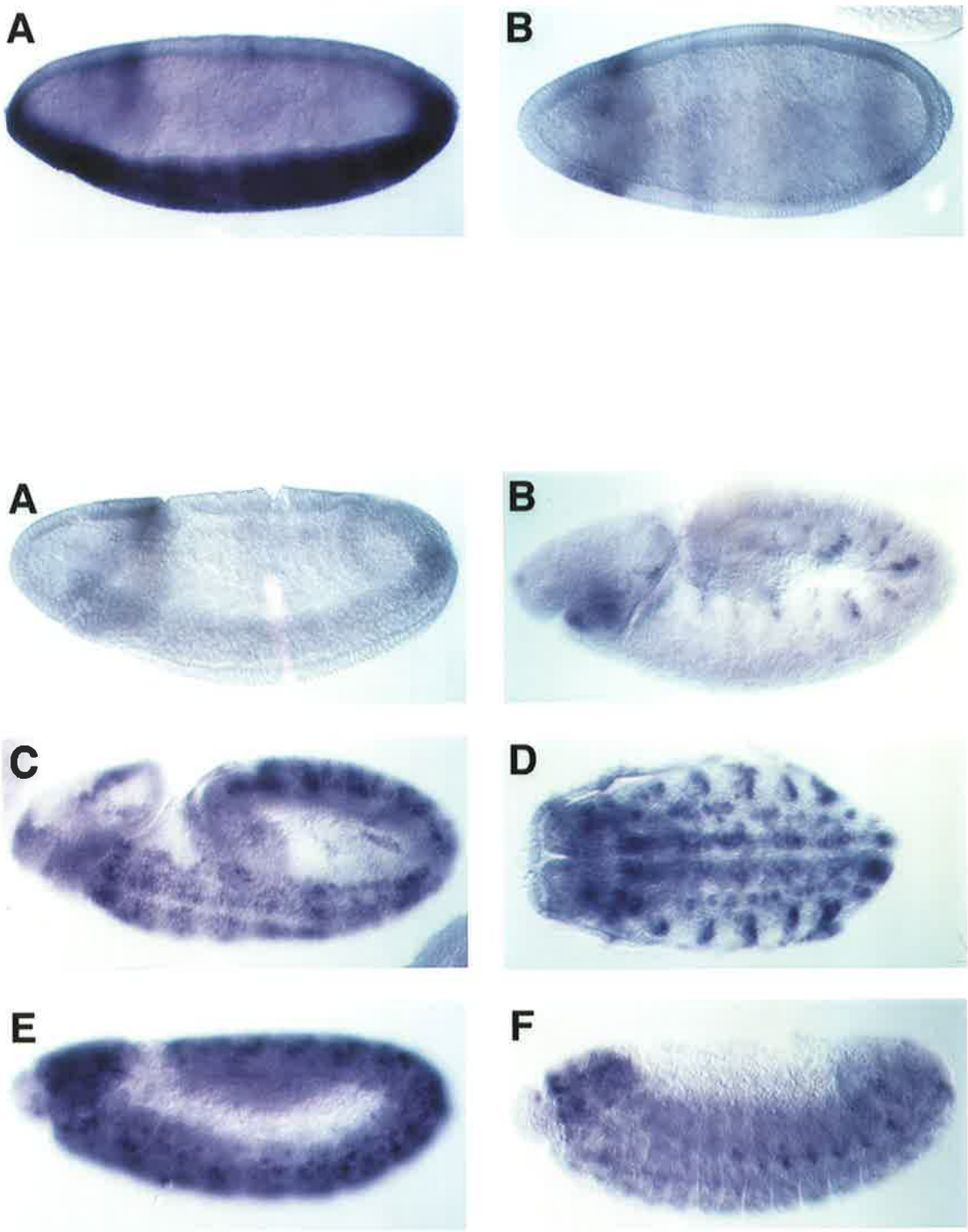
Both the 45bp fragment from the *zen* promoter and the 500bp *stg* promoter fragment, that gave artefactual mesodermal expression, were compared to look for similar sequences that may have been acting in combination with the *ry* enhancer elements. The program Signal Scan 4.0 (D. S. Prestridge 1996 in press) was used to search for transcription factor binding sites. The only similarity identified was the presence of a TATA box in each. This suggests either, that there is some TATA box specificity whereby the *ry* enhancer element can interact with the TATA box of *stg* and *zen* but not that of *hsp70*, or that the presence of an additional TATA box may increase the level of transcription achieved from the *ry* enhancer element. This second option may be the most likely given that some

Figure 3-3

A comparison of the early *lacZ* mRNA expression patterns obtained with the 0.5kb fragment in the vectors HZ50PL and HZ50PL*w*. **A**, a lateral view of an embryo carrying the HZ(K)B0.5 construct, just prior to cellularisation. A strong band of staining is apparent along the ventral surface that extends around the anterior and posterior poles. **B**, a lateral view of an embryo carrying the HZ*w*(K)B0.5 construct, also just prior to cellularisation. This time no staining is observed along the ventral surface but faint broad stripes are visible along the anterior to posterior axis. These stripes are also visible, in addition to the strong ventral expression, in the HZ(K)B0.5 embryo.

Figure 3-4

Further *lacZ* mRNA expression patterns obtained in embryos transgenic for the construct HZ*w*(K)B0.5. All views are lateral unless described otherwise. **A**, the early pattern of stripes along the anterior to posterior axis fades as germband extension begins. **B**, no further expression is detected until the end of rapid germband elongation (stage 8) when a series of lateral stripes appear along the germband. There is also some expression in the head at this stage. **C**, during slow germband elongation (stage 9), while the lateral stripes are still visible, expression begins in the ventral neurogenic region in what appears to be a subset of domains M and N. **D**, a dorsal view of the germband portion of a similarly staged embryo which clearly shows the expression in the ventral neurogenic region. **E**, expression continues in the ventral neurogenic region and also becomes apparent in the developing brain prior to germband shortening (stage 11). **F**, following germband shortening (stage 11) expression remains strong in the brain lobes and is also visible, to a lesser degree, in the ventral nerve cord and PNS.


A
B
A
B
C
D
E
F

line specific mesodermal expression was observed with the HZ50PL vector alone (see section 3-2.1) suggesting that the *hsp70* TATA box alone may be sufficient to drive the *ry* enhancers in some cases, possibly depending on the insertion site of the vector. Alternatively, some other unidentifiable feature(s) within the *stg* and *zen* regulatory regions, but not that of *hsp70* bring about the mesodermal transcription.

3-3 CHARACTERISATION OF THE *string* GENOMIC REGION

3-3.1 Background

Cloning and characterisation of the *stg* transcript revealed a putative TATA box about 420bp upstream of the most likely translation initiation site (B. A. Edgar and P. H. O'Farrell 1989). RNase protection and primer extension assays defined a transcription start site 30bp downstream of this TATA box (Wigley, O'Keefe, and Saint, unpublished observations). However, the RNase protection assay also revealed the presence of a larger product in the 0-2 hour sample suggesting the presence of another, maternal specific, transcription start site upstream of the defined TATA box. This is supported by the Northern analysis performed by Edgar and O'Farrell (1989) where a 3.0kb transcript was detected in maternal RNA samples as well as the 2.8kb transcript (see section 1-2.2) (B. A. Edgar and P. H. O'Farrell 1989). The primer extension data did not, however, reveal this larger product, suggesting that this maternal transcription start site may be quite some distance upstream of the defined transcription start site, or it may involve alternative splicing. Another putative TATA box has since been identified, about 470bp upstream of the first (see figure 4-11) but a further transcription start site has not been defined. Comparisons of the genomic and cDNA sequences revealed a single intron of 0.9kb, 50bp downstream of the translation start site (see figure 3-5).

3-3.2 Mapping of upstream genomic clones

A series of overlapping phage clones were isolated as part of a chromosome walk that was initiated at *stg* and extended upstream of the transcript (R.Tearle, unpublished observations). These phage clones were then used to subclone various promoter fragments for construct generation. A precise restriction map, covering the first 28kb upstream of *stg*, has been generated as particular promoter regions were analysed in detail (figure 3-5).

3-3.3 Mapping of downstream genomic clones

Genomic fragments that rescued only a small number of *stg* domains (B. Edgar, personal communication) suggested that the *stg* regulatory sequences were spread over a large distance. For this reason, a cosmid library was screened using a probe from within the *stg* transcript to obtain genomic DNA that extended downstream of the transcript. Cosmids extending at least 30kb downstream of the transcript were isolated using this probe. No cosmids were obtained that extended further upstream than about -8kb in several independent screenings, indicating the absence of genomic clones spanning this region in the library (figure 3-5).

3-4 EXTENDED PROMOTER ANALYSIS

3-4.1 Background

The initial promoter analysis concentrated on the first 8.7kb upstream of the *stg* transcript. In addition to the contiguous constructs described in section 3-1, discrete fragments from within the 8.7kb were also cloned into HZ50PL and tested for enhancer activity (P.Wigley, figure 3-6). None of these fragments spanned the first 500bp which gave the artefactual mesodermal pattern or the reiterated lateral patches and the late neural expression. In fact, none of them drove expression of the ß-galactosidase reporter gene.

3-4.2 An enhancer search

The finding that the *stg* enhancers are spread over a large distance (B. Edgar, personal communication) suggested that a broader approach was necessary to identify the *stg* regulatory regions. Consequently, large overlapping fragments from the *stg* promoter spanning up to -28kb and down to +22kb were used to generate constructs in HZ50PL. This work was largely performed by P.Wigley and is illustrated in figure 3-6. No expression patterns were detected during the time of the post-blastoderm divisions 14 to 16. However, two late expression patterns were detected in germband retracted embryos. The 2.6kb EcoRI fragment at about -15kb and the 8.8kb SalI fragment that partially overlaps this 2.6 and extends further upstream, showed expression in the dorsal vessel and ventral nerve cord, and strong PNS expression was observed within the 4.3kb fragment at about -25kb (data not shown).

3-5 DISCUSSION

The most obvious reason for the lack of enhancer activity detected using the *lacZ* reporter gene constructions is that the enhancers that drive the patterns of *stg* transcription are located outside of the region tested thus far. However, it is now known, from the work of B. Edgar, that at least some of the domain specific enhancers are located in this region (see section 4-1). This fact emphasises the requirement for using alternative strategies to confirm any results obtained using artificial constructions of this kind. In particular, genomic rescue and the creation of small P element deletions provide information about enhancers in an environment that more closely resembles the normal situation.

There are a number of possibilities to explain the almost complete lack of enhancer activity detected using the *lacZ* reporter gene strategy. The vector HZ50PL has been successfully used to study the regulation of the pair-rule gene *ftz* (Y. Hiromi and W. J. Gehring 1987), suggesting that it is able to function as an enhancer trap. It is possible however, that *stg* regulation is more complex than that of *ftz*, with many more regulatory inputs, both positive and negative, being required to direct expression in a particular domain. In this situation it may not be possible to separate regulatory fragments without disrupting some enhancer sequences. If repressor sequences are moved it is possible that they will have a global effect. Alternatively, if a key activator sequence is deleted expression will be diminished.

It is also possible that the method of ß-galactosidase detection used may not be sensitive enough to reflect *stg* transcription, which is very dynamic. If the transcription of *lacZ* is in any way diminished,

Figure 3-5

A genomic map of the upstream and downstream regions surrounding the *stg* transcript. The transcript is identified and the intron is shown by the dashed lines that join the two exons, drawn as unfilled boxes. The top half of the diagram shows the *stg* transcript and the upstream region that has been mapped to about -29kb using a series of overlapping λ phage clones that extend from the *stg* transcript up to this point. The dashed lines on the ends of the λ phage clones *stg* B, *stg* D, and *Dr*L13 indicate that the exact end point of the clone has not been defined. The bottom half of the diagram shows the *stg* transcript again and the map of the downstream region to about -23kb, which was determined from a series of overlapping cosmid clones. Cos -10a represents the furthest upstream cosmid that was obtained from the library. The arrow on the downstream end of this clone indicates that it extends beyond this map. Restriction sites mapped are: EcoRI (R), SalI (Sal), XbaI (X), BamHI (B, upstream only), PstI (P, upstream only), HindIII (H, upstream only), SacI (S, within the transcript and up to -12kb), SpeI (Sp, within the transcript and up to -12kb) and ClaI (C, only upstream of -12kb). There are no KpnI sites or NotI sites within the region of this map. Sites that are italicised and marked by a dashed line are alternative positions for a single site.

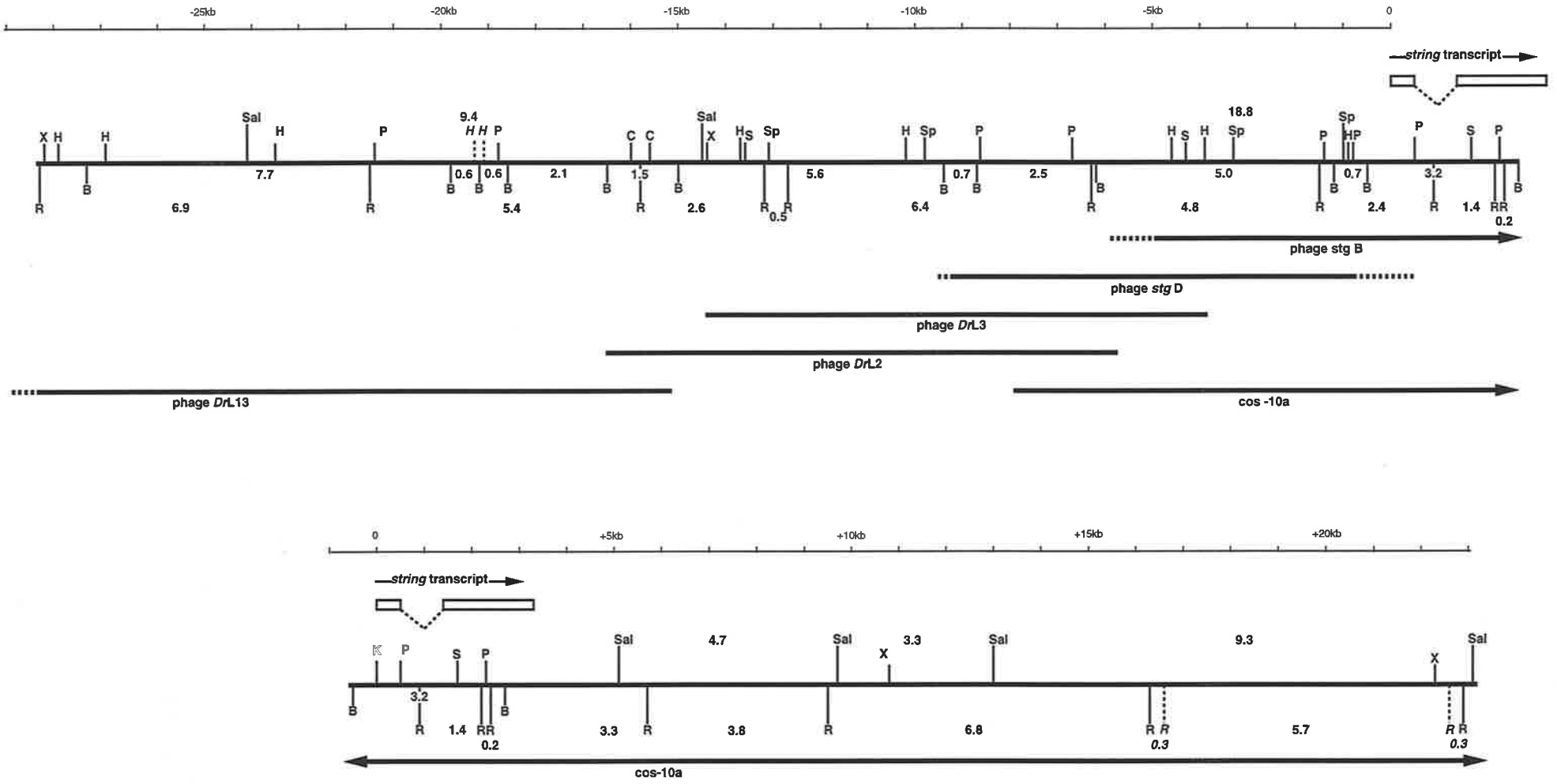
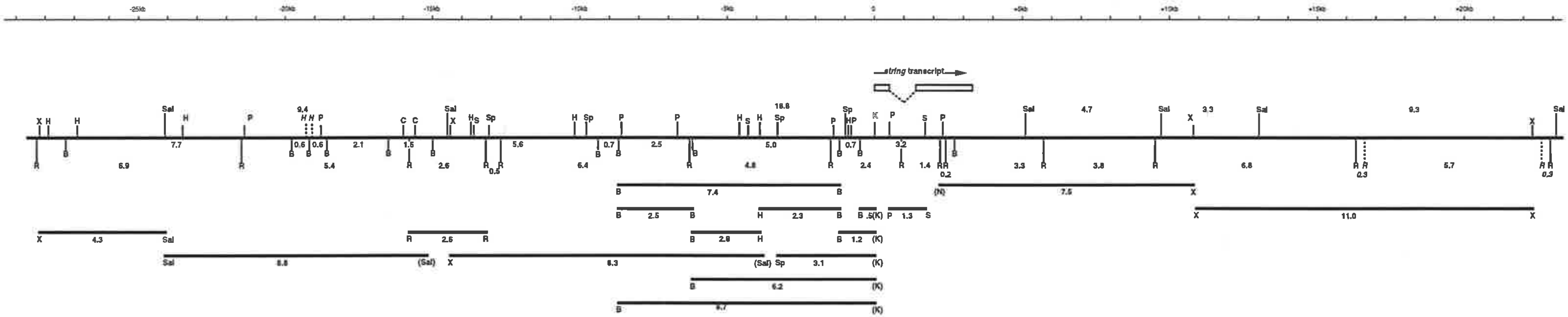

-25kb
-20kb
-15kb
-10kb
-5kb
0
string transcript
X H H Sal H P 9.4 H H P C C Sal X H S Sp H Sp P P H S H Sp 18.8 P Sp H P P S P
R B 6.9 7.7 R B B B 5.4 0.6 0.6 2.1 B R 1.5 B 2.6 R R 0.5 5.6 B 6.4 0.7 B 2.5 R B 4.8 5.0 R B B 2.4 0.7 R 1.4 R R 0.2 3.2 B
phage stg B
phage stg D
phage DrL3
phage DrL2
phage DrL13
cos -10a
0
+5kb
+10kb
+15kb
+20kb
string transcript
K P S P Sal 4.7 Sal X 3.3 Sal 9.3 X Sal
B 3.2 R 1.4 R R 0.2 B 3.3 R 3.8 R 6.8 R R 0.3 5.7 R R 0.3
cos-10a

Figure 3-6

A map showing the constructs generated to cover the regions surrounding the *stg* transcript. The map and restriction sites are those described in figure 3-5A except that the introduced KpnI site (K), in the TATA box at -30bp is shown here. Each construct is represented by a solid bar with the restriction sites bounding the fragment shown, as well as the size of each fragment. Restriction sites within brackets indicate that the site is not present in the genomic DNA, rather it exists within the polylinker of the λ clone used to isolate the fragment.


-25kb
-20kb
-15kb
-10kb
-5kb
0
+5kb
+10kb
+15kb
+20kb
string transcript

relative to normal *stg* transcription, then there may be insufficient protein produced to be detected by this method.

The few late patterns of expression that were detected in this search, may or may not reflect real patterns of *stg* transcription. For example, the enhancers that activate *lacZ* transcription in the PNS which are located about 25kb upstream of the *stg* transcript, may reflect the expression of another unrelated gene. Other transcripts have been identified in the region surrounding *stg*; the *pathless* gene is involved in neural development and is located just 3' to the *stg* transcript, and an un-named transcript with no detectable zygotic expression pattern was found within the 6.9kb EcoRI fragment between -21 and -29kb (B. A. Edgar *et al.*, 1994). Alternatively, these patterns could be artefactual, as was the mesodermal expression. These possibilities again emphasise the importance of complementary methods for determining if the patterns of expression that have been identified (ie. CNS, PNS and dorsal vessel) reflect real patterns of *stg* transcription.

The fact that enhancers that are known to be present within the regions tested fail to drive ß-galactosidase expression casts doubt on the putative regulatory elements that have been detected.

CHAPTER 4: FURTHER PROMOTER ANALYSIS

4-1 BACKGROUND

The work of Edgar *et al.*, (1994) showed, by genomic rescue and local P element deletions, that enhancers for some *stg* expression domains are located within the region previously tested for enhancer activity, detailed in chapter 3 (B. A. Edgar *et al.*, 1994). The following is a summary of that work. A large deletion, AR5, which was generated by P element mediated excision, removes genomic sequences upstream of about 2.1kb to an unidentified point more than -28kb upstream of the *stg* transcript. In this strain, the vast majority of the mitotic domains are missing, leaving *stg* expression in only domains 8, 10, and 15. This suggests that *stg* domains 8, 10, and 15 require enhancer sequences that are located outside of this deficiency. Two partial *stg* transgenes that span the *stg* transcription unit and varying degrees of upstream DNA further define enhancer regions for some domains. The 10.5kb transgene extends approximately 6kb upstream of the *stg* transcript and can rescue *stg* expression in domains 16, 17, 21 and N, as well as in domains 8, 10, and 15. The 15.3kb transgene extends a further 4.8kb upstream and as a result, also rescues domains; 1, 2, 3, 20, 23, and 24. The absence of the enhancers for domains; 4,5,6,7,9,11,12,14,19,25 and M within these transgene regions suggests that they are located further upstream of the 15.3kb transgene (ie., greater than 10.5kb upstream of the *stg* transcript, see figure 4-1). However, it is possible that some domains have very disparate regulatory regions, that also require sequences downstream of the transgenes.

4-2 INCLUSION OF PROXIMAL SEQUENCES IN CONSTRUCTS

4-2.1 Introduction

The apparent abundance of domain specific enhancers that failed to be identified in the reporter gene constructs discussed in chapter 3, indicated that there was a serious problem with our method of detecting the *stg* enhancers using this approach. I decided to re-test the regions already covered, this time including some proximal *stg* sequences, as it seemed feasible that there may be basal sequences located close to the *stg* transcription unit that would be required to achieve correct transcriptional activation of *stg* and therefore also of the *lacZ* reporter gene.

The 500bp fragment, from the *in vitro* mutagenised KpnI site in the *stg* TATA box to the upstream BamHI site, was inserted between the promoter fragment to be tested and the *hsp70/lacZ* fusion gene. The identification of some cycle 14 domain N and M expression as well as a cycle 15 domain within this fragment suggested that it may contain necessary basal sequences. Also a fragment spanning the intron in *stg* was tested, in a similar manner, given its proximity to the beginning of the *stg* transcript. This fragment has been previously tested for enhancer activity, as described in chapter 3, and found not to generate any expression that reflected a mitotic domain type pattern.

4-2.2 Construct generation

The region up to -8.7kb had already been tested as a contiguous construct, HZ(K)B8.7, which included the 500bp fragment (see section 3-1), however, I decided to retest it, this time in the HZ50PL*w* vector.

Figure 4-1

A genomic map showing the regional assignment (A to E) of the cycle 14 mitotic domain enhancers as determined by Edgar *et al.* (1994). The P element deletion AR5 begins at the boundary between regions C and D and extends beyond the upstream limit of the genomic map. In this strain, *stg* is expressed only in mitotic domains 8, 10, and 15, which places the enhancers for these domains either downstream of the C to D boundary or upstream of the 5' break in the the AR5 deletion. The 10.5kb transgene which contains the genomic DNA of regions C and D expresses *stg* in domains 8, 10, and 15 as well as 16, 17, 21 and N. This indicates that domains 8, 10, and 15 must be within the region marked D and the domains 16, 17, 21 and N must be within the region marked C. The 15.3kb transgene which contains the genomic DNA corresponding to B, C and D, expresses *stg* in the domains of the 10.5kb transgene as well as in domains 1, 2, 3, 20, 23, and 24. This indicates that the enhancers for these new domains must be located within the region termed B. One larger transgene which has the same 5' end as the 15.3kb but extends further downstream of the transcript does not show expression in any further mitotic domains. On this basis, as well as the AR5 deletion result, no mitotic domains were assigned to the region termed E. The remaining domains were assigned to region A given the absence of their expression in the transgenes and the deletion strain AR5. Edgar *et al.*, (1994) do not list all of the unassigned domains to region A for reasons that are unclear. The genomic map is the same as that described in figure 3-5.

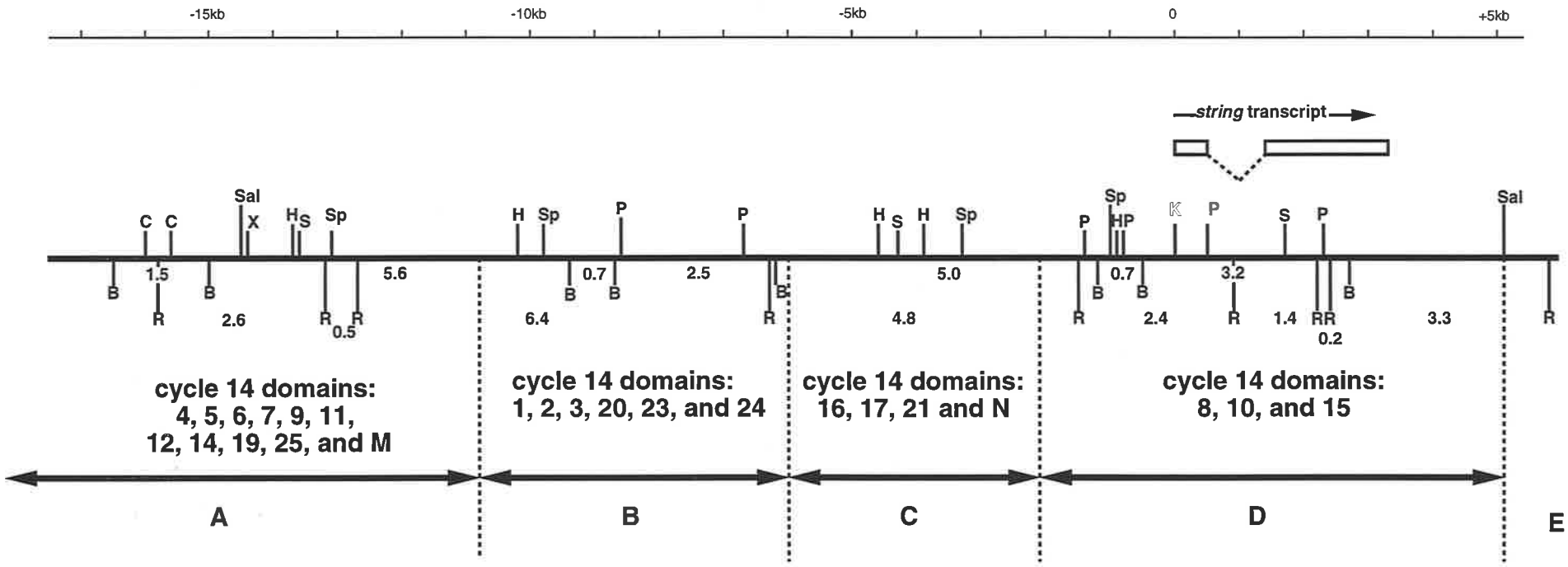
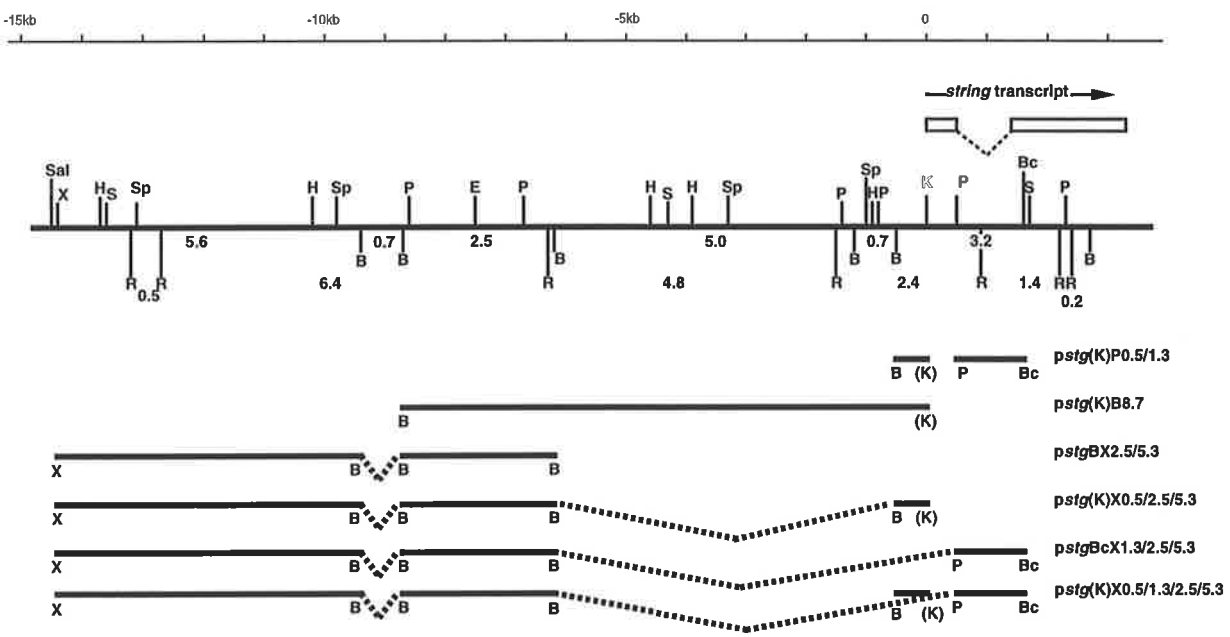

-15kb
-10kb
-5kb
0
+5kb
string transcript
C C
Sal
X
H S
Sp
H Sp
P
P
H S H Sp
P
Sp
HP
K
P
S
P
Sal
B
R
1.5
2.6
0.5
5.6
6.4
0.7
2.5
4.8
5.0
0.7
2.4
3.2
1.4
0.2
3.3
cycle 14 domains:
4, 5, 6, 7, 9, 11,
12, 14, 19, 25, and M
cycle 14 domains:
1, 2, 3, 20, 23, and 24
cycle 14 domains:
16, 17, 21 and N
cycle 14 domains:
8, 10, and 15
A
B
C
D
E

Figure 4-2

A schematic diagram showing the genomic regions that were included in the constructs described in section 4-2.2. The genomic map is as described in figure 3-5 but also shows the BclI site (Bc), situated just downstream of the intron, and the *in vitro* mutagenised KpnI site (K) that replaces the TATA box at -30bp. Constructs are named as the pBluescript clones prior to insertion into the vector HZ50PL*w*, hence the terminology p*stg*. The letters refer to the genomic restriction enzyme sites that are at the boundaries of the clones. and the numbers refer to the size of the fragments. Note the missing 0.7kb BamHI fragment from between the 2.5 and 5.3kb fragments. The construct p*stg*(K)P0.5/1.3 places the intron region on the upstream side of the 500bp proximal frament and the construct p*stg*(K)X0.5/1.3/2.5/5.3 places the intron region between the 500bp and the 2.5kb fragments while maintaining the same relative orientation of the intron as in the other constructs.


-15kb
-10kb
-5kb
0
string transcript
Sal
X
H
S
Sp
H
Sp
P
E
P
H
S
H
Sp
Sp
P
H
P
K
P
Bc
S
P
5.6
0.7
2.5
5.0
0.7
3.2
B
B
B
B
B
R
R
0.5
6.4
R
4.8
R
2.4
R
1.4
R
R
0.2
pstg(K)P0.5/1.3
B (K) P Bc
pstg(K)B8.7
B (K)
pstgBX2.5/5.3
X B B B
pstg(K)X0.5/2.5/5.3
X B B B B (K)
pstgBcX1.3/2.5/5.3
X B B B P Bc
pstg(K)X0.5/1.3/2.5/5.3
X B B B B (K) P Bc

The main test fragment for the new constructs was a 7.7kb fragment extending from the BamHI site at -6kb to the XbaI site at -14.5kb. A 5.3kb XbaI to BamHI clone derived from the phage clone *Dr*L3 and a 2.5kb BamHI clone derived from phage clone *stg*D, were joined to make the 7.7kb BamHI to XbaI fragment (see figure 3-5). Unfortunately, due to an error in the restriction map of this region, the 0.7kb BamHI fragment from within this region was inadvertently left out, hence the name 2.5/5.3 rather than 7.7 (figure 4-2). This fragment still contained significant overlap with the HZ(K)B8.7 construct (as the BamHI 2.5kb is present in both), in case enhancers present at or near -8.7kb were previously disrupted.

The clone p*stg*(K)X0.5/2.5/5.3 was generated by inserting the 5.3kb XbaI to BamHI fragment into the clone p*stg*(K)B0.5, XbaI to BamHI, to create p*stg*(K)X0.5/5.3. Into this clone was inserted the 2.5kb BamHI fragment, making p*stg*(K)X0.5/2.5/5.3 (figure 4-2).

As a control, the same upstream construct was also generated without the 500bp fragment. This time the 2.5kb BamHI fragment was inserted into the BamHI site of p*stg*BX5.3 to make p*stg* BX2.5/5.3 (figure 4-2).

A 1.3kb PstI to BclI fragment was used as the intron containing fragment, this was isolated from within the intron clone HZPS1.3 (see figure 3-6) using BclI and BamHI. The BclI site cuts just inside the SacI (S) site and creates an end that is compatible with BamHI, this does not delete any of the intron sequence. The BamHI site is in the polylinker of the clone HZPS1.3 which cuts just outside of the genomic PstI site. This 1.3kb intron fragment was inserted into the BamHI site of p*stg*BX5.3 to make p*stg*BcX1.3/5.3, the 2.5kb BamHI was then inserted into the only remaining BamHI site to make p*stg*BcX1.3/2.5/5.3. The intron was kept in the same orientation, with respect to the upstream sequences, even though it was put upstream of the transcript (figure 4-2). There was no particular reason for this other than it was unclear as to which way would be correct, and this way simplified the cloning steps.

The final constructs in this series included both the intron and 500bp sequences. The construct p*stg*(K)B0.5/1.3/2.5/5.3 was made by inserting the BamHI to BclI 1.3kb intron fragment into the BamHI site of p*stg*(K)B0.5 to make p*stg*(K)P0.5/1.3. The following cloning steps, to make p*stg*(K)X0.5/1.3/2.5/5.3 (figure 4-2), were the same as described above. The p*stg*(K)P0.5/1.3 clone was also tested for enhancer activity in the absence of the upstream fragment (figure 4-2).

Each of the above clones:
p*stg*(K)X0.5/2.5/5.3
p*stg* BX2.5/5.3
p*stg*BcX1.3/2.5/5.3
p*stg*(K)B0.5/1.3/2.5/5.3
p*stg*(K)P0.5/1.3
p*stg*(K)B8.7

was inserted into the transformation vector HZ50PL*w*, XbaI to KpnI, and P element mediated transformation was used to generate transformant flies. For each construct, the expression pattern of at least three independent lines was tested to ensure that the pattern observed was not dependent upon the position of the P element insertion.

4-2.3 Analysis of expression patterns

(i)Background

To avoid possible problems associated with the ß-galactosidase detection as discussed in section 3-5, two different detection methods were utilized in the analysis of these constructs. A rabbit polyclonal anti-ß-galactosidase antibody serum was used to enhance the detection of the protein. By this method, the anti-ß-galactosidase antibodies, once bound to the ß-galactosidase, can be detected via a secondary anti-rabbit antibody which is conjugated to biotin. The biotin can then be detected using streptavidin linked to alkaline phosphatase, which catalyses a colour substrate reaction. Each of these steps amplifies the original signal and therefore provides a much more sensitive detection method than the ß-galactosidase enzymatic activity assay used previously.

The other strategy employed was to detect the *lacZ* RNA, rather than the protein product, by *in situ* hybridization using a labelled antisense RNA probe. This method directly monitors *lacZ* transcription and therefore avoids any problems that may be associated with the translation of this foreign protein in *Drosophila*.

Both of these methods proved to be useful in assessing patterns of expression. The antibody staining gave particularly strong signals with low backgrounds while the *in situ* hybridisations identified the transcript about 20 minutes earlier, developmentally, than the protein could be detected. With the rapid movement of tissues during gastrulation and germband extension this difference in timing was significant.

(ii) The proximal fragments alone

The construct HZ*w*(K)P0.5/1.3 was compared to the previously made HZ*w*(K)B0.5 (see section 3-2.3), to determine whether any additional expression patterns were produced by the addition of the intron sequences.

No early expression patterns were detected consistently in the HZ*w*(K)P0.5/1.3 construct. The later expression pattern of this construct detected the lateral stripes that may correspond to a cycle 15 equivalent of domain 11 as was seen in HZ*w*(K)B0.5 (figure 4-3A&B), also expression in the ventral neurogenic region was detected, however, this was significantly stronger in HZ*w*(K)P0.5/1.3 by the time of germband retraction (stage 12)(figure 4-3C&D). Late neural expression was also apparent, this was also significantly stronger than in the 500bp construct (figure 4-3E&F). Whether these neural patterns are just much stronger in the HZ*w*(K)P0.5/1.3 construct or, reflect expression in different cells is difficult to determine.

Although no consistent early expression patterns were detected in the HZ*w*(K)P0.5/1.3 construct, one line did show an additional pattern that was not obvious in any of the other lines tested. The *lacZ* reporter gene was transcriptionally activated in this line, towards the end of rapid germband elongation (stage 8), in what is most likely to be domain 25 (figure 4-4A&B).

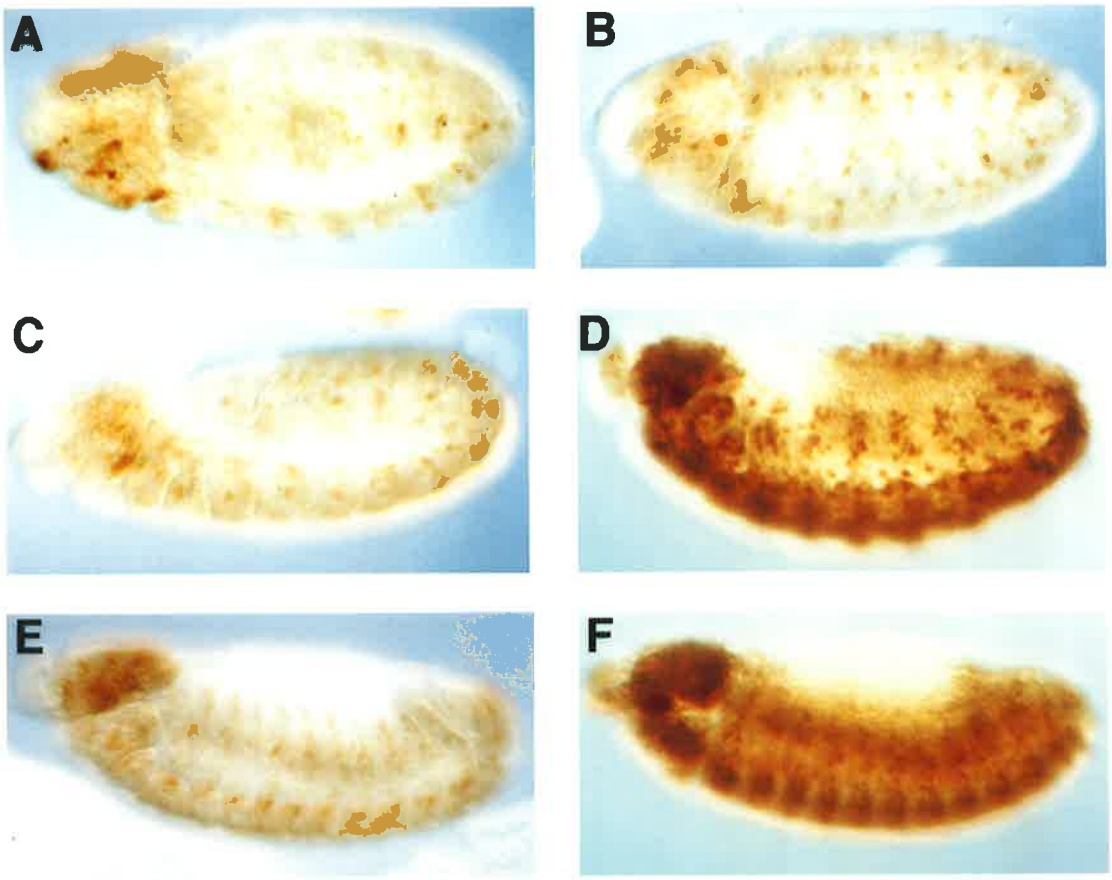
Interestingly, one line of the construct HZ*w*(K)B0.5 also showed an additional expression pattern during rapid germband elongation, in the mesectoderm, which correlates with cycle 14, domain 14 (figure 4-4C&D). The presence of these two expression patterns in only one line each, suggests that they may result from fortuitous enhancers that are located nearby to the insertion site of the particular construct. However, their domain like nature makes it possible that domain-specific elements, which are on their own are insufficient to drive expression, are able to function in the chromosomal context

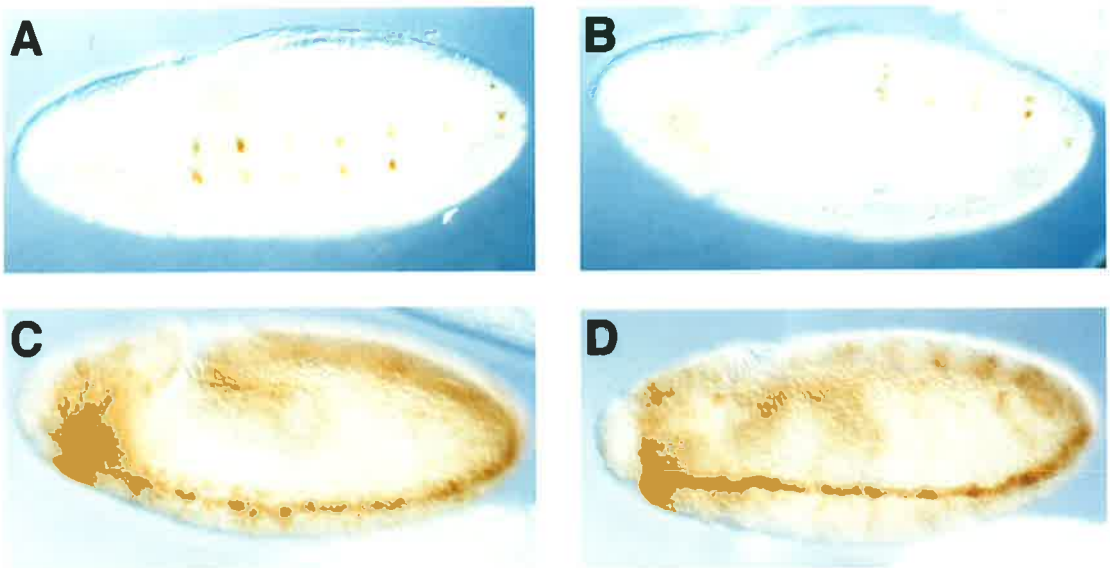
Figure 4-3

A comparison of the expression patterns of HZ*w*(K)B0.5 (A, C,& E) and HZ*w*(K)P0.5/1.3 (B, D,& F) detected by ß-galactosidase antibody staining. All panels show lateral views with anterior to the left and dorsal up unless stated otherwise. **A** (slightly dorso-lateral) and **B**, embryos during slow germband elongation (stage 9) showing the reiterated lateral stripes along the germband, that may correspond to a cycle 15 equivalent of domain 11, and staining in the ventral neurogenic region (faint in A) that represents a subset of cells of domain M and N during cycle 14. **C**, at the beginning of germband shortening (stage11) expression is faint but visible in the ventral neurogenic region. **D**, a germband shortening embryo (stage 12) showing significantly stronger expression in all of the developing neural tissues than is seen in HZ*w*(K)0.5. **E** and **F**, towards the end of germband shortening (stage 13), the intensity of staining in the constuct HZ*w*(K)P0.5/1.3 is far greater than in HZ*w*(K)B0.5.

Figure 4-4

The unique expression patterns detected by ß-galactosidase antibody staining in a single independent line of HZ*w*(K)P0.5/1.3 (A & B) and HZ*w*(K)B0.5 (C & D). All panels show lateral views with anterior to the left and dorsal up unless described otherwise. **A**, a ventral view and **B**, embryos during rapid germband elongation (stage 8) showing staining in what appears to be cycle 14, domain 25. **C** and **D**, embryos during rapid germband elongation (stage 8) showing staining that correlates with cycle 14, domain 14.


A
B
C
D
E
F


A
B
C
D

of these lines. Furthermore, the characteristic later patterns of expression seen in these constructs (described above) are unaffected.

(iii) The upstream fragment alone

The construct HZ*w*BX2.5/5.3 was tested for any expression that was not detected in the previous round of constructs that spanned this region (detailed in section 3-4.2). No early patterns were detected. During germband retraction (stage 12), a distinct, reiterated pattern of cells along the germband begin to express (figure 4-5A), and as retraction proceeds, it can be seen that these cells form part of the developing PNS. Weak expression in this subset of PNS cells continues throughout germband retraction and disappears during stage 14, a few cells in the head region, that do not appear to be part of the PNS, still express *lacZ* at this stage (figure 4-5C&E). Presumably expression with this upstream region was not previously identified because of the lack of sensitivity associated with the ß-galactosidase detection method, although it is possible that the fragment used in the construct HZ8.3 (see figure 3-6), which completely covers the 2.5/5.3, also contains some sequences that prevented this transcriptional activation.

(iv) Combining the proximal and upstream fragments

The intron was tested for basal elements that are required to act in combination with upstream enhancers in the construct HZ*w*BcX1.3/2.5/5.3. RNA *in situ* hybridisations revealed the same, reiterated lateral pattern of expression in the developing PNS, that was observed in the construct HZ*w*BX2.5/5.3 (figure 4-5B compared with A), as well as some expression in the CNS. The later neural pattern showed slightly stronger expression in the PNS and expression in the CNS (figure 4-5D&F). This subtle change in the expression pattern upon addition of the intron fragment may be a result of providing proximal intron sequences that contain necessary basal elements. However, without re-testing the intron fragment alone for faint neural expression, particularly in the CNS, this cannot be determined.

In contrast to the subtle effects of the intron sequence, the addition of the proximal 500bp fragment to the 2.5/5.3 did result in a dramatically different expression pattern when compared to the HZ*w*BX2.5/5.3 construct. A consistent early pattern of expression was obtained in the construct HZ*w*(K)B0.5/2.5/5.3. Using RNA *in situ* hybridization, this was first visible just prior to gastrulation (about 180 minutes after egg deposition (AED) at 25^{o}C) and became significantly stronger during gastrulation. Expression was detected in a region just anterior to the cephalic furrow, and in another region, more anterior and dorsally, corresponding to domains 1 and 2 respectively (figure 4-6A). As germband extension progressed, these domains remained visible and became strong and discrete again by the end of rapid germband elongation (stage 8), suggesting that they are also transcriptionally activated during cycle 15 (figure 4-6B,C&D).

In addition to these two early domains, a distinct pattern of expression was observed in the ventral neurogenic region during slow germband elongation HZ*w*BcX1.3/2.5/5.3 which appears to represent a subset of the cycle 14 cells of domains N (iVN) and M (mVN)(figure 4-6D&E). These cells appear to be the early dividing neuroblasts, the timing of which, is also regulated by *stg* transcription (K. Weigmann and C. F. Lehner 1995). As development proceeds, an increasing number of cells in this ventral neurogenic region express *lacZ* (figure 4-6F), suggesting that some of the later dividing neuroblasts of the cycle 14 may also be represented. Expression in the VN continues throughout the

window of neural cell proliferation (figure 4-6G to J), suggesting that subsequent division cycles in domains N and M are also represented in this construct.

During gnathal and clypeolabral lobe formation (stage 10), while expression is still visible in the VN, a reiterated pattern of lateral patches in the dorsal ectoderm, corresponding to the tracheal placodes becomes apparent (figure 4-6G&H). The expression in these cells precedes invagination of the tracheal pits and has been estimated to be an early cycle 16 domain (B. A. Edgar *et al.,* 1994), as the 16th mitosis takes place during this invagination (V. Hartenstein *et al.,* 1994).

These data indicate that the 500 bp fragment contains sequences that are required in combination with the 2.5/5.3 for expression in some cycle 14 domains, and that the 1.3kb intron fragment does not. The possibility remained, however, that *stg* transcription in other domains may require both the 500bp and intron proximal sequences. This was tested in the construct HZ*w*(K)B0.5/1.3/2.5/5.3. No further cycle 14 domains or later patterns were identified in this construct that were not observed in the construct HZ*w*(K)B0.5/2.5/5.3 (figure 4-7 compared to 4-6, although expression in the HZ*w*(K)B0.5/1.3/2.5/5.3 line shown is stronger than in the HZ*w*(K)B0.5/2.5/5 line). However, some subtle variations between the neural patterns of these two constructs may only be detected by undertaking further analysis of these expression patterns.

(v) Retesting the 8.7kb construct

Somewhat surprisingly, when the construct HZ*w*(K)B8.7 was tested, using RNA *in situ* hybridization, expression was detected in cycle 14 domains 1, 2, and 21 as well as in the tracheal placodes and the ventral neurogenic region (figure 4-8). These early domains were not previously identified by ß-galactosidase staining of the construct HZ(K)B8.7 (see section 3-1).

This raises the question of why these expression patterns were not originally detected. In fact, the HZ(K)B8.7 embryos did show these expression patterns but the artefactual mesodermal expression in this construct, which was initially thought to be domain 10, lead to their mis-assignment to cycle 15. This "domain 10" expression was visible well before that of domains 1 and 2 and it was therefore assumed that the domain 1 and 2 patterns must have been part of cycle 15. The domain 21 pattern was then also assigned to cycle 15 given its relative timing of expression to domains 1 and 2. Another factor in the mis-assignment resulted from the use of the ß-galactosidase assay to detect patterns of expression rather than the *lacZ* mRNA *in situ* hybridization. The, approximately, 20 minute delay between *lacZ* transcription and its subsequent translation makes a significant difference to the stage at which these patterns can be observed. The removal of the artefactual expression in the construct HZ*w*(K)B8.7 and the use of the RNA *in situ* hybridization technique made it clear that these were in fact cycle 14 domains.

Domain 21 occurs as reiterated blocks of cells along the lateral VN that are visible during rapid germband elongation (figure 4-8B). According to Foe (1989), these blocks are not present in the abdominal segments of domain N (V. E. Foe 1989), Hartenstein *et al.*,(1994) however, suggest that they are present in all segments of the lVN (V. Hartenstein *et al.,* 1994). If the pattern observed in these constructs does correspond to domain 21, then it appears that this domain is present in the abdominal as well as thoracic segments.

Expression in the VN, in the construct HZ*w*(K)B8.7, is first visible during gnathal and clypeolabral lobe formation (stage 10), at the same time as the tracheal placodes become visible (figure 4-8C to E). This is significantly later than the expression seen in the construct HZ*w*(K)B0.5/2.5/5.3, suggesting

Figure 4-5

A comparison of *lacZ* mRNA expression patterns in the constructs HZ*w*BX2.5/5.3 (A,C&E) and HZ*w*BcX1.3/2.5/5.3 (B,D&F). All views are lateral. **A**, during germband shortening (stage 12) in a HZ*w*BX2.5/5.3 embryo, expression begins in the cells of the developing PNS. **B**, at a similar stage in a HZ*w*BcX1.3/2.5/5.3 embryo, staining can be seen in these same cells, as well as in the developing CNS. **C**, by stage 13, the faint expression in the cells of the PNS can be clearly distinguished in a HZ*w*BX2.5/5.3 embryo while in a HZ*w*BcX1.3/2.5/5.3 embryo (late stage 12, early stage 13, **D**), expression is visible in these same PNS cells as well as in the CNS. **E** and **F**, by stage 14, expression has disappeared from the PNS in both constructs leaving only a few expressing cells in the head region in HZ*w*BX2.5/5.3 while HZ*w*BcX1.3/2.5/5.3 continues to express *lacZ* in the cells of the ventral nerve cord and brain.

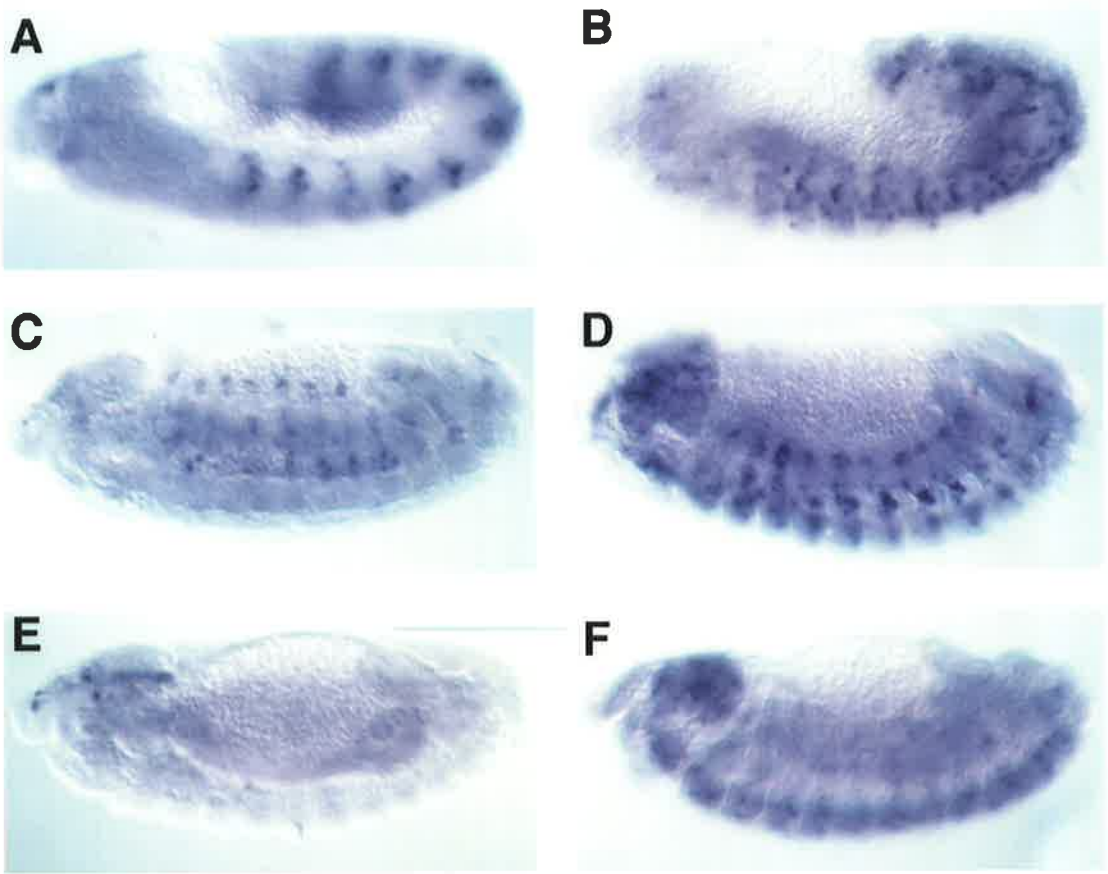

A
B
C
D
E
F

Figure 4-6

lacZ mRNA expression patterns in a strain carrying the construct HZ*w*(K)X0.5/2.5/5.3. All views are lateral unless described otherwise. **A**, expression is first visible prior to gastrulation (stage 5) in the regions corresponding to domains 1 and 2. Faint dorsal stripes are also apparent along the embryo similar to those seen in the construct HZ*w*(K)B0.5 (compare to figure 3-3B). **B**, as germband extension begins (stage 7) the intensity of staining in these tissues has increased and the size of domain 1 suggests that it has undergone mitosis. Domain 2 has probably also divided but with the formation of the cephalic furrow, some of it has been drawn into the interior of the embryo. **C**, expression in domains 1 and 2 is still visible towards the end of rapid germband elongation (stage 8). **D**, the domain 1 and 2 patterns persist during slow germband elongation (stage 9) suggesting that these enhancers are also responsible for the cycle 15 mitosis in these domains. Expression also begins in the ventral neurogenic region. **E**, a ventral view of an embryo at the same stage as in D, showing the distinct cells within the region of domains N and M (probably the neuroblasts) that are expressing *lacZ*. **F**, a ventro-lateral view of a slightly older embryo than that in E (late stage 9), showing additional cells from within domains N and M also expressing *lacZ*. **G** and **H** (dorsal view), early in gnathal and clypeolabral lobe formation (stage 10), the pattern of expression in the ventral neurogenic region has become more complex and expression now begins in the tracheal placodes. **I**, expression is still visible in the ventral neurogenic region as germband shortening begins (late stage 11), as well as in the developing PNS and brain lobes. **J**, in the germband retracted embryo (stage 14) expression can be seen in the tissues of the central and peripheral nervous system.

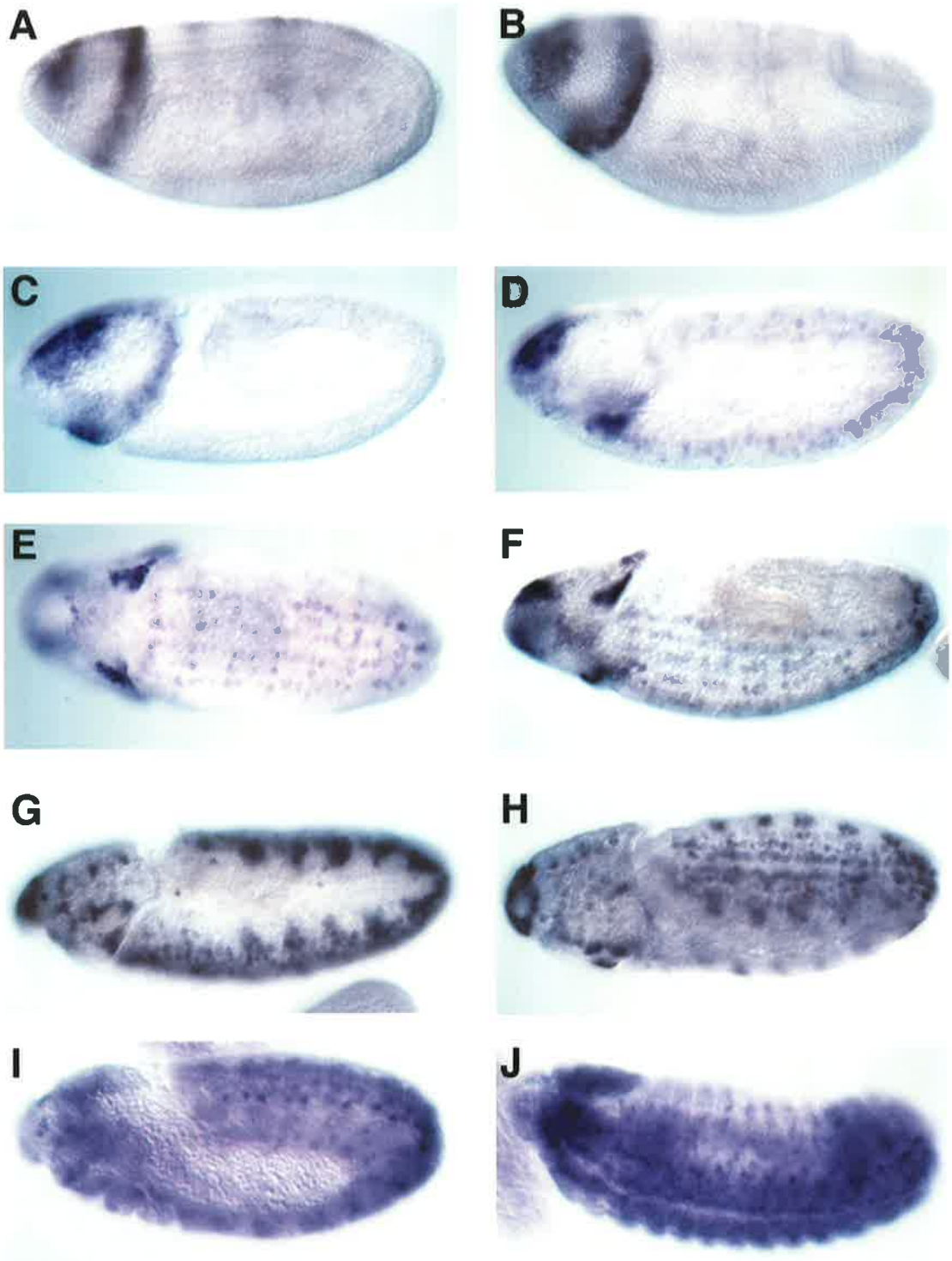

A
B
C
D
E
F
G
H
I
J

Figure 4-7

lacZ mRNA expression patterns in a strain carrying the construct HZ*w*(K)X0.5/1.3/2.5/5.3. All views are lateral unless described otherwise. **A**, during rapid germband extension (stage 8), expression can be seen in domains 1 and 2. **B**, during slow germband elongation (stage 9) expression can be seen in domains 1 and 2 as well as in the ventral neurogenic region in the same pattern as seen in HZ*w*(K)B0.5/2.5/5.3 (figure 4-6E). **C** and **D**, show slightly later stages during slow germband elongation, as more cells in the ventral neurogenic region express *lacZ*. **E**, a ventral view of an embryo similarly staged to that in D. **F** and **G**, are ventral views during gnathal and clypeolabral lobe formation (stage 10), showing the increasingly complex expression in the ventral neurogenic region and expression in the tracheal placodes. **H**, a germband retracted embryo (stage 14), showing strong expression in the tissues of the central and peripheral nervous system.

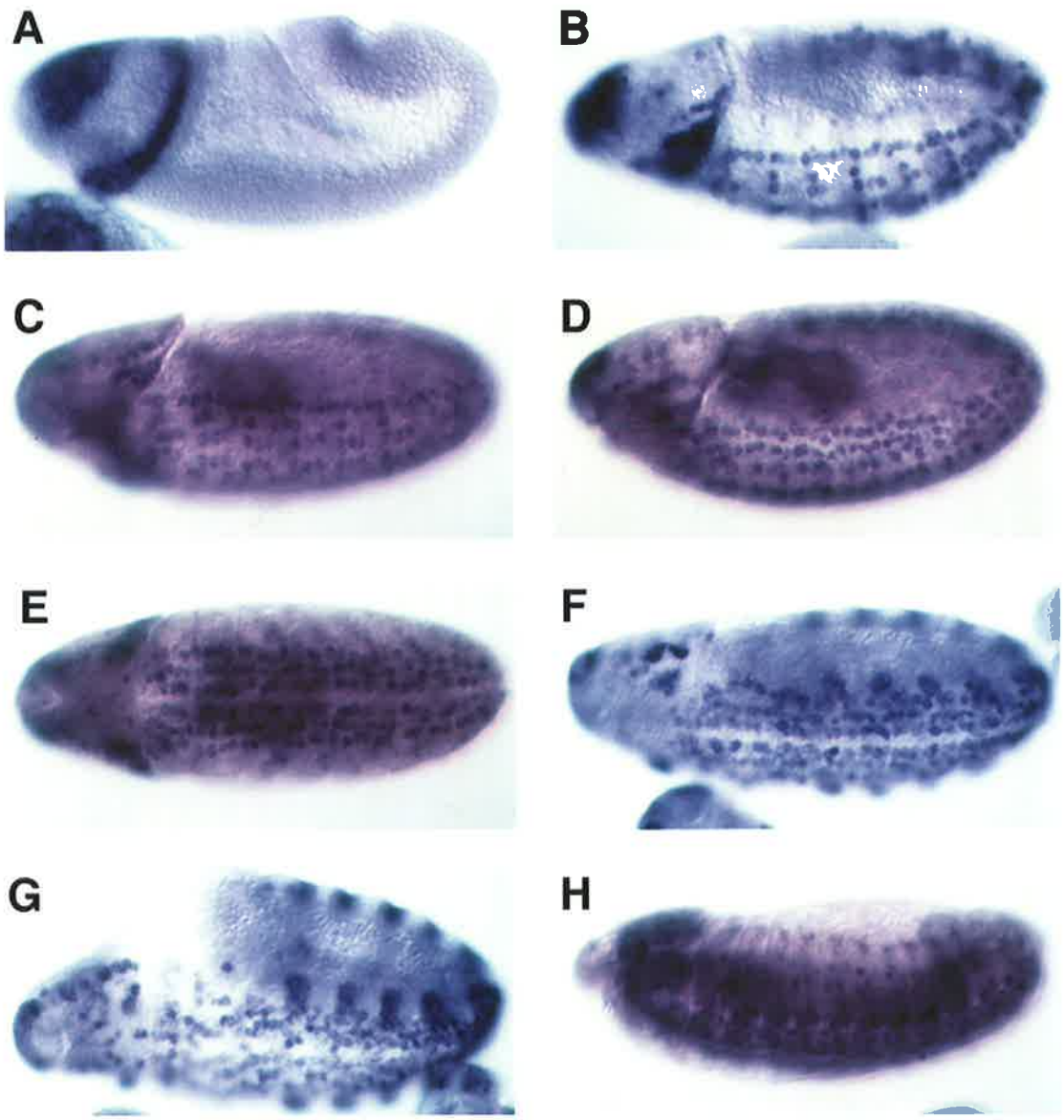
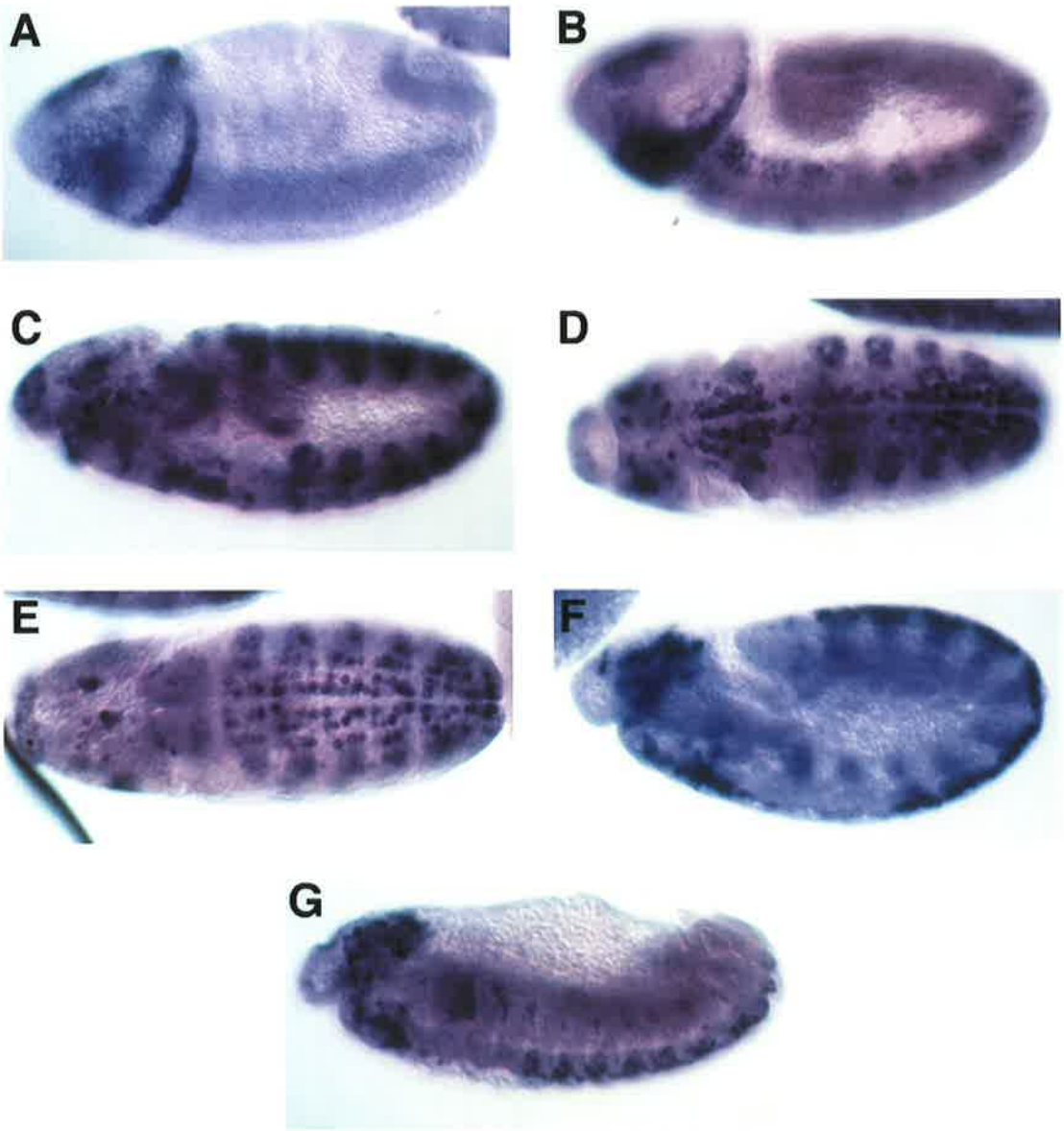

A
B
C
D
E
F
G
H

Figure 4-8

lacZ mRNA expression patterns in a strain carrying the construct HZ*w*(K)8.7. All views are lateral unless described otherwise. **A**, during gastrulation (stage 7) expression can still be seen in domains 1 and 2. **B**, as rapid germband elongation proceeds (stage 8) expression also becomes visible in reiterated blocks along the germband which correlate with domain 21 of cycle 14. **C**, **D** (ventral), and **E** (dorsal), during gnathal and clypeolabral lobe formation (stage 10), expression in the ventral neurogenic region and the tracheal placodes begins almost simultaneously. This VN expression may represent a subset of the cells of domain M during cycle 14 and domain N during cycle 15. In D, the reduced VN expression in the thoracic segments can be seen. **F**, prior to germband shortening (stage 11), some expression can still be seen in the tracheal pits (out of plane of focus) and in the ventral nerve cord, expression is still absent in the thoracic region of the VN. **G**, towards the end of germband shortening (stage 12) expression can be seen in the brain, part of the ventral nerve cord and in a very small subset of the PNS.


A
B
C
D
E
F
G

of these lines. Furthermore, the characteristic later patterns of expression seen in these constructs (described above) are unaffected.

(iii) The upstream fragment alone

The construct HZ*w*BX2.5/5.3 was tested for any expression that was not detected in the previous round of constructs that spanned this region (detailed in section 3-4.2). No early patterns were detected. During germband retraction (stage 12), a distinct, reiterated pattern of cells along the germband begin to express (figure 4-5A), and as retraction proceeds, it can be seen that these cells form part of the developing PNS. Weak expression in this subset of PNS cells continues throughout germband retraction and disappears during stage 14, a few cells in the head region, that do not appear to be part of the PNS, still express *lacZ* at this stage (figure 4-5C&E). Presumably expression with this upstream region was not previously identified because of the lack of sensitivity associated with the ß-galactosidase detection method, although it is possible that the fragment used in the construct HZ8.3 (see figure 3-6), which completely covers the 2.5/5.3, also contains some sequences that prevented this transcriptional activation.

(iv) Combining the proximal and upstream fragments

The intron was tested for basal elements that are required to act in combination with upstream enhancers in the construct HZ*w*BcX1.3/2.5/5.3. RNA *in situ* hybridisations revealed the same, reiterated lateral pattern of expression in the developing PNS, that was observed in the construct HZ*w*BX2.5/5.3 (figure 4-5B compared with A), as well as some expression in the CNS. The later neural pattern showed slightly stronger expression in the PNS and expression in the CNS (figure 4-5D&F). This subtle change in the expression pattern upon addition of the intron fragment may be a result of providing proximal intron sequences that contain necessary basal elements. However, without re-testing the intron fragment alone for faint neural expression, particularly in the CNS, this cannot be determined.

In contrast to the subtle effects of the intron sequence, the addition of the proximal 500bp fragment to the 2.5/5.3 did result in a dramatically different expression pattern when compared to the HZ*w*BX2.5/5.3 construct. A consistent early pattern of expression was obtained in the construct HZ*w*(K)B0.5/2.5/5.3. Using RNA *in situ* hybridization, this was first visible just prior to gastrulation (about 180 minutes after egg deposition (AED) at 25^{o}C) and became significantly stronger during gastrulation. Expression was detected in a region just anterior to the cephalic furrow, and in another region, more anterior and dorsally, corresponding to domains 1 and 2 respectively (figure 4-6A). As germband extension progressed, these domains remained visible and became strong and discrete again by the end of rapid germband elongation (stage 8), suggesting that they are also transcriptionally activated during cycle 15 (figure 4-6B,C&D).

In addition to these two early domains, a distinct pattern of expression was observed in the ventral neurogenic region during slow germband elongation HZ*w*BcX1.3/2.5/5.3 which appears to represent a subset of the cycle 14 cells of domains N (iVN) and M (mVN)(figure 4-6D&E). These cells appear to be the early dividing neuroblasts, the timing of which, is also regulated by *stg* transcription (K. Weigmann and C. F. Lehner 1995). As development proceeds, an increasing number of cells in this ventral neurogenic region express *lacZ* (figure 4-6F), suggesting that some of the later dividing neuroblasts of the cycle 14 may also be represented. Expression in the VN continues throughout the

window of neural cell proliferation (figure 4-6G to J), suggesting that subsequent division cycles in

while expression is still visible in the VN, a

Pages 42 & 43 are duplicated

n, corresponding to the tracheal placodes

these cells precedes invagination of the tracheal pits and has been estimated to be an early cycle 16 domain (B. A. Edgar *et al.*, 1994), as the 16th mitosis takes place during this invagination (V. Hartenstein *et al.*, 1994).

These data indicate that the 500 bp fragment contains sequences that are required in combination with the 2.5/5.3 for expression in some cycle 14 domains, and that the 1.3kb intron fragment does not. The possibility remained, however, that *stg* transcription in other domains may require both the 500bp and intron proximal sequences. This was tested in the construct HZ*w*(K)B0.5/1.3/2.5/5.3. No further cycle 14 domains or later patterns were identified in this construct that were not observed in the construct HZ*w*(K)B0.5/2.5/5.3 (figure 4-7 compared to 4-6, although expression in the HZ*w*(K)B0.5/1.3/2.5/5.3 line shown is stronger than in the HZ*w*(K)B0.5/2.5/5 line). However, some subtle variations between the neural patterns of these two constructs may only be detected by undertaking further analysis of these expression patterns.

(v) Retesting the 8.7kb construct

Somewhat surprisingly, when the construct HZ*w*(K)B8.7 was tested, using RNA *in situ* hybridization, expression was detected in cycle 14 domains 1, 2, and 21 as well as in the tracheal placodes and the ventral neurogenic region (figure 4-8). These early domains were not previously identified by ß-galactosidase staining of the construct HZ(K)B8.7 (see section 3-1).

This raises the question of why these expression patterns were not originally detected. In fact, the HZ(K)B8.7 embryos did show these expression patterns but the artefactual mesodermal expression in this construct, which was initially thought to be domain 10, lead to their mis-assignment to cycle 15. This "domain 10" expression was visible well before that of domains 1 and 2 and it was therefore assumed that the domain 1 and 2 patterns must have been part of cycle 15. The domain 21 pattern was then also assigned to cycle 15 given its relative timing of expression to domains 1 and 2. Another factor in the mis-assignment resulted from the use of the ß-galactosidase assay to detect patterns of expression rather than the *lacZ* mRNA *in situ* hybridization. The, approximately, 20 minute delay between *lacZ* transcription and its subsequent translation makes a significant difference to the stage at which these patterns can be observed. The removal of the artefactual expression in the construct HZ*w*(K)B8.7 and the use of the RNA *in situ* hybridization technique made it clear that these were in fact cycle 14 domains.

Domain 21 occurs as reiterated blocks of cells along the lateral VN that are visible during rapid germband elongation (figure 4-8B). According to Foe (1989), these blocks are not present in the abdominal segments of domain N (V. E. Foe 1989), Hartenstein *et al.*,(1994) however, suggest that they are present in all segments of the lVN (V. Hartenstein *et al.*, 1994). If the pattern observed in these constructs does correspond to domain 21, then it appears that this domain is present in the abdominal as well as thoracic segments.

Expression in the VN, in the construct HZ*w*(K)B8.7, is first visible during gnathal and clypeolabral lobe formation (stage 10), at the same time as the tracheal placodes become visible (figure 4-8C to E). This is significantly later than the expression seen in the construct HZ*w*(K)B0.5/2.5/5.3, suggesting

that a different and later subset of domains M and N are represented here. It is also possible that the more lateral VN expression, presumably within domain N, may be part of cycle 15 rather than 14, since most of the cells in this domain divide during slow germband elongation (stage 9). This VN expression may represent the same subsets that are seen in the construct HZ*w*(K)B0.5.

One interesting feature of this construct is the absence of CNS expression specifically in the thoracic segments. This is first evident when expression begins in the VN during slow germband elongation (figure 4-8D) and can be clearly discerned, later in development, as a gap in expression along the ventral nerve cord (figure 4-8F&G). Expression at this late stage is also apparent in the brain, and a very limited amount is also detectable within the PNS (figure 4-8H).

]

4-2.4 The role of proximal sequences in *string* transcription

	cycle 14	cycles 15-16	late neural
0.5	(domain 14) M subset?	domain 11? M & N subsets?	weak CNS weak PNS
0.5/1.3	(domain 25) M subset?	domain 11 M & N subsets?	strong CNS strong PNS
2.5/5.3			very weak PNS
1.3/2.5/5.3			weak PNS some CNS
0.5/2.5/5.3	domains 1 & 2 N and M neuroblasts	domains 1 & 2 N and M neuroblasts? tracheal placodes	CNS PNS
0.5/1.3/2.5/5.3	domains 1 & 2 N and M neuroblasts	domains 1 & 2 N and M neuroblasts? tracheal placodes	CNS PNS
8.7	domains 1, 2, & 21 M subset	domains 1 & 2 M &N subsets?	partial CNS very little PNS

Table 4-1 A summary of expression patterns.
This table summarises the information generated from the constructs tested in section 4-2.3, for more detailed information please refer to this section. Each construct is listed by its component fragments. Expression patterns have been divided into those corresponding to cycle 14 domains, domains of cycles 15 and 16, and later neural patterns. A question mark indicates some degree of uncertainty about the domain allocation, and brackets indicate uncertainty about the presence of a domain.

From the expression patterns described in section 4-2.3 and the table 4-1, it is clear that combining proximal fragments with upstream fragments can have a dramatic effect on the pattern of transcriptional activation. In particular, the presence of the proximal 500bp fragment is required for transcriptional activation in many distinct patterns.

In contrast to this, the 1.3kb intron appears to have little effect on patterns of transcriptional activation in other fragments. In constructs that included the intron sequence, slight differences were detected in the late neural patterns which suggest that this sequence is having some effect on expression. The intron fragment alone has been previously tested as the construct HZ1.3, described in section 3-4.2, and shows no expression. However, it is possible that a faint expression pattern

could have been missed, using the ß-galactosidase assay detection method, and that this pattern is now being detected in the constructs containing the intron sequence. The only intron construct that did show a distinct increase in the neural expression pattern was HZ*w*(K)P0.5/1.3, which had strong PNS and CNS expression, whereas in HZ*w*(K)B0.5 both were relatively weak. However, the presence of the 500bp proximal fragment in this construct suggests that it may actually be this 500bp sequence that is required, as in the other constructs, for the transcriptional activation of enhancers present within the intron fragment.

The additional patterns that are observed when upstream fragments are added to the 500bp proximal fragments are not present when these fragments are tested alone. It is therefore the combination of these upstream sequences and the 500bp fragment that generate the new patterns of expression. The common requirement for the 500bp fragment and its proximal location suggest that it may contain basal sequences with which the domain specific enhancers, present in the upstream fragments, must interact to bring about transcriptional activation.

4-2.5 Further constructs using the 500bp proximal fragment

(i) Background

The identification of some cycle 14 mitotic domains in constructs that included the proximal 500bp fragment suggested that it would be worthwhile to retest other upstream fragments that had previously been thought not to contain domain specific enhancers (see section 3-4.2). Two upstream fragments were selected such that, in addition to the previous constructs, the complete genomic region up to -17kb of the *stg* transcript would be retested. The first of these spanned the 0.7kb fragment that was missing from within the 2.5/5.3 constructs (see figure 4-9) and extended half way into the 2.5kb BamHI fragment that gave expression in domains 1 and 2 and the tracheal placodes. The other fragment extended a further 2.1kb upstream of the 5.3 (see figure 4-9).

(ii) Construct generation

A 2.3kb EagI to SpeI fragment was generated that covered the 0.7kb fragment. This was subcloned from the larger clone p*stg*RSp3.8, which was isolated from the phage *Dr*L3, by digesting with EagI which cuts at the internal EagI site and also in the polylinker just outside of the SpeI site. This made the clone p*stg*ESp2.3. The 500bp fragment was released from the clone p*stg*(K)B0.5 using KpnI and EcoRI (a polylinker site) and inserted into the polylinker of p*stg*ESp2.3 to make the construct p*stg*(K)Sp0.5/2.3. This construct was then released using KpnI and a partial NotI digest (to avoid a repeated NotI polylinker site between the 0.5 and 2.3 fragments) and then inserted into the HZ50PL*w* vector KpnI to NotI.

A 2.8kb HindIII to SacI fragment was isolated from the phage clone *Dr*L2 which overlapped with the previous construct 2.5/5.3 by about 0.7kb and extended a further 2.1kb upstream. The SacI site is an artificial vector site that coincides with the upstream end of the phage *Dr*L2 and is not present within the *stg* sequence. This SacI site was blunted using T4 DNA polymerase to remove the 3' overhang before digesting with HindIII and inserting into the clone p*stg*(K)Rsa0.5, SmaI to HindIII, making the construct p*stg*(KS)0.5/2.8. The clone p*stg*(K)Rsa0.5 was derived from p*stg*(K)B0.5 by releasing the 0.5kb fragment with KpnI and RsaI, which cuts 16bp inside the BamHI site (see figure 4-11). This fragment was then inserted into pBST KpnI to HincII to increase the number of possible cloning sites

Figure 4-9

A schematic diagram showing the genomic regions that were included in constructs described in section 4-2.5 and 4-3.2. The genomic map is the same as described in figure 3-5 but also shows an EagI site (E) at about -7.5kb and the *in vitro* mutagenised KpnI site (K) that replaces the TATA box at -30bp. Constructs are named in the same manner as described in figure 4-2.

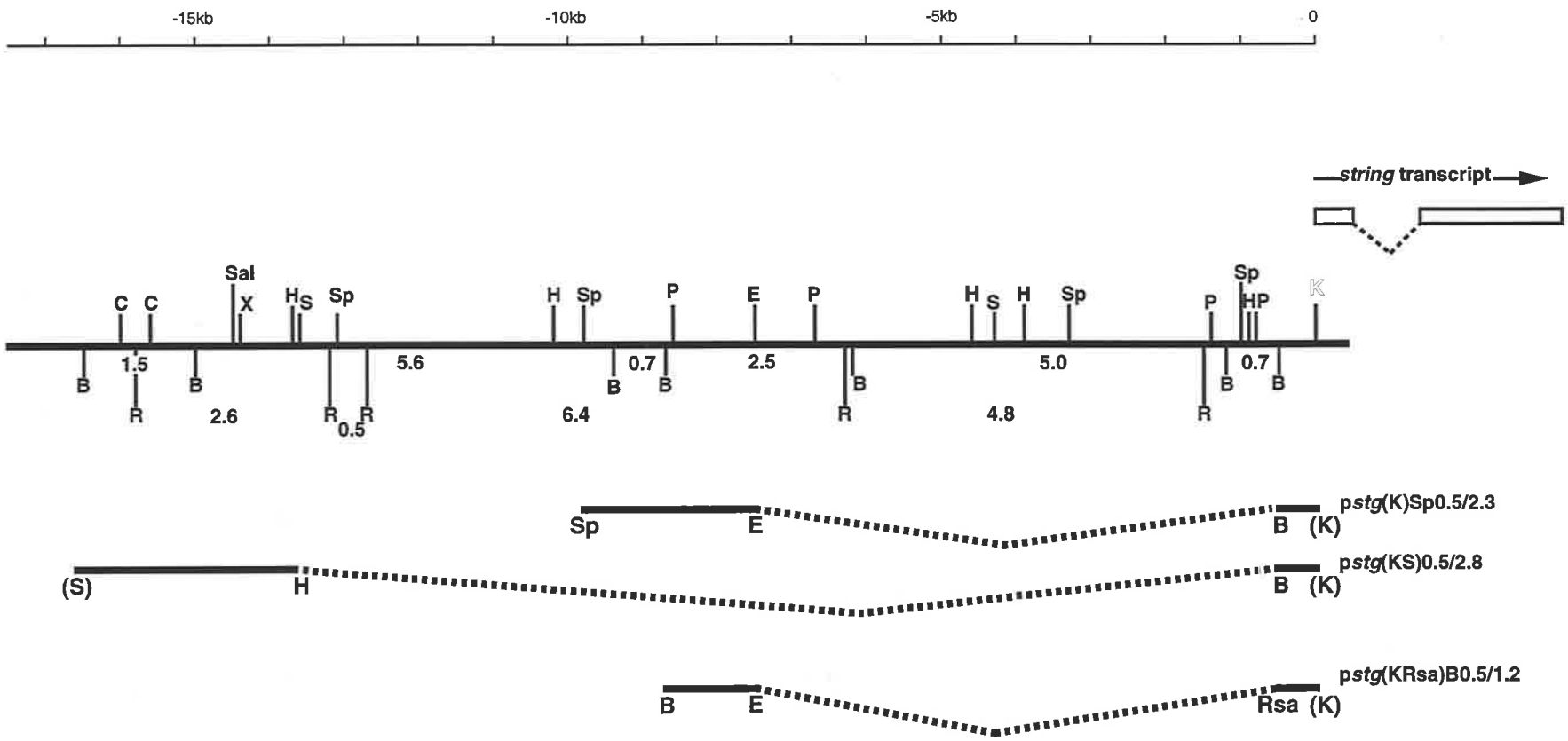
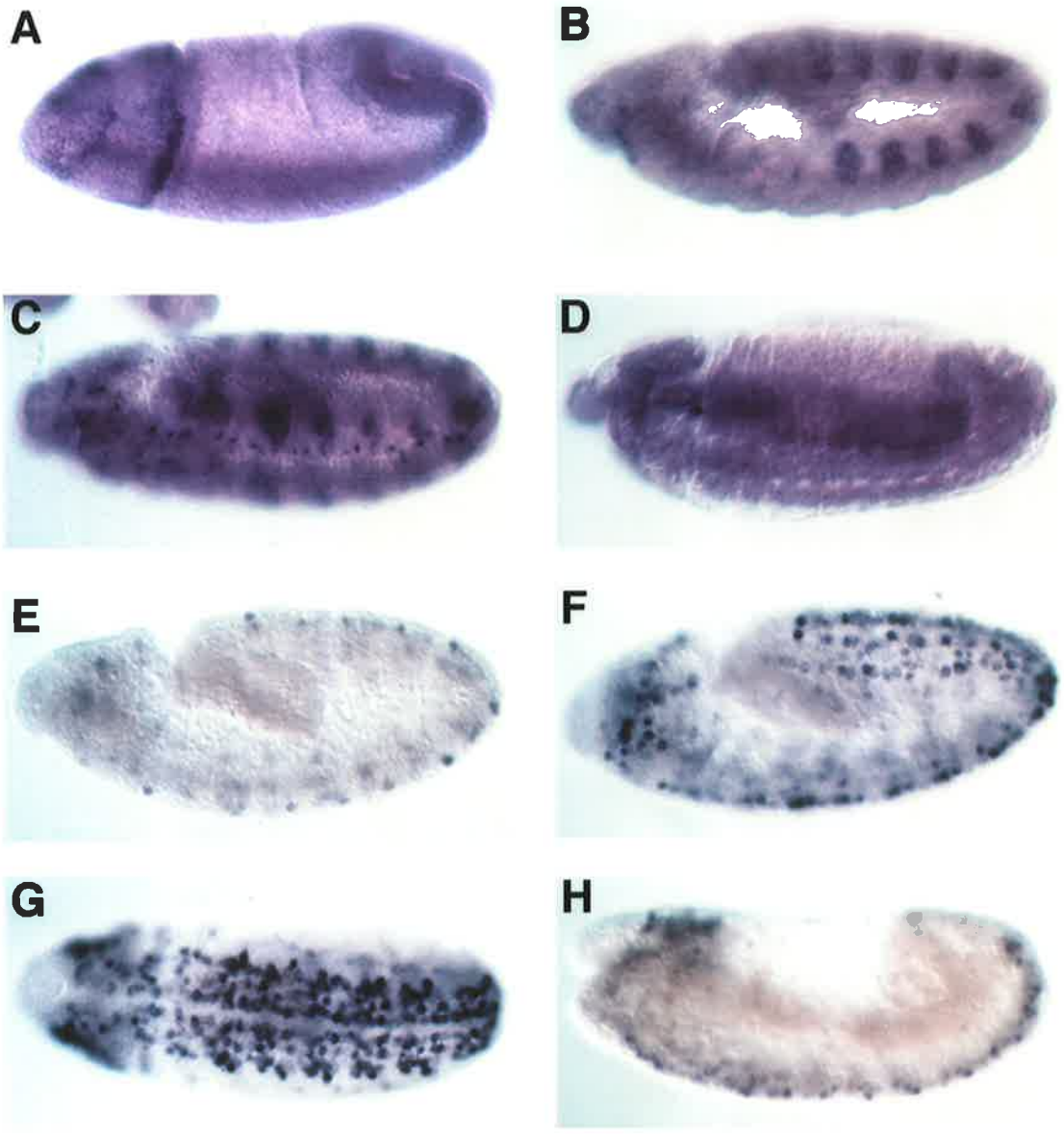

-15kb
-10kb
-5kb
0
string transcript
C C
Sal
X
H S
Sp
H Sp
P
E
P
H S H Sp
P
Sp
H P
K
B
1.5
R
B
2.6
R
R
0.5
5.6
B
0.7
B
6.4
2.5
B
R
5.0
R
B
0.7
B
B
pstg(K)Sp0.5/2.3
Sp
E
B (K)
pstg(KS)0.5/2.8
(S)
H
B (K)
pstg(KRsa)B0.5/1.2
B
E
Rsa (K)

Figure 4-10

lacZ mRNA expression patterns in strains carrying the construct HZ*w*(K)Sp0.5/2.3 (A, B, C, & D) or HZ*w*(KS)0.5/2.8 (E, F, G, & H). All views are lateral unless described otherwise. HZ*w*(K)Sp0.5/2.3: **A**, expression in domains 1 and 2 during early germband extension (stage 7). **B**, expression in the tracheal placodes during gnathal and clypeolabral lobe formation (stage 10). **C**, a slightly later embryo that in B (also stage 10), also showing additional expression in an unidentified set of small cells within the VN. **D**, in a germband shortened embryo (stage 14) there is no specific neural expression detectable. HZ*w*(KS)0.5/2.8: **E**, expression is first detectable at the beginning of slow germband elongation (stage 9) in a subset of domain N or M in cycle 14. **F** and **G** (ventral view), early in stage 10, expression increases in the ventral neurogenic regions. **H**, towards the end of germband retraction (stage 12), expression is visible within a few cells of the ventral nerve cord and brain, this pattern is clearly different to that seen in other constructs.


A
B
C
D
E
F
G
H

upstream of the 0.5kb. The 0.5/2.8 construct was then released using KpnI and NotI and inserted into the HZ50PL*w* vector.

(iii) Analysis of expression patterns

The construct HZ*w*(K)Sp0.5/2.3 showed early expression in domains 1 and 2, and later in the tracheal placodes as was seen in the constructs HZ*w*(K)B8.7 and HZ*w*(K)X 0.5/2.5/5.3 (figure 4-10A&B). Faint expression is also seen in a few cells within the ventral neurogenic region prior to germband shortening, and no late neural expression was apparent (figure 4-10C&D).

The HZ*w*(K)S0.5/2.8 construct revealed no early mitotic domain patterns. Expression was first detected in this construct early in slow germband elongation in the ventral neurogenic region, in what appears to be a different subset of domain N and M to that observed in either the HZ*w*(K)B8.7 or the HZ*w*(K)X 0.5/2.5/5.3 constructs (figure 4-10E,F&G). Late neural expression was present in only a few cells of the CNS (figure 4-10H).

It should be noted that a slightly smaller proximal fragment was used in this construct (ie. missing 16bp at the upstream end of the 500bp). However, other constructs using this proximal fragment were still found to express in domains 1 and 2 when the appropriate upstream fragment was inserted (see section 4-3.3), suggesting that this slightly smaller fragment is still sufficient to achieve transcriptional activation.

4-3 ANALYSIS OF THE 500bp PROXIMAL FRAGMENT

4-3.1 Background

The requirement for the 500bp proximal fragment in combination with upstream regions to achieve expression in domains 1, 2, and 21, as well as some later patterns suggests that there are important basal sequences located within this fragment. As a means to further define these sequences a deletion series of the 500bp was generated. These were then tested for their ability to transcriptionally activate in domains 1 and 2.

4-3.2 Construct generation

A 1.2kb BamHI to EagI fragment was subcloned from within the clone p*stg*ESp2.3 (see 4-2.5(ii)) by digesting with BamHI which cuts within the clone and in the polylinker just outside of the EagI site. This fragment is common to the constructs p*stg*(K)B8.7, p*stg*(K)X0.5/2.5/5.3, and p*stg*(K)Sp0.5/2.3 which all gave expression in domains 1 and 2 and the tracheal placodes (see figure 4-13). This 1.2kb fragment was cloned into the polylinker BamHI site of p*stg*(K)Rsa0.5 to make p*stg*(KRsa)B0.5/1.2 (figure 4-9).

The sequence was available for the 500bp fragment (L. O'Keefe, personal communication) which allowed deletions of this fragment to be generated by making use of the complete restriction map. A unique AvaI restriction site at -426bp allowed the isolation of a 396bp *stg* sequence that extended from the introduced KpnI site in the *stg* TATA box up to the AvaI site (figure 4-11). This was done by digesting the clone p*stg*(K)Rsa0.5 with AvaI and using Klenow to end fill the 5' overhang (which regenerates the sequence to -425bp) and then digesting with KpnI to release the 396bp fragment. This was then subcloned into pBST, KpnI to HincII, to make the clone p*stg*(K)A396.

A unique NdeI site at -310bp was used to remove the upstream portion of the 500bp (figure 4-11). The clone p*stg*(K)Rsa0.5 was digested with NdeI and then Klenow was used to end fill the 5'

overhang (which regenerates the sequence to -308bp). The presence of no other NdeI sites in this clone then allowed the upstream region to be 'dropped out' using the polylinker site EcoRV and the clone was then religated to make p*stg*(K)N279.

An MboII restriction site -201bp allowed the isolation of a 179bp fragment that extended downstream of this point to the KpnI site (figure 4-11). The clone p*stg*(K)Rsa0.5 was digested with MboII and KpnI and the 179bp fragment was isolated and subcloned into pBST, KpnI to HincII, to make p*stg*(K)M179.

The 1.2kb EagI to BamHI fragment, isolated as a BamHI fragment from p*stg*ESp2.3 (see above), was subcloned into the BamHI polylinker site of each of the above clones, making the series:

p*stg*(KA)AB396/1.2

p*stg*(KN)B279/1.2

p*stg*(KM)B179/1.2.

These clones were then inserted into HZ50PL*w* using KpnI and an XbaI partial digest to avoid repeated polylinker sites within the clones.

4-3.3 Analysis of deletion constructs

The constructs described in section 4-3.2 were analysed for expression patterns, particularly the presence or absence of the domain 1 and 2 patterns. Firstly, expression in the construct HZ*w*(KRsa)B0.5/1.2 carrying the 16bp deletion from the upstream end of the 500bp was analysed. This construct still showed expression in domains 1 and 2 as well as the later expression in the tracheal placodes, however neural expression was not detectable (figure 4-12A,C&E). This shows that the critical basal sequences for mitotic domain patterns are not located in this small region.

The constructs p*stg*(K)BA396/1.2, p*stg*(K)BN279/1.2, and p*stg*(K)BM179/1.2 failed to express the *lacZ* reporter gene at any stage of development (figure 4-12B,D&F). This was not due to any technical problems, as side by side positive controls detected expression in other constructs . This indicates that there are critical sequences located between -509 and -425 that are required for transcriptional activation in domains 1 and 2.

A closer look at the sequence of this 84bp region revealed no identifiable sequences apart from a TATA box at -501bp (see figure 4-11). It seems possible, then, that this TATA box may be the critical element within the 500bp fragment, as the other *stg* TATA box at -30bp is specifically mutated in all of the constructs (see figure 4-11). However, it is also possible that other unidentifiable sequences within this 84bp fragment are responsible for the transcriptional activation in domains 1 and 2.

4-4 DISCUSSION

4-4.1 The proximal 500bp

Many transcriptional analyses that have used a *lacZ* reporter gene approach have used minimal promoter sequences from the gene of interest, fused to *lacZ*, rather than the *hsp70* minimal promoter. However, the *hsp70* sequence has been found to be adequate for the promoter analysis of many genes including;*dpp* (J.-D. Huang *et al.,* 1993), *eve* (S. Small *et al.,* 1991), and *ftz* (Y. Hiromi and W. J. Gehring 1987), which is why it was utilized in the initial enhancer analysis of *stg*.

The requirement for the proximal 500bp fragment, or at least the most 5' portion of this fragment in the constructs described here leads one to question why the *hsp70* promoter region, provided in the

Figure 4-11

Sequence of the proximal 500bp region showing the restriction sites used to generate the deletions of this region. The exact position of the *in vitro* mutagenised KpnI site is also shown. In the constructs described in this chapter, the *stg* sequence thus begins at -31bp. Arrows indicate the position to which the genomic sequence is maintained when using these restriction sites (the MboII enzyme cuts 7bp upstream of its recognition sequence).

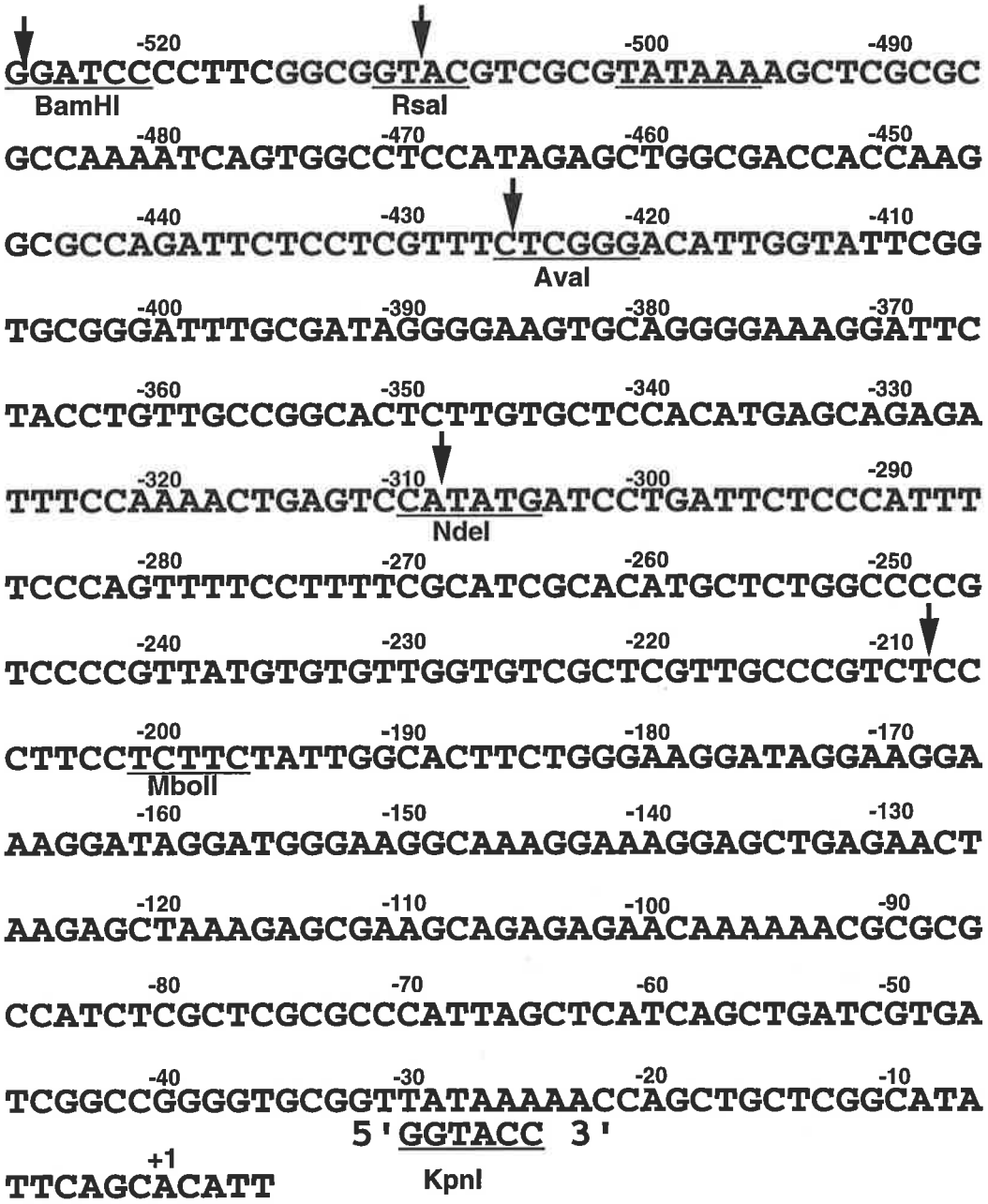
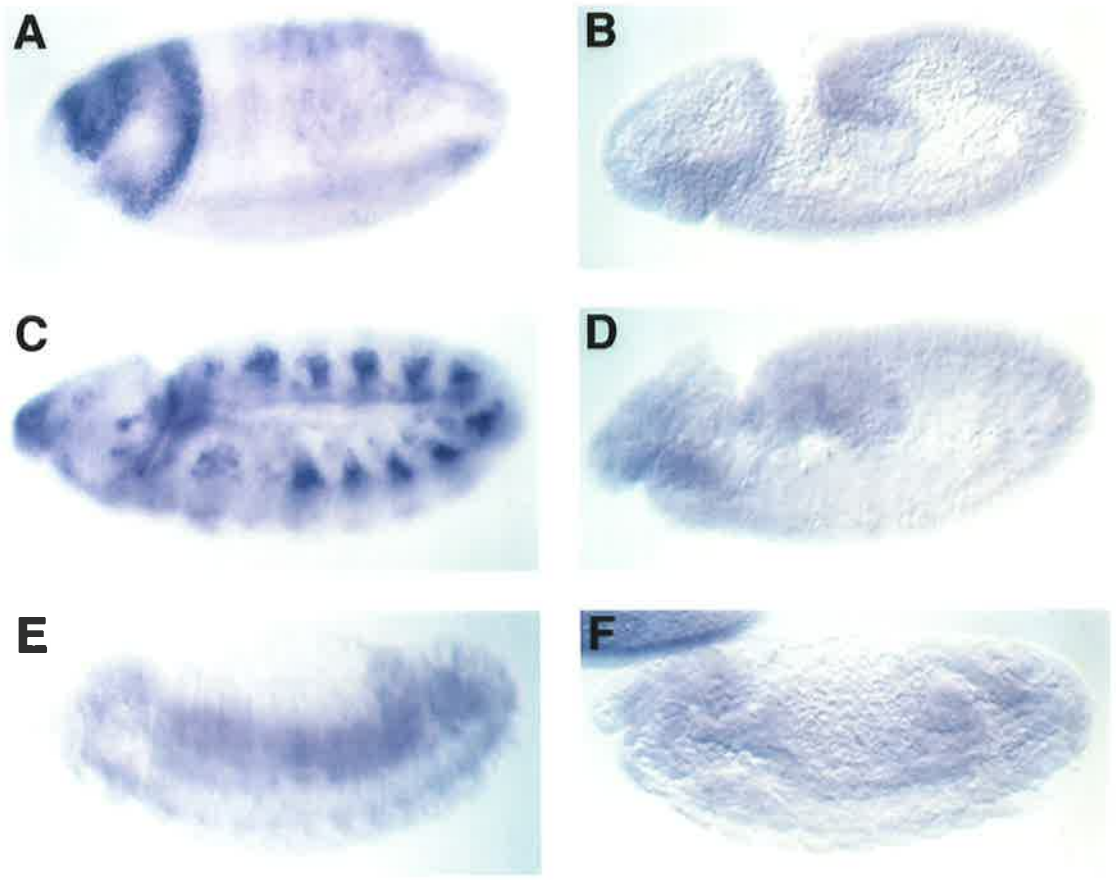

-520
-500
-490
GGATCCCCTTCGGCGGTACGTCGCGTATAAAAGCTCGCGC
BamHI
RsaI
-480
-470
-460
-450
GCCAAAATCAGTGGCCTCCATAGAGCTGGCGACCACCAAG
-440
-430
-420
-410
GCGCCAGATTCTCCTCGTTTCTCGGGACATTGGTATTCGG
AvaI
-400
-390
-380
-370
TGCGGGATTTGCGATAGGGGAAGTGCAGGGGAAAGGATTC
-360
-350
-340
-330
TACCTGTTGCCGGCACTCTTGTGCTCCACATGAGCAGAGA
-320
-310
-300
-290
TTTCCAAAACTGAGTCCATATGATCCTGATTCTCCCATTT
NdeI
-280
-270
-260
-250
TCCCAGTTTTCCTTTTCGCATCGCACATGCTCTGGCCCCG
-240
-230
-220
-210
TCCCCGTTATGTGTGTTGGTGTCGCTCGTTGCCCGTCTCC
-200
-190
-180
-170
CTTCCTCTTCTATTGGCACTTCTGGGAAGGATAGGAAGGA
MboII
-160
-150
-140
-130
AAGGATAGGATGGGAAGGCAAAGGAAAGGAGCTGAGAACT
-120
-110
-100
-90
AAGAGCTAAAGAGCGAAGCAGAGAGAACAAAAAACGCGCG
-80
-70
-60
-50
CCATCTCGCTCGCGCCCATTAGCTCATCAGCTGATCGTGA
-40
-30
-20
-10
TCGGCCGGGGTGCGGTTATAAAAACCAGCTGCTCGGCATA
5'GGTACC 3'
+1
TTCAGCACATT
KpnI

Figure 4-12

A comparison of *lacZ* mRNA expression patterns in strains carrying the construct HZ*w*(K)Rsa0.5 (A, C, & E) or HZ*w*(K)A396 (B, D, & F). **A** and **B**, expression in domains 1 and 2 is apparent during germband extension only in HZ*w*(K)Rsa0.5. **C** and **D**, during stage 10, expression in the tracheal placodes is apparent only in HZ*w*(K)Rsa0.5. **E** and **F**, germband retracted embryos carrying either construct show no late neural expression.


A
B
C
D
E
F

lacZ fusion gene, is not sufficient to direct *stg* transcription. Sequencing of the *hsp70/lacZ* fusion gene in the vector HZ50PL*w* has confirmed that the promoter sequences are intact (M. Clarkson, personal communication), ruling out the possibility that sequence alterations prevent this basal promoter from functioning. The most likely explanation therefore appears to be that there is specificity in the *stg* promoter that is not present in the *hsp70* promoter. This would not be the first example of this kind. The transcripts *decapentaplegic (dpp)* and a neighbouring gene *out at first (oaf)* have been shown to have independent regulation whereby the enhancers for *dpp* do not activate *oaf*, even though they are located closer to the *oaf* promoter than that of *dpp*. Merli *et al.*,(1996), showed that this independent regulation was achieved by promoter specificity. By replacing, *in vivo*, the *oaf* promoter (from -52 to +47bp) with the *hsp70* promoter (-50 to +65bp) they found that the *oaf* transcript was now responsive to the *dpp* enhancers (C. Merli *et al.*, 1996).

The mechanism of promoter specificity is still unknown and no distinguishing features have been identified between the promoters of *dpp* and *oaf*. However, it seems likely that information within these basal promoter regions of less than 100bp each, allows or prevents the initiation of transcription. Presumably a similar difference may exist between the promoters of *stg* and *hsp70*.

4-4.2 Summary of enhancer regions

(i) Enhancers for cycle 14, 15 and 16 patterns

The addition of the 500bp proximal fragment allowed the identification of several enhancer regions that result in expression in patterns that are likely to reflect *stg* transcription. By comparing these expression patterns and the extent of overlap between different constructs it is possible to define particular expression patterns to specific genomic regions.

The 500bp proximal fragment has its own distinct expression pattern, which suggests that *stg* enhancers are present, within this region, for a subset of domains M and N, for cycle 14 and the later cycles, and the cycle 15 domain 11 (see figure 4-13). However, to observe these expression patterns, a significantly longer detection time was required which suggests that these patterns are only weakly expressed. Whether these patterns are maintained upon the addition of other fragments to the 500bp is difficult to determine given the shorter detection times that were used when other, stronger patterns were detected. These expression domains are still apparent in the construct HZ*w*(K)P0.5/1.3, as this construct has no unique expression patterns until late in development. However, it is also possible that the constructs made using HZ*w*(K)Rsa0.5 rather than HZ*w*(K)B0.5 will show less of this 500bp specific pattern due to the deletion of 16bp (see 4-2.5(ii)).

The constructs HZ*w*(K)B8.7 and HZ*w*(K)X0.5/2.5/5.3 have a 2.5kb BamHI fragment in common, and since both of these constructs are clearly expressed in domains 1 and 2 and in the tracheal placodes it seems likely that the enhancers for these expression domains will be located within this 2.5kb fragment. The remainder of HZ*w*(K)B8.7 covers the region downstream of this 2.5kb and includes the 500bp fragment. The unique expression of this construct in domain 21 suggests that its enhancers will be located within the 5.7kb that sits between the 500bp and the 2.5kb (see figure 4-13). The distinct cycle 14 domain N and M expression that is observed in HZ*w*(K)X0.5/2.5/5.3 suggests that these enhancers are located within the 5.3kb region that is unique to this construct (see figure 4-13).

The presence of no new patterns in the construct HZ*w*(K)Sp0.5/2.3 suggests that there are no enhancer sequences located in the new 0.7kb fragment within this construct. However, the fact that

expression in domains 1 and 2 and the tracheal placodes was maintained in this construct further defined the enhancers for these regions to the upstream 1.2kb of the 2.5kb fragment (see figure 4-13). The absence of the domain N and M expression, seen in HZ*w*(K)X0.5/2.5/5.3, further confirmed that this expression derives from the 5.3 fragment (figure 4-13), but not within the downstream most 0.4kb that was covered in HZ*w*(K)Sp0.5/2.3.

The construct HZ*w*(KS)0.5/2.8 also overlapped with the 5.3 fragment by about 0.7 kb but showed a different early expression pattern in domains N and M. This suggested that either different N and M enhancers are located within the 2.8kb and 5.3kb, outside of the region of overlap or, expression in HZ*w*(KS)0.5/2.8 is a subset of that seen in HZ*w*(K)X0.5/2.5/5.3 and the enhancers for this subset lie within the 0.7kb region of overlap.

(ii) Enhancers for late neural expression

Defining the enhancer regions for the different neural patterns that have been detected with the constructs described in this chapter is not as straight forward as for the earlier expression patterns. This is because it is impossible to confidently identify patterns in different constructs as the same or different, without undertaking a much more detailed analysis to identify specific neural cells. It is however, obvious that there are neural specific regulatory regions strewn throughout this *stg* upstream region that are presumably responsible for activating *stg* transcription specifically in these tissues.

The 500bp fragment shows weak CNS and PNS expression. The presence of CNS expression in all other constructs would therefore suggest that the CNS enhancers are located within the 500bp fragment. However, the lack of CNS expression in the construct HZ*w*(K)Sp0.5/2.3 suggests that maybe not all of the CNS patterns observed derive solely from the 500bp proximal fragment. Given that expression from the 500bp fragment is relatively weak it seems possible that there may be other enhancers within the constructs HZ*w*(K)B8.7 and HZ*w*(K)X0.5/2.5/5.3 which generate the strong CNS expression patterns seen in these constructs. This would place a CNS enhancer within the 5.3kb region of the construct HZ*w*(K)X0.5/2.5/5.3 and a partial CNS enhancer (one that does not activate in the thoracic segments) within the 8.2kb of the construct HZ*w*(K)B8.7 (figure 4-11).
A different CNS pattern is observed in the construct HZ*w*(K)S0.5/2.8 than is seen in HZ*w*(K)0.5, suggesting that this pattern may be in part driven by enhancer elements present within the 2.8kb fragment.

The construct HZ*w*(K)X0.5/2.5/5.3 shows strong expression in the PNS while most other constructs show either very limited or no expression here. This suggests that enhancers for expression in the PNS lie in a 5.3kb region that is present only in this construct. Presumably, some weak PNS enhancers also exist within the 500bp fragment.

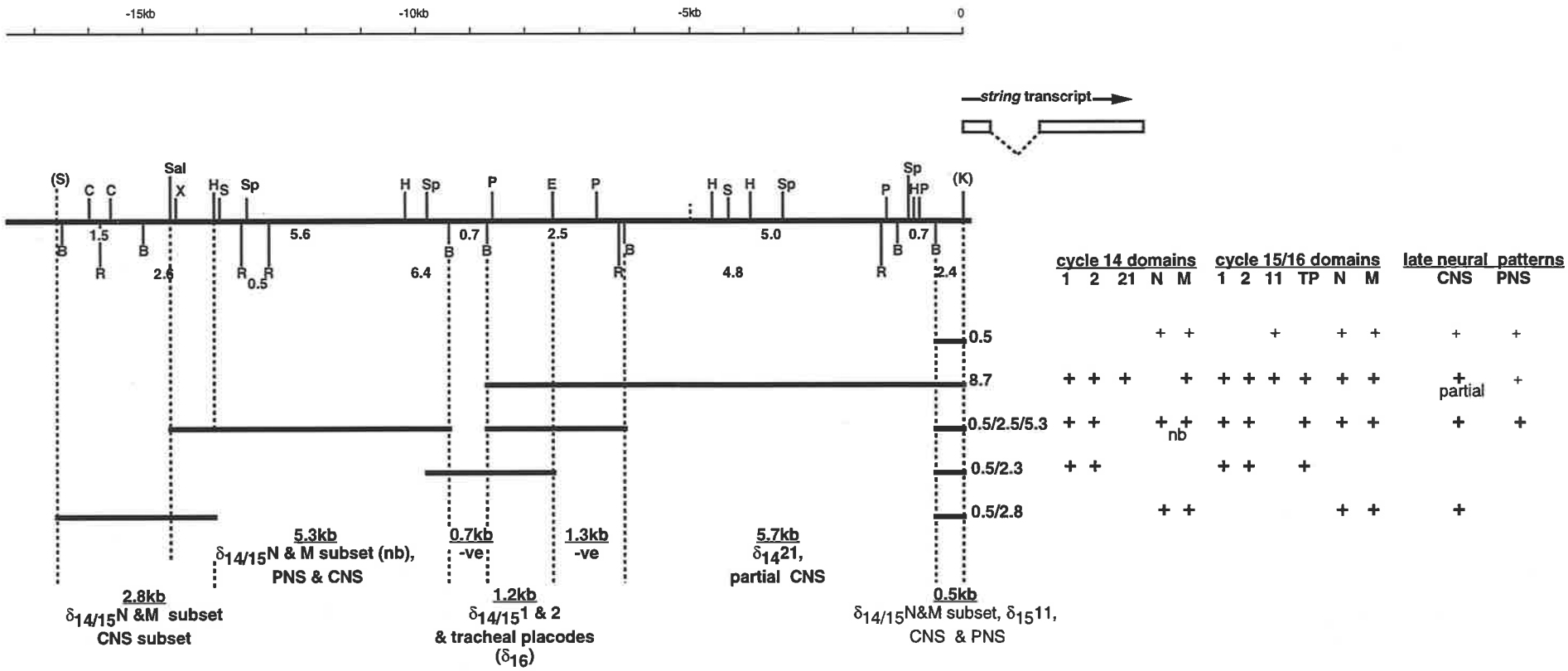
The strong CNS and PNS expression observed in the construct HZ*w*(K)P0.5/1.3 indicates that there may also be enhancers for these expression patterns within the intron fragment.

4-4.3 The search for mitotic domain enhancer regions

The identification of some cycle 14 *stg* domains, in particular 1, 2, and 21 suggests that by using the *stg* basal promoter it is possible to identify domain specific enhancers. Further, that these domain specific enhancers are located within the regions defined by P element deletions and *stg* transgenes

Figure 4-13

A schematic diagram summarising the assignment of enhancers to the *stg* regulatory region. The genomic map is as described in figure 3-5 but also shows an EagI site (E) at about -7.5kb and the *in vitro* mutagenised KpnI site (K) that replaces the TATA box at -30bp. Sites in brackets represent non-genomic sites. The table on the right summarises the expression patterns obtained with the various constructs listed. Underneath the map, bold lines define possible regional locations of the listed enhancers. Bold lettering represents strong staining and fine lettering, weak staining.


-15kb
-10kb
-5kb
0
string transcript
(S)
C C
Sal
X
H S
Sp
H Sp
P
E
P
H S H Sp
P
Sp
HP
(K)
1.5
5.6
0.7
2.5
5.0
0.7
2.6
0.5
6.4
4.8
2.4
0.5
8.7
0.5/2.5/5.3
0.5/2.3
0.5/2.8
5.3kb
δ14/15N & M subset (nb), PNS & CNS
0.7kb -ve
1.3kb -ve
5.7kb
δ1421, partial CNS
2.8kb
δ14/15N &M subset CNS subset
1.2kb
δ14/151 & 2 & tracheal placodes (δ16)
0.5kb
δ14/15N&M subset, δ1511, CNS & PNS
cycle 14 domains
1 2 21 N M
cycle 15/16 domains
1 2 11 TP N M
late neural patterns
CNS PNS
nb
partial

(B. A. Edgar *et al.*, 1994)(see figure 4-1) suggests that these are the enhancer regions that drive *stg* transcription in these domains.

Within the region covered by the constructs described in this chapter it was anticipated, given the work of Edgar *et al.* (1994), that enhancers should also have been located for cycle 14 domains: 3, 16, 17, 20, 23, and 24. None of these were identified and there are several possible explanations for this. Firstly, it is possible that some domains have been incorrectly assigned within this region. Secondly, the proximal promoter region used for these constructs may not be sufficient to achieve the full expression pattern if the different domain specific enhancers require the formation of distinct complexes at the transcription start site. If this were the case then the inclusion of larger proximal sequences may result in the detection of more patterns of expression. Constructs have been made using a larger proximal fragment, from +39bp to -731bp (see section 5-3.2), but these give very similar results to those described here (D. Lehman and B. Edgar, personal communication). Further, the construct HZ*w*(K)B8.7 represents a contiguous stretch of genomic sequence from -30bp to -8.7kb which suggests that additional upstream sequences do not alter the pattern of transcriptional activation. Downstream sequences however, have hardly been tested. About the first 400bp of the *stg* transcript is non-coding so it is possible that there are sequences within this region that are necessary for *stg* transcription in some domains. It is also, still possible that the intron sequences are important but that their proximity and orientation with respect to the transcription start site must be maintained for them to function.

Another possible, but related, explanation for the identification of so few mitotic domain enhancer regions is that the *stg* gene may be very sensitive to disruption. Removing portions of the gene from its native situation by making *lacZ* fusions and then inserting these randomly within the genome may disrupt or repress enhancers located within the fragments being tested. This may occur if boundary elements are involved in defining the regulatory regions that can activate or repress *stg* transcription.

Such elements could also explain the expression in domains 14 and 25 each in only one independent line of a construct. Although a likely explanation for these patterns is that they are driven by non-*stg* enhancers that are located near to the insertion site of that particular construct, it is also possible that they reflect real *stg* enhancers. If this were the case then these independent lines may have fortuitously inserted in positions that do not disrupt the activation of transcription from these enhancers. According to the P element deletion and transgene studies, these enhancer regions are located upstream of -10kb, which places them outside of the constructs HZ*w*(K)B0.5 and HZ*w*(K)P0.5/1.3 in which they were possibly identified. However, these domains were assigned only by the absence of their expression in either of the transgenes and in the P element deletion AR5 (see figure 4-1). It therefore seems possible that if these enhancers are easily disrupted, when removed from their native surroundings, that this might also occur in the transgenes and in the large deletion AR5.

It may be difficult to address the question of which, if any, of the above explanations is correct. Possibly, the generation of many small P element deletions within the *stg* promoter would allow the location of more domain enhancers by causing less disruption. However, this approach would still assign enhancer regions by the absence rather than the presence of expression in a particular domain.

CHAPTER 5: REGULATION OF DOMAIN 2 *string* EXPRESSION

Rather than pursue the full range of *stg* enhancers that are involved in regulating the mitotic domain pattern any further I decided to focus on the regulation of one of the domains already located. By doing this I hoped to determine if the patterning genes can directly regulate *stg* transcription.

5-1 DOMAIN 2

5-1.1 Background

Mitosis in this bilateral domain occurs 71 minutes into cycle 14 (at 25C), at which time the embryo has just completed gastrulation, and rapid germband elongation is about to begin (see figure 1-1). Domain 2 occupies the region immediately anterior to the cephalic fold and extends in a stripe from the cells adjacent to the ventral furrow to the dorsal surface ((V. E. Foe 1989), see figure 1-2). This stripe is broader on the ventral surface, and tapers away as it extends dorsally. This domain is partly obscured as the most posterior portion of it becomes internalised by the cephalic fold.

stg expression in this same domain occurs approximately 25 minutes before mitosis (B. A. Edgar and P. H. O'Farrell 1989) which is around the time that cellularisation is completed. The region of domain 2 expression therefore looks quite different because neither the ventral furrow nor the cephalic fold have formed. *stg* is expressed just anterior to the presumptive cephalic furrow, in a stripe which tapers from ventral to dorsal (figure 1-4A). The ventral boundary of this stripe coincides with the presumptive mesectoderm and dorsally, the stripe tapers away just before it reaches the dorsal midline.

5-1.2 Analysis of domain 2 *string* expression in mutant backgrounds

As an initial attempt to define the regulators of domain 2, the expression pattern of *stg* in this domain was analysed in mutant backgrounds to identify possible regulators. Candidate genes were chosen by their expression pattern in wildtype embryos.

(i) *buttonhead*

The head specific segmentation gene *buttonhead (btd)* is expressed prior to cellularisation, immediately anterior to the presumptive cephalic furrow, in a contiguous stripe that is wider ventrally than it is dorsally (figure 5-1A). During cellularisation the dorsal region of this stripe tapers even further and another region of expression, termed the head spot, becomes visible (figure 5-1B) (E. Wimmer *et al.,* 1993). By comparing embryos where gastrulation has begun, it is clear that the *btd* expression just anterior to the cephalic furrow overlaps that of domain 2 *stg*. Although it is not possible to tell by this comparison whether the *btd* boundaries exactly correlate with any domain 2 *stg* boundaries, their relative position with respect to the cephalic furrow suggests that at least the posterior boundary is common to both. The head spot does not appear to define domain 1, although it is possible that it overlaps domain 1 *stg* expression.

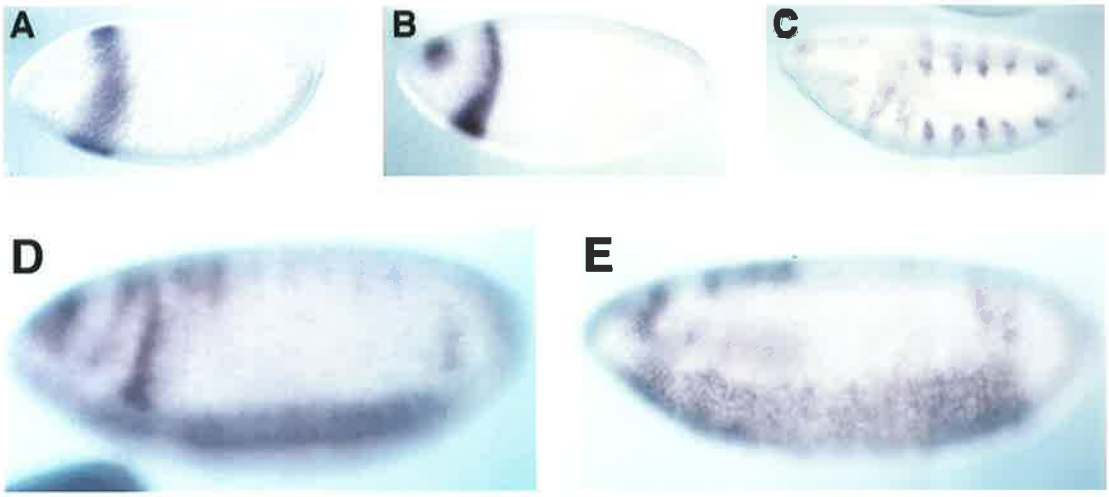
Interestingly, *btd* also has a distinct later expression pattern that appears to include the tracheal placodes during slow germband elongation (figure 5-1C). The isolation of the domain 1, 2 and tracheal placode enhancers within the 1.2kb region and the similarities observed in the *btd* expression patterns make it a likely candidate, not just for the regulation of domain 2 expression, but also for that of the tracheal placodes.

Figure 5-1

btd is required for domain 2 expression. All views are lateral. **A**, in the syncytial blastoderm, *btd* mRNA expression forms a broad stripe in the posterior of the head, **B**, by the completion of cellularization this stripe has narrowed dorsally and the head spot is visible. **C**, much later, during slow germband elongation (stage 9) *btd* is also expressed in the tracheal placodes. **D**, domain 2 *stg* mRNA expression is apparent just prior to and during gastrulation, **E**, in a btd^{XG} homozygote at the same stage of development, domain 2 expression is absent while domain 1 appears to be unaffected.

Figure 5-2

ems is involved in domain 2 expression. All views are lateral. **A**, in the syncytial blastoderm, *ems* mRNA expression forms a broad stripe just anterior to but overlapping that of *btd*, **B**, during cellularization this stripe narrows dorsally and expression on the ventral surface disappears. **C**, domain 2 *stg* mRNA expression is apparent just prior to and during gastrulation, **D**, in an ems^{7D} mutant embryo, at the same stage of development, domain 2 *stg* expression is not apparent, although other early domains are visible in both C and D.


A
B
C
D
E

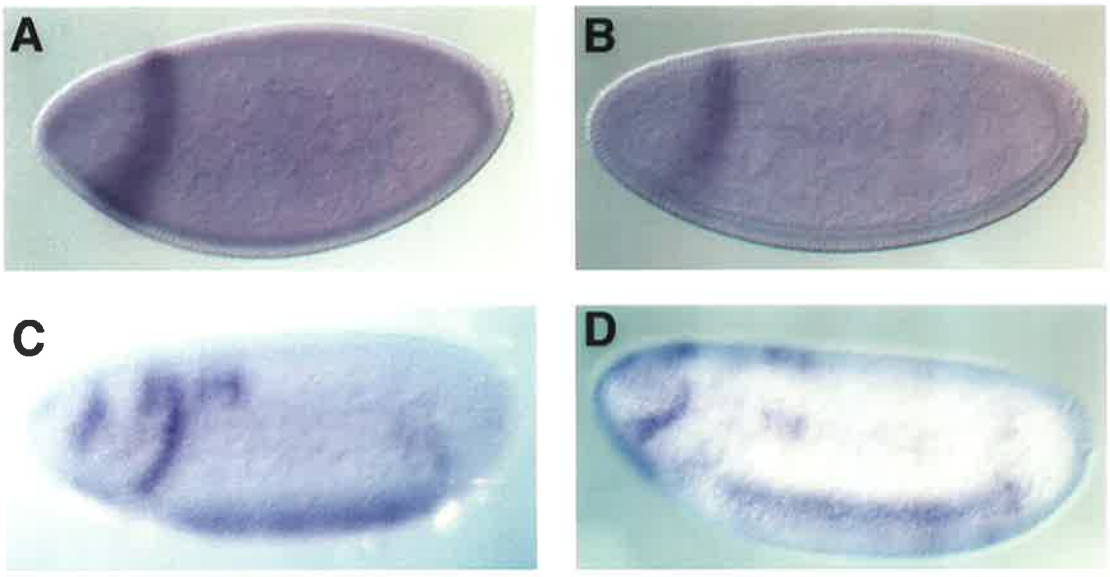

A
B
C
D

Figure 5-3

twi and *sna* both alter domain 2 expression. All views are lateral unless described otherwise. **A**, during cellularization, *twi* mRNA is expressed in a broad stripe along the ventral surface of the embryo which fades out at the the lateral borders (taken from Leptin, 1991). **B**, at the same stage, *sna* mRNA is also expressed in a broad stripe however, its lateral border is sharp. **C**, in a ventro-lateral view, the ventral gap between the bilateral regions of domain is apparent, **D**, a twi^{1} homozygote embryo exhibits a narrower ventral gap. **E**, just prior to gastrulation, the ventral boundary of domain 2 *stg* expression can identified, **F**, a sna^{18} homozygote at the same stage, showing continuous domain 2 expression across the ventral surface.

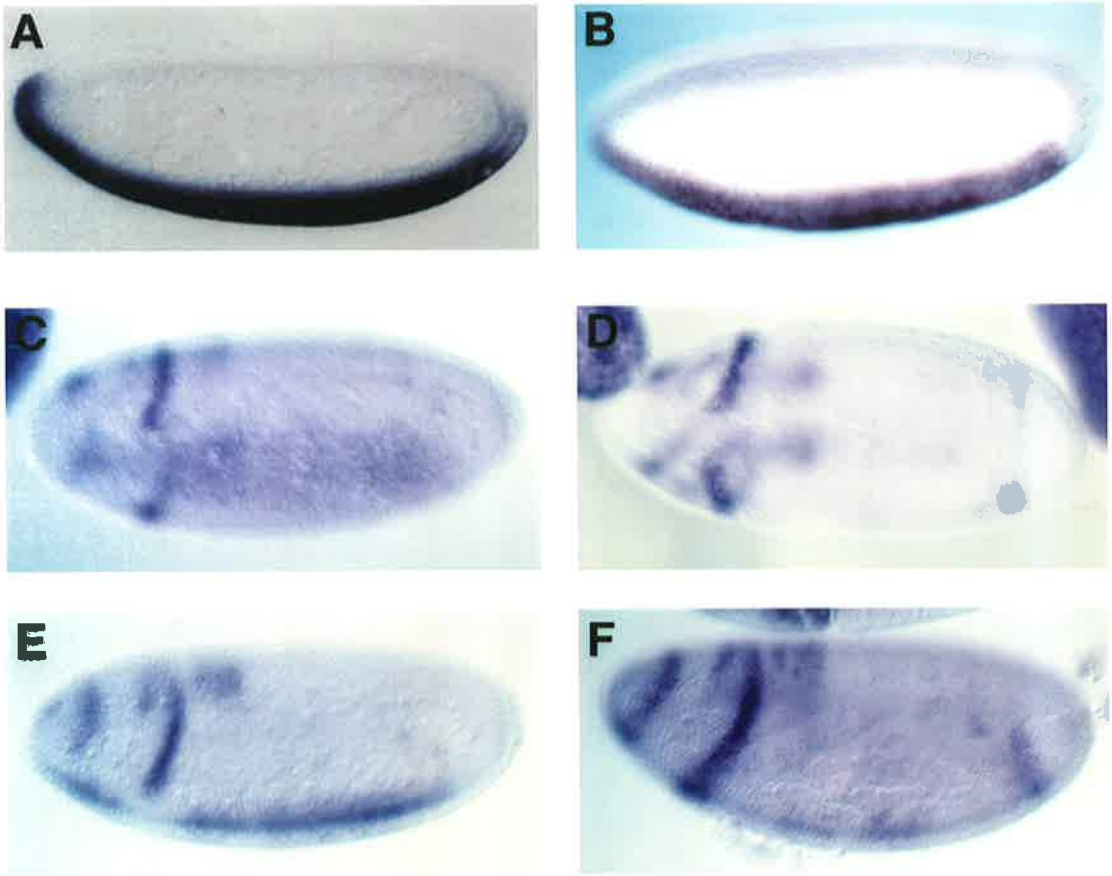

A
B
C
D
E
F

Figure 5-4.1

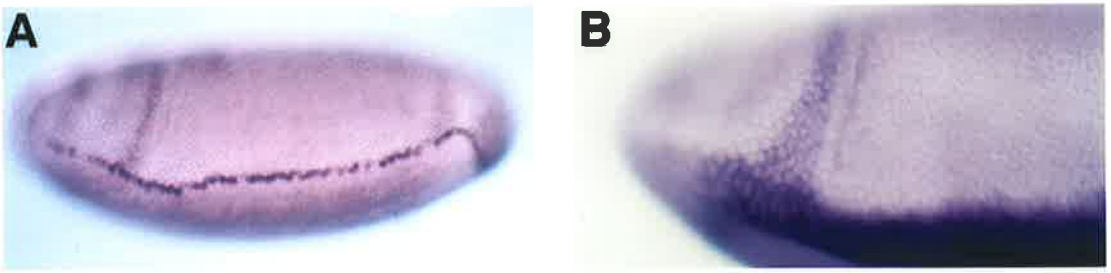
A comparison of *sim*, *sna* and domain 2 *stg* expression patterns by double *in situ* hybridization, all with digoxygenin labelled probes. All views are lateral unless described otherwise. **A**, at cellularization, *sim* expression, which delineates domain 14, is coincident with the ventral boundary of domain 2 *stg* expression. However, it is not possible to tell whether these regions overlap or if domain 2 stops at the dorsal side of the *sim* expressing region. **B**, also at cellularization, this time shown at double the magnification, there is no apparent gap between the region of *sna* mRNA expression and that of *stg* domain 2.

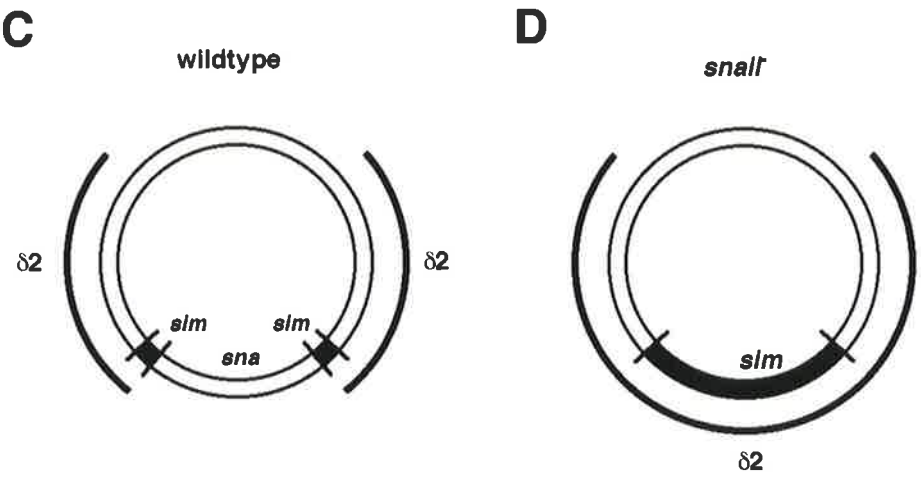
Figure 5-4.2

A schematic diagram showing the regions of *sim* and *sna* expression and domain 2 (δ2) in a wildtype (**C**) and a *sna* mutant embryo (**D**).

Figure 5-4.3

The mRNA expression patterns of *col* and *sim* at cellular blastoderm (**E**), showing the similarity between the region of *col* expression and that of *stg* domain 2 as reported by Crozatier *et al.*, (1996) (compare to figure 5-4.1A).


A
B


C
wildtype
δ2
δ2
sim
sim
sna
D
snail
sim
δ2


E

region constructs have also shown a similar pattern of dorsal expression (D.Lehman and B. Edgar, personal communication). It is also possible that not all of the regulatory regions required for correct domain 2 expression are present. However, the similarity between the HZ*w*(K)B0.5/1.2 *lacZ* and wildtype *stg* patterns of transcription suggests that this fragment contains the major domain 2 enhancer elements. On this basis sequence analysis followed by database searching was initiated as an alternative means of identifying factors that bind to the domain 2 enhancer region.

5-2.2 Sequence generation

Initially, end sequence of the EagI to BamHI 1.2kb fragment was generated using the clones p*stg*EB1.2 and p*stg*EP1.1 and the pBST polylinker primers KS and SK. The clone p*stg*EB1.2 was made in the same way as p*stg*(KRsa)B0.5/1.2, detailed in section 4-3.2, except that the 1.2kb BamHI fragment was cloned directly into pBST. p*stg*EP1.1 was then made from this clone by digesting with PstI which cuts about 100bp inside the genomic BamHI site and also in the polylinker just outside this same BamHI site, the clone was then religated such that this small PstI fragment was 'dropped out'. The end sequence generated from these clones was sufficient to cover the full length of the 1.2kb fragment in one strand. To sequence the other strand and thus confirm the end sequence, the clone p*stg*EB1.2 was digested with Sau3A and the resulting fragments, of which the largest was 352bp, were "shot gun" cloned into the BamHI site of pBST. Random sequencing of these clones using the KS and SK primers was then undertaken to obtain the complete sequence (figure 5-5).

5-2.3 Sequence analysis

The program Signal Scan 4.0 which contains a compilation of known consensus binding sequences, mostly for transcription factors, was used to identify homologies between the *stg* 1.2kb fragment and such sequences (D. S. Prestridge 1996 in press). It was hoped that this would reveal some clues about the factors that bind to the 1.2kb fragment.

Many of the homologies identified in this search are expected not to have any meaning as the context of the elements is not taken into consideration. It is also expected that not all consensus sequences will be detected even if they are represented within the database, as only exact fits are identified.

Within the myriad of consensus binding sequences identified in the 1.2kb fragment, three separate sequences identified the mammalian transcription factor Sp1 and showed homology with several independently determined consensus binding sequences for this factor in some cases. This suggested that these sequences may represent real Sp1 binding sites.

Until quite recently no Sp1 homologues had been identified in *Drosophila*. However the finding that *btd* encodes a *Drosophila* Sp1 homologue (E. Wimmer *et al.,* 1993), which is capable of binding to a consensus Sp1 binding site, suggested that the identified Sp1 signal sequences could reflect binding sites for the Btd protein.

5-2.4 Further sequencing and analysis

The location of these Sp1 sites also suggested that sequence downstream of the 1.2kb may contain more consensus binding sites. One of the Sp1 binding sites was within the most upstream 200bp of the 1.2kb while the other two were within 12 bp from the EagI site at the downstream end of the clone (see figure 5-5). This EagI site is within a very G/C rich region of sequence which suggested that there may be other Sp1 consensus binding sites just downstream of this restriction site. On this basis the sequence was extended in the downstream direction.

The clone p*stg*B2.5 (see section 4-2.2) was used to subclone the 1.3kb EagI to BamHI fragment that extends downstream of the 1.2kb. End sequencing of this clone extended the sequence for another 200bp downstream of the EagI site and revealed another Sp1 consensus binding site within the next 10 bp downstream of the EagI site. No further Sp1 consensus sequences were detected within the 200bp fragment. The sequence now extended for 1.4kb from the BamHI site at -8.7kb to a XhoI site at -7.3kb, with three consensus Sp1 sites surrounding the EagI site and a single site much further upstream of this point (figure 5-5 and 5-6A&B).

To ensure that no small EagI fragments were omitted in the above sequencing strategy, the clone p*stg*B2.5, derived from phage *stg*D, was sequenced using two *stg* primers (#1 and #2) that were generated flanking the G/C rich region surrounding the EagI site (see figure 5-5). This confirmed that there was a single EagI site within this region.

Interestingly, a single base pair change was detected in the sequence of this clone which completely deleted one of the Sp1 consensus sequences within the G/C rich region. The previous sequence, generated using the clone p*stg*EB1.2 was derived from the phage *Dr*L3. This polymorphism was confirmed upon re sequencing of this clone with the primers *stg*#1 and a polylinker primer (figure 5-6B&C).

Both of the above mentioned phage clones were isolated form the same genomic library, made in the vector EMBL-3 SP6/T7 (Clontech), which used the strain Canton S as the source of DNA. Whether this altered base pair represents an actual polymorphism in the Canton S population or a base substitution that occurred during the preparation or amplification of the library is impossible to be certain of. As a means of determining if either sequence was more likely to represent that of wildtype, a cosmid clone which was derived from the isogenic strain, iso-1, was sequenced. The clone -10a extends just beyond the EagI site (figure 3-5), and as there are no EcoRI sites within the cosmid vector, an EcoRI digest followed by religation allowed the majority of the clone to be "dropped out" of the vector. The resulting clone cos-10aR was then sequenced using the *stg* primers (#1 and #2). This clone contained the T rather than the G, which means that it does not have the additional Sp1 consensus sequence (figure 5-6C). From this point onwards, all clones and constructs were made using this sequence even though it contains one less consensus Sp1 site.

Consensus binding sequences for the Sna protein have been identified in three independent studies, none of which were identified using Signal Scan 4.0. This is not so surprising given the differences between the three consensus sequences. A footprinting study of the *sim* promoter, which has been found to be ventrally repressed by *Sna*, identified a 14bp consensus that was present nine times within a 2.8kb *sim* promoter fragment (Y. Kasai *et al.,* 1992)(figure 5-7A). A different consensus was obtained when a *rho* promoter fragment was footprinted and four Sna binding sites were identified (Y. T. Ip *et al.,* 1992a)(figure 5-7B). Finally, an optimal binding site screen using a pool of random oligonucleotides and recombinant Sna protein, followed by several rounds of

Figure 5-5

Nucleotide sequence of the 1.4kb BamHI to XhoI fragment that gives expression in domains1, 2 and the tracheal placodes. Restriction enzyme sites used for sequencing and subsequent cloning steps are underlined, as are the sequences used to generate primers *stg* #1 and #2 with the 5' to 3' direction of the primer marked by an arrow. Identified Sp1 consensus binding sites are marked by shading of the core consensus region GGC, and one identified core sequence for a Sna binding site consensus is also identified by speckling. The BamHI site marks the upstream end of this fragment.

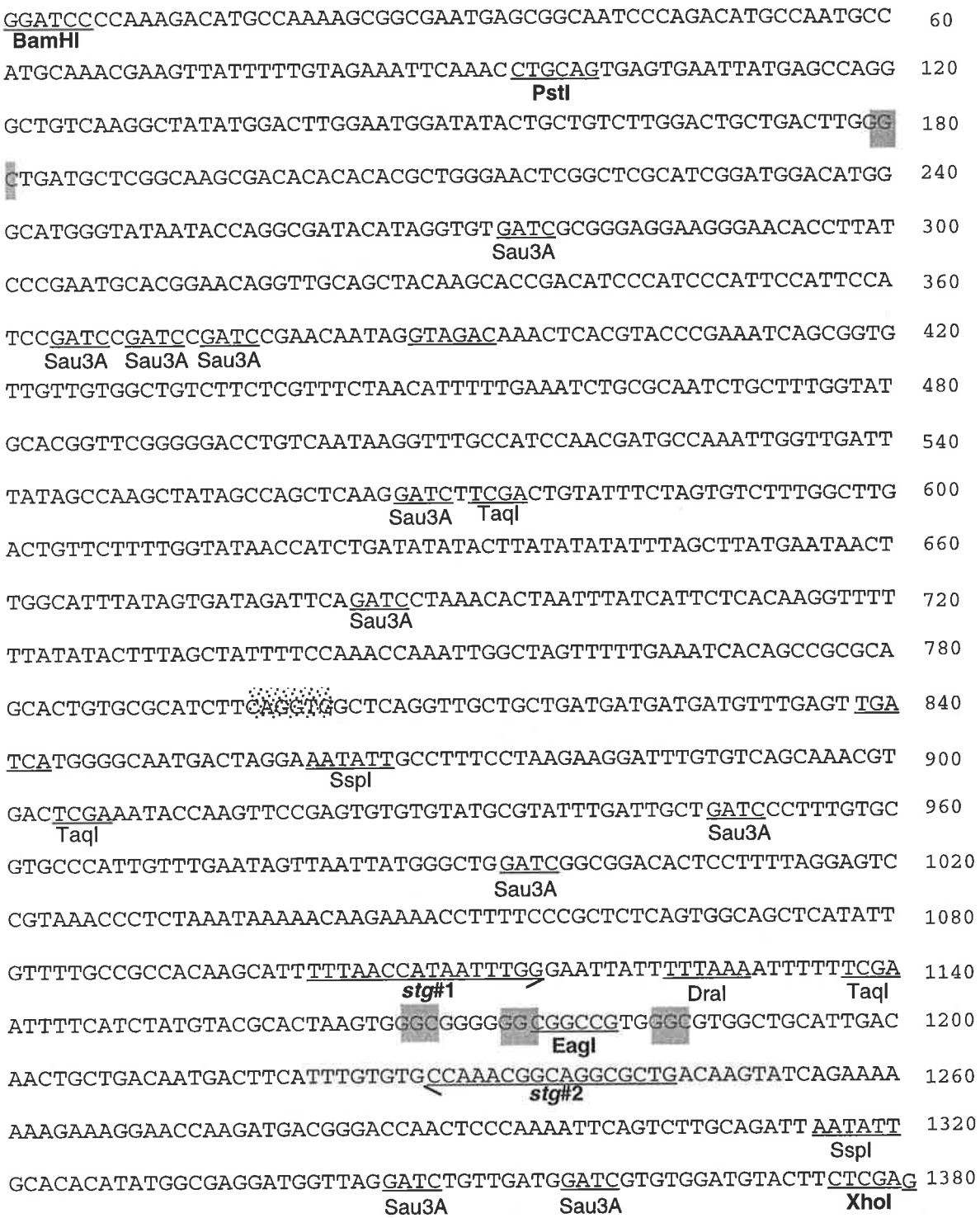
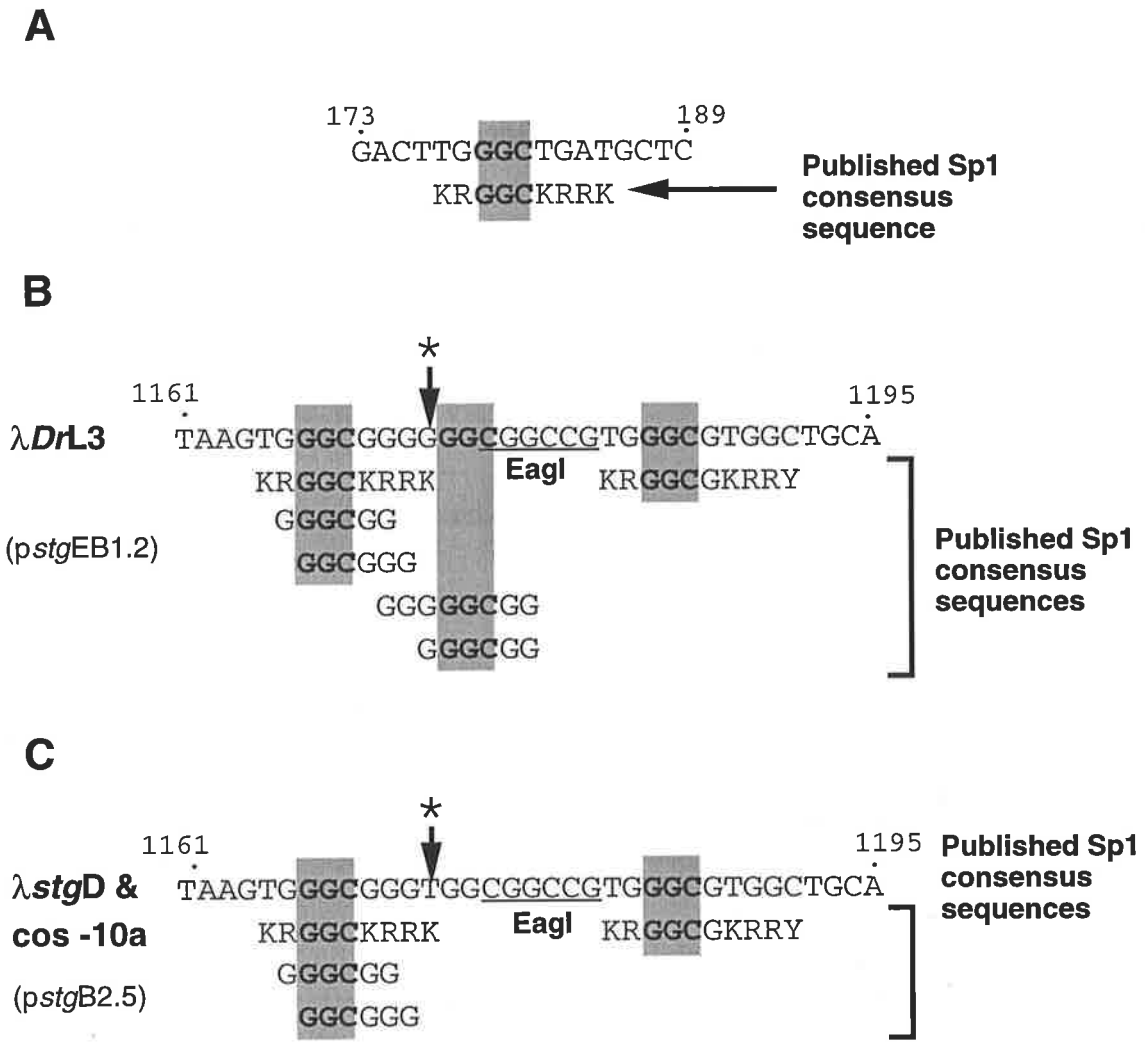

BamHI
PstI
Sau3A
Sau3A Sau3A Sau3A
Sau3A TaqI
Sau3A
SspI
TaqI
Sau3A
Sau3A
stg#1
DraI
TaqI
EagI
stg#2
SspI
Sau3A Sau3A
XhoI

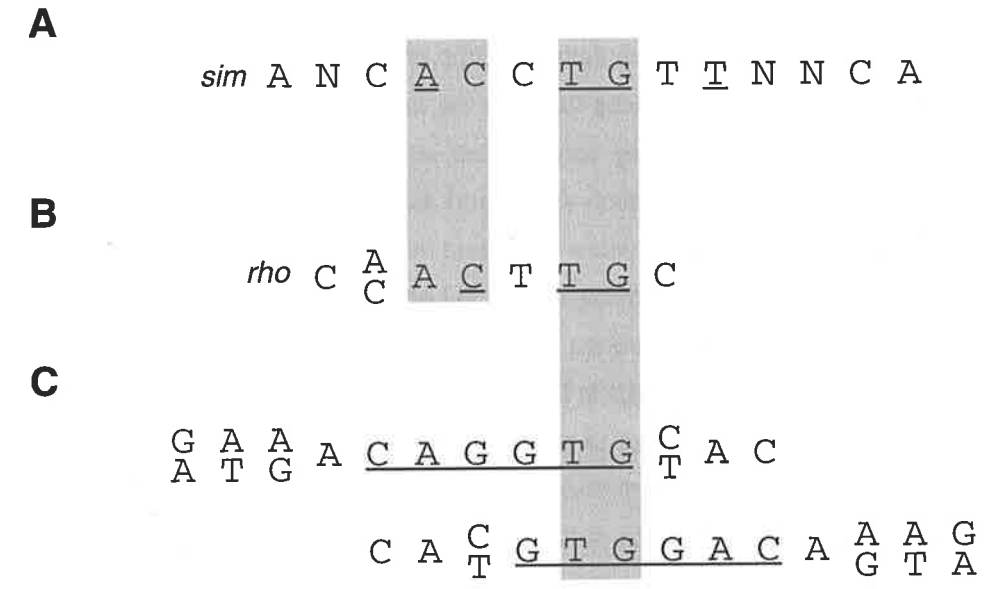
Figure 5-6

Sp1 consensus sequences identified within the 1.4kb BamHI to XhoI fragment. **A** shows the upstream site identified from 177 to 185bp and **B** shows the sites identified surrounding the EagI site from 1165 to 1191bp in the phage clone *Dr*L3. Shading identifies the core consensus region GGC. The nucleotide code is K=G or T, R=G or A and Y=C or T. **C** shows the sequence surrounding the EagI site in the phage clone *stg*D and the cosmid clone -10a with the altered base pair marked by an asterisk. As a result of this base pair change, this sequence contains one less consensus Sp1 binding site than that of the phage clone *Dr*L3.

Figure 5-7

A summary of the Sna consensus binding sites identified in three separate studies. **A** shows the consensus derived from the *sim* promoter study. Underlined bases denote those that were common to all sites identified for a particular study. **B** shows the consensus derived from the *rho* promoter study, and **C** shows the consensus obtained from the optimal binding site screen. Alignment of these sequences, in this figure, was based upon the two residues, TG, which are found in all sequences. The optimal site consensus is fitted in both orientations as this method of binding site identification cannot determine directionality.


A
173
189
GACTTGGGCTGATGCTC
KRGGCKRRK
Published Sp1 consensus sequence
B
1161
1195
λDrL3
TAAGTGGGCGGGGGGCGGCCGTGGGCGTGGCTGCA
EagI
KRGGCKRRK
KRGGCGKRRY
GGGCGG
GGCGGG
GGGGGCGG
GGGCGG
(pstgEB1.2)
Published Sp1 consensus sequences
C
1161
1195
λstgD & cos -10a
TAAGTGGGCGGGTGGCGGCCGTGGGCGTGGCTGCA
EagI
KRGGCKRRK
KRGGCGKRRY
GGGCGG
GGCGGG
(pstgB2.5)
Published Sp1 consensus sequences


A
sim A N C A C C T G T T N N C A
B
rho C A/C A C T T G C
C
G A A/T A/G A C A G G T G C/T A C
C A C/T G T G G A C A A/G A/T G/A

selective PCR amplification determined yet another consensus sequence (V. Mauhin *et al.*, 1993)(figure 5-7C).

The possibility existed that no Sna consensus sequences were identified using the database screen because these sequences were not present in the Signal Scan database. However, searching specifically for these Sna consensus sequences in the 1.4kb still failed to identify any complete matches. The 6bp core sequence, CAGGTG, of the optimal binding site consensus was identified 800bp into the 1.4kb fragment (see figure 5-5), and several other close matches to the *sim* consensus were identified across the sequence. Whether any of these reflect true Sna binding sites was not determined.

5-3 CONSTRUCTS TO DEFINE THE DOMAIN 2 ENHANCERS

5-3.1 Background

As described in section 5-2.1, domain 2 expression in the HZ*w*(K)B0.5/1.2, HZ*w*(K)Sp0.5/2.3 and HZ*w*(K)X0.5/2.5/5.3 constructs was more diffuse than in HZ*w*(K)B8.7 (figure 5-8). This suggested that there may be some enhancers for domain 2 located downstream of the 1.2kb region (figure 5-9A). Given that the sequence analysis also suggests that sequences just downstream of the 1.2kb fragment may be important, a construct using the entire 1.4kb fragment was generated (shown in figure 5-5). Further constructs were also made to determine the importance of these identified Sp1 binding sites

To determine the significance of the putative Btd binding sites, two constructs were made based on the sequence information described in section 5-2. These constructs subdivided the 1.4kb region just upstream of the major Sp1 cluster (figure 5-9B) and it was hoped that this would further define the regulatory sequences required for domain 2.

5-3.2 Construct generation

(i) The 1.4kb construct

The 1.4kb BamHI to XhoI fragment was inserted into a vector called p*stg*-ß (obtained from B.Edgar). This vector is similar to HZ50PL*w* except that it contains a 700bp proximal upstream fragment rather than the 500bp and there are no *hsp70* sequences fused to the *lacZ*. The *stg* sequences at the 5' end are sufficient to initiate transcription as they extend from +39 to -731 and at the 3' end *stg* sequences are used to provide the polyadenylation signal and also instability sequences that are present in the 3' region of the *stg* transcript. This vector was used because it appeared to give slightly more distinct expression patterns (personal communication D.Lehman and B.Edgar), presumably because of the additional 200bp of proximal sequence.

The only available cloning site in the p*stg*-ß vector is EcoRI, so, to insert the 1.4kb fragment into this vector several subcloning steps were incorporated to provide EcoRI polylinker sites on either side of this fragment. The 1.4kb BamHI to XhoI fragment was isolated from p*stg*B2.5 and inserted into the vector pBM20, BamHI to SalI (which is compatible with XhoI), making the clone pBM*stg*BX1.4. This placed the BamHI end of the fragment next to a polylinker EcoRI site, and 200bp further into the vector was a PvuII site. Flanking the XhoI/SalI end of the fragment there was a polylinker EcoRV site. A 1.6kb PvuII to EcoRV fragment was then released from pBM*stg*BX1.4 and inserted into the SmaI site of pBST, to make p*stg*BX1.4+pBM200. By selecting the correct orientation of this clone, the EcoRI

site in the pBST polylinker was positioned on the opposite side to the pBM EcoRI site which allowed the 1.4kb insert to be released using EcoRI and cloned into the p*stg*-ß vector to make p*stg*-ßXB1.4 (figure 5-9B).

(ii) Subdividing the putative Btd binding sites

An 1150bp BamHI to DraI fragment was isolated from within the clone p*stg*EB1.2 which contained all but the 50bp closest to the EagI site (figure 5-9B). This was subcloned into p*stg*(K)Rsa0.5, EcoRV to BamHI, to make the clone p*stg*(KRsa)B0.5/1150.

A 200bp DraI to SspI fragment was isolated to span the Sp1 cluster (figure 5-9B). To generate this clone, a 460bp SspI fragment that spanned the EagI site was isolated from within p*stg*B2.5. This was subcloned into the EcoRV site of pBST and the orientation was selected such that the 200bp DraI to SspI fragment could be isolated by digesting with DraI and HindIII (a polylinker site). This fragment was then inserted into the EcoRV site of p*stg*(K)Rsa0.5 to make p*stg*(KRsa)D0.5/200. Both of the above clones were then inserted into HZ50PL*w*, XbaI to KpnI.

5-3.3 Analysis of constructs

(i) p*stg*-ßXB1.4

lacZ mRNA *in situ* hybridisations on embryos carrying the p*stg*-ßXB1.4 construct showed strong and distinct expression in domains 1 and 2 prior to gastrulation, with only very slight background expression (figure 5-10A&B). However, it is possible that there are no further enhancers within the 200bp region and that it is the proximity of the 1.2kb fragment to the 500bp fragment that results in the more distinct expression. The construct HZ*w*(K)B8.7, which also shows a distinct expression pattern, also has more distance between these two regions. The p*stg*-ßXB1.4 construct contains an extra 400bp that is not present in the HZ*w*(K)B0.5/1.2 construct (by using the 1.4kb and the 700bp proximal fragment from the vector p*stg*ß). This 400bp may contain additional enhancers which resolve further the expression pattern or it may just act as a spacer fragment which allows the enhancers activate transcription more effectively. However, this construct does show that the 1.4kb region is sufficient to achieve a domain 2 expression pattern that very closely resembles that of *stg* expression in domain 2.

(ii) HZ*w*(KRsa)B0.5/1150 and HZ*w*(KRsa)D0.5/200

It was anticipated that the construct HZ*w*(KRsa)D0.5/200 may have contained the major domain 2 regulatory regions, given the identification of the putative Btd binding sites. On this basis it was expected that this construct would show stronger domain 2 expression than HZ*w*(KRsa)B0.5/1150. However, the opposite result was observed.

lacZ mRNA expression in the construct HZ*w*(KRsa)B0.5/1150 was apparent in domains 1 and 2 and the tracheal placodes (figure 5-11A,C&E), much as was observed in HZ*w*(KRsa)B0.5/1.2. The construct HZ*w*(KRsa)D0.5/200 showed broad expression in the region of the cephalic furrow prior to gastrulation (figure 5-11B), but by the time germband extension had begun, no expression was visible in domains 1 or 2 (figure 5-11D). Later during slow germband extension, very weak expression was observed in the tracheal placodes (figure 5-11F). This result suggests that there are sequences that are more important for the expression of *stg* in domains 1 and 2, and the tracheal placodes than the identified Sp1 binding sites within the 200bp region.

Figure 5-9

A schematic diagram depicting the 1.4kb BamHI to XhoI fragment with the relevant restriction sites shown and the putative Sp1 consensus binding sites marked by arrows. Panel **A** compares the constructs; HZ*w*(K)B0.5/1.2, HZ*w*(K)X0.5/2.5/5.3, HZ*w*(K)Sp0.5/2.3 and HZw(K)B8.7 and panel **B** identifies the fragments from the 1.4kb that were used to make the constructs; p*stg*-ßXB1.4, HZ*w*(KRsa)B0.5/1150, and HZ*w*(KRsa)D0.5/200.

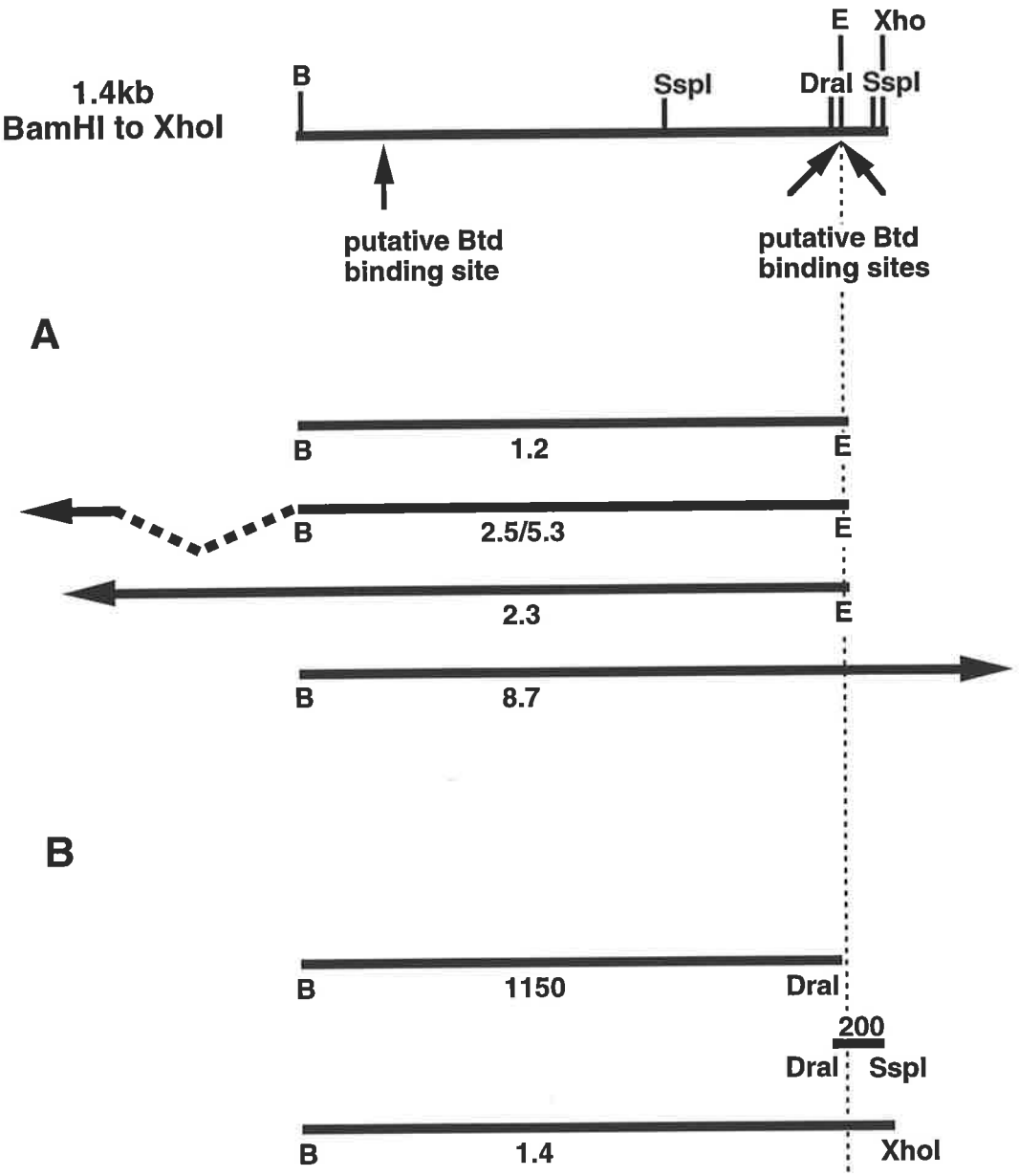
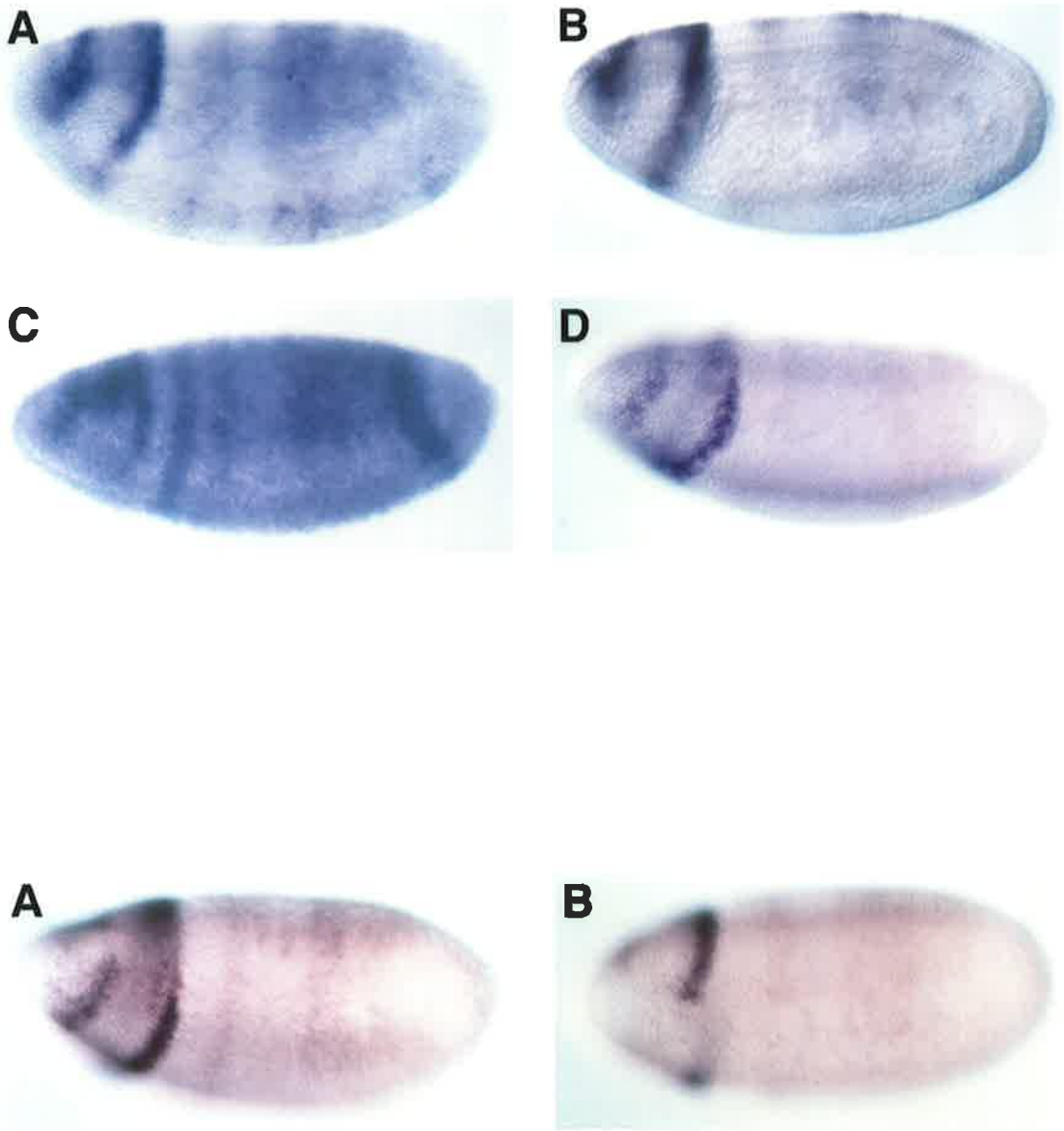

1.4kb
BamHI to XhoI
B
SspI
E
Xho
DraI
SspI
putative Btd
binding site
putative Btd
binding sites
A
B
1.2
E
B
2.5/5.3
E
2.3
E
B
8.7
B
B
1150
DraI
200
DraI
SspI
B
1.4
XhoI

Figure 5-8

lacZ mRNA expression in domain 2 as detected by *in situ* hybridization prior to gastrulation. All views are lateral. The constructs HZ*w*(K)B0.5/1.2 (**A**), HZ*w*(K)X0.5/2.5/5.3 (**B**), and HZ*w*(K)Sp0.5/2.3 (**C**) are all expressed in domains 1 and 2. At this early stage, however, expression is diffuse and there is some background associated with these constructs. In contrast, expression in the construct HZw(K)B8.7 (**D**) is much sharper and the background expression is minimal.

Figure 5-10

lacZ mRNA expression in domains 1 and 2 in a strain carrying the construct p*stg*-βXB1.4, as detected by *in situ* hybridization, prior to gastrulation. **A**, a dorso-lateral and **B**, a ventral view showing the distinct nature of the expression domains and only very slight background expression.


A
B
C
D
A
B

One small spot of expression was also observed on the most anterior tip of embryos carrying the HZ*w*(KRsa)D0.5/200 construct towards the end of rapid germband elongation (figure 5-11D&F). This suggested that a previously unidentified enhancer element may be present within the 200bp fragment. However, as this expression was not detected in the construct p*stg*-ßXB1.4, it seems more likely that this spot is artefactual.

5-3.4 Mutation of the putative Btd binding sites

(i) Background

The above results suggest that the Sp1 sites identified may not play a major role in domain 2 regulation. To test this further, a construct containing the 1.4kb fragment, with the identified Sp1 sites specifically mutated, was generated.

(ii) *in vitro* mutagenesis and construct generation

Three mutagenic primers were synthesised to alter the sequence of the 1.4kb region such that each of the three identified consensus Sp1 sequences were destroyed and a restriction enzyme site inserted in its place.

A clone derived from p*stg*B2.5 (section 4-2.2) was made to generate the single stranded DNA, complementary to the mutagenic oligonucleotide sequences. A 2.3kb BamHI to EcoRI fragment was isolated from p*stg*B2.5 and subcloned into pBST (KS+), BamHI to EcoRI, making p*stg*RB2.3. The most upstream Sp1 consensus sequence was then mutated using the oligo 178/179 (figure 5-12A) which alters nucleotides 178 and 179 and creates a unique PvuII site. The resulting clone, p*stg*RB(Pvu)2.3, was then used for the next mutagenesis step, which mutated the Sp1 site just 5' of the EagI site and inserted an ApaLI site by altering nucleotides 1167 and 1168 (figure 5-12B) to make p*stg*RB(Pvu,Apa5')2.3. Concurrently, the Sp1 site just 3' to the EagI site was mutated using the original p*stg*RB2.3 clone. This site was mutated by altering nucleotides 1184 and 1185 which also incorporated an ApaLI site, making the clone p*stg*RB(Apa3')2.3 (figure 5-12C).

As part of the cloning strategy to generate the 1.4kb mutant construct, it was then necessary to insert the 1.2kb BamHI to EagI fragment, carrying the PvuII and 5'ApaLI sites, into a derivative of the vector pBM20. This vector was missing the polylinker region from The XbaI site to the EcoRV site. This was achieved by digesting pBM20 with XbaI and EcoRV, isolating the vector and then forcing the religation of the incompatible ends, making pBM(ΔXRV). The clone p*stg*RB(Pvu,Apa5')2.3 was digested with BamHI and EagI to release the doubly mutated 1.2kb fragment which was then inserted into pBM(ΔXRV) to make pBM(ΔXRV)*stg*EB(Pvu,Apa5')1.2. Into this clone, the 200bp EagI to XhoI fragment from p*stg*RB(Apa3')2.3 was inserted to make pBM(ΔXRV)*stg*XB(Pvu,Apa5'&3')1.4. This clone was sequenced using the *stg* primers #1 and #2 (figure 5-5) and the reverse sequencing primer, within the polylinker of the pBM vector, to ensure that the *in vitro* mutagenesis had incorporated only the expected base substitutions. The sequence surrounding the PvuII mutation identified a 13bp insertion that was presumably due to secondary structure within primer (results not shown), however, on the basis that the consensus Sp1 sequence was still destroyed and no further sites were generated it was decided to continue with this clone for the mutant construct.

The final step involved isolating the triply mutant 1.4kb fragment BamHI to HindIII (a polylinker site) and inserting it into p*stg*(K)Rsa0.5, making the clone p*stg*(K)B(Pvu,Apa5'&3')0.5/1.4. This construct

was then inserted into HZ50PL*w*, KpnI to XbaI and the use of the pBM(ΔXRV) vector meant that no partial digest was required.

(iii) Analysis of the mutant construct

lacZ mRNA *in situ* hybridisations performed on strains carrying the construct HZ*w*(K)B(Pvu,Apa5'&3')0.5/1.4 revealed no significant change in the expression pattern of any of the patterns associated with this fragment. Expression in domains 1, 2, and the tracheal placodes was still present and looked unchanged from the expression seen in HZ*w*(K)B0.5/1.2 (data not shown).

This result again implies that the identified consensus Sp1 sites do not play a significant role in the regulation of domain 2 or, for that matter, in domain 1 or the tracheal placodes.

5-3.5 Mutant analysis of the construct p*stg*-ßXB1.4

(i) Background

From the initial experiments, looking at *stg* expression in different mutant backgrounds, it is clear that Btd is required for expression in domain 2. However, it is possible that Btd does not interact with the 1.4kb region to achieve this, either because the domain 2 pattern observed in my constructs does not truly reflect domain 2 *stg* expression, or because Btd does not interact directly with the domain 2 regulatory region. To test whether the first possibility was true, the expression of *lacZ* in the construct p*stg*-ßXB1.4 was compared in *btd, ems*, and *sna* mutant backgrounds.

(ii) *buttonhead*

To obtain btd^{XG} mutant embryos that were also carrying the p*stg*-ßXB1.4 construct, a mass cross of virgins, heterozygous for btd^{XG} and males, homozygous on the third chromosome for the construct p*stg*-ßXB1.4 was set up. In the progeny that resulted from this cross, all of the embryos were heterozygous for the p*stg*-ßXB1.4 construct and half of the male embryos also carried the btd^{XG} allele and were therefore *btd* mutants. From this cross, 0 to 5 hour old embryos were collected and *lacZ* mRNA *in situ* hybridisations were used to reveal the expression pattern of the construct. All of the embryos expressed *lacZ* in domain 1, but in one quarter, expression in domain 2 was absent (figure 5-13B compared to A). This result confirms that the role of Btd in domain 2 *stg* activation does require the defined 1.4kb fragment.

(iii) *empty spiracles*

As *ems* is located on an autosome, two generations were required to obtain embryos that were homozygous for ems^{7D} and also carried p*stg*-ßXB1.4. Virgins, homozygous on the X chromosome for the construct p*stg*-ßXB1.4, were crossed to males heterozygous for ems^{7D}, over the balancer third chromosome TM3Sb. All of the progeny resulting from this cross carried one copy of the p*stg*-ßXB1.4 construct. Half of the progeny also carried the balancer chromosome TM3Sb over a wildtype chromosome and were discarded by selecting flies with "stubble" bristles. The other half, that did not have "stubble" bristles, and therefore were heterozygous for the ems^{7D} allele over a wildtype chromosome, were retained. By self crossing these flies, carrying one copy of the p*stg*-ßXB1.4 construct and being heterozygous for the ems^{7D} allele, three quarters of the embryos laid still carried at least one copy of p*stg*-ßXB1.4, and one quarter of these were also homozygous for the ems^{7D} allele.

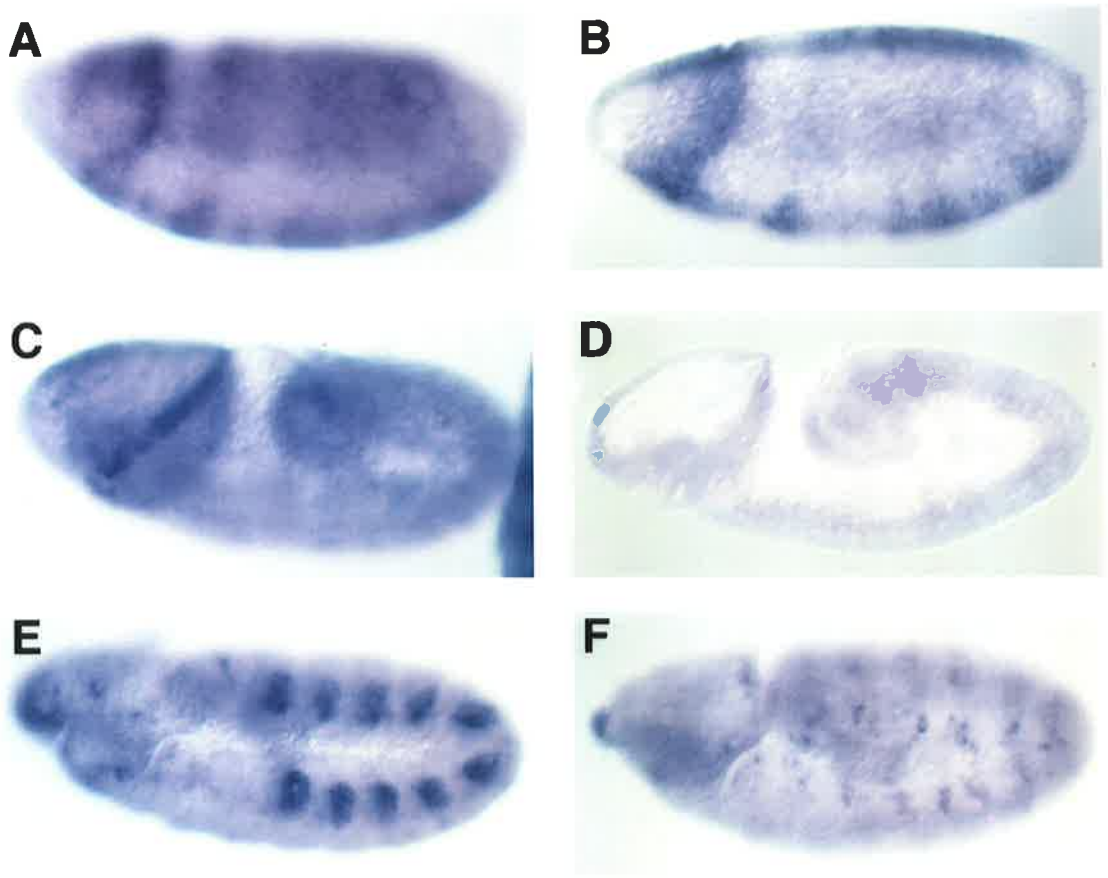
lacZ mRNA *in situ* hybridisations to progeny, staged between 0 and 5 hours AED, failed to reveal any subtle difference within the population of embryos observed (data not shown). This may be because

Figure 5-11

lacZ mRNA expression patterns in the constructs HZ*w*(KRsa)B0.5/1150 (A,C&E) and HZ*w*(KRsa)D0.5/200 (B,D&F). All views are lateral. **A**, prior to gastrulation, expression in HZ*w*(KRsa)B0.5/1150 can be seen in domains 1 and 2 which is somewhat diffuse and background expression is also evident. **B**, at the same stage, in the HZ*w*(KRsa)D0.5/200 construct a broad and vey diffuse band can be seen in the region of domain 2 and the background expression is very high. **C**, during germband elongation, expression can be seen in domains 1 and 2 in HZ*w*(KRsa)B0.5/1150 while in **D**, HZ*w*(KRsa)D0.5/200, the only expression detectable is in a small spot at the very anterior of the head. During slow germband elongation, strong expression can be seen in the tracheal placodes in HZ*w*(KRsa)B0.5/1150 (**E**) while in HZ*w*(KRsa)D0.5/200 (**F**) only very weak expression is detected.

Figure 5-13

lacZ expression in the construct p*stg*-ßXB1.4 in wildtype (A), *btd (B)* and *sna (C)* embryos. All views are lateral. **A**, just prior to gastrulation, the construct p*stg*-ßXB1.4, in a wildtype embryo, clearly shows expression in domains 1 and 2, with the ventral boundary of domain 2 being visible. **B**, the same construct in a btd^{XG} homozygote expresses in domain 1 but not in domain 2. The patch of dorsal expression near the presumptive cephalic furrow is also present in the wildtype embryo (A). **C**, in a sna^{18} homozygote expression from this construct is continuous across the ventral surface in domain 2. Domain 1 also appears to extend further ventrally.


A
B
C
D
E
F

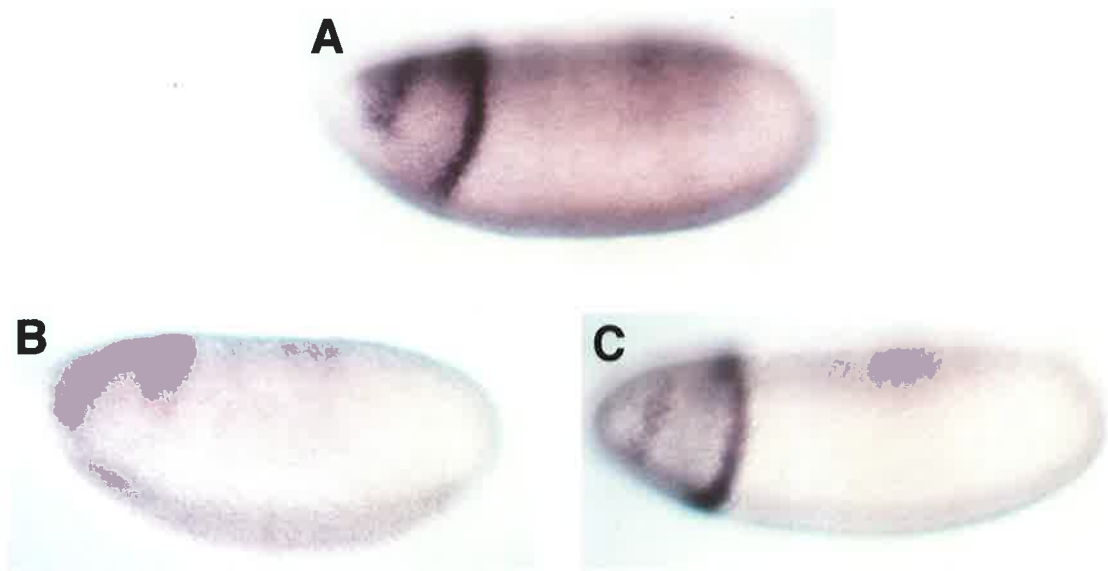
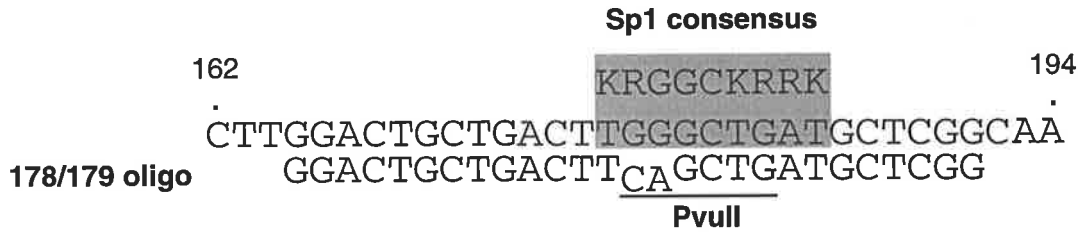
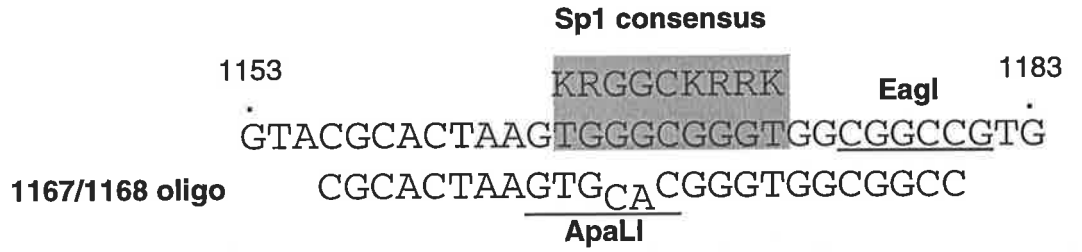
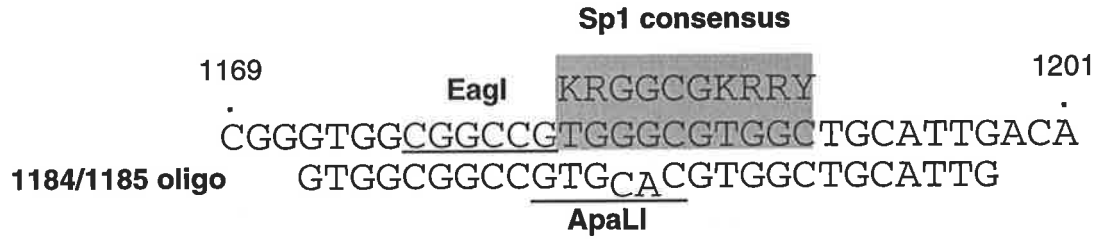

A
B
C

Figure 5-12

Sequences of oligonucleotides synthesised to mutate the identified Sp1 consensus sequences within the 1.4kb *stg* sequence. Consensus sequences are defined by the shaded boxes. **A**, a PvuII site was generated by altering two base pairs within the Sp1 consensus at 177 to 185bp using the oligo 178/179. **B**, an ApaLI site was generated within the Sp1 consensus at 1165 to 1173bp using the oligo 1167/1168. **C**, the final Sp1 site at 1182 to 1191bp was also mutated to generate an ApaLI site using the oligo 1184/1185. The *stg* sequence shown in this figure is of the same strand as was synthesised in the mutant oligonucleotides. When the mutagenesis reactions were performed, the opposite stand was used as the template. The nucleotide code is K=G or T, R=G or A and Y=C or T.


Sp1 consensus
162
KRGGCKRRK
194
CTTGGACTGCTGACTTGGGCTGATGCTCGGCAA
178/179 oligo
GGACTGCTGACTTCAGCTGATGCTCGG
PvuII


Sp1 consensus
1153
KRGGCKRRK
EagI
1183
GTACGCACTAAGTGGGCGGGTGGCGGCCGTG
1167/1168 oligo
CGCACTAAGTGCACGGGTGGCGGCC
ApaLI


Sp1 consensus
1169
EagI
KRGGCGKRRY
1201
CGGGTGGCGGCCGTGGGCGTGGCTGCATTGACA
1184/1185 oligo
GTGGCGGCCGTGCACGTGGCTGCATTG
ApaLI

it was more difficult to determine whether the domain 2 expression was delayed when there are no later mitotic domain patterns to precisely determine the age of the embryo. This experiment was therefore inconclusive.

(iv) *snail*

The crosses required to obtain *sna* homozygous embryos carrying the p*stg*-ßXB1.4 construct were much the same as for *ems*. sna^{18} heterozygous males, balanced over the second chromosome SM6a, were crossed to virgins, homozygous on the X chromosome for the construct p*stg*-ßXB1.4 virgins. The *sna* heterozygotes were selected from the progeny by the absence of the curly wing marker gene (*Cy*) carried on the SM6a balancer. Self crossing these flies resulted in the same proportions of *sna* homozygous embryos carrying the p*stg*-ßXB1.4 construct as in the *ems* crosses described above.

lacZ mRNA *in situ* hybridisations to progeny, staged between 0 and 5 hours AED, revealed that approximately one quarter of the embryos that expressed *lacZ* showed an expansion of domain 2 such that it was continuous across the ventral surface of the embryo (figure 5-13C compared to A). This demonstrates that as well as Btd, Sna is also required for the correct transcriptional activation of the 1.4kb fragment in domain 2. The expression of domain 1 was also extended further ventrally than in wildtype embryos (figure 5-13C), also implicating Sna in the regulation of this domain via the 1.4kb fragment.

5-4 *In vitro* BINDING STUDIES

5-4.1 Background

The mutant analyses described above strongly implicate both Btd and Sna in the regulation of domain 2. As a means of determining whether these proteins interact directly with the 1.4kb *stg* fragment, *in vitro* binding studies were performed. As the identified Sp1 binding sites appear not to be critical for transcriptional activation in domain 2, it is also of interest to determine whether Btd is binding elsewhere within the 1.4kb region.

5-4.2 Expression of fusion proteins

To obtain Btd and Sna protein extracts, clones fused in frame to the inducible Glutathione-S-transferase peptide leader sequence in the vector pGEX (D. B. Smith and K. S. Johnson 1988) were used. The construct pGEX-Btd, obtained from E. Wimmer, contained the zinc finger region of the *btd* cDNA (E. Wimmer *et al.*, 1993). The pGEX-Sna construct was generated by inserting a 1.0kb BamHI (from codon 150) to EcoRI (within the 3' sequence of the vector pNB40 (N. H. Brown and F. C. Kafatos 1988)) fragment from the *sna* cDNA clone (obtained from M.Leptin) into the vector pGEX-2T. This fragment includes the zinc finger region of the protein.

Cultures containing either pGEX alone, pGEX-Btd or pGEX-Sna were induced, harvested and analysed on denaturing polyacrylamide gels to determine levels of expression of fusion proteins. The GST alone extract contained the expected protein band at approximately 25kD which represents the glutathione-S-transferase (GST) product (figure 5-14A). The GST-Btd extract contained a novel protein band at approximately 50kD while the GST-Sna extract contained a novel protein band at approximately 52kD (figure 5-14A). The sizes of these fusion proteins confirmed the predicted sizes, as determined from the amino acid sequences. As the GST-Btd and GST-Sna proteins were insoluble

(data not shown), denaturation and resolubilisation were required to obtain the extracts that were used for subsequent DNA binding experiments, the GST alone protein was also treated in an identical manner for use as a negative control (figure 5-14B). Using the known concentration of the size standards, estimates of the amount of fusion protein present per µl of extract following solubilisation were ; 50-100ng for GST alone, 5-10ng for GST-Btd and 1ng for GST-Sna.

5-4.3 Band shifts using the 1.4kb fragment

The 1.4kb BamHI to XhoI fragment was isolated from the clone pBM*stg*BX1.4 and digested with Sau3A, which generated 10 fragments ranging in size from 12bp to 352bp (figure 5-15A). These fragments were then end filled with Klenow using α^{32}P labelled dATP and sized on a non denaturing polyacrylamide gel using labelled, standard size markers for comparison. Where the sizing was ambiguous, as a result of the differing base composition of the standards and the 1.4kb fragment, further digests were performed to ensure that each fragment was correctly identified.

The labelled 1.4kb Sau3A digest was then incubated with varying amounts of each of the extracts; GST, GST-Btd, and GST-Sna before being resolved on a non-denaturing polyacrylamide gel. Both the GST-Btd and the GST-Sna extracts were able to shift specifically bands from within the 1.4kb.

In the GST-Btd reactions, 2µl of crude extract (approximately 10 to 20ng) was sufficient to show strong shifting in the 352bp fragment that contains the Sp1 cluster and the 272bp fragment that contains the other Sp1 consensus sequence (figure 5-15B). By adding more GST-Btd extract (approximately 20 to 40ng), the 193bp, 155bp, 117bp and 109bp fragments also showed some shifting (figure 5-15B). The 46bp doublet and 12bp fragments shifted non-specifically, suggesting that some protein common to all extracts was binding to these small fragments (figure 5-15B). Further purification of these fusion proteins would be required to determine whether the Btd fusion binds specifically to these fragments.

In the GST-Sna reactions, band shifting was not apparent until at least 3ng of the fusion protein were added. At this point the 352bp and the 272bp fragments showed some shifting which became more apparent in the presence of 4ng of extract (figure 5-15C). No other bands were observed to shift with this extract. However, it was not possible to add more than 4ng of this fusion protein given its low concentration in the extract. The non-specific shifting in the 46bp doublet and 12bp fragments was also observed in this experiment (5-15C).

5-4.4 Oligo shifts

To define further whether the Btd fusion protein was binding to the SP1 consensus sequences, a double stranded 33bp oligonucleotide was synthesised to span the Sp1 sites surrounding the EagI site (figure 5-16A). This was phosphorylated with T4 polynucleotide kinase using γ^{32}P labelled dATP and then incubated with the GST alone or the GST-Btd extract.

The Btd fusion protein was able to bind specifically to this *stg* oligonucleotide. The addition of non-specific competitor DNA did not alter the observed band shift however, the addition of increasing amounts of the unlabelled *stg* oligonucleotide was able to compete for the fusion protein and prevent the observed band shift (figure 5-16B). To test whether the consensus Sp1 sites were responsible for the observed shift, a mutant form of the double stranded oligonucleotide was generated in which 3 bp were altered (figure 5-16A). The middle G in the core of each of the two consensus Sp1 binding

Figure 5-14

Coomassie Blue stained SDS polyacrylamide gels of protein prepared for band shift experiments. **A**, tracks 1 and 6: protein size standards; track 2: GST alone extract showing the induced GST protein at 25kD; track 3: GST-Btd extract with the induced band at approximately 50kD; track 4: GST alone extract (the GST protein has run off the gel); track 5: GST-Sna extract with the induced protein band at approximately 52kD. **B**, extracts after denaturation and solubilisation. Tracks 1 and 6: protein size standards; tracks 2 and 4: 'solubilised' GST protein extracts (the GST protein in track 4 has run off the gel); track 3: solubilised GST-Btd extract; and track 6: solubilised GST-Sna extract (barely visible). Only a very small amount of Sna protein was solubilised using these conditions.

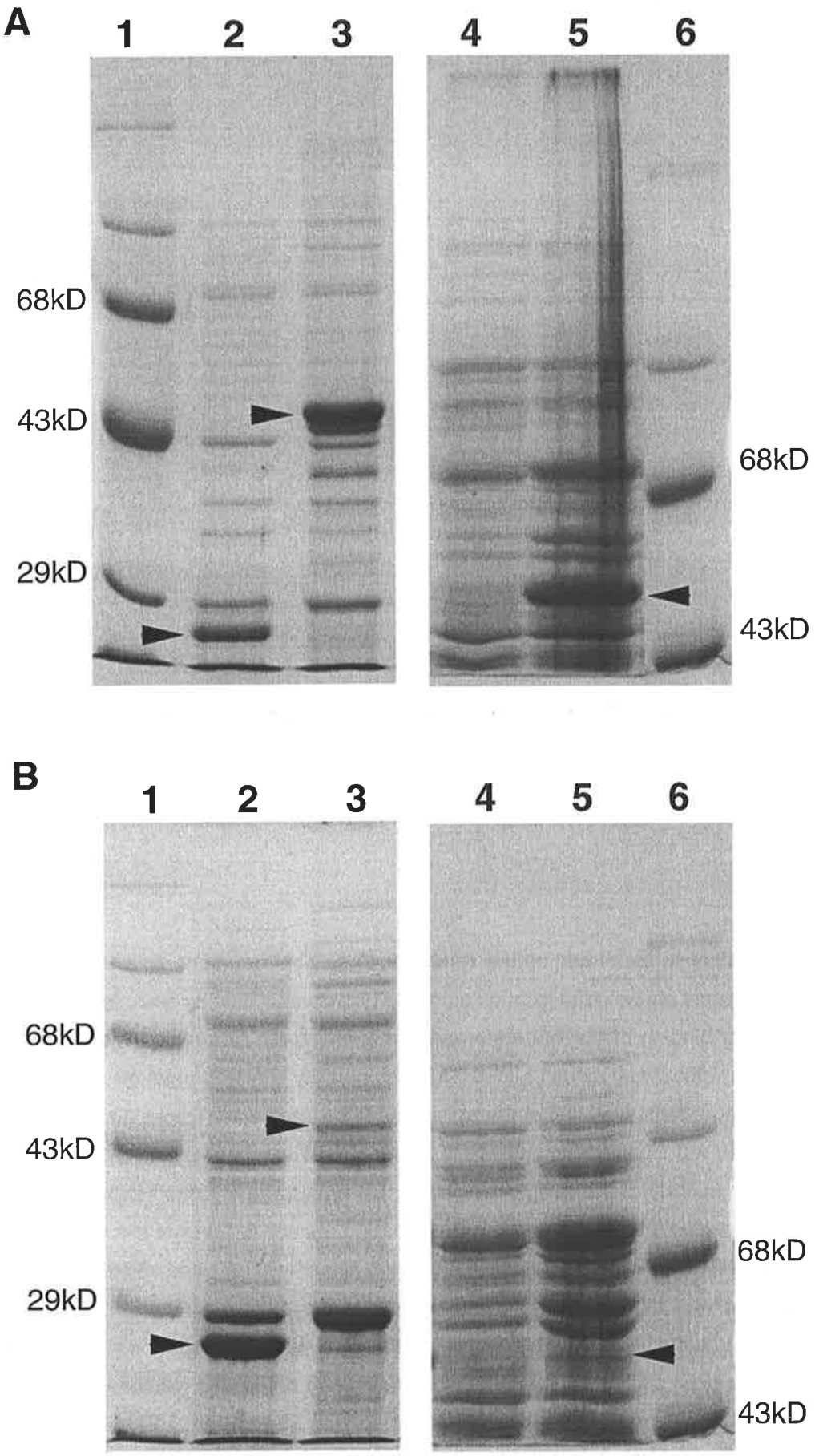

A
1
2
3
4
5
6
68kD
43kD
29kD
68kD
43kD
B
1
2
3
4
5
6
68kD
43kD
29kD
68kD
43kD

Figure 5-15

Band shifting experiments performed on the 1.4kb BamHI to XhoI fragment. **A**, a Sau3A restriction map of the 1.4kb fragment showing the location of the Sp1 consensus binding sites. **B**, a band shifting experiment to test for binding of the Btd fusion protein to the 1.4kb fragment. Track 1, shows the Sau3A digested DNA without any extract added. Note the partially digested products at the top of the gel, resulting in the under-representation of the 352bp and 90bp fragments. Tracks 2 and 4 contain the same digest, incubated with 50 to 100ng and 100 to 200ng of GST extract respectively. The addition of this extract alters the mobility of the 46bp doublet and the 12bp fragment (the 12bp fragment is not shown here). Tracks 3 and 5 were incubated with 5 to 10ng and 10 to 20ng of the GST-Btd extract respectively and in this case the mobility of the 352bp (and its partially digested component 352*) and the 272bp fragments was altered and to a lesser degree the 193bp, 155bp, 117bp and 109bp fragments also showed a mobility shift. The doublet at 46bp and the 12bp fragment also shift with this extract. **C**, is a similar band shifting experiment to test for binding of the Sna fusion to the same DNA. Tracks 1, 3, 5 and 7 contain the digested DNA incubated with 50 to 100ng, 100 to 200ng, 150 to 300ng and 200 to 400ng of the GST extract (respectively). Shifting is again observed in the 46bp doublet and the 12 bp fragment with this extract. Tracks 2, 4, 6, and 8 contain the same DNA incubated with 1, 2, 3, and 4ng of the GST-Sna extract (respectively). The mobility of the 352bp (and 352*) fragment was altered and to a lesser degree the 272bp fragment also showed a mobility shift.

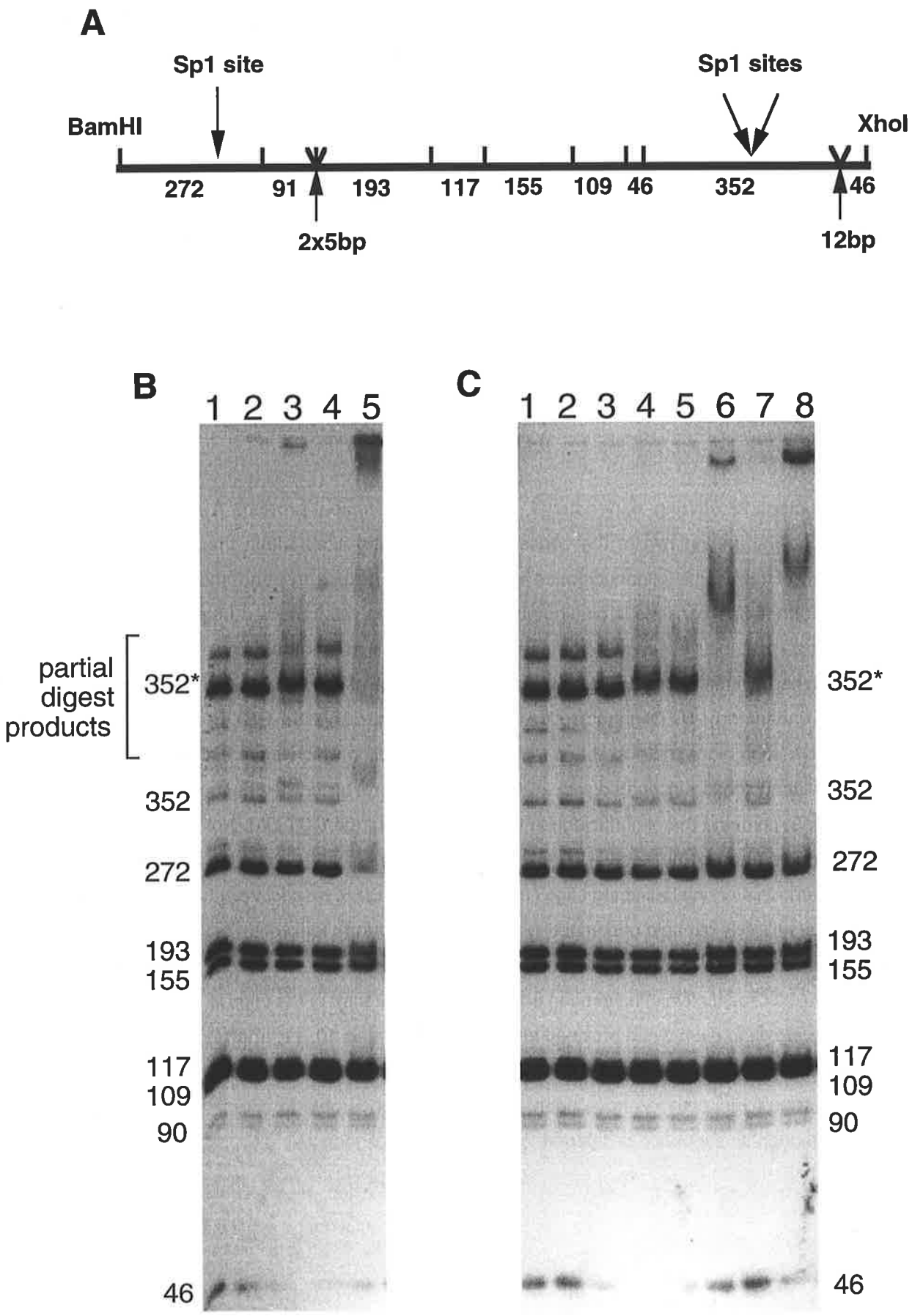

A
Sp1 site
Sp1 sites
BamHI
XhoI
272
91
193
117
155
109
46
352
46
2x5bp
12bp
B
1 2 3 4 5
partial digest products
352*
352
272
193
155
117
109
90
46
C
1 2 3 4 5 6 7 8
352*
352
272
193
155
117
109
90
46

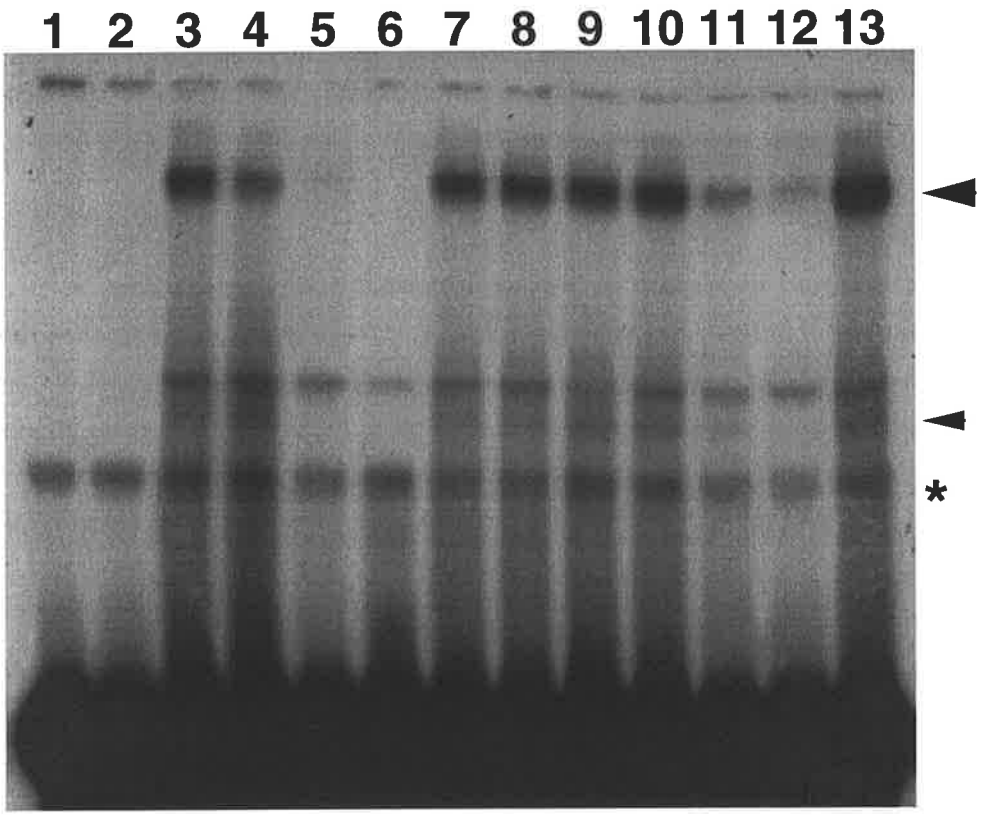
Figure 5-16

Oligonucleotide shifting experiments using the Btd fusion protein. **A**, a double stranded oligonucleotide spanning two of the consensus Sp1 binding sites within the *stg* sequence and a mutant form of this same oligonucleotide were generated to use in band shifting and competition studies. A further Sp1 consensus binding site oligonucleotide was obtained from Promega. Sp1 sites are underlined and their absence in the mutant *stg* oligo results from the single base pair alterations (also underlined) within each site. Note the additional G to A alteration outside of the Sp1 consensus sites (also underlined). **B**, Oligonucleotide shifts performed using the GST-Btd extract and the *stg* oligo. Track 1, oligo alone, note the band (*) above the free oligo which is present in all tracks. Track 2, the same oligonucleotide incubated with 25 to 50ng of GST results in no change in the mobility of the oligo. Track 3, contains the oligonucleotide with 2.5 to 5ng of the GST-Btd. Note the specific mobility shifted products (marked by arrows). Tracks 4, 5, and 6, the same reaction as in track 3 except that increasing amounts of unlabelled stg oligo (10X, 100X, and 1000X respectively) have been added. This unlabelled *stg* oligo is capable of competing with the labelled form for protein binding. Tracks 7, 8, and 9, the same reaction as in track 3 except that increasing amounts of unlabelled mutant *stg* oligo (10X, 100X, and 1000X respectively) have been added. This mutant form of the oligo is unable to compete for the wildtype *stg* oligo even when provided in vast excess. Tracks 10, 11, and 12, the same reaction as in track 3 except that increasing amounts of the unlabelled Sp1 consensus binding site oligo (10X, 100X, and 1000X respectively) have been added. This oligo is able to compete for the wildtype *stg* oligo even though its sequence is different to that of the *stg* oligo.

A

***stg* oligo**	5'GTGGGCGGGTGGCGGCCGTGGGCGTGGCTGCAT3'
mutant *stg* oligo	5'GTGGACGGGTGACGGCCGTGGACGTGGCTGCAT3'
Sp1 consensus binding site oligo	5'ATTCGATCGGGGCGGGGCCGAGC3'

B



sites was changed to an A. One other G within the core region of the third Sp1 consensus, identified in this region of the alternate DNA sequence (see section 5-2.4), was also mistakenly changed. This mutation is therefore not within a consensus Sp1 site (figure 5-16A). Adding increasing amounts of this mutant *stg* oligonucleotide, had no affect upon the binding of the Btd fusion protein to the wildtype *stg* oligo (figure 5-16B). Given the third base substitution in this oligonucleotide, this result is inconclusive in determining whether the Btd binding sites are the same as the Sp1 binding sites. However, it is clear that Btd must bind to the wildtype sequence with high sequence specificity if altering just three single base pairs out of thirty-three is sufficient to prevent binding.

A further test of the ability of the Btd fusion protein to bind to Sp1 sites was performed by using a commercially available Sp1 binding site oligonucleotide (Promega) in competition studies (figure 5-16A). The identical Btd fusion protein has been shown to be capable of binding to this oligonucleotide (E. Wimmer *et al.,* 1993), however, using the conditions described in Wimmer *et al,*(1993), I was unable to repeat this result without observing a similar shift with the GST alone extract (a control not included in their experiment). Under binding conditions subsequently optimised for this experiment, the addition of increasing amounts of this oligonucleotide were able to compete with the wildtype *stg* oligonucleotide for the fusion protein (figure 5-16B). As the sequence of this consensus oligo is quite different, outside of the core GGC, to that of the *stg* oligo, this result does suggest that the Btd and Sp1 binding sites share at least some identity.

5-5 DISCUSSION

5-5.1 Genes involved in regulating *string* expression in domain 2

The genetic analysis of *btd* and *sna* mutants indicate that these genes play important roles in the regulation of domain 2 *stg* expression. Both *stg* expression and *lacZ* expression in flies carrying the p*stg*-ßXB1.4 construct are altered in the absence of either of these gene products. Whether these two genes provide sufficient information to exactly define domain 2 is less certain. Although the early expression pattern of *btd* is far too broad, by the completion of cellularisation the patterns are similar enough to suggest that it may be defining the anterior and posterior borders of this domain. Other genes that may be involved are; *ems*, and *sloppy paired (slp).*

The absence of *ems* appears to have only a subtle effect upon domain 2 expression. The ambiguous nature of this result is substantiated by the work of Edgar *et al.* (1994) who noted that in *ems* mutants domain 2 was narrowed or absent (B. A. Edgar *et al.,* 1994). It seems likely that *ems* does play a role in the regulation of domain 2 although, this role may be to advance the timing of mitosis rather than to define a boundary of expression.

The head specific gap gene *sloppy paired (slp)* is expressed in a broad stripe that appears to span the domain 2 region (M. Crozatier *et al.,* 1996; U. Grossniklaus *et al.,* 1994). The requirement of this gene for normal domain 2 expression has not been tested, however it may be of interest to see if the absence of this gene product alters either the anterior or posterior limits of this domain.

The regulation of the ventral boundary of domain 2 is a particularly interesting problem as it appears, from double *in situ* hybridization experiments, that the boundaries of *sna* and domain 2 *stg* expression touch. This is reasonable given that *sna* is playing a critical role in the ventral repression of domain 2. However, as mentioned in section 5-1.3(i), *sna* expression presumably coincides with domain 10, given its role in mesoderm determination (Y. Kasai *et al.,* 1992; M. Leptin 1991), and

domain 14 sits in a single cell wide line between domains 10 and 2 (V. E. Foe 1989). This suggests that *stg* expression in domain 2 actually spans part of domain 14 which implies that, either domain 14 does not exist in this region or that it regulates the timing of its mitosis in a manner independent of *stg* transcription. In her description of domain 14, Foe (1989), states that she is uncertain about the anterior limit of domain 14. However, she suggests that the narrow domain, domain 15 is contiguous with domain 14 (see figure 1-2) which means that one or the other domain will be present between domains 2 and 10. The possibility that either domain 14 or 15 do not use the timing of *stg* transcription to trigger its mitosis may therefore be the most likely explanation. As discussed in section 1-5.3 it has been shown for domain 14 that other signals, apart from *stg* transcription, are required before mitosis can occur. In a wildtype embryo these signals are thought to be in place prior to *stg* transcription such that *stg* still acts as the rate limiting factor. It is possible however, that this is not the case at least for some regions of domain 14. For example, the phosphorylation state of *stg* , in this small region, may prevent entry into mitosis until a later stage in development.

The other candidate gene for the regulation of *stg* in domain 2 is *col*, which has been shown to delineate this domain precisely (M. Crozatier *et al.,* 1996). However, the finding that *stg* expression is unaffected when *col* antisense is expressed in the embryo suggests that it is not involved in domain 2 regulation but may be regulated in parallel with domain 2. Interestingly, mutant analysis of *col* expression indicates that *btd* is essential for *col* expression and *ems* represses its expression ventrally. The *slp* gene was found not to be involved and expression was not tested in *sna* mutants. The ventrally derepressed phenotype in an *ems* mutant indicates that expression of different genes in the region of domain 2 is achieved in different ways.

5-5.2 Direct regulators of domain 2 *string* expression

The information gained from scanning the 1.4kb sequence for potential binding sites of transcription factors suggests that the Sp1 homologue, Btd, may be directly regulating *stg* transcription via Sp1 consensus binding sites. The importance of these consensus Sp1 sites was tested by examining the regulatory activity of a 200bp fragment that includes two of these consensus binding sites. The observed *lacZ* expression pattern in embryos carrying this construct suggests that this 200bp fragment is not sufficient for the regulation of domain 2. It appears that the upstream 1150bp fragment is more important even though it contains only one Sp1 consensus binding site. This suggests that these Sp1 sequences are not essential, a result that was confirmed by specifically mutating these sites.

It is possible that Btd does not regulate *stg* transcription by interacting directly with *stg* and that an intermediate factor exists. However, the ability of Btd to bind specifically to fragments within the 1.4kb that do not contain an identified Sp1 consensus suggests that some Btd binding sites do not conform to the mammalian Sp1 consensus sequence. A further search of the 1.4kb sequence, using just the 3bp core Sp1 consensus sequence, GGC, identified this sequence 23 times within the 1.4kb fragment. Which if any of these core sequences represent real Btd binding sites remains to be determined, although, it was interesting to note that they were not randomly distributed but were clustered in the 352 and 272bp fragments. The strong shifting observed in these fragments could then be due to the number of binding sites present rather than the previously identified Sp1 consensus sites.

The explanation, that there are multiple Btd binding sites within the 1.4kb, is consistent with the domain 2 enhancer activity observed when the Sp1 sites were specifically mutated. Also, the domain 2 expression seen in the construct HZ*w*(KRsa)B0.5/1150 and not in HZ*w*(KRsa)D0.5/200 may be because there are many more unidentified Btd binding sites within the 1150bp than in the 200bp fragment. However, an alternative explanation for the lack of expression in the 200bp fragment could be that the binding sites that are present within this region are positioned too close to the proximal fragment for correct transcriptional activation to occur.

The issue of whether the additional Sp1 site identified in the phage clone *Dr*L3 results from a library artefact or a real polymorphism may be less significant if there are many Btd binding sites within the 1.4kb region. A polymorphism that alters the number of Btd binding sites from three to four could presumably have a much greater effect upon *stg* transcription than if it increased the number from (for example) ten to eleven.

The importance of the identified Sp1 consensus sequences within the 1.4kb is still unclear. It could be that Btd does not bind to these sites at all, although the oligonucleotide shifting experiments show that it is capable of binding to a consensus Sp1 binding site and that it binds to the 33bp fragment that contains two Sp1 consensus sequences. To address this question directly it will be necessary to do footprinting experiments across this region to exactly define the sequences to which Btd is binding. It is anticipated that these sequences will form a consensus that is distinct but overlapping to that of the mammalian Sp1 binding site consensus.

The band shifting observed with the Sna fusion protein, on the 352bp and 272bp fragments, indicates that Sna may bind directly to repress *stg* transcription in the ventral region of domain 2. Circumstantial evidence, that Sna can act as a short range repressor, also suggests that Sna may be directly involved. Gray *et al.*,(1994) have shown that Sna is able to mediate repression of activators that are within 50 to 100bp of a Sna binding site and they suggest that this mechanism may permit enhancer autonomy (S. Gray *et al.,* 1994). In the context of the *stg* promoter, this would allow Sna to repress domain 2 expression ventrally while not interfering with the activation of *stg* in domain 10, provided that the domain 10 enhancers are more than 100bp away from the domain 2 Sna binding sites. The fact that the P element deletion AR5 removes domain 2 expression but not that of domain 10 (B. A. Edgar *et al.,* 1994),indicates that these enhancers are separated by at least 5kb (see figure 4-1).

Obviously, the identification of Sna binding sites within the 1.4kb fragment by footprinting analysis is necessary to confirm this hypothesis. It would also be of interest to observe the effect of ectopic *sna* expression upon the pattern of *stg* transcription as it would be anticipated that only domains that require Sna to repress their ventral expression, such as domains 2 and 4 would be affected in this situation.

Further to the *in vitro* experiments, it will also be necessary to confirm the identified binding sites for both Btd and Sna *in vivo*. This could be achieved by specifically mutating the identified sites in a *lacZ* construct, depending upon the number of sites within the 1.4kb region.

CHAPTER 6: DISCUSSION

6-1 INTRODUCTION

The aim of the early work described in this thesis was to determine how the complex pattern of *stg* transcription was achieved at such an early stage in development. The possibility that *stg* was directly integrating spatial specific information to define each domain of mitosis implied that it may have independent regulatory regions for each mitotic domain. Alternatively, the timing of *stg* transcription could be dependent upon a common signal that was triggered at different times in the separate domains by particular differentiative events. By testing different *stg* regulatory regions for enhancer activity it was assumed that it would be possible to determine whether one, few or many enhancer regions were involved in *stg* regulation, and therefore decide between the above possibilities.

Within the first 17kb upstream of *stg,* independent enhancers were identified for several domains of *stg* expression suggesting that each domain is regulated independently by spatial specific signals. The further study of the domain 2 regulatory region indicates that, at least in this case, it is the early patterning genes that are providing this developmental information.

6-2 REGULATION OF *string* EXPRESSION IN DOMAIN 2

6-2.1 *string* directly integrates early patterning gene information

The detailed analysis of the regulation of *stg* in domain 2 has revealed that the two early patterning genes, *btd* and *sna* are required and are likely to be directly regulating *stg* transcription via enhancer sequences within the 1.4kb fragment. As *stg* is a conserved component of the cell cycle that is required in nearly all cells of the embryo, this finding represents one of the first examples where the patterning genes actually direct a developmental event. While it is still formally possible that an unidentified 'domain 2 specific' transcription factor is also involved in timing *stg* transcription in domain 2, its does seem that this would be somewhat redundant in light of the role of Btd and Sna.

6-2.2 Further experiments

Although ,Btd and Sna have been shown to bind to the 1.4kb fragment *in vitro*, it is still necessary to identify their exact binding sites and then to verify these sites *in vivo*. Footprinting studies, to determine the exact sequences to which these proteins are binding, followed by specific mutation of these sites to observe the effect upon domain 2 transcription needs to be undertaken.

The identification of the exact Btd and Sna binding sites will determine the importance of the previously identified Sp1 binding sites and may indicate whether Sna is acting as a local repressor of *stg,* as it would be expected that a Sna binding site should be located within 100bp of each Btd binding site (see section 5-5.2). However, it is also possible that this close proximity may not be necessary if Btd, like mammalian Sp1, can form multimers when bound to DNA to bring molecules bound at distant sites close together (W. Su *et al.,* 1991).

The potentially large number of Btd binding sites within the 1.4kb may make it impractical to specifically mutate each of these sites within an expression construct. If this the case, a fragment

containing only a few of the Btd and Sna binding sites could be specifically mutated and then several copies of this region could be included in a construct.

6-3 TRANSCRIPTIONAL REGULATION OF *string* IS COMPLEX

6-3.1 Independent position specific elements direct *string* transcription

Transgene and P element deletion studies performed by B. Edgar *et al.*, (1994) have shown that the *stg* promoter region is very large and that it is composed of separable position specific elements (PSE's) (B. A. Edgar *et al.*, 1994). Although the PSE's defined in this study were large and directed *stg* transcription in groups of domains, it was proposed that smaller, more specific PSE's would exist within these regions. The constructs described in chapter 4 have shown that these smaller PSE's do exist, at least to some degree, with different upstream fragments directing different expression patterns. In particular, the 1.4kb fragment (described in chapter 5) defines the PSE for domain 2 and also an unseparable, but apparently independent, PSE for domain 1. Further, the domain 2 PSE may also direct transcription in the cycle 16 tracheal placode domain, given that Btd is also expressed in the tracheal placodes later in development. This level of transcriptional regulation is similar to that described for the pair-rule genes, where different stripes are regulated by independent enhancer regions to produce a complex pattern of expression (see section 1-4.1 (i)). However, *stg* transcription must incorporate even further complexity, requiring input from both the anterior-posterior and dorsal-ventral patterning cascades to define its domains of expression.

6-3.2 Promoter dissection disrupts some expression domains

Constructs spanning the first 17kb upstream of *stg* revealed enhancers for the cycle 14 domains 1, 2, 21 and subsets of N and M (as described in chapter 4). However, the work of Edgar *et al.*, (1994) indicated that this region also contains enhancers for other cycle 14 domains. Several possible explanations for the diminished number of mitotic domains identified by this method have already been discussed in section 4-4.1, the most likely of which may relate to the removal of relatively small promoter fragments from their 'native' surroundings. This implies that a picture of gene regulation in which enhancer-specific transcription factors interact with the transcription complex is simplistic, especially for a gene with a large regulatory region. It seems likely that higher order regulatory mechanisms, such as boundary elements and chromatin structure, will also be required in the regulation of *stg* transcription.

6-3.3 Promoter specificity in *string* transcription

During the course of identifying enhancers for the domains of *stg* expression it was discovered that a proximal *stg* fragment is required, in combination with the upstream fragments, to obtain patterns using the *hsp70/lacZ* reporter gene, that reflect *stg* transcription. This proximal 500bp fragment extends directly upstream of the identified TATA box at -30bp and a putative, second, TATA box was identified within the distal region of this fragment. The deletion of this distal region, in constructs containing defined *stg* enhancer sequences, resulted in no transcriptional activation. This suggested that the *stg* enhancer activity may specifically require a *stg* promoter, as the *hsp70* promoter did not allow enhancer function.

Promoter specificity has been found to occur at other *Drosophila* loci and it seems likely that it will turn out to be a common mechanism for maintaining the autonomy of gene regulation. The artefactual mesoderm expression, described in section 3-2, which was identified only in the presence of the *stg* basal sequences, may represent an example of promoter specificity between *ry* and *hsp70.* Further, the *zen* promoter (in which the *ry* mesodermal element was first identified (H. J. Doyle *et al.,* 1989)) may also be incompatible with the *hsp70* basal promoter, although to my knowledge, this has not been tested.

Although the means by which promoter specificity is determined is not known, fragments of only 100bp, surrounding the TATA box, are sufficient to determine specificity (C. Merli *et al.,* 1996). It has recently been demonstrated that distinct subunits of TFIID, the TAFs (TATA-binding protein associated factors), can target different activators and basal factors (reviewed in (B. F. Pugh 1996)) suggesting that they may play a role in determining whether a particular transcription factor can interact with a particular transcription complex. Interestingly, several of these TAF's have been cloned from *Drosophila* and dTAF110 has been shown to interact directly with human Sp1, suggesting that it may function as a co-activator between Sp1 and the TFIID complex (T. Hoey *et al.,* 1993). Given this finding it would seem possible that the presence, or absence, of TAF110 in the TFIID complex may generate the promoter specificity by which Btd can activate transcription from the *stg* promoter but not that of *hsp70.*

6-3.4 Identifying further *string* enhancer regions

Unidentified enhancers for domains of *stg* transcription, fall into two classes. Those that should have been identified by the constructs described in chapter 5, and those that probably lie outside of the region tested in chapter 5. To identify more of these *stg* enhancer regions, it may be essential to maintain, as much as possible, their normal genomic context. Small P element deletions in the *stg* upstream region may provide the best available means to achieve this. However, a large number of non-lethal P element insertions scattered within the upstream regulatory region would be required for a thorough analysis.

Possibly, further domains could be identified by working backwards from candidate regulators to identify the binding sites for such genes. This would be similar to the approach undertaken for domain 2 except that much larger DNA regions would need to be screened. A yeast one hybrid system, or affinity chromatography could be used to isolate and identify proteins that bind to particular fragments of upstream *stg* DNA. Sequence analysis, preferably resulting from the *Drosophila* genome sequencing project, could also be used to search large genomic regions for consensus binding sites of candidate regulators for a particular domain. The subsequent *in vivo* verification of any regulatory region defined by these means would obviously be essential.

Although *stg* regulation is a very complex problem, the analysis of further domains will yield general information about the regulators of transcription in particular tissues and could potentially broaden our understanding of transcriptional regulation.

6-3.5 Verification of *string* promoter specificity

The proposal that the proximal fragment contains a *stg* minimal promoter that is specific to *stg* assumes that the putative TATA box within this fragment is functional. Obviously, this must be verified

by transcription start site analysis, such as primer extension and/or PCR analysis before further experiments can be initiated.

To be certain that the *stg* enhancers require this putative TATA box, and surrounding *stg* minimal promoter, rather than some other undefined element within the proximal region (now defined to about 84bp), it will be necessary to test this in *lacZ* expression constructs. This could be achieved by specifically mutating the *stg* TATA, *in vitro*, in the construct p*stg*(KRsa)B0.5/1.2 (figure 4-9), and testing for the presence or absence of *lacZ* expression in domains 1 and 2. An alternative experiment would be to generate a construct that uses the *stg* proximal fragment without the *hsp70* basal sequences, to see if expression is maintained. If it is shown that the *hsp70* sequences are not required, further definition of the *stg* minimal promoter could then be undertaken. Removal of the 84bp region in construct p*stg*(KA)AB396/1.2 (figure 4-12) prevents transcriptional activation, however, this sequence may not sufficient to form the functional *stg* promoter.

Similar experiments to those performed between *dpp* and *oaf* (described in section 4-4.1) could also be attempted between *stg* and its downstream neighbour *pathless (ptl)* (B. A. Edgar *et al.*, 1994) to determine whether *ptl* uses a promoter specific mechanism to prevent *stg* enhancers from activating its transcription.

6-4 THE ROLE OF CELL DIVISION IN DEVELOPMENT

The identification of separable PSE's within the *stg* regulatory region indicates that *stg* is integrating domain specific information to achieve its complex pattern of expression. Further, the *in vitro* binding of Btd and Sna to the domain 2 *stg* enhancer region strongly suggests that these proteins directly regulate domain 2 transcription, however, whether this is true for other domains remains to be seen. This work identifies *stg* as one of the first examples of a 'downstream' or 'target' gene of the patterning transcription factors and emphasises the significance of the regulation of cell proliferation during development by the realisation that the basic developmental program is directly connected to cell cycle progression.

It is interesting to consider how the complex regulation of *stg* might have evolved. Somehow a conserved, ubiquitous cell cycle regulator has become subject to spatial and temporal-specific transcriptional regulation. During the course of evolution *stg* has accumulated the binding sites for numerous transcriptional regulators to effect this. That so many position specific elements exist within the *stg* gene suggests that the selection for this mode of regulation was significant. This supports the idea that the organism has a selective advantage in coordinating cell division with other developmental events in each mitotic domain.

REFERENCES

Alphey, L., J. Jimenez, H. White-Cooper, I. Dawson, P. Nurse, and D.M. Glover.(1992). "*twine*, a *cdc25* Homolog That Functions in the Male and Female Germline of Drosophila." *Cell* . **69** : 977-988.

Amon, A., U. Surana, I. Muroff, and K. Nasmyth.(1992). "Regulation of p34cdc28 tyrosine phosphorylation is not required for entry into mitosis in *S.cerevisiae*." *Nature* . **355** : 368-371.

Arora, K. and C. Nüsslein-Volhard.(1992). "Altered mitotic domains reveal fate map changes in *Drosophila* embryos mutant for zygotic dorsoventral patterning genes." *Development* . **114** : 1003-1024.

Atherton-Fessler, S., F. Lui, B. Gabrielli, M.S. Lee, C.Y. Peng, and H. Piwnica-Worms.(1994). "Cell cycle regulation of the p34cdc2 inhibitory kinases." *Mol. Biol. Cell* . **5** : 989-1001.

Ausubel, S.F., R. Brent, R.E. Kingston, D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl. Current Protocols in Molecular Biology. New York: Wiley, 1994.

Baeuerle, P.A. and D. Baltimore.(1988). "Activation of the DNA-binding activity in an apparently cytoplasmic precursor of the NF-kB transcription factor." *Cell* . **53** : 211-217.

Bier, E., L.Y. Jan, and Y.N. Jan.(1990). "*rhomboid*, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*." *Genes Dev.* . **4** : 190-203.

Bissen, S.T.(1995). "Expression of the cell cycle control gene, *cdc25*, is constitutive in the segmental founder cells but is cell-cycle-regulated in the micromeres of leech embryos." *Development* . **121** : 3035-3043.

Bissen, S.T. and D.A. Weisblat.(1989). "The durations and compositions of cell cycles in embryos of the leech, *Helobdella triserialis*." *Development* . **106** : 105-118.

Boulay, J.L., C. Dennefeld, and A. Alberga.(1987). "The *Drosophila* developmental gene *snail* encodes a protein with nucleic acid binding fingers." *Nature* . **330** : 395-398.

Brown, N.H. and F.C. Kafatos.(1988). "Functional cDNA Libraries from *Drosophila* Embryos." *J. Mol. Biol.* . **203** : 425-437.

Bullock, W.O. and et al.(1987). *BioTechniques* . **5** : 376-378.

Cohen, S.M. and G. Jürgens.(1990). "Mediation of *Drosophila* head development by gap-like segmentation genes." *Nature* . **346** : 482-485.

Coleman, T.R., Z. Tang, and W.G. Dunphy.(1993). "Negative regulation of the Wee1 protein kinase by direct action of the Nim1/Cdr1 mitotic inducer." *Cell* . **73** : 1-20.

Costa, M., E.T. Wilson, and E. Wieschaus.(1994). "A putative cell signal encoded by the *folded gastrulation* gene coordinates cell shape changes during *Drosophila* gastrulation." *Cell* . **76** : 1075-1089.

Courtot, C., C. Fankhauser, V. Simanis, and C.F. Lehner.(1992). "The *Drosophila cdc25* homolog *twine* is required for meiosis." *Development* . **116** : 405-416.

Crozatier, M., D. Valle, L. Dubois, S. Ibnsouda, and A. Vincent.(1996). "*collier*, a novel regulator of *Drosophila* head development, is expressed in a single mitotic domain." *Current Biology* . **6** : 707-718.

Dalton, D., R. Chadwick, and W. McGinnis.(1989). "Expression and embryonic function of *empty spiracles*: a *Drosophila* homeo box gene with two patterning functions on the anterior-posterior axis of the embryo." *Genes Dev.* . **3** : 1940-1956.

Doyle, H.J., R. Kraut, and M. Levine.(1989). "Spatial regulation of *zerknüllt*: a dorsal-ventral patterning gene in *Drosophila*." *Genes Dev.* . **3** : 1518-1533.

Draetta, G., F. Luca, J. Westendorf, L. Brizuela, J. Ruderman, and D. Beach.(1989). "Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF." *Cell* . **56** : 829-838.

Draetta, G., H. Piwnica-Worms, D. Morrison, B. Druker, T. Roberts, and D. Beach.(1988). "Human cdc2 protein kinase is a major cell-cycle regulated tyrosine kinase substrate." *Nature* . **336** : 738-744.

Driever, W. and C. Nüsslein-Volhard.(1988). "A gradient of *bicoid* protein in *Drosophila* embryos." *Cell* . **54** : 83-93.

Driever, W., G. Thoma, and C. Nüsslein-Volhard.(1989). "Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the *bicoid* morphogen." *Nature* . **340** : 363-367.

Dunphy, W.G., L. Brizuela, D. Beach, and J. Newport.(1988). "The *Xenopus* cdc2 protein is a component of MPF, a cytoplasmic regulator of mitosis." *Cell* . **54** : 423-431.

Dunphy, W.G. and A. Kumagai.(1991). "The cdc25 protein contains an intrinsic phosphatase activity." *Cell* . **67** : 189-196.

Edgar, B.A.(1995). "Diversification of cell cycle controls in developing embryos." *Current opinion in Cell Biology* . **7** : 815-824.

Edgar, B.A. and S.A. Datar.(1996). "Zygotic degradation of two maternal *cdc25* mRNAs terminates *Drosophila's* early cell cycle program." *Genes Dev.* . **10** : 1966-1977.

Edgar, B.A., C.P. Kiehle, and G. Schubiger.(1986). "Cell cycle control by the nucleo-cytoplasmic ratio in early *Drosophila* development." *Cell* . **44** : 365-372.

Edgar, B. A., D. A. Lehman, and P. H. O'Farrell.(1994). "Transcriptional regulation of *string (cdc25)* - a link between developmental programming and the cell cycle." *Development* . **120** : 3131-3143.

Edgar, B.A. and P.H. O'Farrell.(1989). "Genetic Control of Cell Division Patterns in the *Drosophila* Embryo." *Cell* . **57** : 177-187.

Edgar, B.A. and P.H. O'Farrell.(1990). "The Three Postblastoderm Cell Cycles of *Drosophila* Embryogenesis Are Regulated in G2 by *string*." *Cell* . **62** : 469-480.

Edgar, B.A., G.M. Odell, and G. Schubiger.(1987). "Cytoarchitecture and the patterning of *fushi tarazu* expression in the *Drosophila* blastoderm." *Genes Dev.* . **1** : 1126-1247.

Edgar, B.A. and G. Schubiger.(1986). "Parameters controlling transcriptional activation during early *Drosophila* development." *Cell* . **44** : 871-877.

Edgar, B.A., F. Sprenger, R.J. Duronio, P. Leopold, and O'Farrell.(1994). "Distinct molecular mechanisms regulate cell cycle timing at successive stages of *Drosophila* embryogenesis." *Genes Dev.* . **8** : 440-452.

Enoch, T., K. Gould, and P. Nurse.(1991). "Mitotic checkpoint control in fission yeast." *Cold Spring Hrabor Symp. Quant. Biol.* . **56** : 409-416.

Fantes, P.(1979). "Epiststic gene interactions in the control of division in fission yeast." *Nature* . **279** : 428-430.

Featherstone, C. and P. Russell.(1991). "Fission yeast p107wee1 mitotic inhibitor is a tyrosine/serine kinase." *Nature* . **349** : 808-811.

Feilotter, H., P. Nurse, and P.G. Young.(1991). "Genetic and molecular analysis of *cdr1/nim1* in *Schizosaccharomyces pombe*." *Genetics* . **127** : 309-318.

Feinberg, A.P. and B. Vogelstein.(1983). "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity." *Anal. Biochem.* . **132** : 6-13.

Fesquet, D., J-C. Labbe, J. Derancourt, J-P. Capony, S. Galas, F. Girard, T. Lorca, J. Shuttleworth, M. Doree, and J-C. Cavadore.(1993). "The *MO15* gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues." *EMBO J.* . **12** : 3111-3121.

Finkelstein, R., D. Smouse, T.M. Capaci, A.C. Spradling, and N. Perrimon.(1990). "The *orthodenticle* gene encodes a novel homoedomain protein involved in the development of the *Drosophila* nervous system and ocellar visual structures." *Genes Dev.* . **4** : 1516-1527.

Fisher, R.P. and D.O. Morgan.(1994). "A novel cyclin associates with MO15/Cdk7 to form the CDK-activating kinase." *Cell* . **78** : 713-724.

Foe, V.E.(1989). "Mitotic domains reveal early commitment of cells in *Drosophila* embryos." *Development* . **107** : 1-22.

Foe, V.E. and B.M. Alberts.(1983). "Studies of nuclear and cytoplasmic behaviour in the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis." *J. Cell Sci.* . **61** : 31-70.

Foe, V.E. and G.M. Odell.(1989). "Mitotic domains partition fly embryos, reflecting early cell biological consequences of determination in progress." *Amer. Zool.,* . **29** : 617-652.

Foe, V.E., G.M. Odell, and B.A. Edgar. "Mitosis and morphogenesis in the *Drosophila* embryo: point and counterpoint." In The Development of *Drosophila melanogaster*, ed. M. Bate and A. Martinez Arias. 149-300. I. Plainview, New York: Cold Spring Harbour Laboratory Press, 1993.

Gao, Q., Y. Wang, and R. Finkelstein.(1996). "*orthodenticle* regulation during embryonic head development in *Drosophila*." *Mech.Dev.* . **56** : 3-15.

Gautier, J., C. Norbury, M. Lohka, P. Nurse, and J. Maller.(1988). "Purified maturation-promoting factor contains the product of a *Xenopus* homolog of the fission yeast cell cycle control gene *cdc2+*." *Cell* . **54** : 433-439.

Gautier, J., M.J. Solomon, R.N. Booher, J.F. Bazan, and M.W. Kirschner.(1991). "cdc25 is a specific tyrosine phosphatase that directly activates p34cdc2." *Cell* . **67** : 197-211.

Gavis, E.R. and R. Lehmann.(1992). "Localization of *nanos* RNA controls embryonic polarity." *Cell* . **71** : 301-313.

Gilmore, T.D. and H.M. Temin.(1986). "Different localization of the product of the v-*rel* oncogene in chicken fibroblasts and spleen cells correlates with transformation by REV-T." *Cell* . **44** : 791-800.

Gould, A. P., R. Y. Lai, M. J. Green, and R. A. White.(1990). "Blocking cell division does not remove the requirement for Polycomb function in Drosophila embryogenesis." *Development* . **110** : 1319-25.

Gould, K.L. and P. Nurse.(1989). "Tyrosine phosphorylation of the fission yeast cdc2+ protein kinase regulates entry into mitosis." *Nature* . **342** : 39-45.

Gray, S., P. Szymanski, and M. Levine.(1994). "Short-range repression permits multiple enhancers to function autonomously within a complex promoter." *Genes & Development* . **8** : 1829-1838.

Grossniklaus, U., K. Cadigan, and W.J. Gehring.(1994). "Three maternal coordinate systems cooperate in the patterning of the *Drosophila* head." *Development* . **120** : 3155-3171.

Grossniklaus, U., R. Kurth Pearson, and W.J. Gehring.(1992). "The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors." *Genes Dev.* . **6** : 1030-1051.

Häcker, U., E. Kaufmann, C. Hatmann, G. Jürgens, W. Knochel, and H. Jäckle.(1995). "The *Drosophila fork head* domain protein crocodile is required for the establishment of head structures." *EMBO J.* . **14** : 5306-5317.

Hagan, I., L. Hayles, and P. Nurse.(1988). "Cloning and sequencing of the cyclin-related *cdc13+* gene and a cytological study of it's role in fission yeast mitosis." *J. Cell Sci.* . **91** : 587-595.

Harlow, E and D Lane. 1st ed., New York: Cold Spring Harbor, 1988.

Hartenstein, P. and J. W. Posakony.(1990). "A dual function of the *Notch* gene in *Drosophila* sensillum development." *Developmental Biology* . **142** : 13-30.

Hartenstein, V. and J.A. Campos-Ortega.(1984). "Early neurogenesis in wildtype *Drosophila melanogaster*." *Roux's Arch. Dev. Biol.* . **193** : 308-325.

Hartenstein, V. and J.A. Campos-Ortega.(1985). "Fate mapping in wildtype *Drosophila melanogaster*. I. The pattern of embryonic cell divisions." *Roux's Arch. Dev. Biol.* . **194** : 181-195.

Hartenstein, V., A. Younossi-Hartenstein, and A. Lekven.(1994). "Delamination and division in the *Drosophila* neurectoderm - spatiotemporal pattern, cytoskeletal dynamics, and common control by neurogenic and segment polarity genes." *Developmental Biology* . **165** : 480-499.

Hiromi, Y. and W.J. Gehring.(1987). "Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*." *Cell* . **50** : 963-974.

Hoch, M., E. Seifert, and H. Jäckle.(1991). "Gene expression mediated by *cis*-acting sequences of the *Krüppel* gene in response to the *Drosophila* morphogens *bicoid* and *hunchback*." *EMBO J.* . **10** : 2267-2278.

Hoey, T. and M. Levine.(1988). "Divergent homeo box proteins recognise similar DNA sequences in *Drosophila*." *Naturc* . **332** : 858-861.

Hoey, T., O.J. Weinzierl, G. Gill, J. Chen, B.D. Dynlacht, and R. Tjian.(1993). "Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators." *Cell* . **72** : 247-260.

Huang, J-D., D.H. Schwyter, J.M. Shirokawa, and A.J. Courey.(1993). "The interplay between multiple enhancer and silencer elements defines the pattern of *decapentaplegic* expression." *Genes & Dev.* . **7** : 694-704.

Ingham, P.W.(1988). "The molecular genetics of embryonic pattern formation in *Drosophila*." *Nature* . **335** : 25-34.

Ip, Y.T., R.E. Park, D. Kosman, E. Bier, and M. Levine.(1992a). "The *dorsal* gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo." *Genes Dev.* . **6** : 1728-1739.

Ip, Y.T., R.E. Park, D. Kosman, K. Yazdanbakhsh, and M. Levine.(1992b). "*dorsal-twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo." *Genes Dev.* . **6** : 1518-1530.

Izumi, T. and J.L. Maller.(1993). "Elimination of the cdc2 phosphorylation sites in the cdc25 phosphataseblocks initiation of M-phase." *Mol. Biol. Cell* . **4** : 1337-1350.

Izumi, T., D.H. Walker, and J.L. Maller.(1992). "Periodic changes in phosphorylation of the Xenopus cdc25 phosphatase regulate its activity." *Mol. Biol. Cell* . **3** : 927-939.

Jiang, J., D. Kosman, Y.T. Ip, and M. Levine.(1991). "The *dorsal* morphogen gradient regulates the mesoderm determinant *twist* in early *Drosophila* embryos." *Genes Dev.* . **5** : 1881-1891.

Jiang, J. and M. Levine.(1993). "Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the *dorsal* gradient morphogen." *Cell* . **72** : 741-752.

Jiang, J., C.A. Rushlow, Q. Zhou, S. Small, and M. Levine.(1992). "Individual *dorsal* morphogen binding sites mediate activation and repression in the *Drosophila* embryo." *EMBO J.* . **11** : 3147-3154.

Jimenez, J., L. Alphey, P. Nurse, and D.M. Glover.(1990). "Complementation of fission yeast *cdc2*^{ts} and *cdc25*^{ts} mutants identifies two cell cycle genes from *Drosophila*: a *cdc2* homologue and *string*." *EMBO J.* . **9** : 3565-3571.

Jürgens, G. and V. Hartenstein. "The terminal regions of the body pattern." In The Development of *Drosophila melanogaster*, ed. M. Bate and A. Martinez Arias. 687-746. 1. Plainview, New York: Cold Spring Harbour Laboratory Press, 1993.

Jürgens, G., E. Wieschaus, C. Nüsslein-Volhard, and H. Kluding.(1984). "Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. II. Zygotic loci on the third chromosome." *Willhelm Roux's Arch. Dev. Biol.* . **193** : 283-295.

Kadonaga, J.T., K. Carner, F.R. Masiarz, and R. Tjian.(1987). "Isolation of cDNA encoding transcription factor and functional analysis of the DNA binding domain." *Cell* . **51** : 1079-1090.

Kakizuka, A., B. Sebastian, U. Borgmeyer, I. Hermans-Borgmeyer, J. Bolado, T. Hunter, M.F. Hoekstra, and R.M. Evans.(1992). "A mouse *cdc25* homologue is differentially and developmentally expressed." *Genes and Development* . **6** : 578-590.

Kalionis, B. and P.H. O'Farrell.(1993). "A universal target sequence is bound *in vitro* by diverse homeodomains." *Mech. Develop.* . **43** : 57-70.

Kane, D.A., R.M. Warga, and C.B. Kimmel.(1992). "Mitotic domains in the early embryo of the zebrafish." *Nature* . **360** : 735-737.

Kasai, Y., J.R. Nambu, P.M. Lieberman, and S.T. Crews.(1992). "Dorsal-ventral patterning in *Drosophila*: DNA binding of snail protein to the *single-minded* gene." *Proc.Nat.Acad.Sci. USA* . **89** : 3414-3418.

King, R.W., P.K. Jackson, and M.W. Kirschner.(1994). "Mitosis in transition." *Cell* . **79** : 563-571.

Knoblich, J.A., K. Sauer, L. Jones, H. Richardson, R. Saint, and C.F. Lehner.(1994). "*Cyclin E* Controls S Phase Progression and Its Down-Regulation during *Drosophila* Embryogenesis Is required for the Arrest of Cell Proliferation." *Cell* . **77** : 107-120.

Kornbluth, S., B. Sebastian, T. Hunter, and J. Newport.(1994). "Membrane localization of the kinase which phosphorylates p34cdc2 on threonine 14." *Mol. Biol. Cell* . **5** : 273-282.

Kraut, R., W. Chia, L.Y. Jan, Y.N. Jan, and J.A. Knoblich.(1996). "Role of *inscuteable* in orienting assymetric cell divisions in *Drosophila*." *Nature* . **383** : 50-55.

Krek, W. and E.A. Nigg.(1991). "Mutations of p34cdc2 phosphorylation sites induce premature mitotic events in HeLa cells: evidence for a double block to p34cdc2 kinase activation in vertebrates." *EMBO J.* . **10** : 3331-3341.

Kumagai, A. and W.G. Dunphy.(1992). "Regulation of the cdc25 protein during the cell cycle in *Xenopus* extracts." *Cell* . **70** : 139-151.

Leptin, M.(1991). "*twist* and *snail* as positive and negative regulators during *Drosophila* mesoderm development." *Genes Dev.* . **5** : 1568-1576.

Leptin, M. and B. Grunwald.(1990). "Cell shape changes during gastrulation in *Drosophila*." *Development* . **110** : 73-84.

Lindsley, D.L. and G.G. Zimm. The Genome of *Drosophila melanogaster*. San Diego: Academic Press, 1992.

Lucibello, F.C., M. Truss, J. Zwicker, F. Ehlert, M. Baeto, and R. Muller.(1995). "Periodic *cdc25C* transcription is mediated by a novel cell cycle-regulated repressor element." *EMBO J.* . **14** : 132-142.

Lundgren, K., N. Walworth, R. Booher, M. Dembski, M. Kirschner, and D. Beach.(1991). "mik1 and wee1 cooperate in the inhibitory tyrosine phosphorlation of cdc2." *Cell* . **64** : 1111-1122.

Matsuoka, M., J.-Y. Kato, R.P. Fisher, D.O. Morgan, and C.J. Sherr.(1994). "Activation of the cyclin-dependent kinase-4 (CDK4) by mouse MO15-associated kinase." *Mol. Cell. Biol.* . **14** : 7265-7275.

Mauhin, V., Y. Lutz, C. Dennefeld, and A. Alberga.(1993). "Definition of the DNA-binding site repertoire for the *Drosophila* transcription factor *snail*." *Nucleic Acids Research* . **21** : 3951-3957.

McKnight, S.L. and O.L. Miller Jr.(1977). "Electron microscopic analysis of chromatin replication in the cellular blastoderm *Drosophila melanogaster* embryo." *Cell* . **12** : 795-804.

Merli, C., D.E. Bergstrom, J.A. Cygan, and R.K. Blackman.(1996). "Promoter specificity mediates the independent regulation of neighbouring genes." *Genes & Dev.* . **10** : 1260-1270.

Millar, J.B.A., C.H. McGowan, A. Bueno, K. Sadhu, R. Jones, H. Richardson, S.I. Reed, and P. Russell.(1991). "*cdc25* M-phase inducer." *Cold Spring Harbor Symposia on Quantitative Biology* . **56 (LVI)** : 577-584.

Moreno, S. and P. Nurse.(1991). "Clues to the action of cdc25 protein." *Nature* . **351** : 194.

Moreno, S., P. Nurse, and P. Russell.(1990). "Regulation of mitosis by cyclic accumulation of p80cdc25 mitotic inducer in fission yeast." *Nature* . **344** : 549-552.

Mueller, P.A., T.R. Coleman, A. Kumagai, and W.G. Dunphy.(1995). "Myt1: a membrane-associated inhibitory kinase that phosphorylates cdc2 on both threonine 14 and tyrosine 15." *Science* . **270** : 86-90.

Murphy, G. and T. Kavanagh.(1988). "Speeding-up the sequencing of double-stranded DNA." *Nucleic Acids Res.* . **16** : 5198.

Murray, A.W. and M.W. Kirschner.(1989). "Cyclin synthesis drives the early embryonic cell cycle." *Nature* . **339** : 275-280.

Myers, F. A., H. Francislang, and S. F. Newbury.(1995). "Degradation of maternal *string* mRNA is controlled by proteins encoded on maternally contributed transcripts." *Mechanisms of Development* . **51** : 217-226.

Nambu, J.R., R.G. Franks, S. Hu, and S.T. Crews.(1990). "The *single-minded* gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells." *Cell* . **63** : 63-75.

Nambu, J.R., J.O. Lweis, Jr. K.A. Wharton, and S.T. Crews.(1991). "The drosophila *single-minded* gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development." *Cell* . **67** : 1157-1167.

Norbury, C., J. Blow, and P. Nurse.(1991). "Regulatory phosphorylation of the p34cdc2 protein kinase in vertebrates." *EMBO J.* . **10** : 3321-3329.

Nurse, P.(1990). "Universal control mechanism regulating onset of M-Phase." *Nature* . **344** : 503-508.

Nurse, P., P. Thuriaux, and K.A. Nasmyth.(1976). "Genetic control of the cell division cycle in the fission yeast Schizosaccharomycespombe." *Mol. Gen. Genet.* . **146** : 167-178.

O'Connell, M.J., A.H. Osmani, N.R. Morris, and S.A. Osmani.(1992). "An extra copy of *nim*E^{cyclinB} elevates pre-MPF levels and partially suppresses mutation of *nimT*^{cdc25} in *Aspergillus nidulans*." *EMBO J.* . **11** : 2139-2149.

O'Farrell, P.H.(1992). "Big genes and little genes and deadlines for transcription." *Nature* . **359** : 366-367.

O'Farrell, P.H., B.A. Edgar, D. Lakich, and C.F. Lehner.(1989). "Directing Cell Division During Development." *Science* . **246** : 635-640.

Pan, D. and A.J. Courey.(1992). "The same *dorsal* binding site mediates both activation and repression in a context-dependent manner." *EMBO J.* . **11** : 1837-1842.

Pankratz, M. and H. Jäckle. "Blastoderm segmentation." In The Development of *Drosophila melanogaster*, ed. M. Bate and A. Martinez Arias. 467-516. 1. Plainview, New York: Cold Spring Harbour Laboratory Press, 1993.

Pankratz, M.J., E. Seifert, N. Gerwin, B. Billi, U. Nauber, and H. Jäckle.(1990). "Gradients of *Krüppel* and *knirps* gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo." *Cell* . **61** : 309-317.

Parker, L.L., S.A. Walter, P.G. Young, and H. Piwnica-Worms.(1993). "Phosphorylation and inactivation of the mitotic inhibitor wee1 by the nim1/cdr1 kinase." *Nature* . **363** : 736-738.

Poon, R.Y.C., K. Yamashita, J.P. Adamczewski, T. Hunt, and J. Shuttleworth.(1993). "The cdc2-related protein $p40^{MO15}$ is the catalytic subunit of a protein kinase that can activate $p33^{cdk2}$ and $p34^{cdc2}$." *EMBO J.* . **12** : 3123-3132.

Prestridge, D.S.(1996 in press). "SIGNAL SCAN 4.0: New databases and sequence formats." *CABIOS* . :

Pugh, B.F.(1996). "Mechanisms of transcription complex assembly." *Current Opinion in Cell Biology* . **8(3)** : 303-311.

Richardson, H.E., L.V. O'Keefe, S.I. Reed, and R. Saint.(1993). "A *Drosophila* G_1-specific *cyclin E* homolog exhibits different modes of expression during embryogenesis." *Development* . **119** : 673-690.

Riddihough, G. and D. Ish-Horowicz.(1991). "Individual stripe regulatory elements in the *Drosophila hairy* promoter respond to maternal, gap, and pair-rule genes." *Genes Dev.* . **5** : 840-854.

Roth, S., D. Stein, and C. Nusslein-Volhard.(1989). "A gradient of nuclear localization of the *dorsal* protein determines dorsoventral pattern in the *Drosophila* embryo." *Cell* . **59** : 1189-1202.

Rothe, M., M. Pehl, H. Taubert, and H. Jäckle.(1992). "Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo." *Nature* . **359** : 156-159.

Rubin, G.M. and A.C. Spradling.(1982). "Genetic transformation of *Drosophila* with transposable element vectors." *Science* . **218** : 348-353.

Rushlow, C. and K. Arora.(1990). *Semin.Cell Biol.* . **1** : 137-149.

Rushlow, C.A., H. Kyuhyung, J.L. Manley, and M. Levine.(1989). "The graded distribution of the *dorsal* morphogen is initiated by selective nuclear transport in *Drosophila*." *Cell* . **59** : 1165-1177.

Russell, P., S. Moreno, and S.I. Reed.(1989). "Conservation of mitotic controls in fission and budding yeasts." *Cell* . **57** : 295-303.

Russell, P. and P. Nurse.(1986). "*cdc25*$^{+}$ functions as an inducer in mitotic control of fission yeast." *Cell* . **45** : 145-153.

Russell, P. and P. Nurse.(1987a). "The mitotic inducer *nim1*$^{+}$ functions in a regulatory network of protein kinase homologs controlling the initiation of mitosis." *Cell* . **49** : 569-576.

Russell, P. and P. Nurse.(1987b). "Negative regulation of mitosis by *wee1*$^{+}$, a gene encoding a protein kinase homologue." *Cell* . **49** : 559-567.

Sadhu, K., S.I. Reed, H. Richardson, and P. Russell.(1990). "Human homolog of fission yeast cdc25 mitotic inducer is predominantly expressed in G2." *Proc. Nat. Acad. Sci. USA* . **87** : 5139-5143.

Sambrook, J., Fritsch, E.F., & Maniatis, T. Molecular Cloning: A laboratory manual. 2nd ed., Cold Spring Harbor Laboratory Press, 1989.

Sanger, F., S. Nicklen, and A.R. Coulsen.(1977). "DNA sequencing with chain-terminating inhibitors." *Proc. Natl Acad. Sci. USA* . **74** : 5463-5467.

Shermoen, A.W. and P.H. O'Farrell.(1991). "Progression of the Cell Cycle through Mitosis Leads to Abortion of Nascent Transcripts." *Cell* . **67** : 303-310.

Simpson-Brose, M., J. Treisman, and C. Desplan.(1994). "Synergy between the *hunchback* and *bicoid* morphogens is required for anterior patterning in *Drosophila*." *Cell* . **78** : 855-865.

Small, S., R. Kraut, T. Hoey, R. Warrior, and Levine. M.(1991). "Transcriptional regulation of a pair-rule stripe in *Drosophila*." *Genes Dev.* . **5** : 827-839.

Smith, A.V. and T.L. Orr-Weaver.(1991). "The regulation of the cell cycle during *Drosophila* embryogenesis: the transition to polyteny." *Development* . **112** : 997-1008.

Smith, D.B. and K.S. Johnson.(1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase." *Gene* . **67** : 31-40.

Smythe, C. and J.W. Newport.(1992). "Coupling of mitosis to the completion of S phase in *Xenopus* occurs via modulation of the tyrosine kinase that phosphorylates p34cdc2." *Cell* . **68** : 787-797.

Solomon, M.J., M. Glotzer, T.H. Lee, M. Phillippe, and M.W. Kirschner.(1990). "Cyclin Activation of $p34^{cdc2}$." *Cell* . **63** : 1013-1024.

Solomon, M.J., J.W. Harper, and J. Shuttleworth.(1993). "CAK, the $p34^{cdc2}$ activating kinase, contains a protein identical or closely related to $p40^{MO15}$." *EMBO J.* . **12** : 3133-3142.

Sonnenblick, B.P.(1950). "The early embryology of *Drosophila melanogaster.*" *Biology of Drosophila (ed. M. Demerec) John Wiley, New York.* . : 62-163.

Sorger, P.K. and A.W. Murray.(1992). "S-phase feedback control in budding yeast independent of tyrosine phosphorylation of p34cdc28." *Nature* . **355** : 365-368.

Steward, R.(1989). "Relocalization of the *dorsal* protein from the cytoplasm to the nucleus correlates with its function." *Cell* . **59** : 1179-1188.

Strausfeld, U., J.C. Labbe, D. Fesquet, J.C. Cavadore, Picard. A., K. Sadhu, P. Russell, and M. Doree.(1991). "Dephosphorylation and activation of a p34cdc2/cyclinB complex *in vitro* by human cdc25 protein." *Nature* . **351** : 242-245.

Su, W., S. Jackson, R. Tjian, and H. Echols.(1991). "DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1." *Genes Dev* . **5** : 820-6.

Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, A. Kienlin, K. Jones, and H. Jäckle.(1987). "Finger protein of novel structure encoded by *hunchback* a second member of the gap class of *Drosophila* segmentation genes." *Nature* . **327** : 383-389.

Tautz, D. and C. Pfeifle.(1989). "A non-radioactive *in situ* hybridisation methods for the localisation of specific RNAs in the *Drosophila* embryo reveals translational control of the segmentation gene *hunchback.*" *Chromosoma* . **948** : 81-85.

Technau, G.M. and J.A. Campos-Ortega.(1986). "Lineage analysis of translated individual cells in embryos of *Drosophila melanogaster.* III. Commitment and proliferative capabilities of pole cells and midgut progenitors." *Roux's Arch. Dev. Biol.* . **195** : 445-454.

Thisse, B., C. Stoetzel, C. Gorostiza-Thisse, and F. Perrin-Schmitt.(1988). "Sequence of the *twist* gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos." *EMBO J.* . **7** : 2175-2183.

Thisse, C., F. Perrin-Schmitt, C. Stoetzel, and B. Thisse.(1991). "Sequence-specific transactivation of the *Drosophila twist* gene by the *dorsal* gene product." *Cell* . **65** : 1191-1201.

Turner, F.R. and A.P. Mahowald.(1976). "Scanning electron microscopy of Drosophila embryogenesis I. The structure of the egg envelopes and the formation of the cellular blastoderm." *Developmental Biology* . **68** : 96-109.

Waldorf, U. and W.J. Gehring.(1992). "*Empty spiracles*, a gap gene containing a homeobox involved in *Drosophila* head development." *EMBO J.* . **11** : 2247-2259.

Weigmann, K. and C.F. Lehner.(1995). "Cell fate specification by *even-skipped* expression in the *Drosophila* nervous system is coupled to cell cycle progression." *Development* . **121** : 3713-3721.

Wickramasinghe, D., S. Becker, M. K. Ernst, J. L. Resnick, J. M. Centanni, L. Tessarollo, L. B. Grabel, and P. J. Donovan.(1995). "Two *cdc25* homologues are differentially expressed during mouse development." *Development* . **121** : 2047-2056.

Wimmer, E., H. Jackle, C. Pfeifle, and S.M. Cohen.(1993). "A *Drosophila* homologue of human Sp1 is a head specific segmentation gene." *Nature* . **366** : 690-694.

Wimmer, E., M. Simpson-Brose, S.M. Cohen, C. Desplan, and H. Jäckle.(1995). "*Trans-* and *cis*-acting requirements for blastodermal expression of the head gap gene *buttonhead*." *Mech.Dev.* . **53** : 235-245.

Wu, L. and P. Russell.(1993). "Nim1 kinase promotes mitosis by inactivating Wee1 tyrosine kinase." *Nature* . **363** : 738-741.

Wu, S. and D. J. Wolgemuth.(1995). "The distinct and developmentally regulated patterns of expression of members of the mouse *cdc25* gene family suggest differential functions during gametogenesis." *Developmental Biology* . **170** : 195-206.

Yanisch-Perron, C., J. Viera, and J. Messing.(1985). *Gene* . **33** : 103-119.