Chapter 4: Dri DNA binding specificity

It would be very rash to call these bodies (bacteriophage) *genes, and yet at present we must confess that there is no distinction known between the genes and them. Hence we cannot categorically deny that perhaps we may be able to grind genes in a mortar and cook them in a beaker after all. Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so.*

- H. J. Müller (1922). American Naturalist **56**: 32-50

Introduction

One of the main objectives of this study of *dead ringer* was to identify its DNA binding domain. Characterisation of the *dri* transcript suggested that it was unlikely to be translated in the frame predicted to generate the DNA-binding activity (see Chapter 3). Consequently it became imperative to establish whether this DNA-binding function was in fact associated with the *dri* product and, if so, to explain how the original clone bk60 might have been isolated if it was expressed in a different frame. The Dri peptide had been found to contain a novel, highly conserved motif that was also seen in a variety of other peptides that were known or predicted to bind DNA (see Chapter 3). It was therefore of great interest to determine whether this region conferred their DNA-binding function and, if so, to determine whether this function had the provocative specificity observed for Bk60.

To test which frame coded for the DNA-binding activity, Glutathione-S-Transferase (GST) fusions with the *dri* cDNA were generated for each frame and at several positions. The original clone for which DNA-binding had been demonstrated was reproduced as a GST fusion identical to the original *lacZ* fusion (a fusion in frame 3 entering *dri* at base 544, called GST1-Bk60). GST fusions with clone Φ10 were generated in all three frames to produce fusions containing most of the coding region, including the ORF expressed by GST1-Bk60. A fusion construct expressing just the conserved region (amino acids 258 to 410) was kindly provided by R.D. Kortschak. As

a positive control, the homeodomain from *engrailed* was also cloned into the GEX expression vector. These fusion proteins were expressed in bacteria and crudely purified by taking the soluble fraction of the bacterial lysate. All of the fusion proteins were readily soluble with the exception of GST1-Bk60 and the almost identical GST2-Φ10 (result not shown).

DNA binding assays

The activity observed for Bk60 bound most strongly to oligomers of the NP consensus Engrailed binding site (Kalionis and O'Farrell, 1993), so a trimer of this sequence (NP_3) was initially used as a probe in gel shift assays for DNA-binding activity. These assays revealed that GST-Bk60 did not bind this probe, but either clone that expressed the conserved region in RF2 gave strong retardation of the NP_3 probe (Fig 4.1). Repeated attempts using varied binding conditions and a longer probe (NP_6) failed to detect any specific DNA-binding activity for GST-Bk60, which could at best retain a small amount of probe non-specifically in the well (eg. Fig 4.1, track 4). In contrast, even small amounts of purified conserved domain fusion on a glutathione-agarose matrix could specifically retain labelled NP_3 (result not shown). Several truncations of the conserved domain were then made to determine how much of the conserved region was required for binding. A construct containing the region conserved with Bright (Fig. 3.6, residues 256 to 410) was capable of binding; shorter constructs (287 to 369 or 307 to 369) showed no binding activity (results not shown). These experiments strongly suggested that the conserved region was a novel DNA binding domain that could specifically bind a homeo domain site.

Although GST-Bk60 could non-specifically aggregate NP_3 to a limited degree, it clearly did not produce the strong specific binding observed by Kalionis and O'Farrell. Instead, this activity was observed from two fusions which expressed the highly conserved motif described in Chapter 3. The simplest explanation for these results was that bk60 had in fact expressed the conserved domain rather than the short RF3 peptide

Figure 4.1 DNA binding by Dri polypeptides

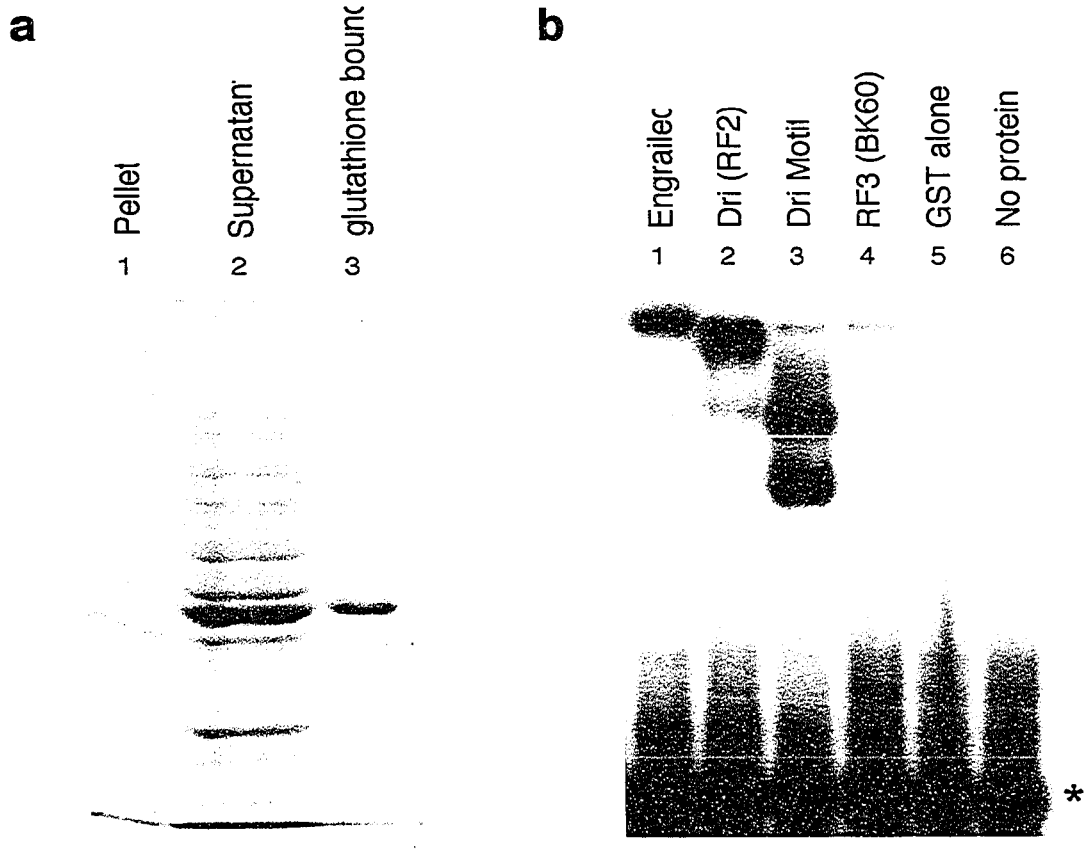


Figure 4.1 Dri protein expressed from reading frame two is able to retard labelled NB oligonucleotide. a) Bacterially expressed frame 2 protein fused t Glutathione-s-Transferase (GST) is shown to be soluble and to bind to a glutathione agarose column: track 1: insoluble pellet following lysis and centrifugation; track 2: supernatant; track 3: protein bound and eluted from a glutathione agarose column. b) Bacterially expressed protein fused to Glutathione-s-Transferase (GST) is shown to retard a trimer of the Engrailed consensus binding site TCAATTAAATGA. Unbound oligo is marked with a asterisk. Strong binding was observed from the second reading frame fusion protein (track 2), in contrast to RF3, which gave only limited retardation in the loading well (track 4). A152 residue fusion containing the widely conserved motif expressed in RF2 was sufficient to confer DNA binding (track 3). An Engrailed homeo domain-GST fusion was used as a positive control for binding (track 1); tracks 5 and 6 show that GST alone does not not bind DNA and that DNA is not retarded in the absence of protein.

predicted by the junction sequence[1]. This could have occurred in two ways. Although the junction sequence of bk60 indicated a RF3 fusion, there may have been a frame shift mutation further upstream in *lacZ* in this clone, so that RF2 was actually produced. A more plausible explanation comes from the observation that the λgt11 vector is prone to slippage, expressing the two other frames at levels as high as one eighth of that seen for the predicted frame (Young and Davis, 1983). Consequently it seems most likely that bk60 was isolated due to the relatively small amounts of RF2 being expressed due to slippage. Any such production of the longer RF2 fusion protein could be detected by Western Blot analysis of bk60 lysogens, although of course this result would not directly contribute to the characterisation of *dri* .

Having established that the protein product of *dead ringer* had DNA binding activity, it was necessary to assess its specificity. A wide variety of homeo domain proteins have been shown to bind related sites with similar preferences, and Bk60 shared these preferences (Kalionis and O'Farrell, 1993). In a preliminary test to confirm that the DNA-binding domain from Dri had a similar specificity, affinity for two related sites (NP_3 and TTA_9) was measured. Like many homeo domains (Kalionis and O'Farrell, 1993), the Dri DNA binding domain bound the NP site more strongly than it bound a poly-TTA site, which contains only the core recognition bases (ATTA) of the NP site (Fig. 4.2). Competition with 1000-fold excess of TTA_9 oligo did not prevent binding to the NP_3 sites, indicating that this conserved domain shows a strong preference for an En consensus homeo domain binding site (Fig. 4.2). The reverse experiment showed that TTA_9 binding was only just detectable with elevated protein levels and was effectively competed by a 10-fold excess of NP_3 oligo (results not shown).

Optimal binding site selection

While binding to the NP and TTA sequences clearly demonstrated the homeo domain-like specificity of the Dri DNA binding domain, the discovery of the NP site as a

[1]The alternative was that two ORFs from the same transcript coded for unrelated DNA binding peptides that bound the same sequence, and that the RF3 peptide could bind as a LacZ, but not GST fusion.

Figure 4.2 The Dri domain specifically binds an En site

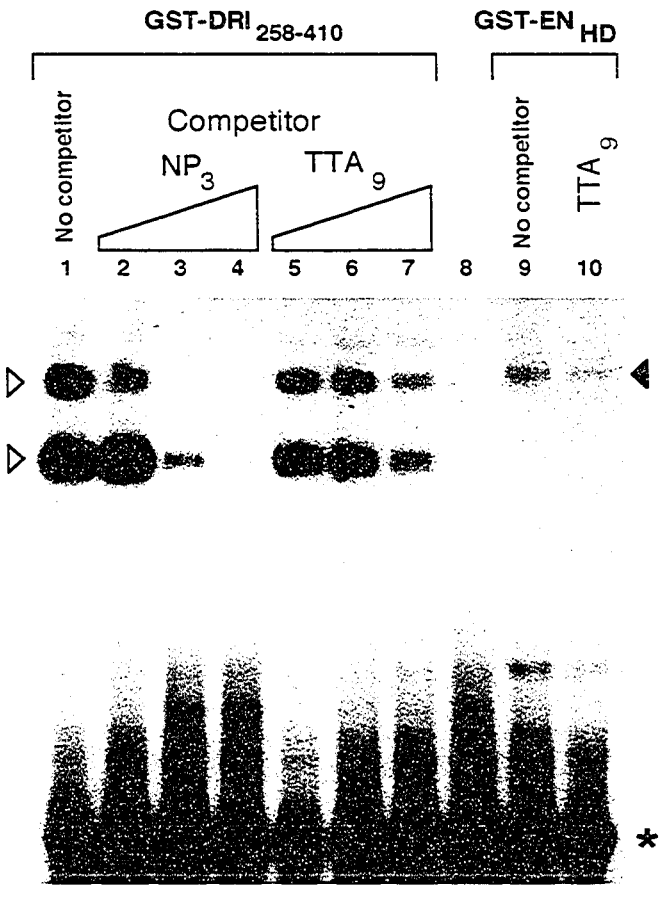


Figure 4.2 The conserved domain from Dri expressed as a GST fusion is able to specifically retard labelled NP_3 DNA. Unbound oligo is marked with an asterisk. All tracks contain 1000-fold excess of non-specific competitor DNA. DNA binding by GST-Dri protein to labelled NP_3 is competed for by increasing amounts of unlabelled specific competitor; either NP_3 :10-fold (track 2), 100-fold (track 3) or 1000-fold excess (track 4), or the variant site TTA_9 : 10-fold (track 5), 100-fold (track 6) or 1000-fold excess (track 7). 1000-fold excess of NP_3 competes away binding (track 4) whereas the same concentration of TTA_9 (track 7) has little effect. Dri-bound oligo migrates at two positions (open arrowheads), suggesting binding of one or two molecules per trimer binding site. The preference of the Engrailed homeo domain for the consensus siteis shown in tracks 9 and 10, where 1000-fold excess of TTA_9 does not eliminate binding to the NP site (closed arrowhead). Track 8 shows that GST alone does not retard the NP site.

potential target was fortuitous (Kalionis and O'Farrell, 1993). Other variants of the homeo domain site may have had higher affinity for Dri, or it might also bind a divergent alternative site, as some homeo domains do (Hoey and Levine, 1988). To test these possibilities, the preferred binding site for Dri was determined by selection and amplification of random oligomers essentially as described by Wilson et al. (Wilson et al., 1993). Briefly, the conserved region fusion protein was bound to glutathione-agarose beads and incubated with a pool of oligomers containing a stretch of 20 random bases. Unbound oligos were washed off, then the remaining pool amplified by PCR. This process was repeated seven times then the remaining oligomers were cloned and sequenced (Fig 4.3A).

The sequences obtained by this process were analysed for the presence of base patterns occurring more frequently than would be expected by chance. An algorithm to detect the frequency of every pattern in the sample (kindly designed by M. Hurd) revealed that the longest strings occurring significantly more often than would be expected by chance were GATTAA and AATTAA. Either or both of these sites were found in 21 of 42 sequences with a probability of $1x10^{-5}$ of this having occurred by chance. The incidence of smaller subsets of this pattern was even more striking with 35/42 containing G/AATTA. No alternative patterns of this length were detected with frequencies placing them above the $p<0.05$ significance cutoff. This statistical analysis indicated that the high incidence of the ATTA site did not arise by chance, implying that incubation with Dri had specifically selected them. Since each sequence could have been selected on the basis of their similarity to A/GATTA, they were aligned to best fit this site (Fig 4.3A). The high level of selection for the purineATTAA consensus is shown in graphical form in Fig. 4.3B. From these data it became obvious that the conserved region from Dri preferentially bound a consensus with striking similarity to the NP En homeo domain binding site.

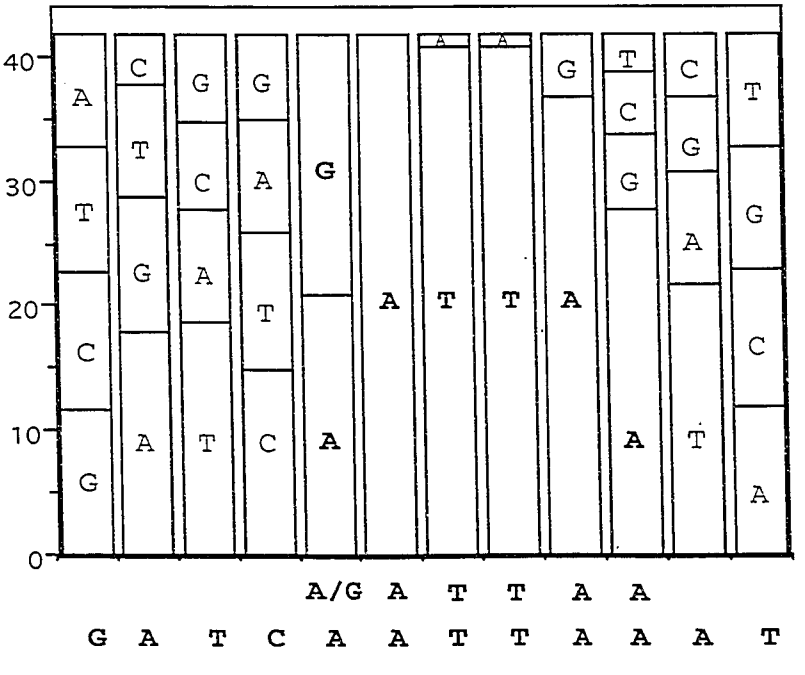
Figure 4.3 DNA-binding site consensus for the $Dri_{258-410}$ peptide. (**A**) Aligned sequences of random oligonucleotides selected by GST-$Dri_{258-410}$ protein binding. The random section of 20 bases plus 6 flanking bases of the primer are shown in each case. These 42 sequences were aligned with the statistical consensus of PuATTAA where Pu stands for either purine base. The homeodomain 'core' binding site ATTA is marked in bold. (**B**) Consensus diagram to indicate the frequency of each base at each position across the aligned region. Positions at which there has been strong selection have the predominant base(s) marked in bold. The consensus binding site for the Dri DNA-binding domain is indicated below the figure, aligned with the NP Engrailed binding site known to specifically bind Dri.

Figure 4.3 The Dri consensus binding site

A

CATCAACGTGATTG**ATTA**GCCCAGTCCCTCAG
CTGAGGCTGATCAAGTGATTAG**ATTA**TTGATG
CATCAATCCTTCGA**ATTA**CATTATACCCTCAG
CATCA**ATTA**ATAAGGTTCAGCTGCACCCTCAG
CTGAGGCGGACGTGCTAATG**ATTA**AATTGATG
CATCAATCCTATTG**ATTA**ATTGGTGCCCTCAG
CATCA**ATTA**ATATCTCCTTGACTGTGCCTCAG
CATCAATCAAACATCA**ATTA**ACATCCCCTCAG
CATCAATGAGGTGA**ATTA**ATTCATGTCCTCAG
CATCAAGAG**ATTA**ATTTCCCCGCCCCCCTCAG
CTGAGG**ATTA**ATCGGGACTTGATTGATTGATG
CATCAAGATCGATCTG**ATTA**CGATCCCCTCAG
CATCA**ATTA**GACCACACGTATCTTGCCCTCAG
TCCATATGATGTTCCAG**ATTA**TGCTGCCTCAG
CTGAGGGGGTGGGCG**ATTA**AATCAGTTTGATG
CATCAATAA**ATTA**GAATTAAAACTCGCCTCAG
CATCAATTTGGCG**ATTA**AAACCCGCCCCTCAG
CATCA**ATTA**AGACCACACGTATCTTGCCTCAG
CTGAGGAATCAAA**ATTA**GTGTGTCAATTGATG
CATGAGGGAAG**ATTA**ATCGTGTTACCTTGATG
CTGAGGCGGGCAACGGAGTGAGTTG**ATT**GATG
CATCA**ATTA**GACCACACGTATCTTTCCCTCAG
CTGAGGTGCAAACA**ATTA**ATAGCTCCCCTCAG
CTGAGGATG**ATTA**ATGTCGCATCCGATTGATG
CTGAGGGGCGATCATAACGTGTTTG**ATT**GATG
CTGAGGCA**AATA**AGATCGGATCAACATTGATG
CTGAGGA**ATTA**ATCTGGAACATCATATGGATA
CTGAGGACGGGAGG**ATTA**ATCACAGCTTGATG
CTGAGGGGAGAG**ATTA**TGATCGGAGATTGATG
CATCA**ATTA**GTAAGGTTTGATTTCCCCCTCAG
CATCAATCTCA**ATTA**ATCGCCTCCCTCCTCAG
CTGAGGGGGGATGCCATA**ATTA**ATGCTTGATG
CATCAATCAGTCG**ATTA**CGTATGTCCCCTCAG
CTGAGGATAAGCGTATCAGGGTCAA**ATT**GATG
CATCAATCCTTCGA**ATTA**CATTATACCCTCAG
CATCAA**ATTA**CAATCGATTTCCCCCTCCTCAG
CTGAGGTCAATTTGATAAGGTTGCG**ATT**GATG
CATCAATG**ATAA**ACTGATTGATCCCGCCTCAG
CATCAATCTGTCGG**ATTA**ACCTGTACCCTCAG
CTGAGGATTA**ATTA**ATCCAGATACGATTGATG
CTGAGGCTATGATAAAGGATCGTTG**ATT**GATG
CATCA**ATTA**ACGGATTAAATCTCGATCCTCAG

B



Because one of the primers used in this selection fortuitously contained a site similar to this (NNTTGA) and because this sequence is also found in NP oligomers (AATTGA) it was possible that many of the oligos were selected by binding this site rather than the consensus obtained from the degenerate region. To test this, gel shift assays were performed using the consensus site and the AATTGA alternative as competitors. The conserved region fusion protein showed a considerably lower, although detectable, affinity for the alternative site since 10-fold more AATTGA trimer than consensus trimer was required to compete away 90% of the protein (Fig. 4.4). Selection of oligos with the consensus site may have been aided by the presence of a potential lower affinity site in the primer, however this site was not required (Fig 4.4 track 1), nor did it prevent selection of a clear consensus from the random bases (Fig 4.3B). Thus it was clear that the Dri conserved region preferentially bound a consensus target site of purineATTAA.

Summary

The result of this optimal binding site screen was consistent with the isolation of bk60 as an NP site binding clone (Kalionis and O'Farrell, 1993) and with the characterisation of the Dri DNA binding domain described above, indicating that the En consensus binding site is among the highest affinity targets of this novel, conserved DNA binding domain. Deletion analysis indicated that the core region of homology alone was insufficient for binding in the assays used. The extended region of homology (identified from the mouse protein Bright) was necessary and sufficient to confer specific binding of the NP site. The DNA binding specificity of the closely related domain from Bright has been recently been identified and contains a core recognition site of purineATa/tAA (Herrscher et al., in press). These results are consistent with the hypothesis that this novel family of DNA binding domains exhibit homeo domain-like specificities.

Figure 4.4 Dri specifically binds its consensus site

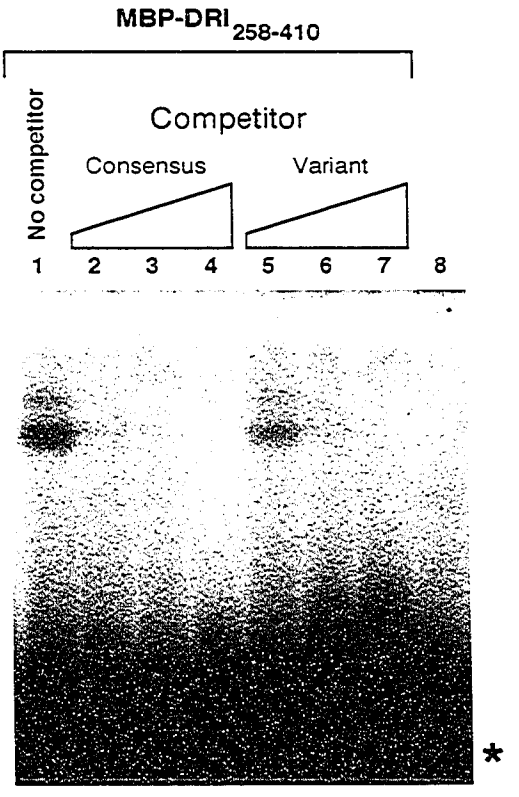


Figure 4.4 The conserved domain from Dri specifically binds the identified consensus sequence. The Dri polypeptide expressed as a maltose binding protein (MBP) fusion can retard a labelled trimer containing the identified consensus $(CCGATTAATCCC)_3$. All tracks contain 1000-fold excess of unlabelled non-specific competitor DNA. Unbound oligo is marked with an asterisk. Binding of MBP-Dri is competed for by increasing amounts of unlabelled specific competitor: either the consensus trimer in 10, 100 or 1000-fold excess (tracks 2-4) or the variant site trimer $(CCGATTGAATCCC)_3$ in 10, 100 or 1000-fold excess (tracks 5-7). 100-fold excess of variant competitor is required to compete to the same level as 10-fold excess of consensus site (compare tracks 2 and 6), indicating a substantial preference for the consensus site. Track 7 shows that MBP alone does not retard the consensus site.

Chapter 5. *dri* expression pattern and mutant phenotype

In a pedigree culture of Drosophila which had been running for nearly a year through a considerable number of generations, a male appeared with white eyes. The normal flies have brilliant red eyes.

The white-eyed male, bred to his red-eyed sisters, produced 1237 red-eyed offspring and 3 white-eyed males.

- T.H. Morgan (1910). Science 32: 120-122

Introduction

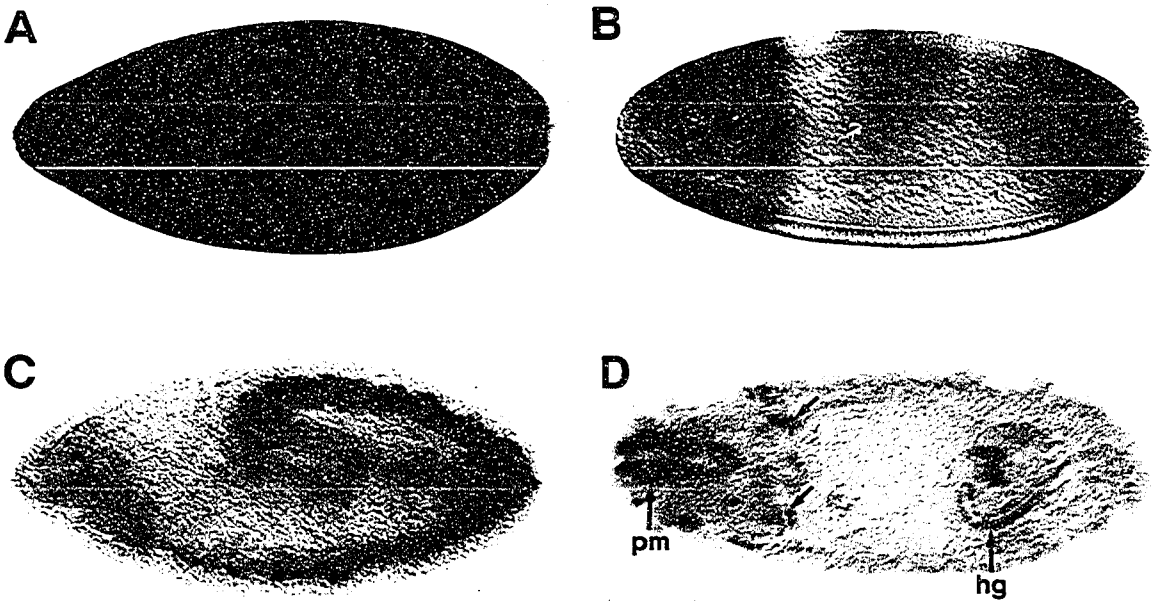
The previous chapters have described the identification and characterisation of a novel DNA binding domain in Dead ringer. Its in vitro specificity indicated that it was likely to target sites similar to those of homeo domain proteins. To have any chance of determining which homeo domains Dri might interact with in vivo and what the outcome of such interactions might be, it was necessary to establish the expression pattern and mutant phenotype of *dri*. Since the pattern of gene expression in *Drosophila* embryonic tissues is relatively well characterised it was anticipated that mutant analysis would also confirm that Dri affected gene regulation and suggest possible developmental roles.

Dri expression patterns during embryogenesis

The expression of *dri* mRNA during embrogenesis was examined by whole mount in situ hybridisation with labelled fragments of *dri* cDNA clones. Dri protein was detected immunochemically with a rat polyclonal antibody raised against a bacterially expressed GST-Dri fusion protein (provided by R.D. Kortschak). Both approaches revealed a dynamic pattern of expression during embryogenesis. Maternal mRNA was found to be distributed throughout the embryo during the syncitial cleavage divisions (Fig. 5.1A), then at cellularisation it was restricted to broad bands at each terminus and one central stripe (Fig. 5.1B). At germ band extension, *dri* mRNA was found

Figure 5.1 In situ localisation of *dri* mRNA during *Drosophila* embryogenesis using a digoxygenin labelled DNA probe. **A**. Stage 2: maternal *dri* mRNA is ubiquitous. **B**. Stage 5: *dri* message is restricted to two terminal bands and a wide dorsal stripe. **C**. Stage 9: *dri* message is found in the mesoderm. **D**. Stage 14: expression is localised to specific organs: the pharyngeal muscles (pm), hindgut (hg) and brain lobes are arrowed.

Figure 5.1 dead ringer mRNA expression



predominantly in the mesoderm (Fig. 5.1C), and at later stages was present in a diverse range of tissues including the hindgut, salivary ducts and pharyngeal muscles (Fig. 5.1D).

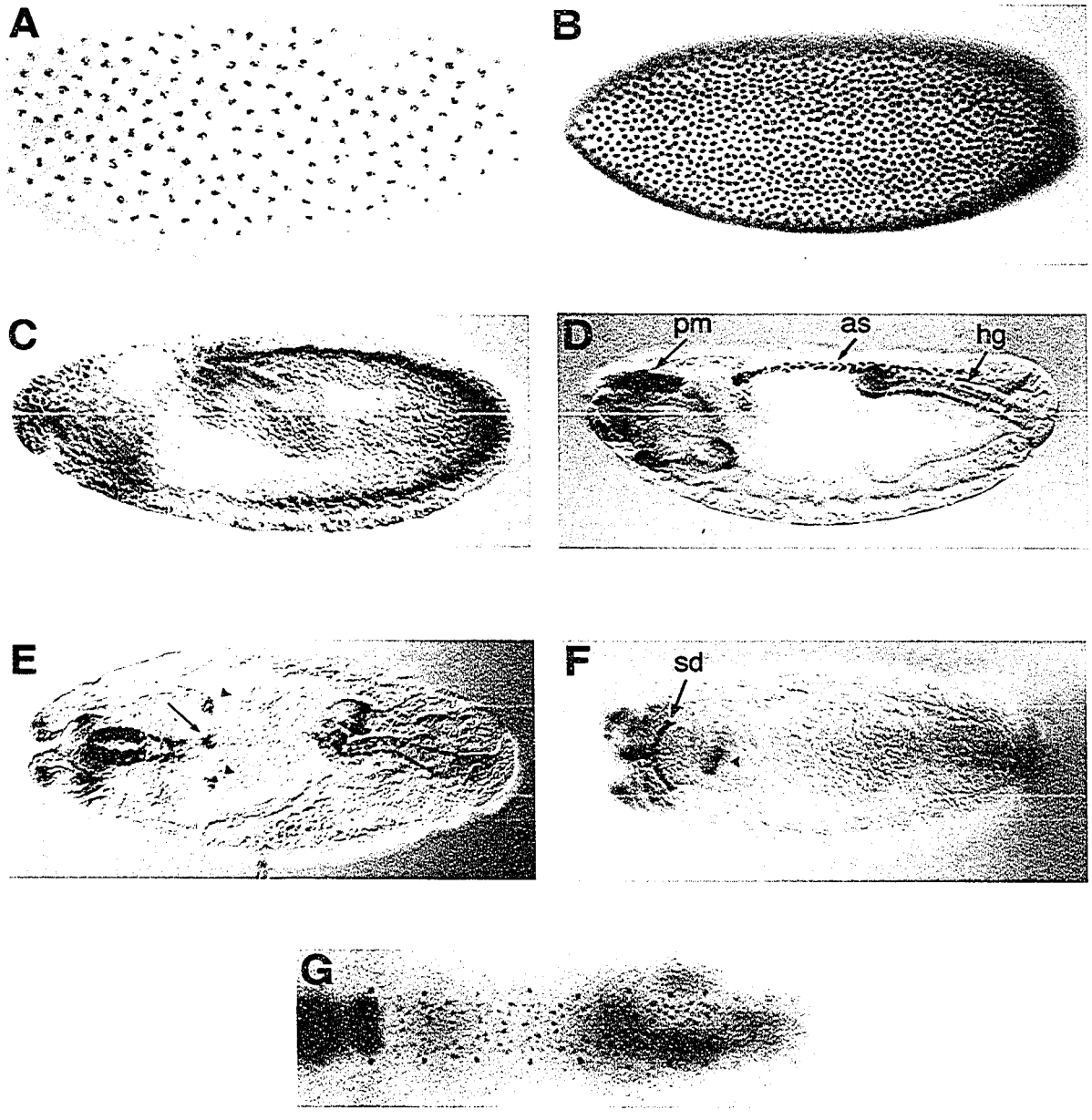
Dri protein was found to be nuclear localised wherever present (Fig. 5.2A). It was found evenly distributed among syncitial nuclei (Fig. 5.2B). The only instance in which the distribution of mRNA and protein differed was at cellularisation, when the stripes of mRNA contrasted with the ubiquitous distribution of protein (compare Fig 5.1B and 5.2B). This presumably reflected persistence of maternal protein after degradation of the maternal mRNA. At germband extension, protein distribution again reflected the mRNA pattern, both appearing primarily in the mesoderm (compare Fig. 5.1C and Fig. 5.2C). After germband retraction, Dri expression could be seen in the pharyngeal muscles, in discrete longitudinal rows of cells in the hindgut epithelium, in the amnioserosa (Fig. 5.2D), the corpora allata of the ring gland, a ring of cells at the midgut-hindgut junction and several cells in or near the mushroom bodies of the brain (Fig. 5.1E). Expression was also observed in all cells of the salivary gland ducts, but not in the salivary glands themselves, in a ring of cells at the foregut-midgut junction (Fig. 5.1F), and in a segmentally repeated pattern down the ventral nerve cord (Fig. 5.1G).

Mutant analysis

The chromosomal location of *dri* had been established at 59E-F by in situ hybridisation to polytenes (B. Kalionis, personal communication). Several strains containing enhancer-trap P-element insertions in this area were stained for LacZ, and one gave a pattern that closely resembled that of *dri* (R. Saint, personal communication). This strain, l(2)02535, was homozygous lethal and was tested for complementation of nearby deletion and insertion strains. These tests indicated that it failed to complement another insertion, l(2)05096, and deletions Df(2R)bw^{s46}, Df(2R)egl^{2}, Df(2R)bw^{5} and Df(2R)*tid*, but could complement Df(2R)*x32*, placing l(2)02535 and l(2)05096 in the chromosomal location 59F1-2 (see Materials). Isolation and mapping of the genomic

Figure 5.2 Immunostainings with polyclonal antibodies to Dri protein. **A.** Stage 2: Dri is nuclearly localised in the syncytium. This image has been enlarged ~1.5x. **B.** Stage 5: Unlike the mRNA pattern (Fig. 5.1B), Dri protein is still ubiquitous at cellularisation. **C.** Stage 10: Dri protein is found in the mesoderm. **D.** Stage 14: Dri expression is seen in the pharyngeal muscles (pm), hindgut (hg) and amnioserosa (as). **E.** Stage 15 dorsal view: Dri expression can be seen in the ring gland (arrow), clypeolabrum, a ring of cells around the hindgut/midgut junction and in the brain lobes (arrowheads). **F.** Stage 15 ventral view: Dri expression can be seen in the salivary ducts (sd) and in a ring of cells around the foregut/midgut junction (arrowhead). **H.** Dri is expressed in a repeated pattern through the central nervous system.

Figure 5.2 Dri protein expression



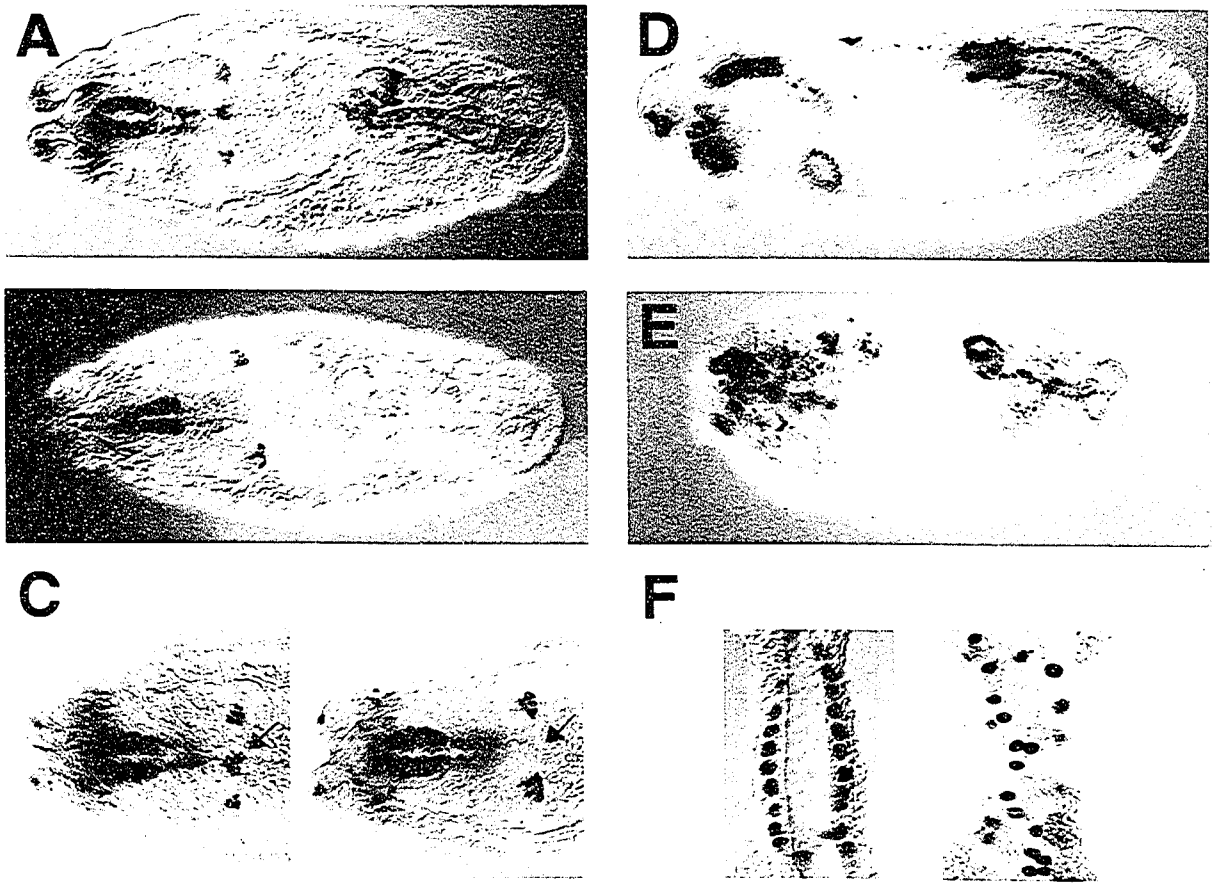
DNA flanking the two P-elements indicated that they were inserted within 1kb on either side of the first exon of *dri* (Fig 5.4C, R. Saint and R.D. Kortschak, person communication). These two non-complementing homozygous lethal insertions are therefore referred to as dri^{P1} (02535) and dri^{P2}(05096) and represent the first mutations known to affect the *dead ringer* gene. The lethality of both of these P-elements could be reverted at a high frequency (>10%) by crossing in a source of transposase (see Methods), confirming that the lethal phenotype was caused by the insertions (results not shown).

To better characterise the enhancers in and around *dri*, the LacZ expression patterns from both enhancer trap insertions were determined by immunostaining. As expected for zygotic expression of Dri (see above), neither strain gave any LacZ expression until germband extension, when P1 gave expression in the mesoderm (results not shown). After germband retraction, both strains gave expression in exactly the tissues which express Dri, for example the ring gland, amnioserosa, pharyngeal muscles, salivary ducts and hindgut stripes (Fig. 5.3A, D). P2 gave very strong expression in the amnioserosa and faint expression elsewhere, whereas P1 gave equivalent levels of expression in all tissues.

The mutant phenotypes of these insertions were determined by crossing in a *wg-lacZ* marked 2nd chromosome balancer, then staining the progeny for LacZ and Dri. Homozygous dri^{P}mutant embryos could then be identified by the lack of the characteristic *wg* striped pattern. Suprisingly, embryos homozygous for either P-insertion gave almost normal staining in the *dri* pattern for both LacZ and Dri until very late in embryogenesis (Fig. 5.3B, D). At stage 15 all of the tissues expressing Dri had formed apparently normally and were still expressing Dri. At stage 16, Dri expression was lost in the ring gland of dri^{P1}mutants, although LacZ expression was retained (Fig. 5.3C). At this stage Dri expression was still seen in the hindgut, but the previously ordered longitudinal lines of expressing cells were now disrupted and their cellular morphology was changed from

Figure 5.3 Immunostaining with polyclonal antibodies to LacZ or Dri on embryos carrying the dri^{P1} insertion. **A.** The expression pattern of the P-element enhancer trap in dri^{P1} heterozygotes at stage 15, dorsal view. **B.** Stage 16 dri^{P1} homozygote: Dri expression is normal except for effects on the ring gland and hindgut. **C.** The ring gland (arrowed) expresses Dri in dri^{P1} heterozygotes (left) but not homozygotes (right). **D.** Lateral view of LacZ enhancer trap expression at Stage 15 in a dri^{P1} homozygote shows no obvious abnormalities. **E.** LacZ expression in a homozygous dri^{P1} embryo late in embryogenesis, showing severe disruption and widespread loss of organ morphology.

Figure 5.3 Enhancer trap expression and phenotype



columnar epithelia to flatter cells with smaller nuclei (Fig. 5.3F). No homozygotes could be identified at stage 17, instead there were many highly disrupted, crumpled embryos with some LacZ staining but very poor organ morphology, suggesting that the lethality observed for these strains took effect very quickly at this stage (compare Fig. 5.3D and E). The cause of death at this stage is unclear, but neural and/or muscular defects are likely possibilites. From these phenotypes it was possible to conclude that *dri* was required to maintain correct hindgut morphology and that embryonic lethality in *dri* mutants was associated with widespread tissue disruption. Since both of these mutants retained Dri expression as detected by polyclonal antibody staining it is likely that they are not amorphic alleles.

It was of particular interest to identify a mutation that disrupted the earlier pattern of *dri* since these tissues (the mesoderm and cellular blastoderm) are genetically much better characterised and might allow detection of transcriptional regulation requiring *dri* function and thus identify a role for Dri. Dri expression was examined in embryos homozygous for Df(2R)*tid* (kindly provided by T. Shandala) since these embryos should contain no zygotically produced Dri and would therefore be expected to have a more extreme phenotype. This revealed that maternally provided protein persisted in the mesoderm during germband extension and in the amnioserosa until dorsal closure (results not shown). This suggested that even a complete zygotic null would not show an early phenotype, and that to assess the early requirement for *dri*, the maternal product would have to be removed. This was to be achieved by using an inducible site specific recombinase to generate homozygous *dri* mutant cells in the maternal germline (Chou and Perrimon, 1992). Neither of the insertions were good candidate mutations for this approach, since they did not eliminate all zygotic *dri* expression and could not be relied upon to remove all maternal expression. Thus it became necessary to generate amorphic *dri* alleles.

Transposon mutagenesis

Two transposable element insertions in *dri* were available, so the most straightforward method for isolating null mutations was to generate local deletions by imprecise excision of the transposons (see Methods). This was carried out with both insertions independantly, by crossing them to a source of transposase, then screening the progeny for loss of the P-element marker *rosy*, to identify transposition events. Approximately 1000 such flies were then individually crossed to dri^{P1} mutants to screen for excision events that removed *dri* function and were therefore lethal over the dri^{P1} chromosome. A total of 135 ry^{-} dri^{-} lines were generated by this process, mostly derived from the P1 insertion since even short 3' deletions from this point would remove the predicted starts of transcription and translation. To identify deletions that removed *dri*, but did not extend into the unknown flanking genes, Southern blot analysis was carried out on genomic DNA from each line using at least two probes from across the *dri* gene. Since both insertions were at the 5' end of the *dri* coding region, probes were chosen to detect deletions extending in the 3' direction.

None of the 90 deletions from P1 showed loss of DNA across the *dri* coding region (Fig. 5.4A and results not shown). All of these deletions were lethal as transheterozygotes with dri^{P1}, and had lost the *ry* marker from the insertion, leaving two possible explanations for the lack of deletions into *dri*. The transposase may have failed to completely excise the insertion in these strains, removing *ry* but leaving enough to retain the insertion phenotype. Alternatively, some feature of the insertion site or P-ends may have caused a strong bias toward 5' deletions, thus removing one or more essential *dri* enhancers, but not affecting the transcribed region. In either case, persistence with this strategy was unlikely to generate a conclusively amorphic *dri* allele.

The P2 insertion was located within the first exon, so deletions 3' from this point could prevent expression of Dri beyond residue 127 (R.D. Kortschak, personal communication). Several deletions from P2 were identified that extended 5' or 3' beyond

Figure 5.4 P-element deletion mutagenesis

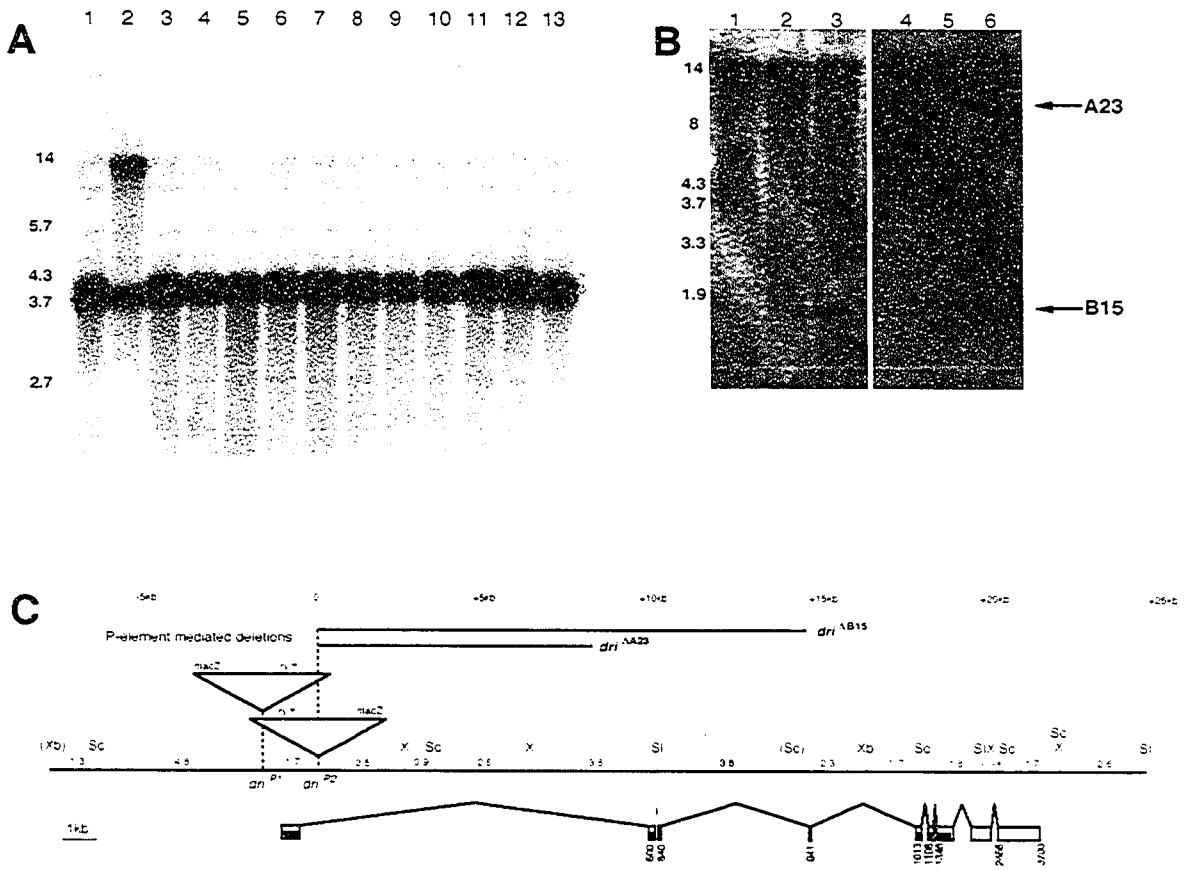


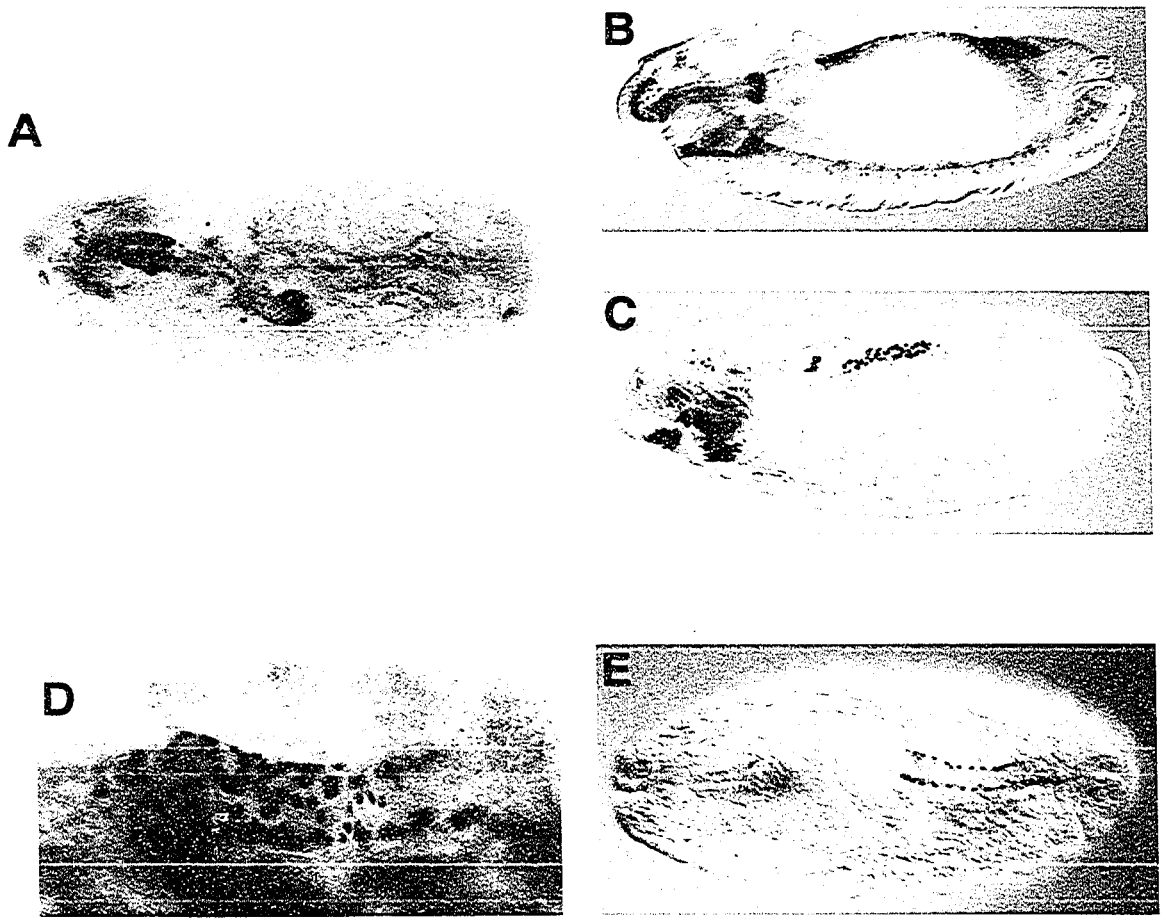
Figure 5.4. Southern analysis of P-element mediated deletions. **A**. Southern blot of genomic DNA from strains from which P1 had been mobilised, probed to identify a 4kb SacI genomic fragment covering the conserved motif in *dead ringer* (top band), and a 3.7kb fragment from *DmcycE* (bottom band) as a loading control. DNA from the parental dri $^{[1]}$ strain was included as a positive control for the presence of two copies of *dri* (track 1). DNA from the deletion strain Df(2R) *tid* was used to show loss of one copy of *dri* (track 2); the balancer chromosome in this strain shows a polymorphic Sac site, so in this strain the probe identifies a ~13kb band. The remaining tracks contain DNA from P1 deletion strains 108, 113, 116, 117, 118, 120, 136, 241, 256, 269 and 312. Quantitation for each strain (not shown) confirmed two intact copies of the *dri* fragment, indicating that mobilisation of the P-element has not deleted as far as the conserved motif in these strains. **B**. Genomic Southern blot of strains from which P2 had been mobilised, probed to show deletion breakpoints in a large XbaI fragment covering the 5' end of *dead ringer*. Tracks 1-6 show DNA from strains B31, B21, B15, A16, A18 and A23 respectively. Strain B15 shows a 1.7kb fragment (track 3) and strain A23 shows an 8.5kb fragment (track 6), localising the deletion breakpoints to the third and first introns respectively. **C**. Genomic map of the *dri* region. The breakpoints of deletions B15 and A23 are indicated relative to the two P-element insertion points and the genomic structure of *dri*, which were characterised by R. D. Kortschak. The dark shaded boxes represent the extent of the *dri* ORF. Restriction sites are abbreviated: Sc - SacI, Sl - SalI, X - XhoI, Xb - XbaI . Sites in parentheses are known to be polymorphic but are present on the P1 chromosome.

the *dri* coding region (results not shown); these were not characterised further because of the possible deletion of neighbouring genes. Only two lines were identified that carried deletions within *dri*: lines A23 and B15. The 3' breakpoints of these two deletions were localised to within the first and third introns respectively (see Fig. 5.4B, C). A23 was discarded since it did not remove any of the *dri* coding region, but B15 was crossed to the lacZ marked balancer to allow identification of its mutant phenotype. Surprisingly, its phenotype was indistinguishable from that of the insertion mutants (Fig. 5.5A and results not shown). Subsequent genomic sequencing indicated that splicing of the first and fourth exons (across the deletion) would retain the correct reading frame, allowing translation of the rest of the protein, including most of the conserved region (D. Kortschak, personal communication). Consequently, this mutation was not definitively null, and was therefore inappropriate for the generation of maternal knockouts. To avoid the expression of at least part of Dri, deletions would have to extend 5' of P2, but not into the next gene, the position of which is unknown. This presents a very small interval for useful deletions (see Fig. 5.4C), so experiments are in progress to generate a *dri* null by the more lengthy process of EMS mutagenesis.

Although it has not been possible to address the role of *dri* early in embryogenesis, the later phenotype allowed examination of *dri* function in at least the hindgut and ring gland. Unfortunately these two tissues are very poorly genetically characterised in the *Drosophila* embryo so no obvious candidates for *dri* regulation were available. Several genes are known to be involved in hindgut morphogenesis (*serpent, tailless, huckebein, T-related gene*), but they all act earlier, affecting formation of the gut (Strecker et al., 1988; Wiegel et al., 1990; Kispert et al., 1994; Reuter, 1994), which occurs normally in *dri* mutants (Figs 5.3D, 5.5A). The hindgut is missing from an early stage in such mutants (eg Fig. 5.5B and C), so it was not possible to assess any regulatory effect on *dri* in this tissue, nor could they be candidates for *dri* regulation. Since *engrailed* is expressed in half of the hindgut (Patel et al., 1989), it was possible that the two lines of Dri expressing cells formed the boundaries of Engrailed expression in a

Figure 5.5 Immunolocalisation of Dri and LacZ expression in several mutant backgrounds. **A.** LacZ expression in a $dri^{\Delta 15}/dri^{P1}$ transheterozygote shows this deletion gives the same hindgut phenotype seen for the insertion homozygotes (cf Fig. 5.2B, F). **B.** Dri expression in a *tll* mutant background showing complete loss of the hindgut. **C.** Dri expression in a *tor* mutant background showing complete loss of the hindgut and most of the head organs. **D.** Dri (black) and Engrailed (brown) localisation in the hindgut of a wild type embryo showing distinct and non-overlapping patterns of expression. **E.** LacZ expression from the enhancer trap in insertion strain p18-13 in the same set of hindgut stripes as seen for Dri.

Figure 5.5 Dri and the embryonic hindgut



process analogous to segmentation. Staining of embryos with antibodies to En and Dri revealed that the En expressing cells were morphologically distinct and none also expressed Dri (Fig. 5.5D), precluding the possibility of a direct interaction.

A better candidate for *dri* regulation was identified fortuitously from the pattern of expression of a randomly inserted enhancer-trap (J. Manak, personal communication). This insertion contained homeo domain binding sites from the *Antennapedia* enhancer region set upstream of *lacZ* and was used to detect *Ultrabithorax* responsiveness (Manak et al., 1994). One insertion, p18-13, did not show the typical *Ubx* pattern in the midgut, but instead strongly stained in two rows of cells in the hindgut (Fig. 5.5E) that resembled those expressing Dri (Fig 5.3D). Double stainings confirmed that Dri and LacZ were being expressed in the same cells in this line (result not shown). Experiments are in progress to determine whether p18-13 expression is affected in homozygous $dri^{\Delta B15}$ embryos (see Discussion). A variety of other hindgut-expressing enhancer traps (see Materials) were tested, but none were co-expressed with Dri (results not shown).

Summary

The embryonic expression patterns of *dri* message and protein were determined by in situ hybridisation and antibody staining. Dri was found to be nuclearly localised and was expressed in a diverse set of tissues including the developing mesoderm, several gut structures, the salivary ducts and parts of the CNS. Two P-element insertions that expressed LacZ in a subset of the Dri pattern were shown to be non-complementing and gave an embryonic lethal phenotype that could be reverted by transposition. Examination of Dri expression in embryos homozygous for either insertion showed the wild-type pattern until late in embryogenesis. At stage 16 expression was lost in the ring gland and disordered in the hindgut; at later stages homozygous embryos were severely disrupted. Deletion mutations were generated by transposition of both insertions and two internal deletions in *dri* were isolated. These mutations did not completely prevent expression of Dri and gave the same hypomorphic phenotype as the insertions. Several genes with the

potential to genetically interact with *dri* were examined but no evidence for direct interaction has yet been conclusively identified. Thus the exact role of this novel DNA-binding protein in development is not yet known, but it is clearly essential for embryonic development and is specifically required for maintenance of normal hindgut morphology.

Discussion and Future Work

At last he came to a cave with big front door set in it, covered by a lot of seaweed. Prince Amilec knocked, feeling rather nervous.

'The witch will probably be horribly ugly, with three eyes and a wart on the end of her nose,' Amilec said to himself. 'But I mustn't let her see that I'm not completely used to people with three eyes and warts, or she may be offended and not help me.'

Just then the door opened, and there stood a very pretty girl holding a lantern.

'Can I help you?' asked the girl.

'Oh—er, yes. I was looking for the witch,' said Amilec, brushing off the seaweed that had fallen on him when the door opened.

'I am the witch,' said the girl. 'Do come in.'

- Prince Amilec Tanith Lee

Introduction

The preceding results chapters report the characterisation of the *dead ringer* gene, starting from a single partially sequenced cDNA clone that could be expressed to produce an Engrailed site binding protein. This clone was used to isolate and sequence a complete *dri* open reading frame in which a novel, highly conserved motif was identified. Biochemical characterisation of this polypeptide revealed a novel DNA-binding domain with an in vitro specificity that was strikingly similar to that of Q_{50} homeo domains. Progress was made towards identifying in vivo functions of this protein by a description of the *dri* developmental expression pattern and mutant phenotype. The conclusions that can be drawn from this work are broadly categorised into those relating to Dri homology, DNA-binding specificity and *dri* mutant phenotype. In each case, further directions for study will be discussed.

Sequence conservation in *dri*

Possibly the most surprising and provocative result presented in this thesis was the identification of a very widely conserved novel protein sequence motif. The most obvious role for this polypeptide was in DNA-binding, as mentioned in Chapter 3, since

several of the related proteins (Bright, MRF, RBP) were known to bind DNA and none had a known DNA-binding domain (Herrscher et al., in press; Whitson et al., in press; Fattaey et al., 1993). This activity has been confirmed by the experiments detailed in Chapter 4, but there are further connections between these proteins that suggest common functional roles.

Both the MRFs and Bright were isolated in screens designed to identify Matrix Attachment Region (MAR)-binding proteins; further characterisation has shown that they can both bind to a variety of MARs from several organisms (Herrscher et al., in press; Whitson et al., in press). Furthermore, Rb has been shown to localise to the nuclear matrix during G1 phase (Mancini et al., 1994), although the involvement of RBP1 or RBP2 has not been tested. The function of MARs in transcriptional regulation is far from clear, but it has been noted that they are frequently associated with enhancer regions (Cockerill and Garrard, 1986; Stief et al., 1989; Klehr et al., 1991) and specifically with potential homeo domain binding sites (Boulikas, 1993). The possibility that this new family of DNA-binding proteins binds in vivo to MAR-associated sites remains to be tested. From the current evidence it remains equally possible that these proteins have no role in MAR function that can be discriminated from their activity as transcription factors.

The homology with SWI1 is perhaps the most informative, since the function of this gene has been studied for some time in a genetically manipulable organism. Although *SWI1* was first isolated as a gene required for *ADH2* activation (Taguchi and Young, 1987), it has been subsequently shown to be required for the activation of many yeast genes as part of the SWI/SNF complex (Peterson and Herskowitz, 1992; Cairns et al., 1994). This enormous (~2MDa) complex is composed of at least ten proteins including SWI1, SWI2, SWI3, SNF5, SNF6 and SNF11 (Cairns et al., 1994; Treich et al., 1995). There is genetic and biochemical evidence to suggest that the SWI/SNF complex assists specific transcriptional activators by the displacement of chromatin

components such as nucleosomal histones (Hirschhorn et al., 1992; Côté et al., 1994). Although the complex is known to associate with DNA (Côté et al., 1994), neither SWI1 nor any other of the components have yet been shown to provide a DNA-binding function (Peterson and Herskowitz, 1992). From the evidence presented in Chapters 3 and 4 it is likely that the Dri-homologous region of SWI1 is a DNA binding domain and contributes that function to the complex. It will be of great interest to see if the specificity of a SWI1 DNA-binding domain resembles that of Dri and Bright, given the wide range of promoters that require SWI1 for activation and the diversity of activators with which the complex is known to interact (Peterson and Herskowitz, 1992; Yoshinaga et al., 1992). Purification of the Dri DNA binding domain for X-ray crystallography is under way, as structural information is likely to contribute greatly to an understanding of the specificity of these related proteins.

The requirement for a SWI/SNF complex is unlikely to be restricted to yeast, as homologues for several of its components have been identified in humans, mouse and *Drosophila* (reviewed in Carlson and Laurent, 1994). A human homologue of SNF2, BRG1, has been shown to complex with the retinoblastoma protein, cooperating to induce cell cycle arrest (Dunaief et al., 1994). The relationship between Rb and RBP1, RBP2, BRG1, nucleosome displacement, cell cycle arrest and MAR binding is unclear, but obviously of great interest. In *Drosophila*, SWI/SNF homologues can be isolated as part of a 2MDa complex (Dingwall et al., 1995) and in humans the analogous complex has been shown to assist activation by nucleosome disruption (Kwon et al., 1994). No homologues of SWI1 have yet been identified in SWI/SNF complexes in other organisms; the homologies described in Chapter 3 make Dri related proteins obvious candidates to test. The *Drosophila* complex thus far contains homologues of SNF2 and SNF5 (Brahma and Snr1), both of which are trithorax group members (Tamkun et al., 1992; Dingwall et al., 1995). The trithorax group of genes are required to maintain the activated state of many homeotic genes during *Drosophila* development (Kennison, 1993), a role not unrelated to that of the SWI/SNF complex, which is also required for

the expression of diverse loci (Peterson and Herskowitz, 1992). The homeo domain-like specificity of Dri is consistent with it playing a role in a SWI/SNF complex mediating the trithorax group function, but the expression pattern of Dri is too restricted for it to be the sole DNA binding protein in such a process. Western analysis and co-immunoprecipitation experiments can be used to determine whether Dri can form part of a *Drosophila* SWI/SNF complex. That RBP1 and 2, Bright, and SWI1 all participate in gene regulatory protein complexes (Herrscher et al., in press; Fattaey et al., 1993; Cairns et al., 1994) strongly suggests that Dri will do likewise, making protein interaction screens a high priority for future work.

Dri DNA binding specificity

As a homeo domain-site binding protein, Dri makes any complex with which it associates a potential component of the currently elusive mechanism of homeodomain specificity (see Introduction). There is an intriguing precedent to suggest such a Dri-HD interaction: the homeo domain protein Ftz can act as a transcriptional activator from NP binding sites in yeast, but this activity is lost in a *SWI1*⁻ strain (Peterson and Herskowitz, 1992). Although the expression pattern of *dri* in *Drosophila* has not suggested any obvious candidates for homeo domain interactions, it should be noted that at the stage of cellularisation when many homeo domain proteins carry out patterning roles, maternal Dri is present in most nuclei (Fig 5.1). Although *dri* has not been isolated in any of the screens for patterning mutants (such as Nüsslein-Volhard et al., 1984), a role for Dri in these early patterning events cannot be ruled out because its maternal component would mask such a phenotype. Assessment of the potential genetic interactions between Dri and early patterning homeo domain genes must await generation of *dri* maternal germline clones. It has been observed that overexpression of Dri ubiquitously is lethal, but ectopic expression in the developing wing and haltere discs gives a striking blistered wing and down-turned haltere phenotype (R.D. Kortschak and S.G., unpublished results). This is consistent with Dri partially competing for binding with a homeo domain protein such as Apterous, which is required for wing formation and, like Dri, is expressed in the CNS

and ring gland (Bourgouin et al., 1992; Cohen et al., 1992). This possibility could be further explored by examining the effect of *apterous* duplications and deletions on the Dri ectopic expression phenotype.

The homeo domain-like specificity of Dri made the pattern of expression of the p18-13 enhancer trap particularly interesting. This insertion was not designed as an enhancer trap, rather, the Antp promoter regions inserted in front of *lacZ* were expected to lead to LacZ expression wherever Ubx was active (Manak et al., 1994). Most random insertions of this construct gave midgut expression (Manak et al., 1994), but at least three insertions gave expression in the hindgut rather than the midgut (see Materials), indicating that Ubx alone was not sufficient to correctly regulate this promoter. The only difference between these lines should be their insertion point, underlining the reliance on flanking sequences to generate specific regulation by homeo domain proteins. Because the expression pattern from p18-13 overlaps that of Dri, it is possible that Dri is one of the additional factors causing expression in the hindgut rather than the midgut. This is being tested by crossing p18-13 into a $dri^{\Delta 15}$ background - if LacZ expression is affected, the sequences flanking the insertion will be of great interest as a likely target of Dri regulation. It is possible that Dri binding sites in the insertion are also required to give the hindgut pattern; even in this case, the flanking sequences must bind essential co-factors that allow Dri-specific regulation and prevent regulation by Ubx. It is tempting to speculate that this insertion was not entirely random but was targeted to some extent by the presence of potential Dri binding sites in the promoter (Manak et al., 1994) to insert into a Dri regulated genomic location, as has been observed with several 'stripe' enhancer insertions (Kassis et al., 1992).

One drawback of demonstrating genetic interactions is that it is rarely certain that the interaction is direct (see Introduction). In this case, if p18-13 LacZ expression is disrupted in a *dri* mutant, it could result from loss of expression of a target gene downstream of *dri* or from the liberation of factors normally sequestered by Dri. It is not

a trivial task to avoid this problem (eg Sun et al., 1995), but as a first step, the ability of Dri to bind homeo domain sites in vivo can be tested using domain swap experiments. A *rough* construct with its homeo box replaced by that of *deformed*, has been shown to partially rescue *rough*$^-$ mutants (Lockett et al., 1993). If Dri has the same binding specificity in vivo as in vitro, its DNA binding domain should also be able to functionally replace the Rough homeo domain. The identification of an in vitro DNA binding specificity for Dri and even a confirmation of that specificity in vivo are only preliminary steps in addressing the questions of broader significance. These include: does Dri have a role in transcriptional regulation, and if so, how is it mediated, what other factors are involved and to what extent can these results be generalised to interpret the role of related proteins. To specifically address whether Dri affects transcription, Dri will be tethered to a promoter via the Gal4 DNA binding domain. The regulatory effects of this binding on a Hunchback regulated promoter/reporter construct will be determined by immunostaining for the LacZ reporter protein in early embryos (as in Muller, 1995).

These experiments must necessarily be speculative because the biological role of *dri* is not clear from its expression pattern and mutant phenotypic analysis. The pattern of expression includes diverse tissues from both mesoderm and ectoderm with no obvious features in common (see Chapter 5). This implies that, like many homeo domain proteins (Krause et al., 1988; Patel et al., 1989; Cohen et al., 1992), Dri does not have a single role or set of target genes; instead it is likely to form part of a regulatory unit that is reused with variant effects depending on the DNA targets and protein partners available in each cell.

Dri expression pattern and mutant phenotype

Part of the difficulty in assigning a definitive function for *dri* comes from the lack of markers for tissues affected by *dri* mutations. There are ample markers for effects at cellularisation and in the mesoderm, but characterisation at these stages must await generation of maternal germline clones. The failure of the P-element induced deletions to

provide a suitable mutant for this process came as an unexpected set-back; now an EMS mutagenesis is being pursued as a high priority. Since all affected organs are apparently formed normally in the available *dri* mutants, any genes identified as organ formation mutants are likely to be genetically upstream and the only available markers for the downstream effects of *dri* are the relatively uncharacterised enhancer trap lines described in Chapter 5. As a consequence, it may be necessary to use further in vitro approaches to characterise Dri function.

Screens for the genomic DNA targets of several homeo domain proteins have been carried out in yeast and in vitro with varying success (Walter et al., 1994; Gross and Gruss, 1995). Given the concerns about specificity raised in the Introduction, this approach is likely to generate a high background unless the appropriate protein partners for Dri can be included. Work is in progress to characterise potential partners isolated using labelled Dri to screen cDNA expression libraries. A slower, but more reliable method for isolating both those factors with which Dri binds and those which it regulates, is to carry out a mutagenesis and screen for those mutants which interact genetically to enhance or suppress a *dri* phenotype. This approach has become practicable following the observation that overexpression of Dri in the developing eye causes severe roughening (R.D. Kortschak, S.G. and T. Shandala, unpublished results). This rough eye phenotype is being used as the basis for a modifier screen to identify mutants that affect the degree of roughness. Because Dri is normally expressed in the larval eye disc (T. Shandala, personal communication), this screen constitutes the best available option for isolating Dri-interacting proteins with in vivo relevance. Unfortunately, ectopic expression of Dri has invariably resulted in lowered viability and fertility (T. Shandala, personal communication), making this process arduous and time-consuming. The possibility that Dri is part of a Drosophila SWI/SNF complex will also be indirectly tested by determining if *brm* and *snr1* are modifiers. This approach is complementary to the in vitro protein interaction assays, so if any cloned modifiers are isolated they will be checked for Dri binding and vice versa.

Summary

The low ratio of affinity that homeo domain proteins display for specific over non-specific DNA has to some extent explained the rarity with which a single homeo domain binds a unique site to give an independent regulatory effect (see Introduction page 8). It has become apparent that many homeo domains interact with other proteins (p14), not only to convey their transcriptional effect but also to increase their specificity. Such interactions are known to include additional DNA-binding proteins (p14), since identification of a correct target sequence requires either a refinement of the homeo domain's intrinsic specificity or the ability to recognise additional nucleotides. It is too early to conclude that Dead ringer is one such factor that affects homeo domain specificity, but the demonstration that it prefers homeo domain binding sites, and the identification of a related protein in yeast that is required for Ftz trans-activation, both make it a likely candidate.

The characterisation of *dead ringer* described in this thesis has led to the identification of a novel family of DNA binding proteins that share a widely conserved motif. The functions of the members of this family are related in several ways which point to their involvement in important transcriptional regulatory complexes. The *dri* mutant phenotype shows that it is essential for *Drosophila* embryogenesis and the Dri expression pattern implies a variety of developmental functions. Future work will concentrate on identifying Dri-interacting factors both biochemically and genetically, structural analysis of the DNA-binding domain and further characterisation of the *dri* phenotype with the objective of clarifying the mechanisms of specificity in transcriptional regulation.