Molecular biology of the *dead ringer* gene

This thesis is submitted in part fulfilment of the requirements of the PhD in Science by Robert Daniel Kortschak, B. Sc. (Hons.).

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 contains no material previously published or written by another person, except where due reference has been given in the text.

I consent to my thesis being made available for loan and photocopying.

R. Daniel Kortschak

Acknowledgments

Thanks to all the people in and outside of the formal sciences who have intentionally or unwittingly helped me, intellectually, practically and emotionally, throughout my studies. You know who you are, but especially Karen, Julie and Deborah.

Nullius addictus iurare in verba magistri.

- Horace

Summary

The *Drosophila* gene *dead ringer* encodes a member of what has been shown to be a large family of proteins with a theme of DNA association, termed the ARID (A/T-rich interaction domain) family of proteins. Sequence analysis has placed *Dead ringer* in a sub-group, termed the eARID (extended ARID) group, together with the mouse protein Bright, which has a virtually identical DNA-binding specificity. This thesis describes the identification of further genes in the *dead ringer/bright* sub-group and the functional analysis of regions of the Drosophila DRI protein conserved in this group.

Three vertebrate homologues of *dri* were isolated, one from human and two from the zebrafish, *Danio rerio*. Comparison of protein sequences supports the notion that *Drosophila Dead ringer*, mouse Bright, *Caenorhabditis elegans* T23D8.8, the human DRIL1 and paralogous *D. rerio* Dri1 and Dri2 form an orthologous group of proteins within the ARID family, sharing the eARID region of homology flanking the core ARID motif. Analysis of the eARID-protein sequences outside the eARID region indicated the presence of two additional conserved motifs, the REKLES$_\alpha$ and REKLES$_\beta$ domains, unique to the eARID group.

Generation of constructs expressing the conserved domains with specific deletions of highly conserved regions allowed biochemical analysis of the domains' functions. A deletion within the eARID was shown to abolish DNA-binding activity. A self-association activity exhibited by the REKLES domains was not abolished by deletion of the REKLES$_\beta$ domain. *dri* transgenes with deletions corresponding to those characterised biochemically were used to determine the requirements for eARID DNA-binding and REKLES$_\beta$ domain activities *in vivo*. DRI lacking eARID DNA-binding activity acted antimorphically. DRI lacking the REKLES$_\beta$ domain was able to substitute for zygotically expressed DRI.

The identification of novel conserved domains within the eARID group suggests mechanisms through which these proteins may exert their regulatory effect. The generation of an antimorphic *dri* transgene will further provide a tool for analysis of *dri* function in vivo.

Contents

Chapter 1: Introduction

It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us. - Charles Darwin

1.1 "The last refuge of scoundrels"

The foundations of our understanding of gene regulation were laid in the study of the developmental control of the simplest of organisms, the bacteriophages (for a review of the study of λ phage see Ptashne 1987). The general principle that was identified in the study of these systems was the action of diffusible, *trans*-acting factors that bind to DNA at *cis*-regulatory sites to control the activation state of a gene (transcriptional activators are reviewed in Ptashne 1988). In the prokaryote/bacteriophage systems, the low complexity of the genomes and the relatively simple developmental choices and environmental changes facing the organisms allow such basic regulatory systems to function in a reliable manner. The transfer of these ideas to the more complex arena of eukaryotic gene regulation has proved invaluable. However, the much greater level of complexity in developmental choices and genome size has favoured the evolution of more reliable mechanisms for gene regulation. As is often the case in biology, the machinery involved in gene regulation plays other roles, in this case the curation of the genetic material of the eukaryotic organism. The paradigms developed in the study of prokaryotic gene regulation, even extended as they have been to include some of the less complicated aspects of eukaryotic control systems, fail to adequately incorporate the full range of eukaryotic regulatory systems. I will describe in this chapter how machinery involved in regulating gene expression and in maintaining chromosomal integrity are an integrated system that effectively evolved to act as a librarian for the eukaryotic genome, focussing on the information retrieval aspects of chromatin structure.

Chromatin is often introduced as the mechanism used by the eukaryotic cells to allow their genomes to be efficiently packaged into the nucleus, and an obstacle for transcriptional machinery to overcome. However, the fine levels of modulation that are observed in the

modulation of chromatin structure in response to developmental and environmental cues leading to complex gene regulatory effects strongly suggest that a regulatory role for chromatin coevolved with the more commonly cited packaging role. The idea that the packaging of DNA by histones and other chromatin proteins into condensed structures is involved in gene regulation has been disparaged by workers in prokaryotic gene regulation as "the last refuge of scoundrels" (Brent 1986). However, recent work into organisation of genetic material in the nucleus, the nuclear architecture, has shown that it is not enough to merely be scoundrels in our analysis of mechanisms of gene regulation in eukaryotes.

1.2 Nuclear architecture in eukaryotic gene regulation

That the nucleus has a defined internal architecture has been known cytologically for over two hundred years, made evident by the high visibility of the nucleolus (in Miller 1981), subsequently shown to be the site of ribosome biogenesis. The arrangement of ribosomal RNA genes (rDNA) within a nucleolar structure may confer advantages for ribosome biogenesis, for example allowing higher concentrations of RNA polymerase I, ribosomal proteins and other factors required for ribosome synthesis (for reviews, see Hadjiolov 1984; Sommerville 1986). The organisation of rDNA into the nucleolus appears to be a property of rDNA or its RNA products, as insertions of rDNA lead to the formation of ectopic nucleoli in the *Drosophila* nucleus (Karpen *et al.* 1988).

It has become clear from a number of lines of evidence that chromosomal arrangements, similar to those seen in the organisation of the rDNA in the nucleolus, are a more general feature of nuclear architecture.

During the initial development of the *Drosophila* blastoderm the nuclei of the syncytial embryo undergo 13 rapid and parasynchronous nuclear divisions (Zalokar and Erk 1976). The syncytium cellularises at the end of cycle 13, and parasynchrony is broken at mitosis of cycle 14 when patches of cells enter mitosis together at different intervals (Foe 1989). Approximately coincident with these events, zygotic transcription initiates from a

large number of loci (Anderson and Lengyel 1979). It may be anticipated that the significant changes in both cell cycle organisation and the requirements for gene expression occurring at this transition would coincide with changes in nuclear architecture if chromosomal organisation within the nucleus plays a role in either control of gene expression or maintenance of chromosome structure throughout the cell cycle. Indeed, although chromosomes remain unpaired throughout the first 13 syncytial division, chromosomes pair at cycle 14 and chromosomal pairing coincides with movement of active loci to the apical side of the nucleus (Hiroaka *et al.* 1993; Fung *et al.* 1998).

Whether pairing is an independent result of regulatory mechanisms or a device for ensuring appropriate regulation of homologous loci is not known. However, evidence from the analysis of pairing-dependent complementation of alleles, transvection, in *Drosophila*, suggests that homologous pairing is involved in gene regulation in at least some cases (Lewis 1954; Babu and Bhat 1980; Geyer *et al.* 1990; Mathog 1990; Kassis *et al.* 1991; Leiserson *et al.* 1994). Transvection has been proposed to be the result of the interaction between the regulatory enhancers of one homologue with the basal elements of the other, allowing communication between the alleles. Evidence for a complex involved in the interaction between *cis*-enhancers and basal elements has come from the study of the *zeste* gene, a modifier of transvection at a number of loci (Wu and Goldberg 1989; Pirrotta 1990). *zeste* encodes a self-associating DNA-binding protein that has been shown to be localised to at least 60 sites on salivary gland polytene chromosomes and has been shown to bind to promoter and other regulatory DNAs from the *white*, *Ubx* and *dpp* genes *in vitro* (Pirrotta *et al.* 1988; Benson and Pirrotta 1988; Bickel and Pirrotta 1990).

Alleles of a number of members of the Polycomb group (Pc-G), a group of genes involved in stable repression of genes at a variety of loci, interact with the *zeste* modifier of transvection (Wu and Howe 1995; Pirrotta 1997), indicating a link between the pairing affects outlined above and the developmentally important Pc-G gene silencing system. Broadly, two competing models have been proposed to explain how the Pc-G stably inactivate loci. Both models involve the formation of large protein-DNA complexes at the inactivated loci. The

models differ, however, in how the silencing is effected by the multi-protein complex. The compaction model states that Pc-G proteins act by concatamerisation, packaging the chromatin into a highly compacted state and thus rendering the locus inaccessible to transcription factors and other components of transcriptional machinery (Paro 1990). In the competing compartment model, inactivated loci are positioned in domains within the nucleus lacking the factors required for expression of those loci, while active loci are located in domains in which critical factors are present (Schloßherr *et al.* 1994).

Although it is likely that neither model will explain the entire range of effects observed, and that both mechanisms are acting in the stable regulation of gene expression, the compaction model proposed to explain Pc-G silencing cannot be the whole story, as Schloßherr *et al.* (1994) have shown that loci inactivated by the Pc-G do not have significantly reduced accessibility to restriction endonucleases (Some reduction is observed at the *Antp* locus in the eye-antennal disk relative to the wing disk; although statistically significant, the reduction is proposed to be insufficient to account for the observed reduction in *Antp* expression). Additionally, it has been shown that transcription by *Drosophila* RNA polymerase II directed by the yeast GAL4 protein is blocked in an environment silenced by the Pc-G, but transcription by the bacteriophage T7 RNA polymerase is not, further confirming that Pc-G–mediated silencing is not merely a steric obstruction of protein access to silenced loci (McCall and Bender 1996). The compartment model is supported by the observation that the Polycomb protein localises in SL2 culture cells to a small number of loci in the nucleus while binding to a large number (~100) of sites on polytene chromosomes indicating that silenced loci colocalise to specific domains within the nucleus (Messmer *et al.* 1992).

An effect termed position effect variegation (PEV) shares a number of characteristics with silencing mediated by the Pc-G (reviewed in Karpen 1994; Weiler and Broach 1995). PEV is the stochastic and heritable silencing of euchromatic loci positioned near heterochromatic regions by chromosome rearrangement. The mechanism of PEV silencing, like that of Pc-G–mediated silencing, is not clear; the two models proposed for Pc-G silencing

are also used to account for the effects seen in PEV (Paro 1990). Indeed, PEV and Pc-G silencing are thought to be mediated through similar mechanisms. As is the case with Pc-G silencing, *trans*-acting factor accessibility is not lost in regions inactivated by PEV silencing (Schloßherr *et al*. 1994), favouring a model which includes some aspect of nuclear compartmentalisation. One of the models used to study PEV is *brown*^{D}. *bw*^{D} is a null allele of the *bw* gene caused by a large insertion of heterochromatin within the coding region of the gene (Slatis 1955). Interestingly, the *bw*^{D} allele acts dominantly to repress the expression of a wild-type allele on the homologous chromosome, employing a pairing-dependent mechanism reminiscent of the transvection effects described above (Dreeson *et al.* 1991; Henikoff *et al.* 1995). Two elegant studies using *bw*^{D} have given further weight to the involvement of nuclear architecture in PEV by investigating the nuclear arrangement of homologous variegating *bw* loci by fluorescence *in situ* hybridisation (Csink and Henikoff 1996; Dernburg *et al.* 1996). Dernburg *et al*. (1996) examined the localisation of the wild-type and *bw*^{D} alleles of the *bw* locus in larval imaginal disks and neuroblasts. They found that *bw*^{D} homozygous flies showed a strong association of the paired *bw* loci with the centromeric AACAC satellite block on 2R when compared to wild-type. The number of nuclei showing association of paired *bw* loci with centromeric heterochromatin was only slightly lower in *bw*^{D} heterozygous larvae than in homozygous larvae. Analysis of pairing has shown that 95% of *bw* homologous loci are paired in larval nuclei. This observation together with the finding that the level of pairing is identical when comparing wild-type, *bw*^{D}/+ and *bw*^{D} larvae, suggests that pairing of homologous *bw* loci in the heterozygote carries the wild-type homologue to the centromeric heterochromatin with the *bw*^{D} allele. The results obtained by Csink and Henikoff (1996) extend the observations made by Dernburg *et al*. to show that the association of the paired *bw* loci with the centromeric chromatin is reduced both by *Su(bw*^{D}*)5*, a translocation of *bw*^{D} that suppresses *bw*^{D} *trans*-inactivation, and by suppressors of PEV including *Su(var)205*, a loss of function mutation in the gene encoding the heterochromatin protein HP-1. The HP-1 gain of function mutation *E(var)39A* increases the association of paired *Su(bw*^{D}*)5* loci with centromeric chromatin, but paradoxically does not show the same effect with the *bw*^{D} allele. In embryos, Dernburg *et al*. show that *bw*^{D} loci are strongly associated

with the nuclear envelope. A role of HP-1 in this interaction suggested by the results of Csink and Henikoff is consistent with the finding that a human homologue of HP-1 interacts with an integral nuclear envelope protein (Ye and Worman 1996). These results indicate that the nuclear localisation of the *bw* locus to heterochromatin nuclear compartments is dependent on pairing and heterochromatin proteins, firmly tying together the models of transvection, position effect-variegation and nuclear architecture in gene regulation.

Work in yeast has also shown that nuclear architecture is modulated in conjunction with changes in silencing states of chromatin. The silent mating type loci and loci near the chromosomal telomeres are repressed by a complex of proteins including the products of the *SIR* loci and the RAP1 protein (reviewed in Gotta and Gasser 1996; Gasser *et al.* 1998). Histones H3 and H4 have been shown to interact with Sir3 and Sir4 *in vitro*, and disruption of *HHTA1*, *HHF2* (the histone H3 and H4 loci), *SIR2-4* or RAP1 results in deregulation of the silent mating type loci (Thompson *et al.* 1994; Aparicio *et al.* 1991; Hecht *et al.* 1995). The silent telomere complexes have been shown to localise to the nuclear cortex (Gotta *et al.* 1996) through a mechanism dependent on the interaction between SIR3, SIR4 and the amino termini of the histones H3 and H4 (Paladino *et al.* 1993; Hecht *et al.* 1995), which have been shown to be necessary of silent mating type locus repression (Johnson *et al.* 1990; Thompson *et al.* 1994). In contrast to the evidence from *Drosophila* however, the compartmentalisation of silent loci to specific domains happens in association with protection from digestion by restriction endonucleases (Loo and Rine 1996). Maillet *et al.* (1996) have shown that artificial silencing cassettes are dependent on nuclear context. In their assay, silencers were assisted by proximity to telomeres, which are proposed to be acting as pools of Sir proteins, and by artificially elevated expression of Sir proteins, suggesting that silent domains may be required to be located within nuclear regions with high concentrations of these factors, unifying the conflicting models of chromatin compaction and nuclear compartmentalisation.

1.3 Regulation of eukaryotic chromatin structure

Clearly chromatin packaging and nuclear organisation play a role in the regulation of gene expression in eukaryotes. The stable maintenance of gene expression by higher order organisation of genetic material raises the important question of how it can be involved in dynamic gene regulation. During development and in response to environmental factors, patterns of gene expression must change. Fluid gene regulation requires mechanisms to circumvent the systems described above. What mechanisms exist to enable genes regulated at the chromatin level to have their expression modulated?

Studies on *Drosophila* salivary polytene chromosomes show that the arrangement of chromosomes within the nucleus does not change significantly in response to heat-shock or treatment with β-ecdysone, indicating that nuclear architecture is not a target of mechanisms regulating the expression of ecdysteroid or heat-shock responsive genes (Hochstrasser and Sedat 1987). This does not eliminate the possibility that other loci are regulated in this manner, for example homologous histone cluster loci pair and move to the apical side of the nucleus in the *Drosophila* blastoderm at the time when they become active zygotically (Hiroaka *et al*. 1993); nor the possibilities that combinatorial associations of larger numbers of constituent loci or associations with other structures, such as the nuclear envelope (Marshall *et al.* 1996), are utilised in compartmental regulation in the nucleus. Currently there is little evidence for widespread dynamic modulation of nuclear domain architecture, nor are any systems known which specifically alter nuclear domain architecture, although some progress has been made towards identifying possible points of regulation (reviewed in Marshall *et al.* 1997).

However, analysis of a number of systems has shown a general theme of large protein complexes which facilitate the dynamic reorganisation of chromatin structure in response to regulatory needs, influenced by the presence and activation state of transcription factors, the chromatin state in the local environment and possibly by broader nuclear architecture. The following sections will detail the variety of chromatin modulating 'machines', their effects on

nucleosomal chromatin structure and inputs in the control of chromatin remodelling. The possible affects of these systems on nuclear domain architecture will then be explored.

1.3.1 The SWI/SNF and Brahma complexes

Although transcriptional activators and repressors have been shown to act to modulate the levels of transcription from eukaryotic promoters *in vitro*, the range of effective modulation achieved on naked DNA is far from that observed *in vivo* (Kornberg and Lorch 1995). The arrangement of DNA into nucleosomes has often been postulated to be a significant impediment to the initiation and progression of eukaryotic transcription by blocking access to transcription factors and polymerases, and by retarding the movement of active polymerases. Supporting these suggestions, depletion of histone H4 levels in yeast promotes expression of *CUP1*, *HIS3* and *PHO5* under non-inducing conditions (Han and Grunstein 1988; Durrin *et al.* 1992), and transcription from reconstituted chromatin including histones, activated by the yeast GAL4 protein *in vitro*, is significantly reduced compared to transcription from naked DNA (Mizuguchi *et al.* 1997).

The use of nucleosomal chromatin structure as a repressive mechanism requires an opposing system to remove the transcriptional block imposed by the presence of nucleosomes on the DNA. The yeast SWI/SNF complex was the first identified of a number of 'chromatin remodelling machines' that enable the dynamic modulation of chromatin structure (for reviews, see Burns and Peterson 1997a; Gregory and Hörz 1998; Varga-Weisz and Becker 1998). The members of the SWI/SNF complex were first identified as viable mutations blocking the expression of the *HO* and *SUC2* genes, required for mating type switching and sucrose fermentation respectively. It was subsequently shown that the complex was required more generally for the activation of a wide variety of genes (Peterson and Herskowitz 1992).

The activation role of the SWI/SNF complex was investigated in two studies using a LexA-SNF5 fusion and a lexA operator site. The LexA-SNF5 fusion was able to activate a *GAL1-lacZ* fusion gene downstream of either one or six overlapping *lexA* operator sites, but

not in the absence of a *lexA* operator in *cis* (Laurent *et al.* 1990). However, further studies showed that activation in this assay was dependent on the presence of functional *SNF2* and *SNF6*, two other SWI/SNF encoding genes (Laurent *et al.* 1991). A LexA-SNF2 fusion is able to activate in the same assay, dependent on functional *SNF5* and *SNF6*. These observations indicate that SNF2 and SNF5 activities are interdependent, possibly through the formation of a protein complex. However, they do not distinguish the mode of action of the SNF2 and SNF5 proteins from that of more familiar transcription factors, such as the yeast GAL4 protein (reviewed in Ptashne 1988).

Subsequent studies have shown, however, that the SWI/SNF proteins activate genes in a more indirect manner. An extensive examination of the functions of the SWI/SNF complex by Côté *et al*. (1994) showed that the complex contains 10 subunits (Varga-Weisz and Becker 1998; Burns and Peterson 1997a) and has DNA-stimulated ATPase activity. Using GAL4 derivatives and nucleosomal DNA with a single GAL4 binding site, they showed GAL4 binding in gel retardation assays was stimulated by SWI/SNF in the presence of hydrolysable ATP. DNaseI footprinting analysis showed that nucleosomes block GAL4 binding, but that the block was removed by SWI/SNF in the presence of ATP. Further, new DNaseI sites appear when nucleosomal DNA is treated with SWI/SNF in the presence of ATP with or without GAL4, showing interaction with the nucleosomal DNA, possibly moving the nucleosomes. Previously the SWI2 proteins had been shown to have DNA-stimulated ATPase activity associated with a DNA helicase-related domain (Laurent *et al.* 1993); Côté *et al*. showed that a SWI/SNF complex containing $SWI2^{K798A}$ which has no ATPase activity, was unable to facilitate GAL4 binding in gel retardation assays as above and did not activate GAL4 dependent transcription.

A number of studies have strengthened the notion that the SWI/SNF complex acts by facilitating the binding of *trans*-acting factors to nucleosomal DNA. SWI/SNF is required for GAL4 activation in nucleosomal DNA, unless the GAL4 DNA binding-site is high affinity or located within a nucleosome free region (Burns and Peterson 1997b). Similarly, although SWI2 and SNF5 are required for *SUC2* expression and opening the *SUC2* promoter to

micrococcal nuclease activity, this can be overcome by halving the number of histone H2A and H2B genes (Hirschhorn *et al.* 1992). Supporting the interaction between nucleosomes and the SWI/SNF complex, mutant histones H3 and H4 expressed from low copy number plasmids partially suppress *swi1* transcription defects, presumably by competing with wild-type H3 and H4 histones for nucleosome formation (Kruger *et al.* 1995).

In a restriction endonuclease access assay with a positioned nucleosome array, SWI/SNF enhances rate of digestion of nucleosome-covered *Sal*I sites in an ATP-dependent reaction (Logie and Peterson 1997). Additional nucleosomal DNA could act as a substrate for SWI/SNF activity indicating that the complex acts catalytically in the remodelling of nucleosomes. Removal of ATP using apyrase after one hour of treatment with SWI/SNF complex resulted in *Sal*I digestion kinetics virtually identical to that seen when ATP was removed with apyrase at time zero, indicating that nucleosome remodelling results in only transient destabilisation of nucleosomes. Others have found that remodelled nucleosomes in nucleosomal arrays are stabilised in the presence of the GAL4 DNA-binding protein (Owen-Hughes *et al.* 1996), indicating that the transient destabilisation of nucleosomal structure can be utilised opportunistically by DNA-binding proteins.

Thus it appears that the SWI/SNF complex acts catalytically to facilitate the movement of nucleosomes in nucleosomal DNA, permitting the access of DNA-binding factors and transcription machinery to a number of loci in the yeast genome.

Protein complexes homologous to SWI/SNF have been identified in humans (Dunaief *et al.* 1994; Wang *et al.* 1996b) and *Drosophila* (Tamkun *et al.* 1992; Elfring *et al.* 1994). The wealth of genetic studies in *Drosophila* make it an excellent system in which to examine the role of SWI/SNF complexes in multicellular organisms.

Evidence that Brahma, the closest known SWI2 homologue in *Drosophila*, functions in a role similar to that of SWI2 in the SWI/SNF complex comes from a number of sources. The DNA-dependent ATPase domain of BRM, substituted for the SWI2 ATPase domain in yeast, has been shown to give partial rescue of *swi2* growth defects and to allow activation of

lacZ reporter-gene transcription, while that of the second SWI2 homologue (ISWI, discussed later) did not (Elfring *et al*. 1994). Protein homology between BRM and SWI2 extends to a number of regions outside the conserved ATPase domain, regions which have been proposed to function as protein interaction domains with other members of a protein complex. Indeed, SWI2 coimmunoprecipitates in a 2MDa complex with SNR1, a *Drosophila* SNF5 homologue (Dingwall *et al.* 1995). Further, studies of BRG1, a human BRM homologue, also found in a 2MDa complex, show that this complex, like the yeast SWI/SNF, facilitates the binding of transcription factors to nucleosomal DNA *in vitro* (Kwon *et al.* 1994; Wang *et al.* 1996a; Wang *et al*. 1996b).

Elfring *et al*. (1998) have conducted a comprehensive examination of the role of *brm* in *Drosophila* development, and an analysis of functionally important regions of the protein. Unlike SWI2 and the other member of the yeast SWI/SNF complex, BRM is essential for viability. Consistent with its essential role in *Drosophila*, BRM is expressed ubiquitously at high levels throughout development. Quantification of BRM copy-number indicates that there are more than 10^5 molecules per nucleus, equivalent to 1 complex per 20 nucleosomes, considerably higher than the estimate of 100-200 SWI/SNF complexes per nucleus in yeast (Côté *et al*. 1994; Cairns *et al.* 1996), corresponding to approximately 1 complex per 300-600 nucleosomes. As *brm* is essential for *Drosophila* viability, the role of *brm* in development was examined by generating somatic *brm* deficient clones in larval tissues. *brm* clones in imaginal disks showed significant reduction in viability. Examination of adults with *brm* clones showed a high percentage of bristle defects, indicating a role for *brm* in peripheral nervous system development. The roles of specific conserved domains within the BRM protein were examined by the generation of targeted mutations. BRM^{K804R}, in which a conserved lysine in the ATP-binding site of the DNA-dependent ATPase domain changed for an arginine acts antimorphically to produce a homeotic phenotype and decreased viability. The equivalent amino acid substitution has been shown to abolish ATPase activity in the yeast SWI2 protein (Laurent *et al*. 1993). Gel filtration analysis showed that BRM^{K804R} is incorporated into the 2MDa complex. The inclusion of the putatively ATPase-defective BRM^{K804R} in the 2MDa complex indicates that ATPase activity is necessary for BRM

complex function. Expression of BRM^{K804R} in the wing imaginal disk causes wing vein loss and PNS defects including ectopic campaniform sensilla, campaniform sensilla loss, duplication or triplication, or transformation into bristles, effects that are similar to those seen in *brm* deficient clones in the adult. A second mutant, lacking domain II ($brm^{\Delta 549\text{-}610}$) failed to rescue *brm* mutant lethality. Gel filtration analysis indicated that domain II is required for inclusion in the BRM complex. A highly conserved domain, the bromodomain was also examined. A *brm* transgene lacking the bromobox ($brm^{\Delta 1446\text{-}1517}$) rescued *brm* homozygous lethality, indicating that the bromodomain is not essential for BRM function.

Although *brm* is required for *Drosophila* viability, it is clear from the observation that *brm* deficiency is not cell lethal in clones that the BRM complex cannot be the only factor involved in chromatin remodelling. Similarly, the viable phenotype exhibited by the *swi/snf* mutants indicates that the SWI/SNF complex is involved in regulating only a subset of genes in *S. cerevisiae*, or that other factors are able to compensate for a lack of SWI/SNF activity.

1.3.2 RSC

As mentioned above, the SWI/SNF complex is not required for viability, raising the question of whether nucleosome remodelling is required on a large scale in yeast. However, computer based searches of the *S. cerevisiae* genome have identified homologues of members of the SWI/SNF complex which together form a chromatin remodelling complex (Cairns *et al.* 1996). Thus yeast appears to utilise a complex of chromatin remodelling machineries in the regulation of chromatin structure.

Five unique homologues of members of the SWI/SNF complex were identified by comparison of SWI/SNF proteins with the database of protein sequences predicted by open reading frame analysis of the *S. cerevisiae* genome: Sth1/Nps1p (homologous to Swi2/Snf2p, originally identified by reduced stringency hybridisation Laurent *et al.* 1992), Rsc6p (Swp73p), Rsc8p (Swi3p), Sfh1 (Snf5p) and Ycr052p (Tfg3/Anc1/Swp29p) (Cairns *et al.* 1996). No paralogue of Swi1p, Snf6p or Snf11p was found. Sth1/Nps1p, Rsc6p and Sfh1p

were shown to be components of a 1MDa, 14/15 member complex by Western analysis of the purified complex and coimmunoprecipitation of Sth1/Nps1p and Rsc6p with anti-Sfh1p antibodies (Cairns *et al*. 1996; Cao *et al.* 1997). Mass spectrometry and peptide sequencing confirmed the presence of Sth1/Nps1p and Rsc6p within the complex and further indicated the presence of Rsc8p (Cairns *et al*. 1996). The protein complex, RSC, was shown to have DNA-stimulated ATPase activity and the capacity to remodel the structure of chromatin, dependent on ATP, as assayed by nucleosomal protection from DNaseI digestion. Significant distinctions were made between the SWI/SNF complex and RSC: RSC was estimated to be at least an order of magnitude more abundant in the nucleus, the two complexes were shown not to share the homologous subunits Swi2p/Sth1p, Swp73p/Rsc6p or Swi3p/Rsc8p, and most significantly, yeast lacking Sth1/Nps1p, Rsc6p, Rsc8p or Sfh1 are inviable indicating a crucial role for this complex in normal growth in contrast with the non-essential nature of the SWI/SNF complex (Cairns *et al*. 1996; Cao *et al*. 1997). However, the striking similarities between the SWI/SNF complex and RSC, including the specificity of action and conserved constituents, suggests an underlying similarity of function. The identification of a conserved core group of subunits in chromatin remodelling complexes suggests a strong conservation of the mechanism of chromatin remodelling.

1.3.3 ISWI containing complexes

As mentioned previously, a second divergent SWI2 homologue has been identified in *Drosophila*, ISWI (for imitation switch). Clones encoding ISWI were first identified on the basis of homology to the yeast *SWI2* gene in a reduced stringency hybridisation screen (Elfring *et al*. 1994). The ISWI protein is implicated in the activity of a number of chromatin remodelling complexes that have apparently divergent functions (for reviews, see Varga-Weisz and Becker 1998; Armstrong and Emerson 1998).

i. NURF

Biochemical characterisation of the 4 subunit, 500kDa *Drosophila* nucleosome remodelling factor (NURF) identified ISWI as the NURF ATPase analogous to BRM in the BRM complex (Mizuguchi *et al*. 1997; Tsukiyama *et al.* 1995). NURF exhibits nucleosome disruption in the presence of GAGA factor and hydrolysable ATP in restriction endonuclease accessibility and DNaseI footprinting assays. In contrast to the DNA-stimulated ATPase activity of SWI/SNF (Côté *et al*. 1994), NURF ATPase activity is stimulated by nucleosomal, but not naked DNA. NURF ATPase activity has been shown to be reduced by proteolytic removal of histone tails and competition with histone tail–GST fusions proteins, suggesting that nucleosomal histone tails direct the activity of the NURF complex (Georgel *et al.* 1997).

An *in vitro* transcription assay using reconstituted chromatin with GAL4 DNA binding sites upstream of a minimal adenovirus E4 promoter has shown that transcriptional activity and nucleosome remodelling as assayed by micrococcal nuclease digestion, are dependent on GAL4 binding, and addition of NURF and ATP prior to the addition of *Drosophila* nuclear transcription extracts (Mizuguchi *et al*. 1997). In this assay, yeast SWI/SNF and another ISWI containing complex, CHRAC (discussed below), were unable to activate transcription in the same manner. Removal of NURF from the nucleosome assembly reaction by treatment with sarkosyl before addition of the nuclear transcription extracts did not affect transcription levels or nucleosome remodelling, indicating that nucleosome remodelling by NURF is stable. However, gel shift assays showed that GAL4 remained bound to the DNA, suggesting that the GAL4 protein may protect the GAL4 biding-sites from reoccupation by nucleosomes. Mizuguchi *et al*. (1997) comment that the high levels of chromatin remodelling activity attributable to NURF, indicate that NURF is the main chromatin remodelling activity in *Drosophila* nuclear extracts. The number of NURF complexes per nucleosome is estimated to be 1 NURF per 20 nucleosomes (Tsukiyama *et al*. 1995), the same as that estimated for the BRM complex (Elfring *et al*. 1998).

ii. CHRAC and ACF

Two other ISWI containing nucleosome remodelling complexes have been isolated biochemically, namely the chromatin-accessibility complex (CHRAC, Varga-Weisz *et al.* 1997) and the ATP-utilising chromatin assembly and remodelling factor (ACF, Ito *et al.* 1997). CHRAC is a 670kDa complex with 5 subunits including topoisomerase II in addition to ISWI, while ACF has 4 subunits and a mass of 220kDa. The presence of topoisomerase II in CHRAC has been postulated to either direct the activity of the topoisomerase to chromatin, enhancing its function, or alternatively topoisomerase II may direct CHRAC to specific chromosomal sites. A common feature of CHRAC and ACF is that both have been shown to convert irregular chromatin into regular arrays of evenly spaced nucleosomes in an ATP dependent manner, an activity that NURF lacks. The chromatin assembly role suggested by the homogenisation of nucleosome spacing by these factors is supported by a reconstituted chromatin assembly system using ACF and dNAP-1, a *Drosophila* histone chaperone protein, in which ACF was able to assemble chromatin with regularly spaced nucleosomes with the same structure as that generated with standard *Drosophila* chromatin assembly nuclear extracts (Ito *et al*. 1997). The role of dNAP-1 in the chromatin assembly system above appears to be as a donor of core histones and not as an integral part of the nucleosome remodelling complex as nucleosome spacing is modulated by ACF in the absence of dNAP-1. In addition to its role in chromatin assembly, ACF is also able to remodel chromatin in the vicinity of a promoter and activate transcription under the influence of the GAL4 transcription factor in the same manner as that observed with the NURF complex (Ito *et al*. 1997), activities which CHRAC lacks in at least one assay (Mizuguchi *et al*. 1997).

The existence of multiple chromatin remodelling complexes incorporating a common catalytic component, but having divergently targeted activities, suggests that the common function of nucleosome remodelling is used in a variety of roles, possibly reflecting distinct regulatory roles in nuclear architecture.

1.3.4 Histone acetyltransferase complexes and histone deacetylases

A theme that emerges from a comparison of systems involved in the stable maintenance of gene expression and the dynamic modulation of nucleosomal chromatin structure is the central role of histone amino-terminal tails in these processes (also reviewed in Luger and Richmond 1998). Contrary to the long held belief that the highly basic histone tails are responsible for increasing histone binding-affinity for nucleosomal DNA, a high resolution crystal structure of the nucleosome indicates that the regions responsible for most histone tail effects do not contact the DNA (Luger *et al.* 1997). It has thus been suggested that these regions mediate inter-nucleosomal interactions (Luger *et al.* 1997). Evidence also exists for an interaction between histone tails and the *Drosophila* NURF (Georgel *et al.* 1997), indicating a role in interactions with other components of chromatin and associated proteins.

Two classes of regulatory molecules which act through modification of histone tails are the histone acetyltransferases (HATs) and the histone deacetylases (HDAs) (for reviews, see Grunstein 1997; Tsukiyama and Wu 1997). Histones are reversibly acetylated at conserved lysine residues in the histone amino-terminal tails, changing the charge of the lysine from positive to neutral and so altering potential protein interactions. One such possible interaction is between the amino-terminal tail of histone H4 and a conserved, highly negatively charged region in the histone H2A/H2B dimer, an interaction which is proposed to mediate interactions between neighbouring nucleosomes, stabilising higher order chromatin structure (Luger *et al.* 1997). Thus, the observed hyperacetylation of euchromatic histone H4 tails relative to those in heterochromatin may facilitate the disruption of chromatin structure by reducing the strength of this interaction (Olins *et al.* 1991; Jeppesen *et al.* 1992; O'Neill and Turner 1995).

i. SAGA

The transcription factor Gcn5p, was originally identified as a HAT by homology to a catalytically active HAT from *Tetrahymena* (Brownell *et al.* 1996). Recombinant Gcn5p has been shown to have HAT activity, but is unlike *GCN5*-dependent HAT activity from yeast

extracts in that it is unable to acetylate nucleosomal histones, suggesting that other factors may be modifying the activity of the Gcn5p protein (Grant *et al.* 1997). Gcn5p and Ada2p copurify with a HAT activity capable of acetylating nucleosomal histones H2B and H3, in two complexes, either with (SAGA, 1.8MDa) or without (AGA, 0.8MDa) Spt20p and Spt7p. The SAGA and AGA HAT activities were shown to be dependent on *GCN5*, suggesting that Gcn5p is the catalytic subunit of these nucleosomal HAT complexes. The HAT activity of Gcn5p has been shown to be necessary for *HIS3* gene activation by a mutational study of the HAT domain of Gcn5p protein (Kuo *et al.* 1998).

A connection between genes encoding members of the SAGA complex and the regulation of genes through chromatin structure is made by the observation that *hht2-3* and *hhf2-7*, mutations in histone genes that suppress SWI/SNF defects, also suppress *gcn5* loss of *HO-lacZ* expression (Pollard and Paterson 1997). A role for SAGA in gene regulation through chromatin structure is supported by the observations that SAGA-encoding genes are required for expression of the SWI/SNF dependent reporter gene, *HO-lacZ* (Pollard and Paterson 1997) and are synthetically lethal in conjunction with *swi2/snf2* (Roberts and Winston 1997). Further, *gcn5* deletions and Gcn5p HAT domain mutants exhibit modified nucleosomal structure compared to wild-type as assayed by DNaseI digestion and restriction endonuclease protection (Gregory *et al.* 1998).

ii. HDAs

Appropriate regulation of gene expression is achieved by a counter-acting histone deacetylase activity which has been shown to be present in a number of transcription repressor complexes (Armstrong and Emerson 1998). The link between HDA activity and transcriptional repression was formalised by the biochemical isolation of an HDA protein, HDAC1, which was shown to be 60% identical to the yeast Rpd3p transcriptional repressor (Taunton *et al.* 1996). Human HDAC1 has been shown to exist in complexes with the repressor protein mSin3 (Zhang *et al.* 1997; Hassig *et al.* 1997) and the Mad1 repressor protein whose repression activity is blocked by the HDA inhibitor, trichostatin A (Laherty *et al.* 1997). These initial findings support the notion of histone deacetylation as a repressive

mechanism acting through chromatin structure. The situation is complicated by the isolation of a *Drosophila* homologue of yeast Rpd3, dRPD3. A *dRPD3* mutant showed an increased degree of *w* repression in the w^{m4h} position-effect variegating strain (De Robertis *et al.* 1996), strengthening the role of histone deacetylation in chromatin regulation, but confounding the direction of control by suggesting that dRPD3 is, in this case, an activator.

1.4 Communication between regulatory factors

It is clear from the sections above that a wide variety of factors are involved in the regulation of gene expression at the chromatin level, raising the question of how different factors direct their activities to the appropriate chromosomal domains. The answer to this question reflects the complexity of chromatin regulatory systems, as a variety of mechanisms have been shown to act to define targets for these regulatory systems. However, the diversity of mechanisms that have been analysed can be broadly grouped in to the categories of opportunistic or directed regulation.

The targeting of silencing by SIR and Pc-G complexes describe above, is readily explained by the existence of specific binding-sites within loci silenced by these complexes, although Pc-G mediated silencing has additional levels of target regulation based on the activation states of the target loci (for reviews, see Lustig 1998; Pirrotta 1997). Thus these two systems appear to be follow the directed regulation model described above, as do systems involved in modification of histone acetylation state.

SIR mediated silencing is dependent on Rap1p, a protein that has been shown to bind to *cis*-acting sequences in telomeres and at the silent mating-type loci (Wright *et al.* 1992; Shore *et al.* 1987). The Rap1p carboxy-terminus has been found to interact with Sir3p and Sir4p in a two hybrid system (Moretti *et al.* 1994). A significant role in the targeting of SIR-mediated silencing by Rap1p is indicated by the demonstration of *SIR*-dependent silencing at a locus containing GAL4 binding-sites using Rap1p carboxy terminus–GAL4 DNA-binding domain fusion protein (Buck and Shore 1995). Similar results have been obtained using

fusions of Sir3p and Sir4p with heterologous DNA-binding domains (either LexA or GAL4 Lustig *et al.* 1996; Marcand *et al.* 1996). Additionally, telomeric silencing is restored in a *rap1* mutant with GAL4 binding-sites located telomerically when a Sir3p– and Sir4p–GAL4 DNA-binding domain is expressed (Marcand *et al*. 1996). Together, these results suggest that Rap1p acts to recruit the SIR complex to SIR silenced loci.

Pc-G mediated silencing bears a striking resemblance to SIR mediated silencing with respect to target locus determination, although an extra layer of regulation is added in the form of a dependence on the transcriptional activation state of potential target loci. A region of the *Ubx* promoter, termed the PRE, has been shown to confer Pc-G dependent repression on a lacZ reporter gene(Chan *et al.* 1994). Insertion of this element at a new site introduces a new Polycomb complex binding site, as indicated by the binding of the Posterior sex combs protein to polytene chromosomes at the site of the insertion . PREs are a common feature of a number of genes regulated by the Pc-G (reviewed in Pirrotta 1997). A PRE from the *engrailed* gene identified as a Pc-G- and pairing-sensitive silencer in a *w* expression construct (Kassis *et al*. 1991; Kassis 1994) has been shown to bind Pc-G proteins in cross-linking studies (Strutt and Paro 1997). A Pc-G protein, Pleiohomeotic (PHO), has been shown to bind to this *engrailed* PRE (Brown *et al.* 1998). In experiments analogous to those described above for Rap1p/Sir3/4p, it appears that the stable Pc-G mediated silencing of a *lacZ* reporter gene can be established by transient localisation of Polycomb or Polycomblike proteins to the promoter of a target gene (Müller 1995, S. O'Connell, T. McGrath and R. Saint, unpublished results). In these experiments, Polycomb- and Polycomblike-GAL4 proteins were targeted to GAL4 binding sites in the regulatory region of a *lacZ* reporter driven by a promoter containing element from the *Ubx* gene. Repression was dependent on endogenous *Pc* and was maintained even when the Pc-G fusion protein was no longer present. These results indicate that tethering members of the Pc-G complex to DNA enables the formation of a stable silencing complex. It is likely that the PHO protein acts to target the Pc-G complex to Pc-G responsive elements, in a manner analogous to Rap1p targeting of SIR silencing, although this is yet to be tested.

Similarly, the targeting of histone acetylation and deacetylation can be explained in part by the observation that a number of site-specific DNA-binding transcription factors are able to bind HATs or HDAs (for reviews, see Tsukiyama and Wu 1997; Wu 1997; Armstrong and Emerson 1998).

The situation is more complicated when considering the SWI/SNF complex. No member of the SWI/SNF complex has been shown to possess DNA-binding activity nor has there been shown to exist a consensus binding-sequence for the SWI/SNF complex in loci known to be regulated by the complex (Peterson and Herskowitz 1992). However, Quinn *et al.* (1996) have shown that purified SWI/SNF complex binds to DNAs forming a four-way junction. The structure of DNA in a four-way junction has been proposed to resemble the crossed stucture of DNA leaving the nucleosome (Peterson 1996), suggesting that targeting of the SWI/SNF complex is to nucleosomal DNA in general and not to specific loci in the genome as is the case with Pc-G and SIR mediated silencing, and the modification of histone acetylation state. This raises the question of how the SWI/SNF complex has specific nucleosomal remodelling activity in response to the presence of transcription factors, that is "How is remodelling targeted?" Two complementary models have been put forward; the first, termed direction, is that transcription factor binding stimulates the activity of the SWI/SNF complex, while the second, termed opportunism, suggests that nucleosomal remodelling is an ongoing, undirected process that enables transcription factors to bind opportunistically to DNA with transiently disrupted nucleosomal structure (Varga-Weisz and Becker 1998). Both models have support.

The finding that the SWI/SNF complex is able to modify the DNase I digestion pattern of nucleosomal DNA in the absence of a DNA-binding protein indicates that the complex does not require a specific targeting activity (Côté *et al*. 1994). However, SWI/SNF-disrupted nucleosomal structure has been observed to be readily reversible in the absence of a DNA-binding protein (Logie and Peterson 1997), but stabilised in the presence of a GAL4 DNA-binding protein (Owen-Hughes *et al*. 1996). Thus, the SWI/SNF complex appears to be able to facilitate the opportunistic binding of DNA-binding protein.

In contrast to the observations of Côté *et al.* (1994), the DNA-bound thyroid hormone receptor has not been shown to cause significant chromatin remodelling in the absence of thyroid hormone (Wong *et al.* 1997). However, extensive remodelling was caused by the addition of thyroid hormone. In a similar situation, the glucocorticoid receptor, which has been shown to bind to nucleosomal DNA without causing nucleosome remodelling, stimulates yeast SWI/SNF activity when bound to the glucocorticoid response element (Li and Wrange 1995; Östlund Farrants *et al.* 1997). This result, in conjunction with the observation that Swi3p coimmunoprecipitates with glucocorticoid receptor in a *SWI1/SWI2* dependent manner (Yoshinaga *et al.* 1992), suggests that the glucocorticoid receptor actively targets the SWI/SNF complex to remodel nucleosomes when the receptor is bound to DNA. Consistent with the results obtained by Wong *et al.* (1997), both GAL4 and a GAL4-estrogen receptor-VP16 fusion showed nucleosomal remodelling *in vivo* under conditions in which the proteins would be in a transcriptionally activating state, but not in non-activating conditions (Stafford and Morse 1997). The finding that GAL4 only stimulates chromatin remodelling in the activated state can be reconciled with *in vitro* studies showing remodelling in the presence of DNA-bound GAL4 by considering that, *in vitro*, GAL4 lacks its associated *in vivo* repressor activities. A targeting model similar to that described by Wong *et al.* (1997) and Stafford and Morse (1997), involving the stimulation of chromatin remodelling by heat-shock factor, has been proposed to account for heat-shock induced chromatin remodelling in front of a stalled RNA polymerase II at the *hsp70* locus (Brown and Kingston 1997).

Targeting of chromatin remodelling in another setting has been suggested by the demonstration of an association of the SWI/SNF complex with RNA polymerase II by Wilson *et al.* (1996). Such a complex may be utilised to remodel chromatin in advance of the polymerase, enabling transcriptional elongation, and facilitating the form of regulation proposed by Brown and Kingston (1997). However, the association of the SWI/SNF complex with RNA polymerase II has been questioned by Cairns *et al.* (1996)

The examples above show that the role of transcription factors in regulating gene expression is clearly greater than proposed in the traditional model of transcriptional

activation, in which transcription factors were proposed to recruit RNA polymerase. The diversity of roles played by DNA-binding transcription factors supports the multidomain model of transcription factor structure (Ptashne 1988), but suggests that the degree of multifunctionality is probably greater than previously thought.

1.5 Chromatin – nuclear domain architecture interactions

Two aspects are apparent from the discussion of nuclear architecture in eukaryotic gene regulation above, namely, chromosomal pairing or clustering and association with the nuclear envelope (NE). The generality of these phenomena is indicated by the variety of organisms and developmental systems exhibiting them (Tartof and Henikoff 1991; Aramayo and Metzenberg 1996; LaSalle and Lalande 1996; Csink and Henikoff 1996; Georgiev and Corces 1995; Gotta *et al*. 1996; Dernburg *et al*. 1996; Marshall *et al*. 1996; Taniura *et al.* 1995; Ye and Worman 1996). Interesting questions arising from the generality of these phenomena such as: What roles do they play in nuclear organisation and gene expression? and How do they interact with each other and with other gene regulatory systems? I will briefly address possible roles in gene expression and then discuss, with reference to commonalities in mechanisms, how interactions between DNA-binding factors, chromatin regulatory factors and nuclear architecture may give rise to an integrated system of gene regulation and genetic storage.

1.5.1 Roles of chromatin – nuclear domain architecture interactions in gene expression

The information storage and retrieval model outlined in the first paragraph of this chapter proposes that chromatin structure is intimately involved in the maintenance and regulation of genetic information, acting to integrate the functions of information storage and retrieval. The evidence presented above suggests that chromosomal pairing or clustering and

association with the NE are involved in the maintenance and regulation of genetic information within the nucleus through the mechanisms discussed above. At the level of information retrieval, pairing and NE association are likely to be involved in the appropriate and efficient coordination of gene regulation. Chromosomal pairing in humans has been shown at loci that are subject to parental imprinting (LaSalle and Lalande 1996), possibly allowing communication between homologues to ensure appropriate imprinting establishment and maintenance. Similarly, the z^a allele has been shown to cause decoupling of transcription rates from homologous *Ubx* loci (Goldsborough and Kornberg 1996). Pairing of homologous loci and clustering of loci at the NE may also permit sharing of limited regulatory and translational resources. The observed clustering of silenced loci *in vivo* may allow efficient utilisation of silencing factors (Messmer *et al*. 1992; Dernburg *et al*. 1996; Gotta *et al*. 1996). Although clusters of Pc-G regulated loci have not been shown to associate with the NE (Messmer *et al*. 1992), yeast telomeres (and possibly the silent mating type loci) and the bw^D locus in conjunction with centromeric heterochromatin have been shown to exhibit strong NE association (Gotta *et al*. 1996; Dernburg *et al*. 1996), suggesting that this nuclear domain may be used to store inactive genetic material. Contrasting with the previous examples, but consistent with a role in ensuring efficient use of translational resources, pairing has been shown to be required for strong expression of the *Ubx* gene as asynapsis due to rearrangement reduces transcription from *Ubx* (Goldsborough and Kornberg 1996).

However, although pairing has been shown to play a role in the regulation of transcription of a number of genes, it is clearly not essential for viability in general, as flies heterozygous for multiply rearranged chromosomes, known to disrupt pairing, do not exhibit significantly reduced viability (Lindsley and Zimm 1992).

1.5.2 Implications of chromatin – nuclear domain architecture interactions

The modes of chromosomal organisation seen in the systems described in this chapter share common mechanisms involving interactions of DNA with multiprotein complexes

(reviewed in Pirrotta 1990; Wu 1993; Henikoff 1997; Cook 1997; Marshall *et al*. 1997). Indeed, pairing effects such as *trans* inactivation have been suggested to result from Pc-G–mediated clustering (Sigrist and Pirrotta 1997) and *z*-mediated pairing effects such as transvection may reflect a limited form of clustering, as Z is thought to exist as large aggregates *in vivo* (Bickel and Pirrotta 1990). A molecular role of the two complexes (and of other factors) appears to be to facilitate the close physical association of coregulated loci, permitting the efficient regulation of those loci as described above. This supposition is supported by a number of lines of evidence.

Both the Pc-G and Z have been shown to form large complexes *in vivo* (Franke *et al.* 1992; Bickel and Pirrotta 1990). Further, at least one member of the Pc-G has been shown to self-associate (S. O'Connell pers. comm.). Pleiohomeotic, a member of the Pc-G, and Z have DNA-binding activity (Brown *et al*. 1998; Benson and Pirrotta 1988), conferring on their respective complexes the capacity to bind DNA. Additionally, both complexes have been shown to bind non-uniformly to polytene chromosomes at a larger number of sites (Pirrotta *et al.* 1988; Messmer *et al*. 1992; Franke *et al*. 1992; Lonie *et al.* 1994). The self-association and DNA-binding activities of the complexes support the idea that they are involved in mediating associations between chromosomal loci *in vivo*.

In addition to the observation that Pc-G binding loci cluster into small numbers of foci (Messmer *et al*. 1992), studies with *P*-element constructs carrying PREs have shown preferential insertion at sites corresponding to Pc protein binding sites, suggesting that Pc-G–complex mediates an interaction between the PREs present in the *P*-element construct and endogenous PREs (Fauvarque and Dura 1993; Chan *et al*. 1994). The "homing" of PRE containing *P*-elements gives credence to the model that the Pc-G is able to form a protein-bridge between PREs. Interestingly, the favoured insertion of *P*-elements into Pc binding sites also argues against the model in which Pc-G mediated silencing is a result of dense packing of chromatin, as this would be predicted to reduce the chance of insertion at silent sites, and further suggests that Pc-G silencing is effected by clustering of coregulated loci to enable storage of silent loci in silent domains within the nucleus. Clustering of heterochromatic, and

possibly also Pc-G silenced regions may increase avidity of binding to the nuclear envelope through factors such as nucleosomes and HP1, both of which have been shown to associate with nuclear lamins (Taniura *et al*. 1995; Ye and Worman 1996).

Homologous pairing along the length of chromosomes has been suggested to be mediated by protein interactions at a large number of loci in the manner of a zipper (Gemkow *et al.* 1998). However, to generate the observed level of specificity of pairing, the molecular interactions necessary to mediate the associations between coregulated loci are likely to require a greater level of heterogeneity than would be provided by only two types of protein bridge. Thus to extend the model, the pairing of homologous chromosomes has also been proposed to be mediated by transcription factors (Pirrotta 1990), potentially explaining why *z* is not essential for viability (Goldberg *et al.* 1989). Consistent with this idea, the modification of transvection effects by *z* is accompanied by destabilisation, but not complete loss of Bithorax-Complex pairing (Gemkow *et al*. 1998), suggesting that other factors, such as the Pc-G or transcription factors, can mediate pairing and, further, that transvection effects must be more sensitive to homolgous association than is visible pairing. Other evidence for the involvement of transcription factors in homologous pairing comes from the finding that the intergenic promoter regions of the rRNA gene repeats are required for pairing between the X and Y chromosomes in *Drosophila* (McKee *et al.* 1992).

A consideration of factors involved in pairing/clustering and remodelling together leads to some interesting implications of the chromatin and nuclear architecture gene regulation model. In the discussion of eukaryotic chromatin structure regulation through nucleosome remodelling and histone modification, it became evident that the accessibility of DNA-binding transcription factors to DNA is affected by the state of chromatin condensation and nucleosome mobility and that transcription factor binding feeds back onto the chromatin condensation state. If transcription factors are involved in nuclear architecture, as has been proposed (Pirrotta 1990), then genome regulation at the nuclear architecture level is dependent on chromatin structure modulation and, similarly, accessibility of transcription

factors to nuclear domains is likely to feed into chromatin condensation state regulation through targeting of chromatin remodelling by transcription factors.

In summary, DNA-binding transcription factors affect chromatin structure and nuclear organisation. Conversely, chromatin structure and nuclear organisation affect DNA-binding transcription factor accessibility. Consequently, nuclear organisation and chromatin structure indirectly have reciprocal affects on each other through DNA-binding transcription factors. Thus, DNA-binding transcription factors are a central component of gene regulation, not only in the traditional model, but also in the model of gene regulation in which nuclear architecture is a key factor in the control of gene expression. To paraphrase Darwin (1876), the nucleus is "an entangled bank" in which classical regulatory factors and the "elaborately constructed forms" of chromatin structure and nuclear architecture are "dependent on each other in so complex a manner".

A crucial ramification of the proposal that transcription factors are central to nuclear regulation is that the protein-protein interaction effector domains of these proteins are likely to be intimately involved in nuclear regulation. Evidence of this exists for modulation of chromatin structure by transcription activation domains discussed above (Stafford and Morse 1997). Although no evidence exists for a similar role in regulation of nuclear architecture, it is an obvious implication of self-interaction domains associated with transcription factors which would allow the formation of protein bridges between homologous binding sites as has been suggested for Z (Bickel and Pirrotta 1990). A mechanism such as this has been proposed to explain the observation that gene regulatory regions associate with the nuclear matrix (Cockerill and Gerrard 1986; Gasser and Laemmli 1986); large scale aggregates of transcription factors with bound DNA would cosediment with the scaffold due to the large size of the aggregate (Bickel and Pirrotta 1990).

1.6 Dead ringer and the ARID family of proteins

A group of proteins that may play a central role in transcriptional regulation at both the nuclear architecture level and at the level of transcriptional machinery recruitment is the ARID (A/T-rich interaction domain) family of proteins. The ARID proteins share a novel, evolutionarily conserved DNA-binding domain and have been shown to be involved in a number of developmental and regulatory systems. The ARID-family of DNA-binding proteins are encoded by a number of partially characterised genes of known or predicted regulatory significance, including the *Drosophila dead ringer* (*dri*) (Gregory *et al*. 1996) and *eyelid* (Treisman *et al*. 1997) genes, yeast *SWI1* (O'Hara *et al*. 1988), the mammalian *bright* (Herrscher *et al*. 1995), *jumonji* (Takeuchi *et al*. 1995), *SmcX* (Agulnik *et al*. 1994b) and *SmcY* (Agulnik *et al*. 1994a) genes and the genes encoding the modulator response factors, MRF1 and MRF2 (Huang *et al.* 1996, K. Itakura pers. comm.), and the Retinoblastoma binding factors, RBP1 and RBP2 (Fattaey *et al*. 1993). Currently, five members of the family have been shown to bind directly to DNA *in vitro*, DRI, Bright, MRF1 and 2, and Eyelid (Gregory *et al*. 1996; Herrscher *et al*. 1995; Huang *et al*. 1996, K. Itakura pers. comm., J. Treisman, pers.comm.). Two of them bind specifically to a site similar to the Engrailed-homeodomain consensus binding-site (Gregory *et al*. 1996; Herrscher *et al*. 1995). In the cases of DRI and Bright, the DNA-binding activity has been shown to be mediated by the conserved domain (Gregory *et al*. 1996; Herrscher *et al*. 1995) which is predicted to form a helix-turn-helix motif similar to that of the HMG box that has been implicated in a number of chromatin regulatory systems (Gregory 1996; for reviews of the HMG-1 box family, see Landsman and Bustin 1993; Baxevanis and Landsman 1995). One member, the mouse Bright protein, has been shown to bind as a tetramer through the minor groove of the DNA, causing bending of the DNA (Herrscher *et al*. 1995). Another member, SWI1, is a component of the SWI/SNF complex which has been shown to bind DNA, also through the minor groove (Peterson and Herskowitz 1992; Quinn *et al*. 1996). However, no direct DNA-binding activity has been shown for SWI1 alone.

A theme that arises from functional comparison of the members of the ARID family is an association of the proteins with chromatin regulation in one way or another: ARID proteins have been shown to bind to matrix attachment or DNaseI hypersensitive regions (Herrscher *et al*. 1995; Huang *et al*. 1996, K. Itakura pers. comm.), to be involved in chromatin remodelling (Peterson and Herskowitz 1992) and to be expressed from the inactive X chromosome in mammals (Wu *et al.* 1994; Agulnik *et al*. 1994b). Although circumstantial, the association of ARID proteins with chromatin and gene regulation is suggestive of a involvement of ARID proteins in chromatin structure or regulation. The variation in DNA-binding specificity within the group suggests that the roles played by ARID proteins are likely to be divergent within the group.

Consistent with the functional difference, proteins in the ARID group show significant variation in sequence, maintaining only five invariant residues (figure 1.1). Grouping members on the basis of sequence identity allows clear sub-groups to be identified (figure 1.2). Notably, the only members shown to possess site-specific DNA-binding activity clearly group together to the exclusion of the other members of the family, suggesting that site-specificity evolved by modification of the standard ARID to the Dead ringer/Bright-like ARID.

1.6.1 DRI-like proteins

The two characterised DRI-like ARID proteins, DRI and Bright share an extended region of homology (here termed the eARID) outside and including the core ARID shared with the other members of the family (Gregory *et al*. 1996; Herrscher *et al*. 1995). Additionally, the conservation of the core ARID is higher in the eARIDs than between the eARID- and the other ARID-proteins. The strong sequence similarity exhibited by DRI and Bright reflects common functional characters, both proteins bind to DNA in a site-specific manner, recognising the sequence RATTAA (R indicates purine). In addition to the core recognition sequence, Bright has been shown to exhibit preferences for AT rich sequences

Figure 1.1 — Alignment of ARID proteins over the core ARID, generated using the MAP multiple sequence alignment program (Huang 1994) to generate the alignment and the MacBOX program v1.0.8 (a Macintosh implementation of BOXSHADE, M. D. Baron) to generate the shaded figure. Invariant and conserved residues are indicated by white text with a black or dark grey background respectively. Similar residues are indicated by black text with a light grey background. Residues are considered similar for this purpose if they fall into one of the following groups: FYW, ILVM, RK, DE, ST, QN or GA. Additionally, residues not in the groups above are considered similar if a residue in the group, ADEFIKLNPRSVY, is conserved in more than three sequences according to the following scheme: A similar to G; D similar to E or N; E similar to Q; F similar to Y or W; I similar to L,V or M; K similar to H; L similar to V or M; N similar to E or Q; P similar to G, R similar to K or H; S similar to T or A; V similar to I or M; and Y similar to W. The consensus was generated at each position from three or more similar or identical residues. Note that if a consensus is generated for a position on the basis of similar residues with no residue at or above the threshold of three, the consensus residue shown is the residue existing at that position in the first sequence listed.
Sources for the sequences are: Dead ringer (Gregory *et al.* 1996), Bright (Herrscher *et al.* 1995), T23D8.8 (PID g1628238), Eyelid (Treisman *et al.* 1997), RBP1/RBP2 (Fattaey *et al.* 1993), SmcX (Agulnik *et al.* 1994b), SmcY (Agulnik *et al.* 1994a), MRF1 (PID g188684), MRF2 (PID g553592), Jumonji (Takeuchi *et al.* 1995), SWI1 (O'Hara *et al.* 1988) c08b11.3 (PID g576996), r20183 (PID g774817), c01g8.8 (PID g1703595), zk593.4 (PID g1184605), YM8010.06 (PID g854446).

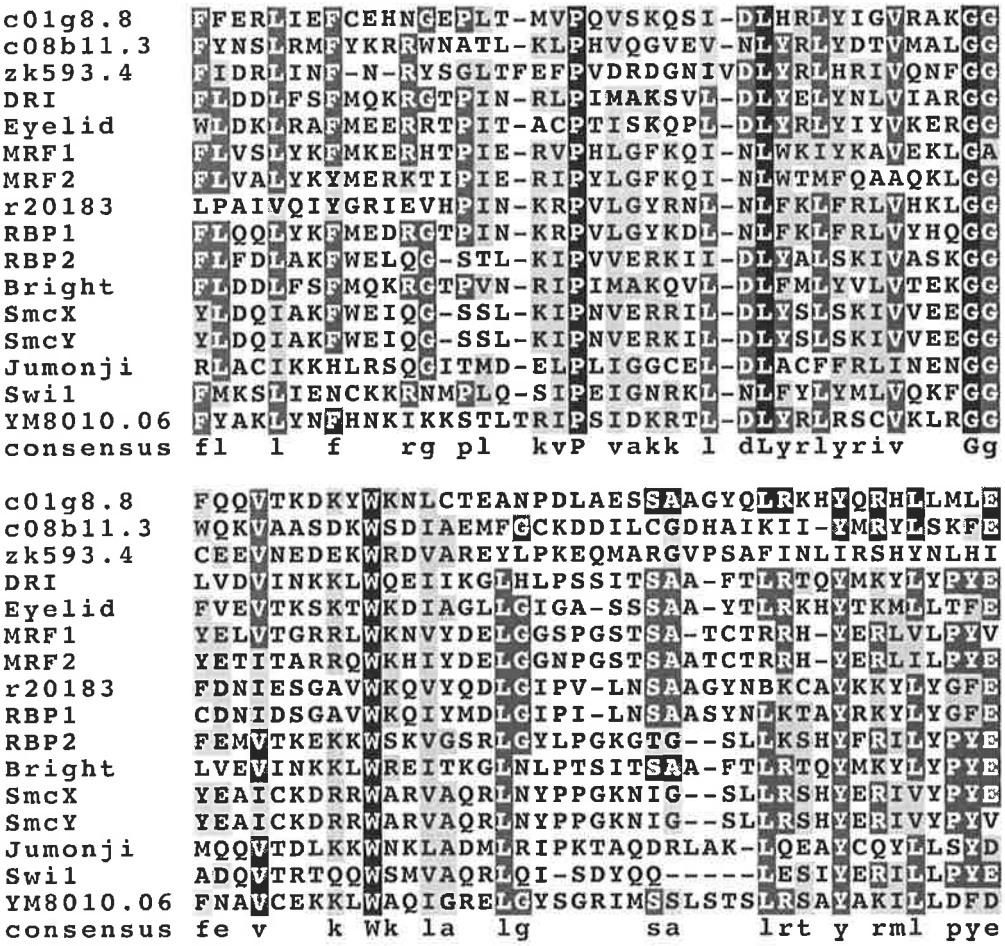
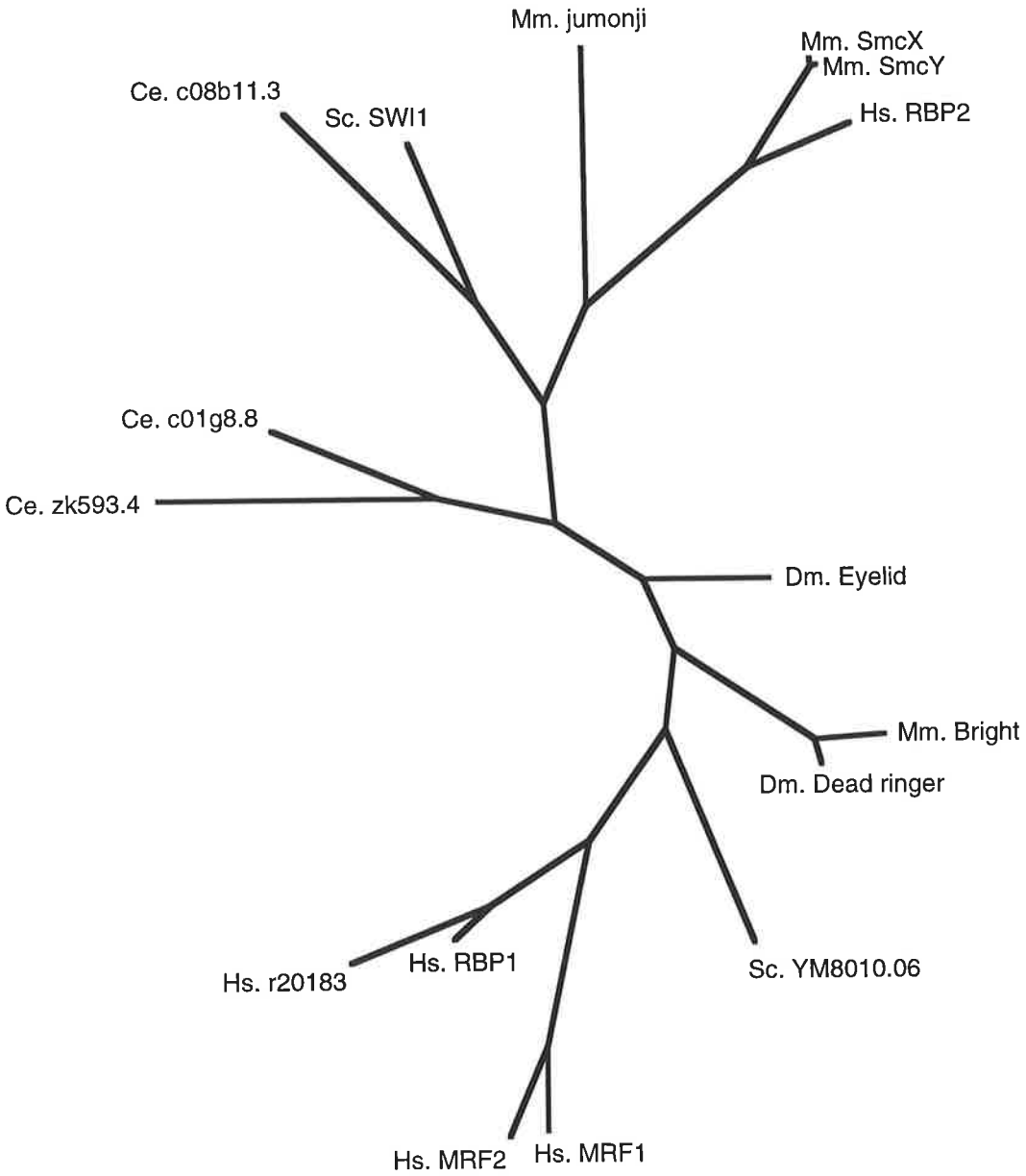

c01g8.8 FFERLIEFCEHNGEPLT-MVPQVSKQSI-DLHRLYIGVRAKGG
c08b11.3 FYNSLRMFYKRRWNATL-KLPHVQGVEV-NLYRLYDTVMALGG
zk593.4 FIDRLINF-N-RYSGLTFEFPVDRDGNIVDLYRLHRIVQNFGG
DRI FLDDLFSFMQKRGTPIN-RLPIMAKSVL-DLYELYNLVIARGG
Eyelid WLDKLRAFMEERRTPIT-ACPTISKQPL-DLYRLYIYVKERGG
MRF1 FLVSLYKFMKERHTPIE-RVPHLGFKQI-NLWKIYKAVEKLGA
MRF2 FLVALYKYMERKTIPIE-RIPYLGFKQI-NLWTMFQAAQKLGG
r20183 LPAIVQIYGRIEVHPIN-KRPVLGYRNL-NLFKLFRLVHKLGG
RBP1 FLQQLYKFMEDRGTPIN-KRPVLGYKDL-NLFKLFRLVYHQGG
RBP2 FLFDLAKFWELQG-STL-KIPVVERKII-DLYALSKIVASKGG
Bright FLDDLFSFMQKRGTPVN-RIPIMAKQVL-DLFMLYVLVTEKGG
SmcX YLDQIAKFWEIQG-SSL-KIPNVERRIL-DLYSLSKIVVEEGG
SmcY YLDQIAKFWEIQG-SSL-KIPNVERKIL-DLYSLSKIVVEEGG
Jumonji RLACIKKHLRSQGITMD-ELPLIGGCEL-DLACFFRLINENGG
Swi1 FMKSLIENCKKRNMPLQ-SIPEIGNRKL-NLFYLYMLVQKFGG
YM8010.06 FYAKLYNFHNKIKKSTLTRIPSIDKRTL-DLYRLRSCVKLRGG
consensus fl l f rg pl kvP vakk l dLyrlyriv Gg
c01g8.8 FQQVTKDKYWKNLCTEANPDLAESSAAGYQLRKHYQRHLLMLE
c08b11.3 WQKVAASDKWSDIAEMFGCKDDILCGDHAIKII-YMRYLSKFE
zk593.4 CEEVNEDEKWRDVAREYLPKEQMARGVPSAFINLIRSHYNLHI
DRI LVDVINKKLWQEIIKGLHLPSSITSAA-FTLRTQYMKYLYPYE
Eyelid FVEVTKSKTWKDIAGLLGIGA-SSSAA-YTLRKHYTKMLLTFE
MRF1 YELVTGRRLWKNVYDELGGSPGSTSA-TCTRRH-YERLVLPYV
MRF2 YETITARRQWKHIYDELGGNPGSTSAATCTRRH-YERLILPYE
r20183 FDNIESGAVWKQVYQDLGIPV-LNSAAGYNBKCAYKKYLYGFE
RBP1 CDNIDSGAVWKQIYMDLGIPI-LNSAASYNLKTAYRKYLYGFE
RBP2 FEMVTKEKKWSKVGSRLGYLPGKGTG--SLLKSHYFRILYPYE
Bright LVEVINKKLWREITKGLNLPTSITSAA-FTLRTQYMKYLYPYE
SmcX YEAICKDRRWARVAQRLNYPPGKNIG--SLLRSHYERIVYPYE
SmcY YEAICKDRRWARVAQRLNYPPGKNIG--SLLRSHYERIVYPYV
Jumonji MQQVTDLKKWNKLADMLRIPKTAQDRLAK-LQEAYCQYLLSYD
Swi1 ADQVTRTQQWSMVAQRLQI-SDYQQ-----LESIYERILLPYE
YM8010.06 FNAVCEKKLWAQIGRELGYSGRIMSSLSTSLRSAYAKILLDFD
consensus fe v k Wk la lg sa lrt y rml pye

Figure 1.2 — Dendrogram of ARID-containing proteins based on amino acid identity generated using the PAUP (Phylogenetic Analysis Using Parsimony 3.1.1, D. L. Swofford) program and the sequence alignment shown in figure 1.1. Branch lengths indicate degree of sequence divergence. Only one member of an orthologous group is shown for simplicity, except in the case of Dead ringer and Bright.


Mm. jumonji
Mm. SmcX
Mm. SmcY
Ce. c08b11.3
Sc. SWI1
Hs. RBP2
Ce. c01g8.8
Ce. zk593.4
Dm. Eyelid
Mm. Bright
Dm. Dead ringer
Hs. r20183
Hs. RBP1
Sc. YM8010.06
Hs. MRF2
Hs. MRF1

flanking the core site (Herrscher *et al*. 1995). Although Gregory *et al*. (1996) do not comment on a similar sequence preference for DRI, examination of their published DRI target sequences suggests that such a preference exists.

The Bright protein was identified on the basis of binding to a regulatory element of B-cell–specific immunoglobulin heavy chain (IgH) gene and has been shown to bind a number of sites throughout the IgH enhancer (Herrscher *et al*. 1995). Consistent with the binding of Bright to these elements, Bright has been shown to activate expression downstream of the IgH enhancer in transient transfection assays. Similarly, DRI has been shown to act as a transcription factor in a multiprotein complex containing Dorsal and Groucho to repress ventral expression of *zerknüllt* during embryonic development (Valentine *et al.* 1998). The observation that Bright is a matrix attachment region (MAR) binding protein (Herrscher *et al*. 1995) along with the other similarities between the two proteins suggests that DRI may also act as an MAR binding protein. This suggestion implicates the eARID proteins in chromatin/nuclear architecture regulation in addition to transcriptional regulation. However, the caveat of Bickel and Pirrotta (1990) mentioned above, regarding MAR binding proteins should be heeded, as Bright has been shown to tetramerise (Herrscher *et al*. 1995), possibly resulting in the assignment of MAR status to Bright-binding regions.

A factor complicating the analysis of the eARID proteins is that no conserved sequence outside the eARID has been identified. Thus the tetramerisation functions of Bright which has been shown to require sequences outside the eARID, may not exist in DRI, or may be mediated by a heterologous domain. All of the deletion constructs expressing DNA-binding activity which were used to determine functional domains within Bright were able to *trans*-activate (Herrscher *et al*. 1995), suggesting that the eARID may be an integrated DNA-binding–*trans*-activation domain. Thus the question that comes out of comparison of the eARID proteins is whether the eARID is the regulatory effector-domain (Ptashne 1988) of the eARID proteins in addition to being the DNA-binding domain, or if there are other regions within the protein which confer effector function

1.7 Comparative molecular biology towards understanding functions of DRI

Comparative biology is the central tenet of model organism based biology; analysis of a small number of organisms with varying degrees of evolutionary divergence enables the elucidation of a more general picture of the system being studied. I will be using comparative molecular biology to examine functional domains associated with the eARID in *Drosophila melanogaster* Dead ringer.

The specific question to be addressed in this thesis is whether the eARID proteins mediate *trans*-regulation through effector domains external to the eARID. This will involve the isolation of cDNA species encoding eARID proteins that are more closely related to the known eARID proteins, reasoning that less conservative motifs will be more recognisable when comparing less diverged sequences. The assumptions made in this analysis are that DRI and Bright are closely related sequences resulting from divergent evolution and that conserved functions are mediated by conserved sequences. The implication of this assumption is that either DRI or Bright may be used as starting points for evolutionary analysis, as results obtained from one source may, with justification, be used in comparisons with the other. With this in mind and for reasons discussed in chapter 3, homologues of DRI and Bright will be isolated from vertebrate rather than arthropod species. Further, the *in vitro* and *in vivo* roles of the eARID and conserved domains identified by analysis of heterologous eARID proteins will be examined. In particular, the *in vivo* significance of the DRI ARID will be examined in the light of the observation that the Swi1 ARID is not required for SWI function (C. Peterson, pers. comm.)

Chapter 2: Genomic structure of the *Drosophila dri* gene

I have yet to see any problem, however complicated, which when you looked at it in the right way, did not become still more complicated. - Poul Anderson

2.1 Characterisation of the *dri* locus towards identifying mutations

Functions of the extended ARID proteins have begun to be characterised. However, the focus of previously published studies has been on *in vitro* DNA-binding specificity, developmental expression and transcription-factor activity (Gregory 1996; Herrscher *et al.* 1995). These studies defined the regions of the proteins able to confer DNA-binding activity and oligomerisation. The functions demonstrated *in vitro* have yet to be shown to be important for the *in vivo* function of the proteins. Indeed, a mutant SWI1 lacking the ARID fails to show any loss of SWI1 function (Craig Peterson, pers. comm.). *In vivo* analysis of mutants lacking specific *in vitro* defined functions or domains would allow the importance of these domains for organismal function to be examined.

As phage-style residue saturation mutagenesis to generate mutants is impractical for *Drosophila* genes, it is necessary to use a rational approach for mutagenesis of specific domains; mutants generated *in vitro* in defined regions of the gene are introduced into the genome to replace the function of the endogenous gene. Although targeted gene replacement mutagenesis in *Drosophila* has been investigated (Gloor *et al.* 1991), this approach has not been shown to be practicable. Thus a targeted mutagenic approach in *Drosophila* requires the complete removal of the endogenous gene in combination with expression from mutated genomic rescue constructs or from heat-shock or GAL4 driven constructs (Brand and Perrimon 1993).

This chapter describes the characterisation of the genomic locus of the *dead ringer* gene, an important step towards the generation of *dri* mutants, and genomic rescue with mutant genomic constructs.

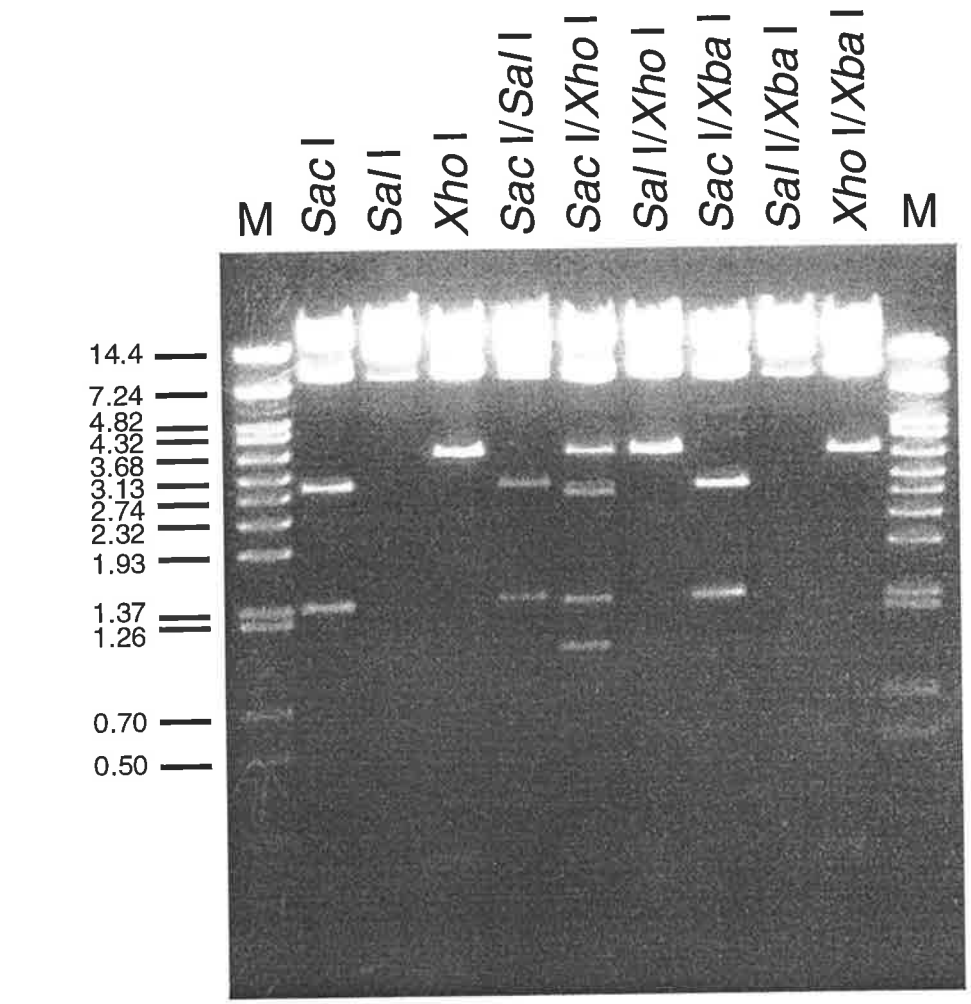
2.2 Cloning and restriction mapping the genomic locus

The genomic region corresponding to the *dri* locus was cloned by screening a *Drosophila* genomic EMBL3 SP6/T7 library (Clontech) with the bk60 *dri* cDNA. This region was then extended further 5' by rescreening with an 800bp *Xho*I fragment at the 5' end of the cloned region. This contig of cloned DNA covering the full transcription unit as defined by the extent of the cloned cDNA (Gregory *et al*. 1996). All clones, λ1.1 and λ5.2 (Gregory *et al*. 1996) and λ4.1b (this study, figure 2.1), were mapped with the enzymes *Sac*I, *Sal*I, *Xho*I and *Xba*I.

2.3 Mapping the genomic structure

In order to define the intron-exon structure of *dri*, fragments spanning the genomic region were subcloned (figure 2.2); the sequence of those fragments showing hybridisation to the cDNA (not shown) was determined by Southern analysis (data not shown). All exons but the first were unambiguously localised on the map by hybridisation (not shown) and sequence analysis (figure 2.3). The genomic sequence shows the *dri* transcription unit to have 12 exons, and to span approximately 20-30kb. The first exon was localised to a 2.5kb *Eco*RI fragment within a 7.7kb *Sac*I/*Xho*I fragment at the 5' end of the contig. The previously published *dri* sequence starts with the pentanucleotide sequence CGGGA (Gregory *et al*. 1996), a sequence not present in the genomic DNA. This pentamer may be the artefactual result of cDNA synthesis over the CAP structure on the *dri* mRNA. The sequence immediately upstream of the transcription start site lacks a consensus TATA box; two motifs resembling the TATA consensus are present(Bucher 1990; Bucher and Trifonov 1986), but changes present in the *dri* motifs have been shown to abolish TATA-binding protein dependent transcription *in vitro* using human and yeast TBP (Wobbe and Struhl 1990). Thus it can be concluded that *dri* utilises a TATA-less promoter.

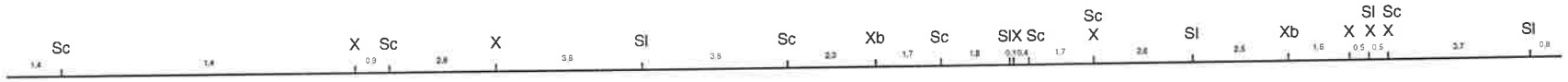
Figure 2.1 — Mapping the genomic clone λ4.1b. Note that the 3.7kb band in the *Sac* I/*Xho* I lane is a result of incomplete digestion by *Sac* I at the right-hand *Sac* I site. The map shown below was derived from the fragment sizes shown in the ethidium bromide stained agarose gel.



Sc - *Sac* I
Sl - *Sal* I
X - *Xho* I
Xb - *Xba* I

1kb

Figure 2.2 — Phage clones and plasmid subclones covering the region of the *dead ringer* locus. The origin of the subclones with respect to the restriction map is indicated by the position of the clone. Cloning sites used are indicated for subclones: E - *Eco* RI; Sc - *Sac* I; Sl - *Sal* I; X - *Xho* I. Bracketed sites are not present in the genomic DNA.



Clones

Subclones

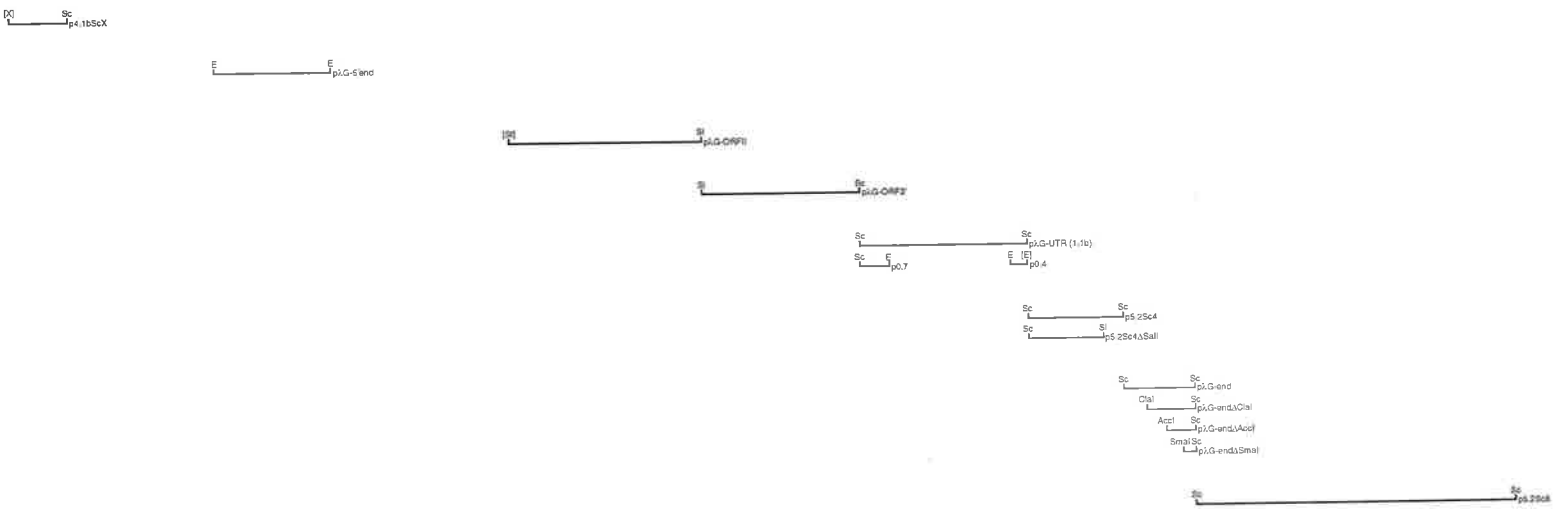


Figure 2.3 — Genomic structure of *dead ringer*. Exonic sequence is shown in uppercase. Intronic sequence is shown in lower case. Two near matches to the TATA consensus are indicated by double underlining (Bucher and Trifonov 1986; Bucher 1990). The initiation and termination codons are underlined. Five nucleotides present at the 5' end of the published cDNA (Gregory *et al*. 1996), but absent from genomic sequence are struck through.

```
      tcgtgtgtgtgtgagtgtatcggaatgagcgaataagaaggcggcgtatt
      gcgcacactctcacacacacacgaatcgggcattgcatagattgcgatta
      caattttaatttaattgaccacttcccccacccccctcccataacaaacga
      tggaacggccatccgattggctgcgaacgagttcggggaacgagtgtgaa
      ttggaacgtgcaattccattcgttcgatttaaacggccgttccggcaaca
      tttgggaatgttgcagcaacttgttgctattgctgtgagatggtgtcggt
      gtcttattcgccagtttcttgttcgctactcagaccgatttgaacttttcg
      agtgaacgagacgagttccgtgcggtcggataacggcttttcaaatggac
      acatcactgtaattagataacgtcgcaatgtggattaataacaattcgtg
      ttctgtgtgtgtgagcgtgtggagcagtaaacttcttgatccataattaa
      agtgagaagcatagaagtcactaggaaccaaatcgcactaaacaaccgca
      agcaattgcaaacaaacagaaaggcataaatgcaagtgaacggttcttaa
      gcgcactctcctgtttctgttttccaaccagaataaagtgcaaagaggat
      cgacgatcacaagtaacggttcgcgctcacgaaaaaaaaaaagaaatcaa
      atgctggccgcaaagtaatgccatatagaagctaaaaacaacacaacaaa
      agtgtgctgtgccaattaaacttcagctaaggaaataacatcaaaagtat
      caaaggtctgcgcaggt

0     CGGGACCTTTGATCAACCGCAACATTCGTTATATTCGTCATATTCGTTAT      50
      ATTCGTGACATTCGAGCAGGGTGATCCAATTCGCCTCAATGCAACTGCGA     100
      GTGCACCCTACTATGGACTGTAGCGGTCGTTCAACCAGCAATATTGAACG     150
      CGATTCCGATCTGGGGGACGATTTG

      gtaagaaaaggaacttgtaacacccaatatgcacctctacagtaacgccc
      gaagttattattattataaaggatt...aactcatttcaaaaaccatctg
      cgtttccttctgtgcag
10.2                                  TCACATGGTGATCGAACGGACGACG     200
      AGATGCGCGACTGCGACTCCGTGGATGGGGAGCATCATCAGCTGAGCGCT     250
      AAGGCGGCGATTGCAGCCCGCCTGAGTCACACAGTTTCCGGCGGCGGTGG     300
      TAGCTTCGCCAGCCCCGAACCGCAGACCGAGCTGCCCCTGAG

      gtgagtgactgctcctgattaatgtcctggaggtgcggaggtgcggtaat
      aatcgctgcaacaatcctcttcccacag

10.4                                                CCATCACC     350
      ATCAACTGCCGCCGAATCATCCGCTCAACGCTCTGGGCAGCTTCATGGGC     400
      ATCGGCGGACTACACAGCATTCCAAATCTCCAGCACAGCGACGTGCTGGA     450
      GAAGCTCAAGATGCAGGTGCGCGACATGAAGGTGGGCCTGATG

      gtgagtgttaaggtgtctatcgagcttgaagatactgtagctactatcat
      acagctaaagtaatctgaagggaagcaattgtatgccaaagtttagacaa
      atctgtcgaggattggatatttatattggaaatacctgtatttacccgca
      g
10.7                                              GAACAGG     500
      ACTATGCTGCCGCAGCACATGCCGCTGCTTTCGGGGCCAATATGCTGCCC     550
      ACGACGATCAGCTCGGGCTTCCCACTGCCCCACAACTCGGTGGCCTTTGG     600
      CCACGTCACATCGGCGCCCAGCGGTGGGAATGGGAGCAGCTACAACGGAG     650
      GCACCACCCCCACGAATAGCTCCAACAGCAATGCGACCACCAATGGAGGC     700
      GGCACAGCCGGACCTGGAGGAACAGGAGGATCGGGCGGCGGAGGAGCAGG     750
      AGGAGGTGGCGGCGGCGGCGGAGGAGTGGGTGGCCACCAGTTCTCGTTCG     800
      CATCACCCACAGCAGCGCCGAGCGGCAAAGAAG

      gttggtatctatctatcgcgtggcaggcggcgatccatttgtgtttaact
      gttgcactcagcccgccactttgcag

11.2                                          CCAATTCCGCATCGAAC     850
      TCGTCGACGTCCAGCGAGGCTTCCAATTCATCGCAGCAGAATAATGGATG     900
      GAGCTTTGAAGAGCAGTTTAAACAAGTCAGACAG

      gttagtagagtgctgggcgaagggaagggagttggaaagtggtggtgtac
      agggggcggggtgcggg...gcttaactggtattgggaaagcgtttgcat
      tgaattaggaaacttaatagcccagtctaatctatacatttctaacactt
      ctaattgcag

15.5                                          CTCTATGAAATCAATG     950
```

Figure 2.3 continued

ATGATCCCAAGCGCAAAGAGTTCCTGGACGACTTGTTCTCGTTTATGCAA 1000
AAGCGCG

gtaagttgcacattgcattgaagtttttcaaggggcgctcatagagttgg
gatggcgggcg...tccccctattcacatttcag

18.7 GAACTCCGATCAATCGGCTGCCGATAATGGCCAAATCGGTGCT 1050
GGATCTCTACGAGCTGTACAATCTGGTGATAGCCCGCGGCGGCTTGGTGG 1100
ATGTTATCAACAAGAAGCTGTGGCAGGAGATCATCAAGGGGCTGCACCTG 1150
CCCTCCAGCATCACCAGTGCCGCCTTCACCCTGCGCACCCA

gtccgtggatattccattcaatccgttcaatccttgcaaactcaaacatc
catcgcggttactgccgacgtgagctatgtcaaggcaaacacaaacagca
a...gagctccctctgtgggatcgtgatttgtttgcctgtcgctttcgta
tttacacgataattcaacttaaacaggacatcaacgtgaagtttgcccct
ttctctctctcccgctgccgcttagagcgtctgtttgccgcgcttattaa
taattttatgcatcttgcag

19.0 ATACATGAA 1200
GTATCTGTACCCGTACGAGTGCGAGAAAAAGAATCTGAGCACGCCGGCGG 1250
AGCTGCAGGCGGCCATCGATGGGAATCGCCGGGAAGGACGCCGCTCCAGC 1300
TACGGCCAGTACGAGGCCATGCACAACCAGATGCCGATG

gtgagtacacaaagcagcagcttggatcagccaagacttgtggtactaat
gtttctgttatggtatgctcctatgctccatag

19.3 ACGCCCATTTC 1350
GCGACCCTCTCTGCCCGGTGGCATGCAGCAAATGTCGCCGCTGGCGCTGG 1400
TCACCCATGCCGCGGTGGCCAACAATCAGCAGGCTCAGGCCGCCGCTGCC 1450
GCCGCCGCAGCTCATCATCGCCTGATGGGCGCTCCCGCCTTTGGCCAGAT 1500
GCCCAATCTGGTCAAGCAGGAGATCGAGAGCCGGATGATGGAGTATCTAC 1550
AGCTGATCCAGGCCAAGAAGGAGCAGGGCATGCCGCCGGTCCTAGGCGGC 1600
AATCATCCCCACCAGCAGCAGCACTCACAGCAGCAGCAGCAGCAGCATCA 1650
CCACCAGCAGCAGCAGCAGCAGTCGCAGCAGCAACATCACCTGCAGCAGC 1700
AGCGCCAGCGATCGCAGAGTCCGGATCTGAGCAAGCATGAGGCACTCAGT 1750
GCTCAGGTGGCCCTGTGGCATATGTATCACAACAACAACAGCCCACCGGG 1800
ATCGGCACACACCTCGCCGCAGCAACG

gtaagctctctcacgatgtgcgacattgatttccgtcctttcgagatttc
gcaattgccaaatgcttacagtggcaaccagtttaattgaacgcactcca
cctggtttttgctttggcagattcataatgttgaaatctgcctaaattca
ggttttacttaattaacaaagtaaatgcatattgaaaatagtaatattta
aaacgaattaagttggaataaattcagagtgctggcgccatctagctaac
acacttctcaggcaataggaacagcgccacctgcttgcgaagcatagctg
tacagcgaagacactaaccatttaccatttctctccattctccactccag

20.2 CGAAGCCCTGAACCTGTCCGACT 1850
CGCCTCCAAATCTCACAAATATCAAACGGGAACGCGAACGAGAACCCACA 1900
CCAGAGCCCGTGGACCAGGATGACAA

gtaagttggctccatacccgcacaataaaatgtcagaaagtcgaggcatt
caatgttcaaactttgcgtaccttatgtttgttaaagaatagattttgtt
ttatccagcttttgtactttaacgtaagattacatttgtgctaaccaaat
tggtttcttcatccag

20.4 ATTTGTGGACCAGCCACCTCCAGC 1950
GAAGCGCGTGGGCAGTGGCCTCCTTCCGCCCGGCTTTCCCGCCAACTTCT 2000
ACCTGAATCCACACAACATGGCCGCTGTGGCAGCAGCTGCGGGATTCCAT 2050
CACCCATCGATGGGCCACCAGCAGGATGCCGCATCCGAGGGCGAACCAGA 2100
GGATGACTACGCTCACGGTGAGCACAATACCACGGGCAACTCGTCCTCCA 2150
TGCACGACGACAGCGAACCGCAGCAGATGAACGGACACCACCACCATCAG 2200
GCCCACCATCTGGACAAGTCCGACGACTCGGCCATTGAGAACTCACCCAC 2250
CACGTCGACCACCACCGGTGGGTCGGTGGGTCATCGTCACAGTTCGCCCG 2300
TTTCCACTAAGAAAAAGGGCGGCGCTAAGCCCCAGAGTGGAGGAAAGGAC 2350

Figure 2.3 continued

CTGCCGACCGAGGACAAGGATGCGTCGTCCAGTGGCAAGCTCAATCCTCT 2400
CGAGACGCTGAGCCTGCTGTCCGGAATGCAGTTTCAAGTGGCACGAAATG 2450

gtgagttcactcgaaattaacttaaagcatttatcacggtagcaagtaag
ataagatcggaaataactgaagtattaaaatcactttcaattgaatagaa
agattatatgtacctcagtagttggattgtagccacccaagcatatgacc
attgaactaatataaaccgtgatcccttccag

21.1

GAACTGGCGATAACGGCGAACCGCAGCTGATTGTCAATCTGGAGCTTAAT 2500
GGCGTCAAGTACTCGGGAGTGCTGGTGGCCAATGTGCCGCTGTCACAGAG 2550
CGAGACTAGGACGAGCTCACCCTGCCACGCCGAAGCCCCGACAGTCGAGG 2600
AGGAGAAGGATGAGGAGGAGGAGGAGGAGGAGGAGCCGAAAGCCGCTGAA 2650
GAGGAATCGCATCGATCGCCAGTTAAACAGGAGAACGAGGATGCCGACCA 2700
GGACATGGAGGGCAGTGAAGTCCTTCTGAACGGAGGCGCTTCGGCGGTGG 2750
GTGGTGCTGGTGCTGGTGTGGGTGTGGGTGTGGGTGTGCCCCTGCTCAAG 2800
GATGCCGTGGTCAGCTAGGAAGGATAACGCGACGAACCGGCATGGAAAGA 2850
AGAGACAAGGAAAGGAATACCAAAGTCATATTTTTCAGCAAAATACAAAA 2900
GCAGTTAAACTCAAACACACACACACACACACAAGCACACAGATGCCGAA 2950
CCTAAATGTACATACACACGCATGCGGCAGGAGCACAGGAGCATCAAGGA 3000
TGTGGACAGGACATGAAAAGCGCATTAGGCGCATCCCAGCATTCGATGTT 3050
CATGCATGCAGCACCATACCCAACCGCCCGAAACTCAAGTGCGAGATGCC 3100
GCCAGTAGAAGATCATGGAATATATATACATTATCGAGAGTGTATATAGG 3150
ATGCGTGTGTGCTCCATTTCCACATTGCATAAAAGACAAGCCAGCGAGCG 3200
TGTAGTTATCGATAGCCATTACTTACATAAGCGAGTGCATAAGCATCACT 3250
TCGGATACATACAACCAATCCAATACCCGTATATATATATGATATGATAT 3300
ATGATATATGAATAGGTTACGATTCCCACACACCAATTTGTTTTTAGCGT 3350
GAAATCTGTTTTATCCTATTACGAACTTATCCTTTTTTTGTGTTAGGCTGC 3400
AAGGCTTTTTGTACATATCAATCAATACTTAGCTCGTAGGCCATATGCTA 3450
TACTATATACTATACTATATATATATATATATATATATATACTATATATA 3500
CTATACTATATCGGAACCTAGTTTGTAAGCACACAAGCACGCATCACACC 3550
GCAGTAATACACATGCATATCCATTATGACCGCATCGTAGACATATTTAG 3600
ACGTAAAATCAACACAAGTAATAAGCATAAATGTATCTGTCGTAAAATTA 3650
AATAGCGCTTTTGGCACCCACACACACACACACACAAAAA 3690

cacacaccacttcgattgccccacagtttctactgctccagagccaagcg
gatggtggcaagccacggcagattgatgagcatattctaagcataagcgt
tcaatttattgaataataaatctatttaattttagttcaatgcgcattcg

2.4 Molecular mapping of lethal *dri P*-element insertions

Two lethal, non-complementing, enhancer-trap *P*-elements, *l(2)02535* and *l(2)05096*, have been localised to 59F1-2, the cytological location of *dri*, and shown to express a *lacZ* reporter gene within the *P*-element in the same developmental pattern as *dri*. These insertions also disrupt expression of *dri* in some tissues (Gregory 1996) . The *l(2)05096* insertion site has been shown previously by restriction site mapping to be 2.9kb 5' of the 5' most *Xho*I site in the contig (Kortschak 1993), within the 7.7kb fragment containing the first *dri* exon (figure 2.4). Sequence obtained from genomic DNA flanking the two *P*-element insertions indicates that *l(2)02535* and *l(2)05096* are inserted 471bp and 439bp upstream of the transcription start site respectively (figure 2.5). These data show that the first exon of the *dri* transcription unit is located 8.8kb 5' of the second exon, showing the transcription unit to be 22kb in size.

The genomic structure of the *dri* locus, including the location of the *P*-elements, is summarised in figure2.6.

2.5 Discussion

The primary objective in mapping the *dri* genomic region was to allow the identification of mutants in the locus and to determine the feasibility of generating genomic constructs for attempting rescue.

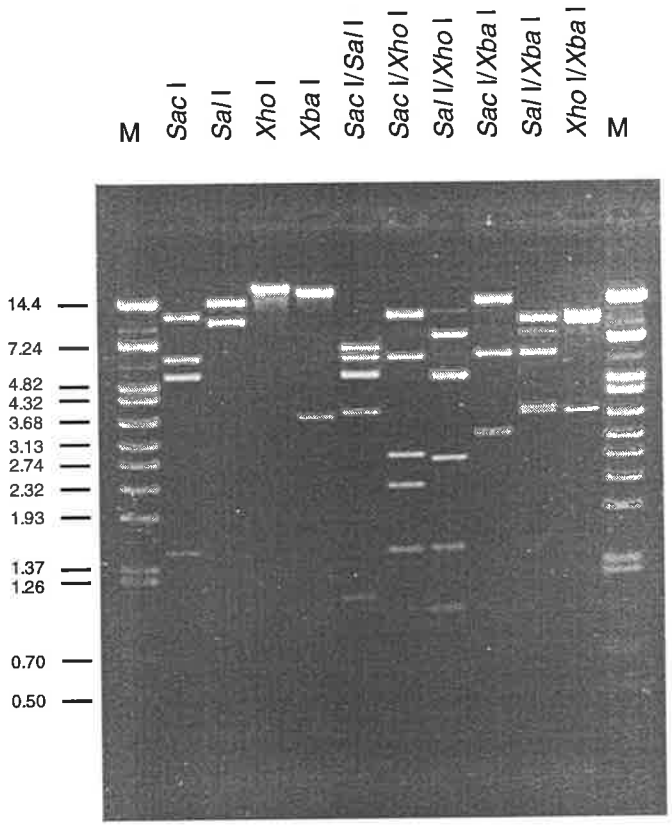
Two partially characterised mutants had previously been mapped to the region (Gregory 1996). I have shown here that these *P*-elements are inserted immediately upstream of the transcription start site of *dri*. Deletions with break-points within the transcription unit, generated by imprecise excision of the *l(2)05096* P-element, show the same phenotype as the two *P*-element insertions, indicating that the gene affected in all cases is *dri* (Gregory 1996). The *P*-element insertions *l(2)02535* and *l(2)05096* will be refered to as dri^7 and dri^8 respectively.

Figure 2.4 — Localisation of the *l(2)05096 P*-element insertion

A. Restriction endonuclease analysis of the p59F^{P2} plasmid isolated by plasmid rescue from the *l(2)05096*/CyO strain (Kortschak 1993). Note that the 7.5kb band in the *Sac* I/*Xho* I lane appears to be the result of incomplete digestion by *Xho* I.

B. Alignment of the restriction map of the *dri* locus with the restriction map of the plasmid p59F^{P2} derived from the gel shown in A.

A


M
Sac I
Sal I
Xho I
Xba I
Sac I/Sal I
Sac I/Xho I
Sal I/Xho I
Sac I/Xba I
Sal I/Xba I
Xho I/Xba I
M
14.4
7.24
4.82
4.32
3.68
3.13
2.74
2.32
1.93
1.37
1.26
0.70
0.50

B

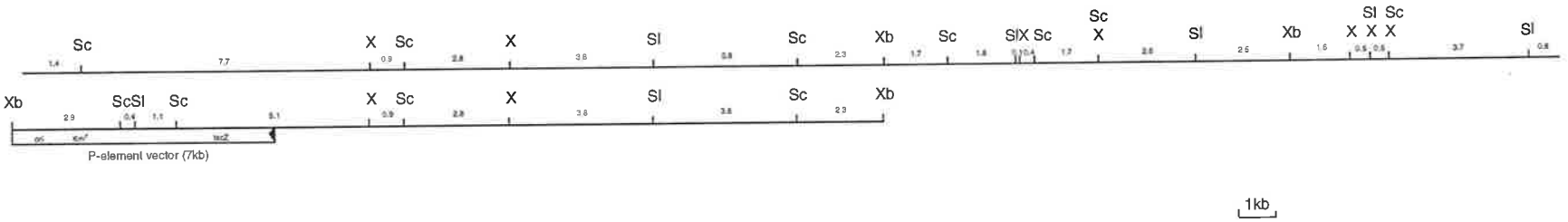

Sc
X
Sc
X
Sl
Sc
Xb
Sc
SlX
Sc
Sc
X
Sl
Xb
Sl
Sc
X X X
Sl
Xb
ScSl
Sc
X
Sc
X
Sl
Sc
Xb
P-element vector (7kb)
1kb

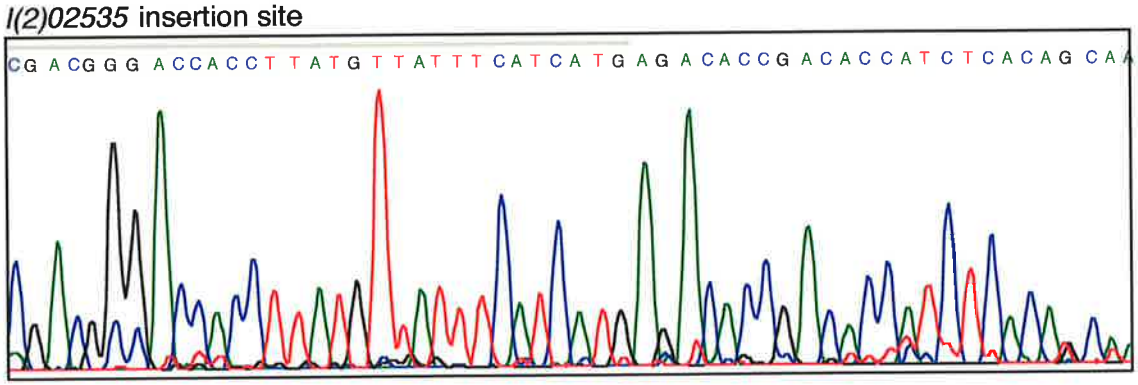
Figure 2.5 — Localisation of *l(2)02535* and *l(2)05096* *P*-element insertions with respect to the transcription initiation site.

A. Sequence traces covering the insertion sites of the two *P*-elements. The grey bar above the sequence trace indicates *P*-element sequence.

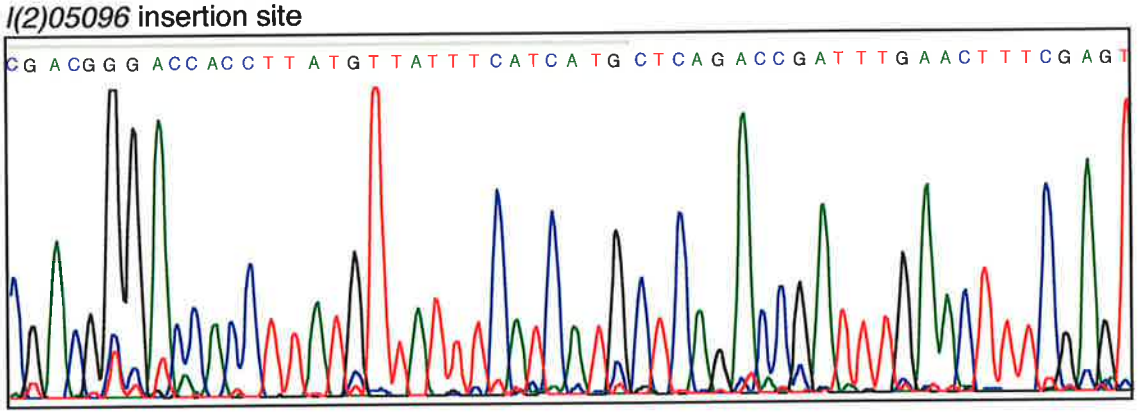
B. Genomic sequence upstream of the transcription initiation site. The locations of the *P*-element insertions are indicated by boxed ∇s. The sequence bracketed with the ∇ is the predicted 8 nucleotide tandem repeat generated by *P*-element insertion (Engels 1989).

A

l(2)02535 insertion site



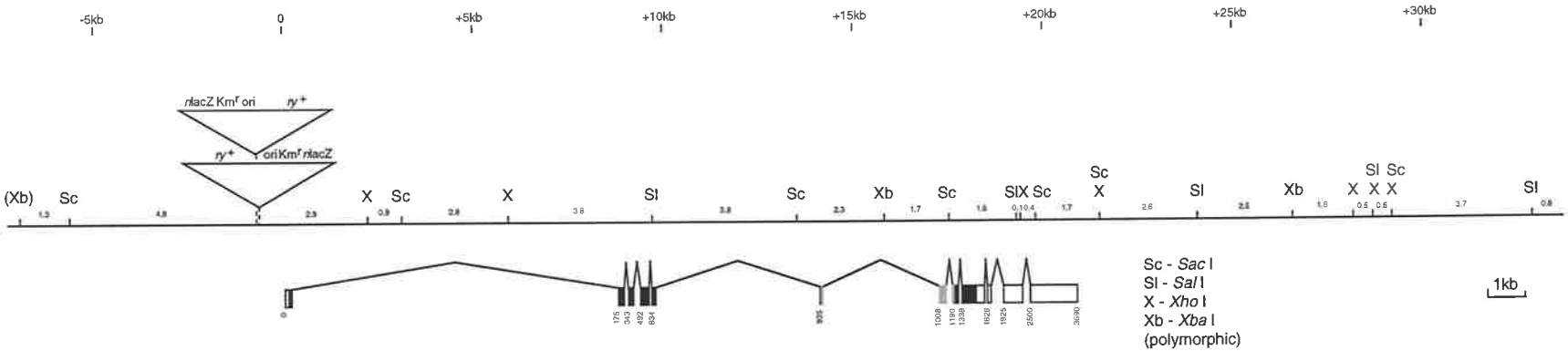
l(2)05096 insertion site



B

```
tcgtgtgtgtgtgagtgtatcggaatgagcgaataagaaggcggcgtatt
gcgcacactctcacacacacacgaatcgggcattgcatagattgcgatta
caattttaatttaattgaccacttcccccacccccctcccataacaaacga
tggaacggccatccgattggctgcgaacgagttcggggaacgagtgtgaa
ttggaacgtgcaattccattcgttcgatttaaacggccgttccggcaaca
tttgggaatgttgcagcaacttgttgctattgctgtgagatggtgtcggt
gtct(∇cggtgtct)tattcgccagtttcttgttcgcta(ctcagaccV)
ctcagaccgatttgaactttcgagtgaacgagacgagttccgtgcggtcg
gataacggcttttcaaatggacacatcactgtaattagataacgtcgcaa
tgtggattaataacaattcgtgttctgtgtgtgtgagcgtgtggagcagt
aaacttcttgatccataattaaagtgagaagcatagaagtcactaggaac
caaatcgcactaaacaaccgcaagcaattgcaaacaaacagaaaggcata
aatgcaagtgaacggttcttaagcgcactctcctgtttctgttttccaac
cagaataaagtgcaaagaggatcgacgatcacaagtaacggttcgcgctc
acgaaaaaaaaaaaagaaatcaaatgctggccgcaaagtaatgccatatag
aagctaaaaacaacacaacaaaagtgtgctgtgccaattaaacttcagct
aaggaaataacatcaaaagtatcaaaggtctgcgcaggt

CGGGACCTTTGATCAACCGCAACATTCGTTATATTCGTCATATTCGTTAT      50
ATTCGTGACATTCGAGCAGGGTGATCCAATTCGCCTCAATGCAACTGCGA     100
GTGCACCCTACTATGGACTGTAGCGGTCGTTCAACCAGCAATATTGAACG     150
CGATTCCGATCTGGGGGACGATTTG

gtaagaaaaggaacttgtaacacccaatatgcacctctacagtaacgccc
gaagttattattattataaaggatt...
```

Figure 2.6 — Genomic map of the *dead ringer* locus showing *P*-element insertions and intronic structure.


-5kb
0
+5kb
+10kb
+15kb
+20kb
+25kb
+30kb
nlacZ Km^r ori
ry+
ry+
ori Km^r nlacZ
(Xb)
Sc
X
Sc
X
Sl
Sc
Xb
Sc
SlX Sc
Sc
X
Sl
Xb
Sl Sc
X X X
Sl
Sc - Sac I
Sl - Sal I
X - Xho I
Xb - Xba I
(polymorphic)
1kb

The map of the *dri* region described here shows that the locus is relatively large, incorporating a transcription unit of 22kb with 12 exons. Additionally, *P*-element mediated deletion experiments mentioned above have generated a large number of non-complementing deletions without loss of regions within the transcription unit (Gregory 1996), suggesting that *cis*-acting sequences outside this region are crucial for *dri* expression. The size of the *dri* locus suggests that the genomic fragment rescue approach to studying the *dri* gene product is unlikely to be feasible due to construct size limitations in *P*-element mediated transformation of *Drosophila*.

Chapter 3: *Dead ringer* homologues in vertebrate species

Repetition is the only form of permanence that nature can achieve. - George Santayana

3.1 Idenfication of function by analysis of evolutionary conservation

Due to the monophyletic origin of metazoan species it is not surprising that apparently divergent organisms share similarities in tissues and organs, and the developmental processes involved in generating them. Although it is estimated that arthropods and vertebrates were evolutionally separated approximately 670 million years ago (Ayala *et al.* 1998) they still share not only conserved genes, but conserved genetic pathways and regulatory networks that are responsible for the development of conserved tissues and organs.

A number of well characterised developmental systems have been analysed in divergent species, allowing us to build up a picture not only of how the system develops in the model organism, but how the system has evolved. A number of examples of conservation of developmental pathways come from systems involved in the development of mesodermally derived tissues in *Drosophila* and their homologues in other species.

Mesodermal development in flies depends on the patterned expression of the bHLH transcription factor Twist (Baylies and Bate 1996) depending in turn on transcriptional activation of *twi* by the dorso-ventral patterning system through Dorsal (Thisse *et al.* 1991; Thisse and Thisse 1992). In the absence of *twi*, flies fail to gastrulate and die (Leptin and Grunewald 1990). However, manipulation of *twi* activity after its requirement for gastrulation has shown that it is also involved in the determination of muscle fate (Baylies and Bate 1996). Homologues of *twi* have been cloned from *Xenopus*, mouse and *Caenorhabditis elegans* (Hopwood *et al.* 1989; Wolf *et al.* 1991; Harfe *et al.* 1998). The functions of the *twist* genes have clearly diverged. For example, although *twi* is required for gastrulation in *Drosophila*, MTwist is not expressed until after this stage in mice (Leptin and Grunewald 1990; Gitelman 1997). However, there are definite similarities in the roles of these genes in mesodermal development. For example, the *C. elegans* TWI homologue directly activates *ceh-24* a

homologue of *tinman*, a target of TWI activation in *Drosophila* and a regulator of heart development (Harfe *et al*. 1998; Yin *et al.* 1997; Azpiazu and Frasch 1993; Bodmer 1993).

Although the appearance of the insect and vertebrate heart is very different, they have been shown to share an initially similar mode of development (for a review see Bodmer and Venkatesh 1998). The *Drosophila* heart is specified by the action of the homeobox gene, *tinman* (Azpiazu and Frasch 1993; Bodmer 1993). A mouse homologue of *tinman*, *Nkx-2.5*, is expressed initially in cardiogenic progenitor cells, in the developing heart throughout embryogenesis and later in the foetal and adult heart (Lints *et al.* 1993). Mice deficient for *Nkx-2.5* develop a beating linear heart tube, but fail to initiate looping morphogenesis and die nine to ten days post coitum (Lyons *et al.* 1995). The existence of other *tinman* homologues expressed in cardiogenic progenitors in the mouse raises the possibility that these genes may be partially rescuing the function of *Nkx-2.5* (Newman and Krieg 1998).

The homology of the insect and vertebrate eyes is similarly disguised by the disparate appearance of these two organs. The presence of a conserved master regulator, Eyeless/PAX-6, strongly indicates that the two visual systems are homologous (Quiring *et al.* 1994; Halder *et al.* 1995). The *ey/pax-6* genes encode homeodomain proteins of the paired class. These proteins have been shown to regulate the expression of eye-specific genes: *rhodopsin 1* in *Drosophila* and the lens ζ- and βB1-crystallin genes in guinea pig and *Xenopus* respectively, and to induce lens formation when expressed ectopically in *Xenopus* (Sheng *et al.* 1997; Sharon-Friling *et al.* 1998; Altman *et al.* 1997).

It is clear from these examples that the use of a variety of model organisms can facilitate the understanding of developmental processes. Although different models appear different superficially, the fundamental nature of the system being investigated is often conserved and brought into contrast by the idiosyncrasies of each organism. I have therefore set out to isolate *dead ringer* homologues from divergent species to identify conserved functions in the *dead ringer* gene.

3.2 Isolation of *dead ringer* homologues

dri has been shown to have a complex pattern of expression throughout *Drosophila* development (Gregory *et al*. 1996, RDK and T. Shandala unpublished data), and is implicated in various developmental processes in its role as a transcriptional regulator (Valentine *et al*. 1998, Shandala et al., in preparation). The tissues in which *dri* is expressed share no obvious features that would suggest a particular role of *dri* in developmental regulation. Either diverse tissues require a common regulator for a shared function, or DRI interacts with different elements and factors in the different tissues to perform varied roles. Currently we are unable to distinguish these two possibilities.

Although a vast number of DNA-binding proteins have thus far been described, these proteins can be grouped into a small number conserved families. The evolutionary economy of domain motifs in protein structure has made the inference of function based on taxonomic grouping of macromolecular structures one of the central paradigms in molecular biology, expanding the original theme of comparative developmental biology in which analysis of the development of disparate organisms has enabled a generalised picture of developmental processes to be acquired.

In order to examine the role of *dri*-like genes in development and test whether *dri* has divergent functions, I have sought to isolate and characterise divergent homologues of *Drosophila dead ringer*. The organisms selected as models for investigating *dri* function in *Drosophila* were *Homo sapiens* and *Danio rerio*. *Homo sapiens* was chosen in the hope that analysis of phenotypic effects of mutants of the *dri* homologue would be made possible by the existence of patients with a genetic disease caused by mutations in the human *dri* homologue, making use of the large database of information that has been collected relating to human genetic disorders (for a review of transcription factors involved in human genetic disease, see Engelkamp and van Heyningen 1996). Complementing this, the well characterised development and ease of study of *D. rerio*, together with the observation that this species has maintained a high number of members in paralogous groups generated by

genome duplication, favoured this species for analysis of conservation of tissue expression patterns and potential separation of functions performed by the *Drosophila dri* gene.

The initial sequences corresponding to *H. sapiens* and *D. rerio* homologues, named *DRIL1* and *dri1* respectively, were isolated using a degenerate PCR technique. Primers designed to be specific for the *dri*-like gene-specific eARID box sequences (see methods, chapter 6) were used to amplify fragments from a HeLa-derived λZAP cDNA library (Clontech) and a random primed total embryonic *D. rerio* cDNA library (J. Campos-Ortega, University of Köln). Two rounds of amplification with an annealing temperature of 37°C were performed: 30 cycles using the ARID 5' outside and ARID 3' outside primers were followed by 30 cycles using the ARID 5' inside and ARID 3' inside primers. The products obtained from degenerate PCR were then used to screen the appropriate libraries by hybridisation using standard techniques.

The HeLa-derived cDNA library used in the initial PCR amplification was screened by plaque hybridisation using stringent conditions (0.1xSSC, 65°C). One clone, pϕ2.2, corresponding to *H. sapiens DRIL1*, 1.3kb in length, was isolated and sequenced. The sequence showed identity to several ESTs from Genbank (accession numbers AA219626, AA232770, AA371391, T49195, T54949, R36384) but none of these sequences extended to the 5' end of the open reading frame. The 5' end was discovered during construction and expansion of a continuous cosmid contig around the *PRTN3* locus by Zimmer *et al*. (1992) and Pilat *et al*. (1994) (figure 3.1). Southern analysis using stringent hybridisation conditions (0.1xSSC, 65°C) indicated that the *DRIL1* sequence is present in the human genome in only one copy (figure 3.2A).

The embryonic *D. rerio* cDNA library used as the PCR template to isolate the original *dri1* clone was screened by plaque hybridisation using stringent conditions (0.1xSSC, 65°C). The library was first screened with the degenerate PCR product. Two clones, p4112 and p2111, were isolated and their nucleotide sequence determined. p4112 was found to correspond to the original *D. rerio dri1* PCR derived cDNA, while p2111was a novel cDNA, termed *dri2*. These clones were then used to rescreen the library to extend the cDNAs. Four

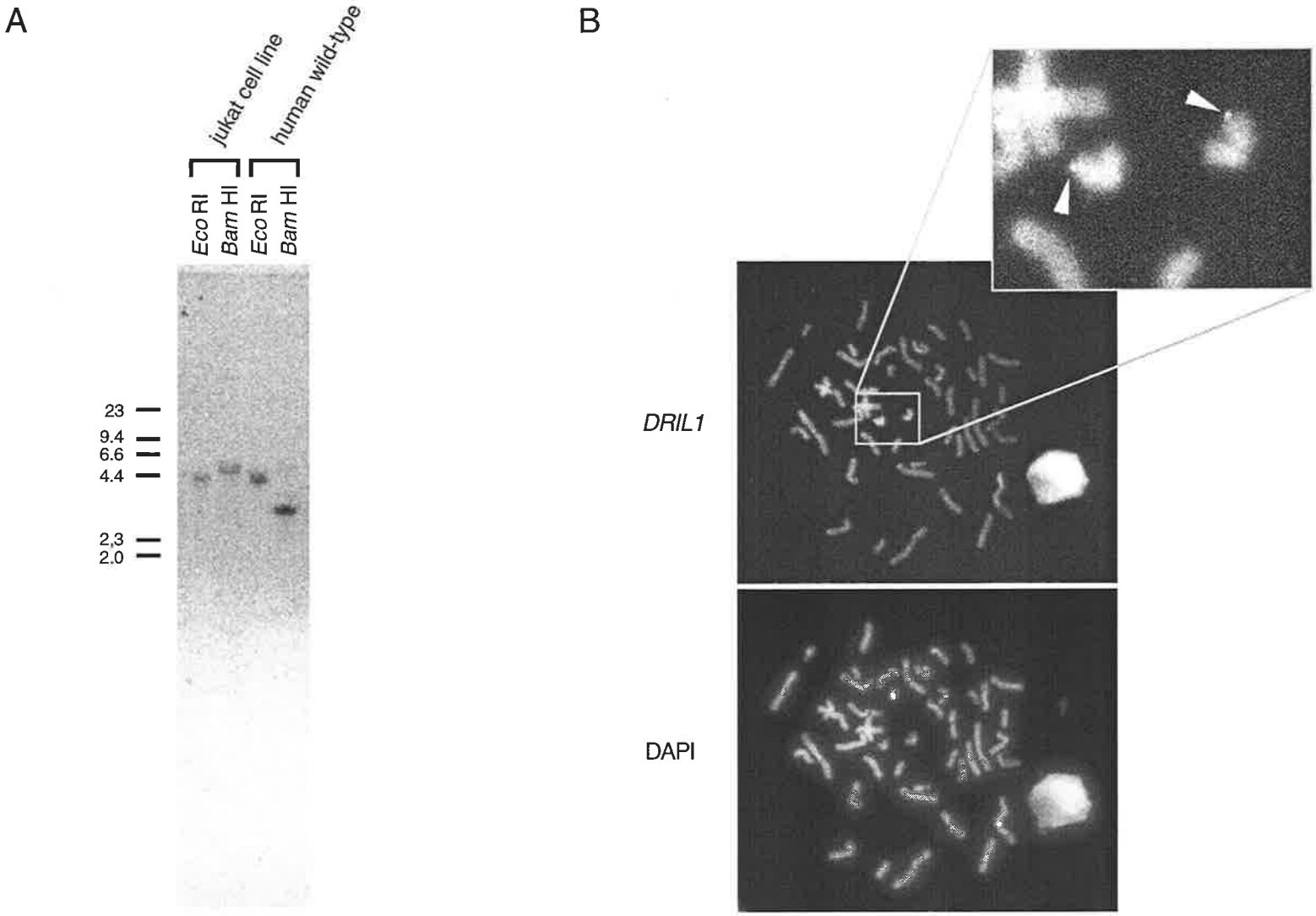
Figure 3.1 — *Homo sapiens dril1* cDNA sequence. The eARID and REKLES domain are shown in bold roman and bold italic type respectively.

```
1                                                                                                   GCATGCCCCATCAGCCTTCA
21   GCTTGAGCCCGGCGGCCCCCGCCCCCGCCCCCTGCCACCCTGCACTGCCCCGGCTCCCCCGCGGCCCCCACGCTGCAGTGCGGCCGGGCC
111  CCCTCCCCGCAGGGGCCGCCCCCGCCGCCCACCCCTAGCGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGCCCGGGCCGCAGGGCC

1    M  K  L  Q  A  V  M  E  T  L  L  Q  R  Q  Q  R  A  R  Q  E  L  E  A  R  Q  Q  L  P  P  D
201  ATGAAACTACAGGCCGTGATGGAGACGCTGTTGCAGCGGCAGCAGCGGGCGCGCCAGGAGCTGGAGGCCCGGCAGCAGCTGCCCCCCGAT

31   P  P  A  A  P  P  G  R  A  R  A  A  P  D  E  D  R  E  P  E  S  A  R  M  Q  R  A  Q  M  A
291  CCCCCTGCTGCACCCCCCGGCCGGGCCCGGGCTGCCCCCGACGAGGACAGAGAGCCCGAGAGTGCCCGGATGCAGCGGGCTCAGATGGCC

61   A  L  A  A  M  R  A  A  A  A  G  L  G  H  P  A  S  P  G  G  S  E  D  G  P  P  G  S  E  E
381  GCACTGGCAGCCATGCGGGCTGCAGCTGCGGGCCTGGGACACCCAGCCAGCCCCGGCGGCTCTGAGGATGGGCCCCCAGGCTCGGAGGAG

91   E  D  A  A  R  E  G  T  P  G  S  P  G  R  G  R  E  G  P  G  E  E  H  F  E  D  M  A  S  D
471  GAGGACGCGGCCCGGGAGGGGACACCGGGCTCACCCGGGCGAGGCAGAGAAGGGCCAGGAGAGGAGCACTTTGAGGACATGGCCTCCGAC

121  E  D  M  K  P  K  W  E  E  E  E  M  E  E  D  L  G  E  D  E  E  E  E  E  E  D  Y  E  D  E
561  GAGGACATGAAGCCCAAATGGGAGGAGGAGGAGATGGAGGAAGACCTCGGGGAGGATGAGGAGGAGGAGGAGGAGGATTACGAGGATGAG

151  E  E  E  E  D  E  E  G  L  G  P  P  G  P  A  S  L  G  T  T  A  L  F  P  R  K  A  Q  P  P
651  GAGGAGGAGGAGGACGAGGAGGGGCTGGGCCCCCCAGGCCCTGCCAGCTTGGGCACCACGGCACTGTTCCCCCGAAAGGCCCAGCCACCC

181  Q  A  F  R  G  D  G  V  P  R  V  L  G  G  Q  E  R  P  G  P  G  P  A  H  P  G  G  A  A  H
741  CAGGCCTTCCGCGGCGATGGCGTTCCCAGGGTGCTGGGGGGCCAGGAGCGGCCGGGGCCTGGCCCTGCCCACCCCGGAGGGGCCGCCCAC

211  V  A  P  Q  L  Q  P  P  D  H  G  D  W  T  Y  E  E  Q  F  K  Q  L  Y  E  L  D  G  D  P  K
831  GTAGCCCCGCAGCTGCAGCCGCCTGACCACGGCGACTGGACTTACGAGGAGCAGTTTAAGCAGCTCTACGAACTCGACGGGGACCCCAAG

241  R  K  E  F  L  D  D  L  F  S  F  M  Q  K  R  G  T  P  V  N  R  I  P  I  M  A  K  Q  V  L
921  AGGAAGGAATTCCTGGATGACTTGTTCAGCTTCATGCAGAAGCGAGGGACACCTGTGAACCGCATCCCCATCATGGCCAAACAGGTCCTT

271  D  L  F  M  L  Y  V  L  V  T  E  K  G  G  L  V  E  V  I  N  K  K  L  W  R  E  I  T  K  G
1011 GACCTGTTCATGCTGTACGTGCTGGTGACGGAGAAGGGCGGCCTCGTGGAGGTCATCAACAAGAAGCTGTGGCGTGAGATCACCAAGGGC

301  L  N  L  P  T  S  I  T  S  A  A  F  T  L  R  T  Q  Y  M  K  Y  L  Y  P  Y  E  C  E  K  R
1101 CTCAACCTGCCCACGTCCATCACCAGTGCAGCCTTCACCCTGCGGACCCAATACATGAAGTACCTGTACCCCTACGAGTGTGAGAAGCGG

331  G  L  S  N  P  N  E  L  Q  A  A  I  D  S  N  R  R  E  G  R  R  Q  S  F  G  G  S  L  F  A
1191 GGCCTCAGTAACCCCAATGAGCTCCAGGCAGCCATAGACAGCAACCGACGGGAGGGCCGGCGCCAGAGCTTTGGTGGCTCCCTCTTTGCC

361  Y  S  P  G  G  A  H  G  M  L  S  S  P  K  L  P  V  S  S  L  G  L  A  A  S  T  N  G  S  S
1281 TACTCGCCAGGCGGGGCACACGGCATGCTCTCCTCACCCAAGCTACCCGTGTCCTCCCTGGGCCTGGCCGCAAGCACCAATGGCAGCTCC

391  I  T  P  A  P  K  I  K  K  E  E  D  S  A  I  P  I  T  V  P  G  R  L  P  V  S  L  A  G  H
1371 ATCACCCCCGCCCCTAAGATCAAGAAAGAGGAGGACTCAGCCATCCCCATCACAGTCCCTGGCCGCCTGCCTGTGTCCCTGGCGGGCCAC

421  P  V  V  A  A  Q  A  A  A  V  Q  A  A  A  A  Q  A  A  V  A  A  Q  A  A  A  L  E  Q  L  R
1461 CCTGTGGTGGCAGCCCAGGCAGCAGCTGTGCAAGCAGCAGCCGCCCAAGCAGCTGTGGCCGCACAGGCAGCTGCCCTGGAACAGCTGCGG

451  E  K  L  E  S  A  E  P  P  E  K  K  M  A  L  V  A  D  E  Q  Q  R  L  M  Q  R  A  L  Q  Q
1551 GAGAAGCTGGAGTCTGCAGAGCCTCCGGAGAAGAAGATGGCCCTGGTGGCCGATGAGCAGCAACGGCTGATGCAACGTGCACTCCAGCAG

481  N  F  L  A  M  A  A  Q  L  P  M  S  I  R  I  N  S  Q  A  S  E  S  R  Q  D  S  A  V  N  L
1641 AACTTCCTGGCCATGGCGGCCCAGCTGCCCATGAGCATTCGGATCAACAGCCAAGCCTCCGAAAGCCGCCAGGACTCTGCTGTGAACCTG

511  T  G  T  N  G  S  N  S  I  S  M  S  V  E  I  N  G  I  M  Y  T  G  V  L  F  A  Q  P  P  A
1731 ACGGGCACCAACGGCAGCAACAGCATCAGCATGTCGGTGGAGATCAACGGCATCATGTACACAGGAGTTCTGTTTGCTCAGCCGCCGGCC

541  P  T  P  T  S  A  P  N  K  G  G  G  G  G  G  S  S  S  N  A  G  G  R  G  G  N  T  G  T
1821 CCCACGCCAACCTCTGCTCCCAACAAAGGAGGCGGCGGCGGCGGCAGCAGCAGCAACGCAGGCGGCCGGGGAGGAAACACCGGAACC

571  S  G  G  Q  A  G  P  A  G  L  S  T  P  S  T  S  T  S  N  N  S  L  P  *
1911 AGCGGCGGCCAGGCTGGGCCAGCGGGGCTGTCCACACCCTCCACATCTACCTCAAATAACTCGTTGCCTTAACCGCATCACTCCCCACCC
2001 GCCACCCACCCTGGAGCCCGCCGGCCTGGGCAGGGGGTCCAGGTGGGCCACACAGGGGCCAGGATGGCGGAAGATACGGGTGGGGAGGGA
2091 AGATATCCAGAAAGGAGCCACAGCTGACGCCAAAAAGAAAAGAAAAAAGATATATATATATATATATATATATATACGTATATATATAAA
2181 GAAGAATTTAATAAAACAGGGGAAAACCAAGGAACACTTGAATTTCTCAGGTTTTGGACATTCAGAGAGATGAATTGTGAGAACAGCAAA
2271 GAAATCCATCAGAAAAACAGAAAGAGGCAGACGTTTCCCAGGGCGTTCAGGCAGCCCTGATGGACCGAAGGCTCTGGTGTCTGGTTTGGC
2361 CCCACAGCAGTGTGGGCCGATCCTGTTTACCTCATACATCCCTGCACTGTGTGTTTTCATTTTTGTCTGCTTTAGTTCTCTTTTATTTTC
2451 TATTCACCACACACTCACCACTCCCAGCTTCTCGTGTCCAGTGAAACCCCTGAACCAAGATCACTGAATTTTTGTTTTTTTTCTTGTTGCT
2541 TTGGGAAATTTTTTTTTTCTCTGTAGGGTTTTTAAGAGGTTTCGGGGGTTTTGTTGTGTAAATATTCTATTTTATTCTTGGGGGGATCAAA
2631 CCTTAGGAAAAGGATATCTATATATCTATATAGCTATATATTTGTGTTCCTTCAGGGAAACTGGTCTTGAAAAAGCAAGAAAAAAAAGCA
2721 AAAAA
```

Figure 3.2 — Southern and fluorescence *in situ* hybridisation analysis of human genomic DNA probed with *dril1*.

A. Southern blot of genomic DNA from the human derived jurkat cell line and human blood (gift of R. Keough), digested with *Eco* RI and *Bam* HI and probed with pϕ2.2.

B. Fluorescence *in situ* hybridisation to a human metaphase spread and DAPI staining of the same spread. FISH and cytology were performed by H. J. Eyre.


A
jukat cell line
human wild-type
Eco RI
Bam HI
Eco RI
Bam HI
23
9.4
6.6
4.4
2,3
2.0
B
DRIL1
DAPI

further clones, p4112.2 and p4112.3 identified by p4112, and p2111.1 and p2111.2 identified by p2111, were isolated and sequenced. The two cDNAs, *dri1* and *dri2* were covered by the clones p4112 and p4112.2, and p2111, p2111.1, p2111.2 and p4112.3 respectively (figures 3.3 and 3.4). That cDNA clones corresponding to both *dri1* and *dri2* were isolated in stringent hybridisation screens using probes specific for *dri1* indicates the extremely high sequence identity between the two genes. Sequence comparisons revealed 78% protein sequence identity over the region encoding the eARID and 87% over the region covered by the PCR derived probe. Although only the *dri2* contig covered the complete ORF as indicated by the presence of three in frame stop codons upstream of the translation initiation codon, both contigs covered the region encoding the eARID and extended to the 3' end of the ORF (figures 3.3 and 3.4). Comparison of non-stringent genomic Southern analysis (2x SSC, 42°C) using the original PCR clone as a probe with stringent genomic Southern analysis (0.1x SSC, 65°C) with clones isolated from the hybridisation screen used as probes, indicates that one of the bands present in the reduced stringency Southern cannot be accounted for by the genes *dri1* or *dri2* (figure 3.5). Thus it is possible that another *dri* homologue is present in *D. rerio*.

3.3 Localisation of the *DRIL1* locus

To determine the chromosomal location of the *DRIL1* gene, fluorescence *in situ* hybridisation was performed in collaboration with the Adelaide Women's and Children's Hospital Cytogenetics Department (Kortschak *et al*. 1998). The 1.3kb *DRIL1* pϕ2.2 cDNA fragment isolated from the HeLa-derived library used in the initial PCR screen was used for this purpose. Clear fluorescent signals on both chromatids were noted on 19p13.3 in twenty metaphases (figure 3.2A FISH picture, Kortschak *et al*. 1998). There were no consistent signals detected from other chromosomal sites. Localisation to 19p13.3 was confirmed by showing physical linkage to the D19S886 marker on 19p13.3 (Kortschak *et al*. 1998).

The marker D19S886 is intimately linked to the Peutz-Jeghers-Syndrome in several large pedigrees (Amos *et al.* 1997) . Peutz-Jeghers Syndrome is an autosomal dominant

Figure 3.3 — *Danio rerio dri1* cDNA sequence. The eARID and REKLES domain are shown in bold roman and bold italic type respectively.

```
   1    L  L  S  G  A  P  G  A  T  V  R  V  K  E  E  P  E  E  E  P  S  P  A  D  R  P  S  D  T
   1 CGTTGCTGTCGGGGGCCCCTGGAGCTACAGTCAGGGTTAAAGAGGAGCCCGAGGAGGAGCCGAGCCCTGCTGACCGGCCCTCAGATACAT

  30 S  P  N  G  P  A  D  W  S  Y  E  E  P  F  K  A  N  G  A  V  S  W  S  D  D  T  D  A  A  R
  91 CTCCAAATGGCCCTGCGGACTGGAGCTATGAGGAGCCATTTAAAGCGAATGGTGCTGTGTCCTGGTCAGATGATACAGACGCAGCTCGCA

  60 S  R  G  E  A  S  R  D  F  A  K  L  Y  E  L  D  S  D  P  Q  R  K  E  F  L  D  D  L  F  T
 181 GCAGAGGTGAAGCTTCCAGAGATTTCGCCAAGTTATACGAACTGGACAGCGACCCGCAGAGGAAGGAGTTTCTGGATGATCTGTTCACAT

  90 F  M  Q  K  R  G  T  P  V  N  R  I  P  I  M  A  K  Q  V  L  D  L  Y  R  L  Y  R  L  V  T
 271 TCATGCAGAAACGAGGGACACCAGTGAACCGGATCCCCATCATGGCGAAGCAGGTTCTGGATCTCTACAGGCTCTACAGGCTGGTGACGG

 120 E  K  G  G  L  V  E  V  I  N  K  K  I  W  R  E  I  T  K  G  L  N  L  P  T  S  I  T  S  A
 361 AGAAGGGAGGCCTGGTGGAGGTCATTAACAAGAAGATCTGGAGGGAGATCACTAAAGGCCTGAACCTGCCCACCTCCATCACCAGTGCTG

 150 A  F  T  L  R  T  Q  Y  M  K  Y  L  Y  P  Y  E  C  E  R  K  A  L  S  S  P  N  E  L  Q  A
 451 CATTCACGCTGCGCACACAGTATATGAAGTATCTGTACCCGTACGAGTGTGAGCGTAAAGCCTTGAGTTCCCCTAATGAGCTTCAGGCTG

 180 A  I  D  S  N  R  R  E  G  R  R  P  S  Y  S  N  T  L  F  H  L  S  P  P  P  G  S  A  P  H
 541 CCATTGACAGTAACCGGCGTGAGGGGCGTCGGCCGAGTTACAGCAACACCCTGTTCCACCTTAGTCCTCCTCCTGGCTCCGCCCCTCACC

 210 L  L  T  S  P  T  M  Q  L  S  A  L  S  A  L  N  G  F  Q  P  T  A  A  A  L  K  K  R  L  E
 631 TCCTGACTTCGCCCACAATGCAGCTGTCGGCTCTGAGTGCCCTGAACGGGTTCCAGCCGACAGCAGCGGCTCTGAAGAAGAGACTGGAAG

 240 D  M  S  P  S  V  F  A  G  R  S  L  P  L  T  P  Q  Q  A  V  C  F  A  R  A  A  T  L  E  Q
 721 ACATGTCTCCGTCTGTGTTTGCGGGCCGCTCTTTGCCGTTGACCCCTCAGCAGGCGGTTTGTTTTGCGCGCGCAGCGACTCTGGAGCAGC

 270 L  R  E  K  L  E  S  E  A  P  E  R  K  L  L  R  L  S  E  E  Q  Q  R  V  L  Q  H  S  T  H
 811 TGAGGGAGAAACTGGAGTCTGAAGCGCCGGAGAGGAAACTGCTGCGTCTGTCAGAGGAACAGCAGAGAGTCCTGCAGCACAGCACACACA

 300 T  N  N  N  T  I  N  T  A  T  A  R  E  E  R  Q  D  L  S  L  S  I  S  S  G  G  S  A  S  I
 901 CCAACAACAACACCATCAACACTGCTACAGCCAGAGAAGAGAGGCAGGATCTGTCCCTCAGCATTTCCTCCGGTGGATCGGCCAGCATCA

 330 S  V  S  V  E  V  N  G  I  V  Y  S  G  N  L  F  A  Q  K  S  V  S  A  A  P  A  G  P  S  G
 991 GTGTGTCAGTGGAAGTCAACGGCATCGTTTACTCAGGGAATCTGTTCGCCCAGAAGTCTGTGTCTGCGGCTCCTGCGGGACCCAGTGGCC

 360 L  F  T  F  S  S  S  A  P  S  Q  S  P  A  S  C  S  S  K  G  R  S  S  V  E  A  S  T  S  G
1081 TGTTCACCTTCTCTAGCTCCGCCCCCAGCCAAAGCCCCGCCTCCTGCTCATCAAAGGGGCGGAGCTCAGTGGAGGCGTCCACCAGCGGCT

 390 S  P  *
1171 CCCCATAGCAGAGCACAGATCAAACTGAGCACAACCAGTACCTCTATCCTCCAGGCCACACCGGCTGAACACACACACACACACACACAC
1261 ACACACACACACACACACACACACACACACACACACACACACACACACACACAGCAGACCTCATTCATCAACCATCAGAGGATCACCAAAAATA
1351 CCCGCATACAATCAGATGGAGCTAAGAACTCAATACTAATGCATGCAGCAAACACCATGCTCATGTGTGTAGACAAATTCAATGCATGCA
1441 AACAAATTCACATGCACACAATATTAATATATATTCACAAAAACAATTCATGTTCACCGACAGCAACG
```

Figure 3.4 — *Danio rerio dri2* cDNA sequence. The eARID and REKLES domain are shown in bold roman and bold italic type respectively. The presence of three in-frame stop codons upstream of the translation initiation codon indicates that the full open reading frame of *dri2* is covered.

1 CGTTGCTGTCGGGGA

16 GGTGAGGTCAGGCGTTTCCCCTTTGGCTGGCCCCAACGTCGACTGTATTTTCCCCCACCGCAAACATCGAGAGGTTTATTATCGTTTTAT

106 CAAAGGGGATATTAAACGTTCGGGTGGATGCACACGTACTCGGTGCGCTGAGCAAAACACGCACGGTGGTTTATCTAGGAAAAGAGAGTG

1 M V D N S S S S K A Q V P S L S Q D S S L A S A F S Q A A F

196 ATGGTGGACAATTCCAGTAGCAGCAAAGCACAAGTGCCCAGCTTGTCTCAGGACAGCAGTCTGGCCAGTGCGTTTTCCCAGGCAGCTTTC

31 A S S A G V K L E A V M E Q L Q R Q Q Q A K L E M E R K E R

286 GCAAGTTCGGCTGGTGTCAAACTGGAGGCAGTTATGGAGCAGCTACAGAGACAACAACAAGCTAAACTGGAGATGGAGCGCAAGGAAAGA

61 H L R E A H I L Y A Q Q L A A Q Q A I M A S A R A Q G A P L

376 CACCTCAGGGAAGCCCACATTCTGTACGCCCAGCAGCTGGCTGCACAGCAAGCCATCATGGCCTCCGCCAGGGCCCAGGGTGCCCCTTTA

91 T P D F Y S K N S R D R A L K G G Q G P L A A R G P V K T A

466 ACCCCTGATTTTTACAGCAAAAACTCCAGAGACAGAGCGCTGAAAGGTGGCCAGGGGCCTCTAGCTGCCAGGGGGCCCGTGAAAACCGCC

121 S D L A H E E D A A D E M D R G R G S E G D E D D D D E M M

556 TCCGACCTAGCACACGAGGAGGACGCTGCTGATGAGATGGACAGAGGAAGAGGCTCTGAGGGGGATGAAGATGATGACGACGAGATGATG

151 D G E D G S E E E E S E G L E F L R K Q S L A L Q Q A A V G

646 GATGGAGAAGATGGGAGTGAAGAAGAGGAAAGCGAAGGCTTGGAGTTCCTCAGGAAGCAAAGTCTCGCGCTTCAGCAAGCTGCTGTCGGA

181 V P P Y S F P V Y T A S P S A A K K R P L S P T T K V K D E

736 GTCCCTCCATATTCGTTCCCCGTCTACACCGCCTCACCCTCTGCTGCTAAAAAACGTCCCTTGTCGCCGACCACTAAAGTGAAAGATGAA

211 P E D S L S D Q Q S F N T P N G M G D **W N L D D S F K Q N G**

826 CCCGAAGACTCCCTCTCTGACCAGCAATCATTCAACACACCAAATGGAATGGGAGATTGGAACCTAGATGACTCATTCAAACAGAATGGA

241 N S S **W N E E G D G G R S R G E A S R D F A K L Y E L D N D**

916 AATTCAAGTTGGAACGAAGAAGGCGATGGTGGGAGAAGCCGAGGTGAGGCTTCCAGAGACTTTGCCAAGCTTTACGAACTGGACAATGAC

271 **P K R K E F L D D L F A Y M Q K R G T P V N R I P I M A K Q**

1006 CCGAAACGAAAGGAGTTCCTCGATGACCTCTTCGCTTACATGCAGAAACGAGGTACACCTGTAAATCGAATCCCCATCATGGCCAAGCAG

301 **V L D L Y M L Y K L V T E K G G L V E V I N K K I W R E I T**

1096 GTGTTGGATCTGTACATGCTGTATAAACTGGTGACTGAGAAGGGAGGCCTGGTAGAGGTGATCAACAAGAAGATCTGGAGGGAGATCACC

331 **K G L N L P T S I T S A A F T L R T Q Y M K Y L Y P Y E C E**

1186 AAAGGCCTGAACCTGCCCACCTCCATCACCAGTGCCGCCTTCACCCTCCGCACACAGTATATGAAGTACCTGTACCCGTACGAATGTGAG

361 **K K G L S S P G E L Q A A I D S N R R E G R** R P G Y S N S L

1276 AAGAAGGGTCTGAGTTCCCCTGGAGAGCTGCAGGCGGCCATCGATAGTAACCGCAGAGAGGGCCGCCGCCCCGGATACAGCAACAGCCTG

391 **Y R F S P** S P S G A A P H L L S P P K M H L P A T G H N E L

1366 TACCGCTTCTCCCCTTCACCCAGCGGAGCAGCCCCTCATCTCCTCTCTCCTCCAAAAATGCACCTTCCCGCCACAGGACACAACGAGCTG

421 Q A T P S P A L K K A T V P E D S P A S M L P A R L P L A L

1456 CAGGCCACCCCCAGTCCTGCCCTGAAGAAAGCCACAGTCCCGGAGGACAGTCCAGCCTCCATGCTGCCAGCTCGTCTCCCCTTAGCCCTG

451 A L G H Q Q Q L A R A A T ***L E Q L R E R L E T G M G A G A V***

1546 GCTCTGGGGCATCAGCAGCAGCTGGCACGTGCCGCCACCCTGGAGCAGCTGAGGGAGAGGCTGGAGACTGGCATGGGTGCCGGCGCTGTG

481 ***N A G P A V V E G P E R K M A R L A E E Q Q R L L Q Q A F Q***

1636 AACGCGGGGCCCGCTGTGGTCGAAGGGCCTGAGAGGAAAATGGCGCGTCTGGCAGAGGAGCAGCAGAGGCTACTGCAGCAGGCCTTCCAA

511 ***Q N L L A M A S Q V N P T N L R M N N A N T R E E K Q D L S***

1726 CAAAACCTGCTGGCCATGGCATCACAGGTGAACCCCACCAACCTGAGGATGAACAACGCCAACACCCGCGAGGAGAAGCAGGACTTGTCT

541 ***L S I S S N G S A S I T V S V E V N G I I Y S G S L Y A Q*** K

1816 CTCAGCATTTCTTCTAATGGCTCTGCCAGCATTACTGTGTCTGTGGAGGTCAACGGGATTATTTACTCAGGCTCCTTGTATGCTCAGAAG

571 S G S V P T A V T A M F P G H T S T L S P G V G S S S S S S

1906 TCTGGATCCGTTCCCACCGCAGTCACCGCCATGTTCCCAGGCCACACTTCCACCCTCAGTCCCGGCGTGGGCTCTTCATCCTCCTCTTCC

601 S S K G P D S V D P A T G G S P *

1996 TCCTCAAAGGGTCCTGACTCTGTGGATCCAGCCACAGGCGGGTCACCCTAACCAGCACCAACACCTCCCAAAACCTGTTACCAAAAACAA

2086 CCCTGCCAGTCCAGCCCAAAACTACAGTCATTTCCAGTGGAGCGCATGAGCCGTTCTGCCCTCTAGAGGTGGTTTGATGAATCACAATCT

2176 GCCTCCAGCACAAGCGCGCAGCTGTGTTTTCCTTCCAGCCAAAAGGAGGGACCACACACACAGACACTTTTACAAACTCCTCAGCTGTTC

2266 AAGCTCGTCACGGACCCCACCGGCTCTTTTCTACATGATGTTTTTTCTCAGGTTTTTAGAGATGTTCGATTCAACAAAATACCTCAGCTG

2356 TCTTAATCTCAGTGCTGTCGTGAGTGTGGGTTTAATCATGCCTTTAATCACGCACCCATAGTCATTCACTCACTAAAGATCTGTGAAGCT

2446 AACCGTGGCATGATCAAATCTACTTCTGCCAAAGTCACATTTTTGTGAGGAGGTTGTGGTTGACAGTCTTTCTTTCTTTCTTTCTCTCTC

2536 TCTCTCTCTCTGGCTTTTGCTGTCTATTAATTCTCTTATTTAAAGTTTGCTGAATTTACCTCATCTGCAGTGAATTTAAAAGACTCAGGA

2626 TCGATTATGATACTGCTGTGTTATTGTTTTGTTTTGTTTTTGGTTTAAAAAACTTCAGGCTTCAATCCCCGTCAAGTGTTGGAAAGGTGG

2716 TTTGGTTAGAAAAGGGGTCAGATTATAATTTTTTTTTTTTTTTATTATTATTGTATTTACTTGATTTTATTTGTTTTCTATTTTCCATTT

2806 GTTTTTTAGCTACTAGATACAGAAATATTAATAAGATTTGAAAGAGTTTTTGAGTTTGAAATTGCATTTTACACAGCTGAGAGTTTCTTT

2896 TTGTTTGTTTATTATTATTATTTAGACATATTTACAGGCAACAGTTTGGTTGAAAGATTTTTCTCTACCTCAGACACAAGACAATGTATG

2986 GTTGAAAGGAAATTATTGTGGACGCGAAGGGTTCGTGTCGCACGTCAATCATTTCTGTGCGTCGGACGATCATGTACAGGCTCGTGATGA

3076 AATGAGAGCCGTAGTAATGTACCATTCTTTAAAACAAACCAGGTCACGGAAAAAAGCAGGATAACATTCGAATCAAACGAAGATTGGATG

3166 GCTTTTTTTTCATTTTGTATTTTTAGTTGTTGAACTTCAGACCTCACATATCTTTCCATATTGGAAAGCGAGTTTCTCAGGAATTTTTTT

3256 TCTTCATAACTTTTAATCTTCAGAACCTAGATGCAGTCTCGCTCCTGAAACGAATGACACTAGACGTAACTCGGTCGCCGTGGGACTGAA

3346 ATAGCTACCTCACTTAGGATTTACCTCGGATGTAGTTCAACTACTCGCAGTACAACACAGAATTACCTCAGACGTGTACTGCTGGCTGTC

3436 TTTTCAAGCGAAAGAAATCACTCAGGTAAAGAATTAGCAAAATACCTCAGAAAAGGCAGAATAGCGTTAATACAGCAGTGCGAGATGAGA

3526 TTGTCGTATGAAGAGTTAGATCTCATATCCATGATGTATTGATTAGTGCAGAAACTGTCTTCCAGTAGCAATGGAGGCTCGTTGGCTTGG

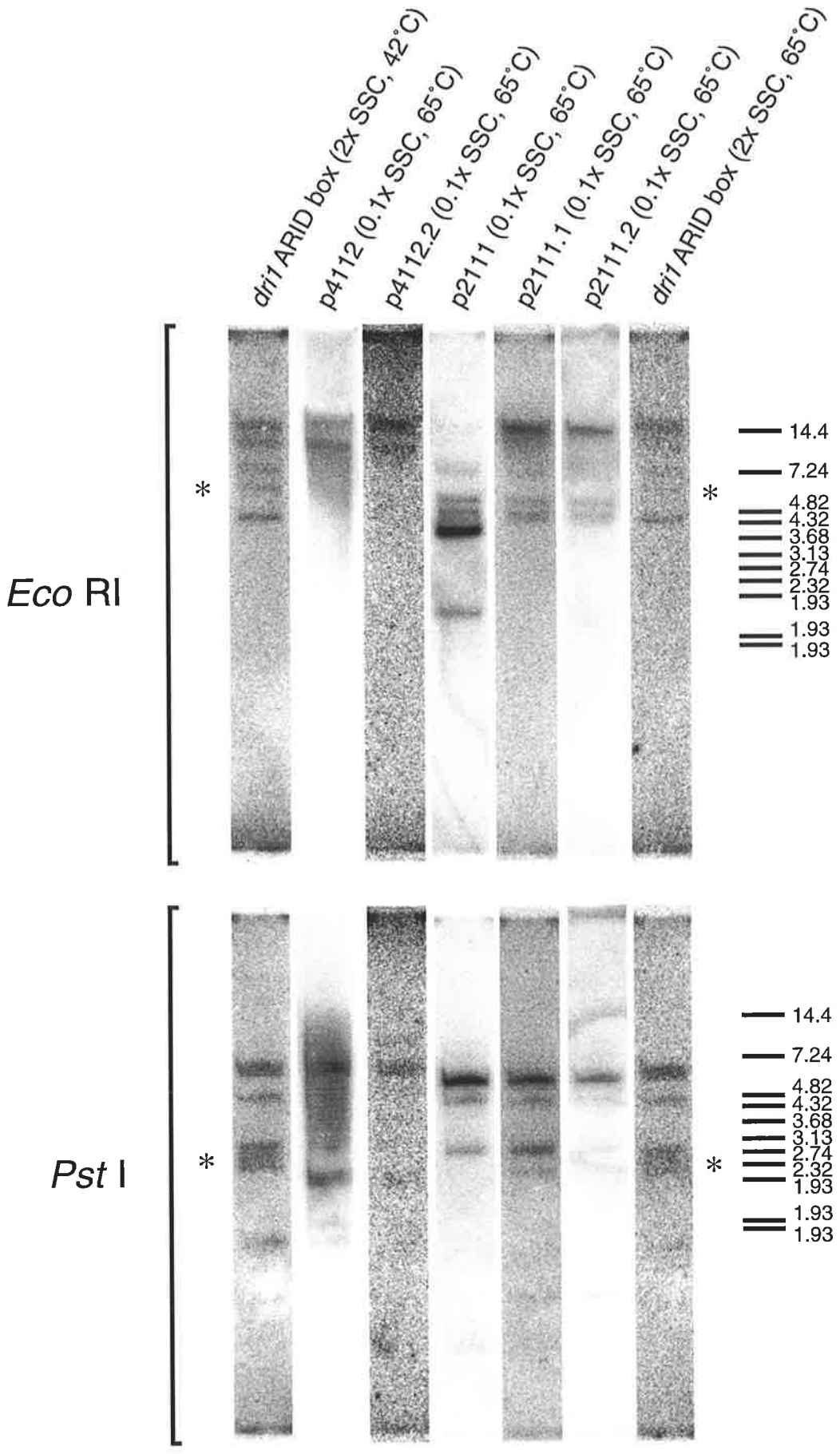
3616 TTATGTTGGTGTAGTCTTGCCGACGTAGCCTCATCAACTGTACATTTTAGCCGTTTATAATCCTGCGTTTTGCTTGACCCTGGACCTGC

3706 TGGGATTCAGACATTTATTAGTTTCTTCGGTTTTACGTTCTATATTCTTCAGATTTCTGCATAGCTAAGAACCAAAGACCATGTGTTTGT

3796 TATCCTCTCTGATGACCTTACTGTATAGACACATGCAACTTTACACAACGAAATTAAAGTCTGATCTGATGCAAAAAAAACCGTTGCTGT

3886 CGCGACAGCAACGGCGACAGCAA

Figure 3.5 — Southern analysis of *Danio rerio dri* paralogues. This figure shows a comparison of low and high stringency hybridisation of a Southern blot of *Eco* RI and *Bam* HI digested *D. rerio* genomic DNA (genomic DNA, gift of M. Lardelli). The filter was probed with the *dri1* degenerate PCR product and washed at low stringency (first and last lanes). High stringency washes were used with probes for *dri1* (p4112 and p4112.2) and *dri2* (p2111, p2111.1 and p2111.2). Bands present at low stringency, but absent at high stringency are marked with an asterisk. Molecular weights are shown to the right in kilobases. Note that the same filter was used to generate autoradiographs at low and high stringencies.


dri1 ARID box (2x SSC, 42°C)
p4112 (0.1x SSC, 65°C)
p4112.2 (0.1x SSC, 65°C)
p2111 (0.1x SSC, 65°C)
p2111.1 (0.1x SSC, 65°C)
p2111.2 (0.1x SSC, 65°C)
dri1 ARID box (2x SSC, 65°C)
Eco RI
Pst I
*
*
14.4
7.24
4.82
4.32
3.68
3.13
2.74
2.32
1.93
1.93
1.93

disorder characterised by mucocutaneous melanin pigmentation and hamartomatous polyps in the gastrointestinal tract (Foley *et al.* 1988). These features could result from developmental irregularities due to a mutation in a developmental regulatory gene. One of the tissues showing the highest level of *DRIL1* message, determined by dot blot analysis, is the colon (Kortschak *et al*. 1998), a tissue significantly affected in Peutz-Jeghers Syndrome sufferers. Furthermore, *Drosophila dri* is expressed in the hindgut (Gregory *et al*. 1996) and *dri* mutants exhibit a disrupted hindgut morphology (Shandala *et al.*, in preparation). Thus *DRIL1* was initially considered a candidate gene for this genetic disorder. However, recently the gene for Peutz-Jeghers Syndrome has been cloned and shown to encode a serine/threonine protein kinase (Hemminki *et al.* 1998; Jenne *et al.* 1998). No other disorders mapping to this region appear to be likely to correspond to *DRIL1*.

3.4 *Danio rerio dead ringer* homologues are expressed in distinct patterns

To examine potential conservation of function of the *D. rerio dead ringer* orthologues with respect to *Drosophila dead ringer* and possible divergence in function of the *D. rerio* paralogues in the fish, I have surveyed the expression patterns of the two *D. rerio* genes.

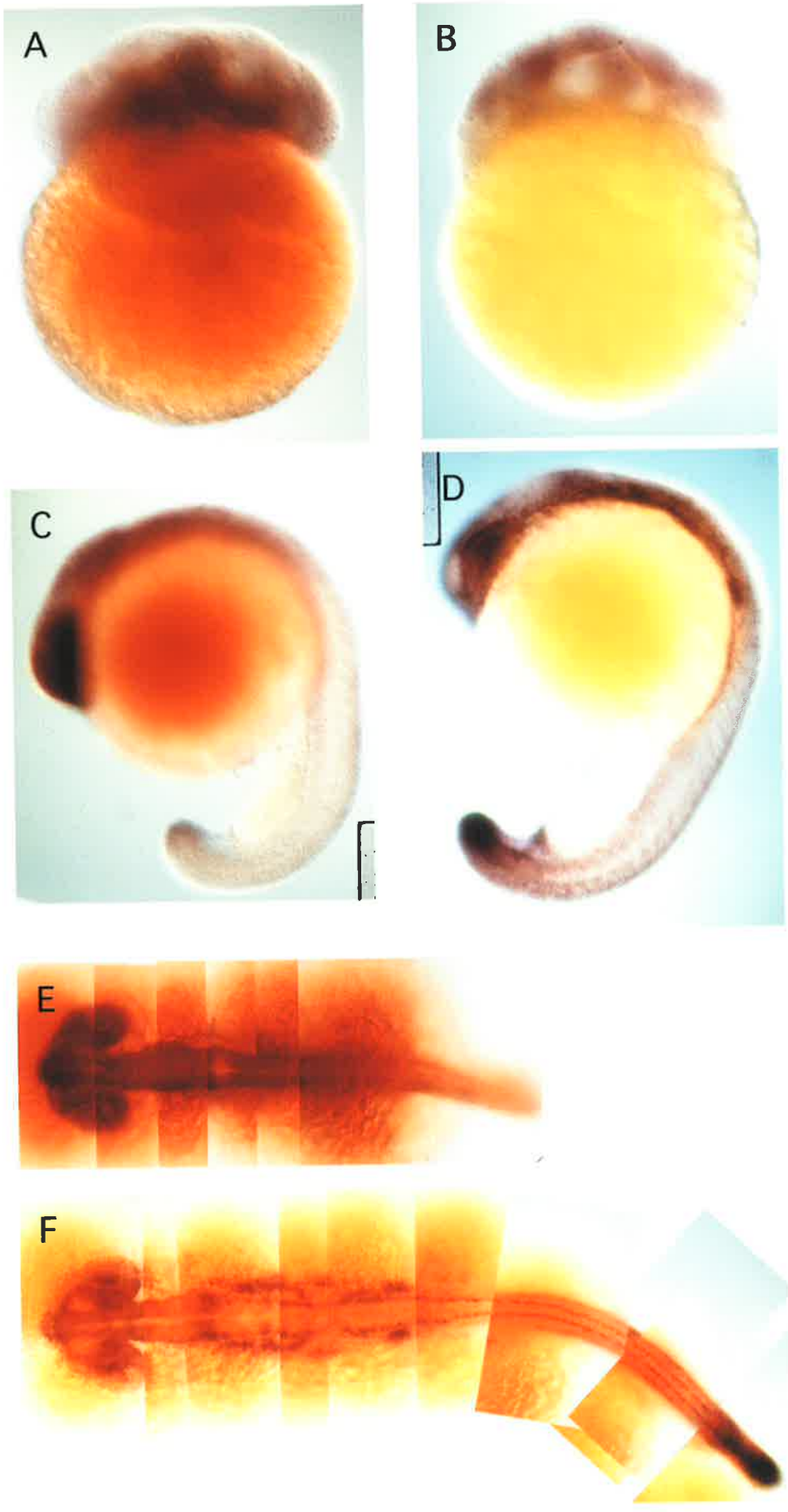
In situ hybridisation analysis using the p4112 and p2111 cDNAs performed in collaboration with M. Lardelli indicates that both *dri1* and *dri2* are maternally deposited in the embryo, as cleavage period embryos show the presence of message (figure 3.6A, B). Consistent with the notion that *dri1* and *dri2* function in homologous tissues to those expressing *dri* in *Drosophila*, the *D. rerio* genes are expressed in mesodermal and neural tissues, tissues which express *dri* in the fly (Gregory *et al*. 1996) (figure 3.6). During tail bud extension *dri1* appears to be expressed in the distal mesoderm of the tail, but not the axial mesoderm, with broad expression in the fore-, mid- and hind-brain, while *dri2* shows expression in the optic stalk, the retinas, midbrain, the cloaca and clusters of neural crest cells. *dri1* was expressed at significantly lower levels than *dri2* as judged by rate of colour development; *dri1* staining was only apparent after 8 hours while *dri2* staining was clearly

Figure 3.6 — Whole mount *in situ* hybridisation of *Danio rerio* with *dri1* and *dri2*
A, C and E show *dri1* (p4112 cDNA) hybridisation. B, D and F show *dri2* (p2111 cDNA) hybridisation.

A and B show cleavage period embryos at 1.5 hours after egg laying (16-cell stage), with hybridisation in the dividing embryonic cells, indicating maternal deposition of *dri1* and *dri2*.

C and D show lateral views of tail bud extending embryos at 18 hours after egg laying.

E and F show dorsal views of the embryos shown in C and D. The axial series was generated by rolling the embryo and taking photographs along the axis. The photographic series was then composed using successive images.


A
B
C
D
E
F

visible within 20 minutes. Hybridisation with sense probes gave no signal in either case when staining was developed for the same length of time (not shown).

The expression patterns of the two *D. rerio* genes do not significantly aid in the analysis of *dri* expressing tissues in the fruit fly. However, it is interesting to note that the two genes are expressed in different patterns, suggesting that the function of the putatively single ancestral *dead ringer/bright* gene has undergone a process of duplication-degeneration-complementation described by Force *et al*. (1998) leading to neofunctionalisation, the gain of a new function, or subfunctionalisation, the separation of previously joined functions .

3.5 *Dead ringer*-like ARID sequence comparisons

Homo sapiens DRIL1 exhibits a high degree of similarity to mouse Bright across the entire open reading frame (79% identity, 83% similarity). The level of conservation increases to 97% identity, 99% similarity over the 141 amino acid DNA binding domain (table 3.1). Note that the extent of the eARID is increased from that described initially in Gregory *et al*. (1996), as the vertebrate members (this study and Herrscher *et al*. 1995) contain a further eight amino acids of similarity (figure 3.8) with some conservation seen in the predicted *C. elegans* protein, T23D8.8, identified in sequence database searches (Wilson *et al.* 1994; Sulston *et al.* 1992). The *C. elegans* protein sequence, T23D8.8, does not contain sequence amino-terminal of the conserved LYE motif in the eARID. Further, the database entry for the protein sequence indicates that the first residue of T23D8.8 is a methionine, contradicting the nucleotide sequence entry which specifies a leucine. Analysis of the nucleotide sequence of *T23D8.8* and comparison with other eARID-encoding genes suggest that this start corresponds to an intron splice site rather than the translation initiation site indicated. DRIL1 is also similar to *Drosophila* DRI across the DNA binding domain (74% identity, 84% similarity) (Gregory *et al*. 1996) (figure 3.8). In addition, the intron-exon structure of the region encoding the ARID is partially conserved between human (Kortschak *et al*. 1998), nematode (accession number: worm_chr_I:8697730-869911, Wilson *et al*. 1994; Sulston *et*

Figure 3.7 — Dendrogram of ARID containing proteins based on amino acid identity within the core ARID generated using the PAUP (Phylogenetic Analysis Using Parsimony 3.1.1, D. L. Swofford) program. Branch lengths indicate degree of sequence divergence. Only one orthologue in an orthologous group is shown for simplicity, except in the case of Dead ringer/Bright related proteins. Sources for the sequences are: Dead ringer (Gregory *et al*. 1996), Bright (Herrscher *et al*. 1995), DRIL1 (this study, Kortschak *et al.* 1998), Dr Dri1/Dr Dri2 (RDK this study), T23D8.8 (PID g1628238), Eyelid (Treisman *et al*. 1997), RBP1/RBP2 (Fattaey *et al*. 1993), SmcX (Agulnik *et al*. 1994b), SmcY (Agulnik *et al*. 1994a), MRF1 (PID g188684), MRF2 (PID g553592), Jumonji (Takeuchi *et al*. 1995), SWI1 (O'Hara *et al*. 1988) c08b11.3 (PID g576996), r20183 (PID g774817), c01g8.8 (PID g1703595), zk593.4 (PID g1184605), YM8010.06 (PID g854446).

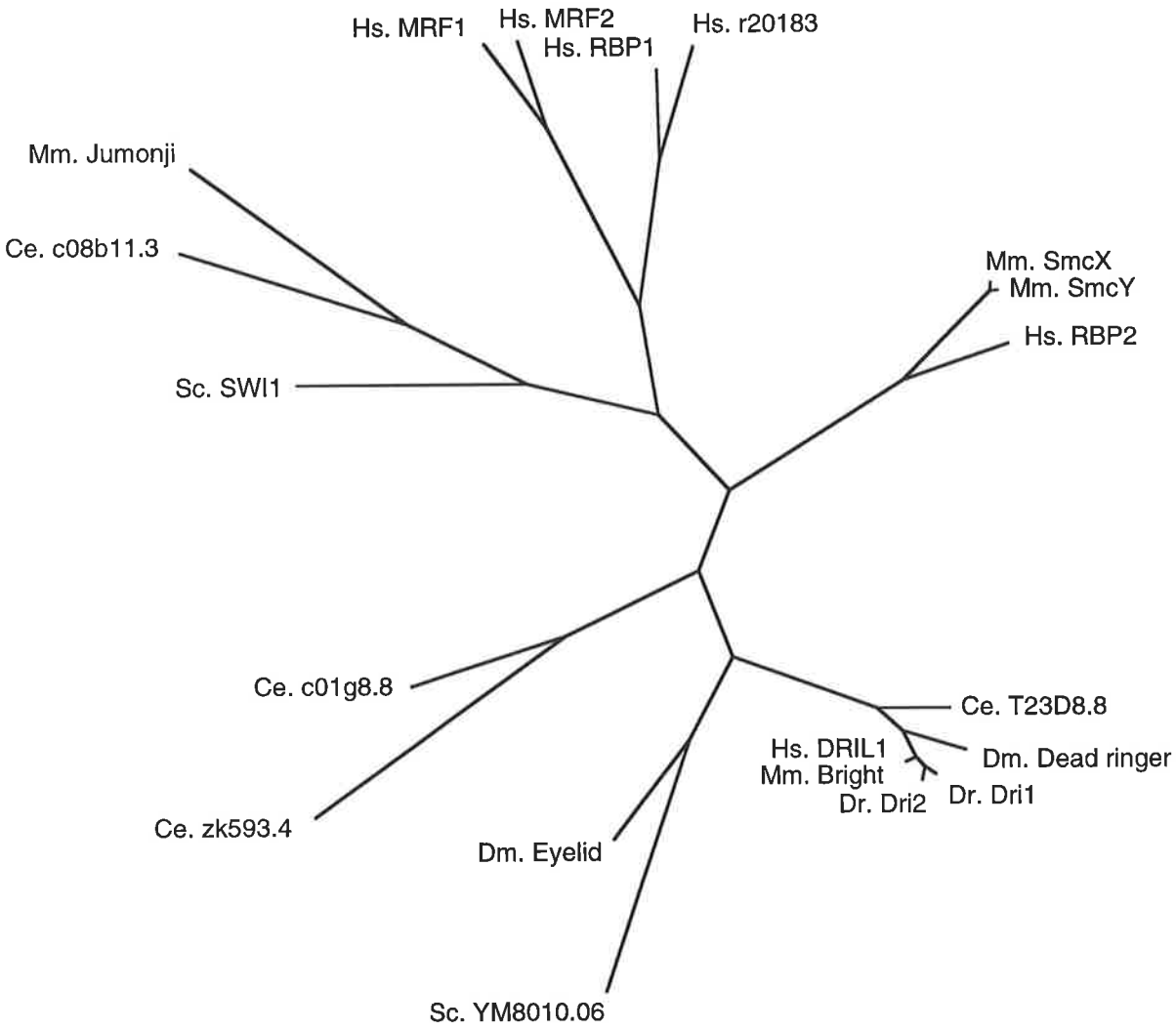
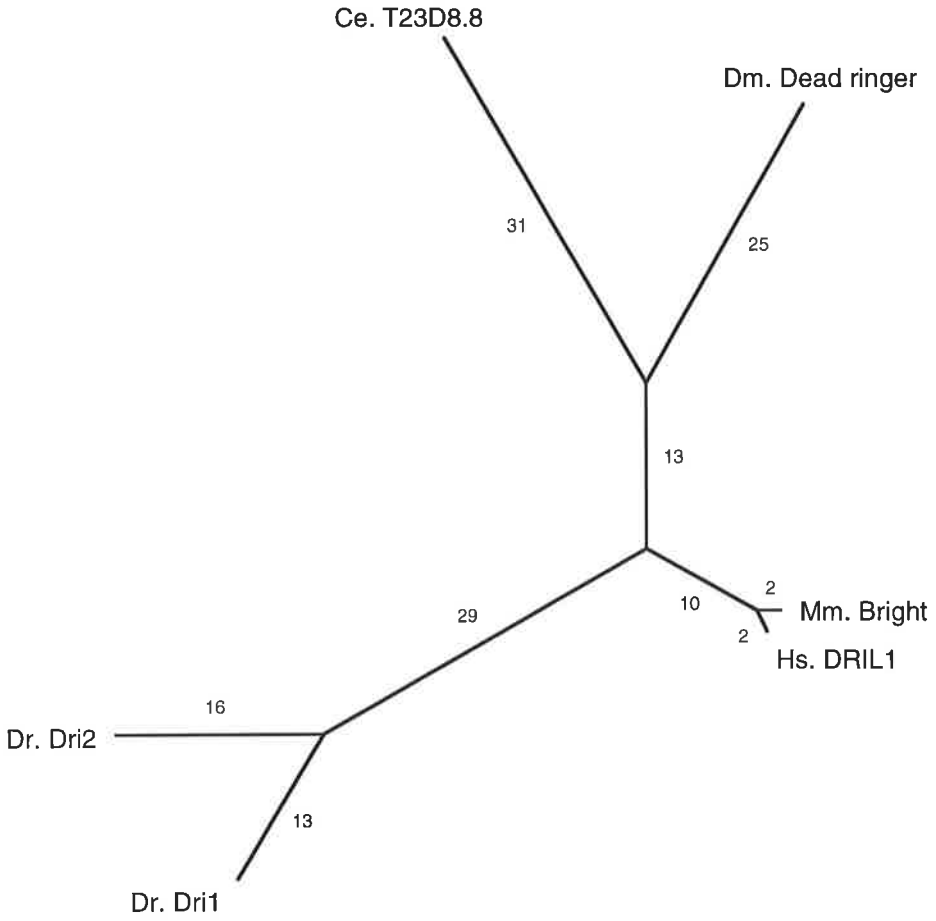

Hs. MRF1
Hs. MRF2
Hs. RBP1
Hs. r20183
Mm. Jumonji
Ce. c08b11.3
Sc. SWI1
Mm. SmcX
Mm. SmcY
Hs. RBP2
Ce. c01g8.8
Ce. zk593.4
Dm. Eyelid
Sc. YM8010.06
Ce. T23D8.8
Hs. DRIL1
Mm. Bright
Dr. Dri2
Dm. Dead ringer
Dr. Dri1

Figure 3.8 — A. Dendrogram of eARID-containing proteins based on eARID amino acid identity generated using the PAUP (Phylogenetic Analysis Using Parsimony 3.1.1, D. L. Swofford) program. Branch lengths indicate degree of sequence divergence.

B. Alignment of eARID proteins over the eARID generated using the MAP multiple sequence alignment program (Huang 1994) to generate the alignment and the MacBOX program v1.0.8 (a Macintosh implementation of BOXSHADE, M. D. Baron) to generate the shaded figure. Invariant and conserved residues are indicated by white text with a black or dark grey background respectively. Similar residues are indicated by black text with a light grey background. Residues are considered similar for this purpose if they fall into one of the following groups: FYW, ILVM, RK, DE, ST, QN or GA. Additionally, residues not in the groups above are considered similar if a residue in the group, ADEFIKLNPRSVY, is conserved in more than three sequences according to the following scheme: A similar to G; D similar to E or N; E similar to Q; F similar to Y or W; I similar to L,V or M; K similar to H; L similar to V or M; N similar to E or Q; P similar to G, R similar to K or H; S similar to T or A; V similar to I or M; and Y similar to W. The consensus was generated at each position from three or more similar or identical residues. Note that if a consensus is generated for a position on the basis of similar residues with no residue at or above the threshold of three, the consensus residue shown is the residue existing at that position in the first sequence listed.
Positions of introns, where known (marked with a black triangle after the name), are indicated by grey triangles. The position of the triangle within the box indicates the position of the intron within the corresponding codon.

Sources for the sequences are: Dead ringer (this study, Gregory *et al*. 1996), Bright (Herrscher *et al*. 1995), DRIL1 (this study, Kortschak *et al*. 1998), Dri1/Dri2 (this study), T23D8.8 (PID g1628238).

A



B

```
T23D8.8      1  ..................................LYELSDDV
DRI        263  WSFEEQFKQVRQ----------------------LYEINDDP
Dri-1       37  WSYEEPFKANGAVSWSDDTDAARSRGEASRDFAKLYELDSDP
Dri-2      132  WNLDDSFKQNGNSSWNEEGDGGRSRGEASRDFAKLYELDNDP
Bright     228  WTFEEQFKQ------------------------LYELDADP
DRIL1      223  WTYEEQFKQ------------------------LYELDGDP
consensus    1  wsfeeqfkq                        LYEld Dp

T23D8.8      9  KRKEWLDDWLNFMHRIGKPVTRIPIMAKQVLDLYELYRLVVQ
DRI        283  KRKEFLDDLFSFMQKRGTPINRLPIMAKSVLDLYELYNLVIA
Dri-1       79  QRKEFLDDLFTFMQKRGTPVNRIPIMAKQVLDLYRLYRLVTE
Dri-2      174  KRKEFLDDLFAYMQKRGTPVNRIPIMAKQVLDLYMLYKLVTE
Bright     245  KRKEFLDDLFSFMQKRGTPVNRIPIMAKQVLDLFMLYVLVTE
DRIL1      240  KRKEFLDDLFSFMQKRGTPVNRIPIMAKQVLDLFMLYVLVTE
consensus   43  kRKEfLDDlfsfMqkrGtPvnRiPIMAKqVLDLymLYrLVte

T23D8.8     51  HGGLVEIINKKLWREITKGLNLPSSITSAAFTLRTQYQKYLY
DRI        325  RGGLVDVINKKLWQEIIKGLHLPSSITSAAFTLRTQYMKYLY
Dri-1      121  KGGLVEVINKKIWREITKGLNLPTSITSAAFTLRTQYMKYLY
Dri-2      216  KGGLVEVINKKIWREITKGLNLPTSITSAAFTLRTQYMKYLY
Bright     287  KGGLVEVINKKLWREITKGLNLPTSITSAAFTLRTQYMKYLY
DRIL1      282  KGGLVEVINKKLWREITKGLNLPTSITSAAFTLRTQYMKYLY
consensus   85  kGGLVevINKKlWrEItKGLnLPtSITSAAFTLRTQYmKYLY

T23D8.8     93  DYECEKEKLSNQSDLQQAIDGNRREAPGRRTAPSFPLPFQLP
DRI        367  PYECEKKNLSTPAELQAAIDGNRRE--GRRSSYGQ-------
Dri-1      163  PYECERKALSSPNELQAAIDSNRRE--GRRPSYSNTLFHLSP
Dri-2      258  PYECEKKGLSSPGELQAAIDSNRRE--GRRPGYSNSLYRFSP
Bright     329  PYECERRGLSSPNELQAAIDSNRRE--GRRQSFGGSLFAYSP
DRIL1      324  PYECEKRGLSNPNELQAAIDSNRRE--GRRQSFGGSLFAYSP
consensus  127  pYECEkkgLSspneLQaAIDsNRRE  GRR sy qslf fsp
```

al. 1992) and *Drosophila*. The location of introns in this region are identical in *Drosophila* and human (Kortschak *et al.* 1998), while the sites of the first two introns within this conserved region are found in the predicted *C. elegans* gene, T23D8.8, the third being absent (figure 3.8). The *D. rerio* proteins also show a high degree of sequence conservation over the DNA-binding domain (table 3.1): 83% identity and 91% similarity between them, while Dri1 and Dri2 show 62% and 61% identity, respectively, compared to DRI, and 72% and 71% identity compared to Bright. The significant drop in sequence identity between Dri1/Dri2, and the other vertebrate eARID proteins is largely a result of the insertion of 25 amino acids near the amino terminus of the domain in both *D. rerio* proteins.

Interestingly, Dri2 has a phenylalanine to tyrosine amino acid change at an invariant position in the other members of the group and a serine/threonine to alanine change immediately amino-terminal of this residue (FLDDLFAYMQKRGT, figure 3.8). Both of these residues lie within the region corresponding to the target of the ARID 5' inside primer, probably preventing primer binding. To test whether these changes were responsible for failure to isolate a *dri2* cDNA by PCR, PCRs were performed using the ARID 5' inside primer in conjunction with other compatible primers and *dri2* cDNAs as a template. No PCR product was obtained from these reactions (data not shown). Dri2 also shows changes in the conserved amino-terminal region of the eARID, which may reflect differences in site-specificity of the protein for its DNA-target (figure 3.8).

Table 3.1 eARID identities and similarities
(upper value is identity, lower value is similarity)

	DRI	Dri1	Dri2	Bright	DRIL1
T23D8.8	59% 73%	54% 66%	54% 64%	63% 74%	65% 74%
DRI		62% 74%	61% 73%	74% 85%	74% 84%
Dri1			83% 91%	72% 78%	72% 78%
Dri2				71% 78%	71% 78%
Bright					97% 99%

As described in chapter 1, sequence analysis of the ARID box family of genes shows the existence of a subfamily of proteins, the extended ARID (eARID) proteins, which have an extended region of similarity. Initially the eARID group was defined by the DRI and Bright proteins. Both DRI and Bright have been shown to exhibit sequence-specific DNA-binding activity (Herrscher *et al*. 1995; Gregory *et al*. 1996). A further four proteins have been identified as members of this group by database searches (T23D8.8) or the analysis of genes isolated by degenerate PCR (DRIL1, Dri1 and Dri2). While the highest similarity between Bright/DRIL1 and DRI is restricted to the DNA binding domain, the similarity is nonetheless striking and of the same order as that observed between the conserved DNA binding domains of homologues of, for example, the homeodomain class of proteins. This relationship is illustrated by a dendrogram generated from the available ARID family protein sequences (figure 3.7). It is clear that the DRI/Bright/DRIL1 sequences group together and with the *C. elegans* T23D8.8, and *D. rerio* Dri1 and Dri2 sequences. No other sequences, including the other characterised *Drosophila* ARID protein, Eyelid, exhibits the same high level of similarity to DRI. Other DNA binding proteins that are considered homologues often exhibit a high level of conservation between DNA binding domains but little or no similarity outside that domain. From these data I conclude that *dri*, *bright*, *DRIL1*, *dri1* and *dri2* form an orthologous group of genes.

3.6 A new protein motif specifically associated with the eARID

Maintenance of sequence similarity in divergent sequences has regularly been used for the identification of functionally important regions in protein or DNA sequences. The use of evolution as an experimental tool for identifying important regions depends on a balance between a high rate of change causing non-critical regions to diverge and strong selection retaining sufficient conservation of the functional region. This balance is easily achieved when a sequence is evolving around an interaction with an essentially invariant structure,

such as DNA, as the system is unable to undergo coevolution of the interacting components. For example histone proteins are conserved in an easily recognisable form in all eukaryotes. The situation is different, however, when the system involves the interaction of two or more less constrained evolving structures, such as members of a protein complex. Thus, a broad range of sequences with different degrees of divergence allow the effective screening for the presence of functionally significant motifs by providing a successive evolutionary sieve for conserved regions. Lower divergence times allow the identification of rapidly evolving motifs while longer periods of divergence highlight the most crucial residues in conserved domains. The identification of sequences orthologous to *dead ringer* with divergence times ranging from approximately 100 million years ago, between humans and mice, to 670 million years ago, between vertebrates and arthropods, would thus allow any domain outside the eARID to be found irrespective of its rate of evolution.

The initial characterisation of the DRI and Bright proteins showed the existence of a single conserved domain, the eARID (Herrscher *et al*. 1995; Gregory *et al*. 1996). However, analysis of the sequences of the extended set of eARID proteins revealed the existence of a second conserved motif immediately carboxy-terminal to the eARID (figure 3.9). This motif has been named the REKLES motif after the near consensus sequence at the N-terminus of the motif in the vertebrate proteins (and for the mistaken notion held by my colleagues of my manner of living).

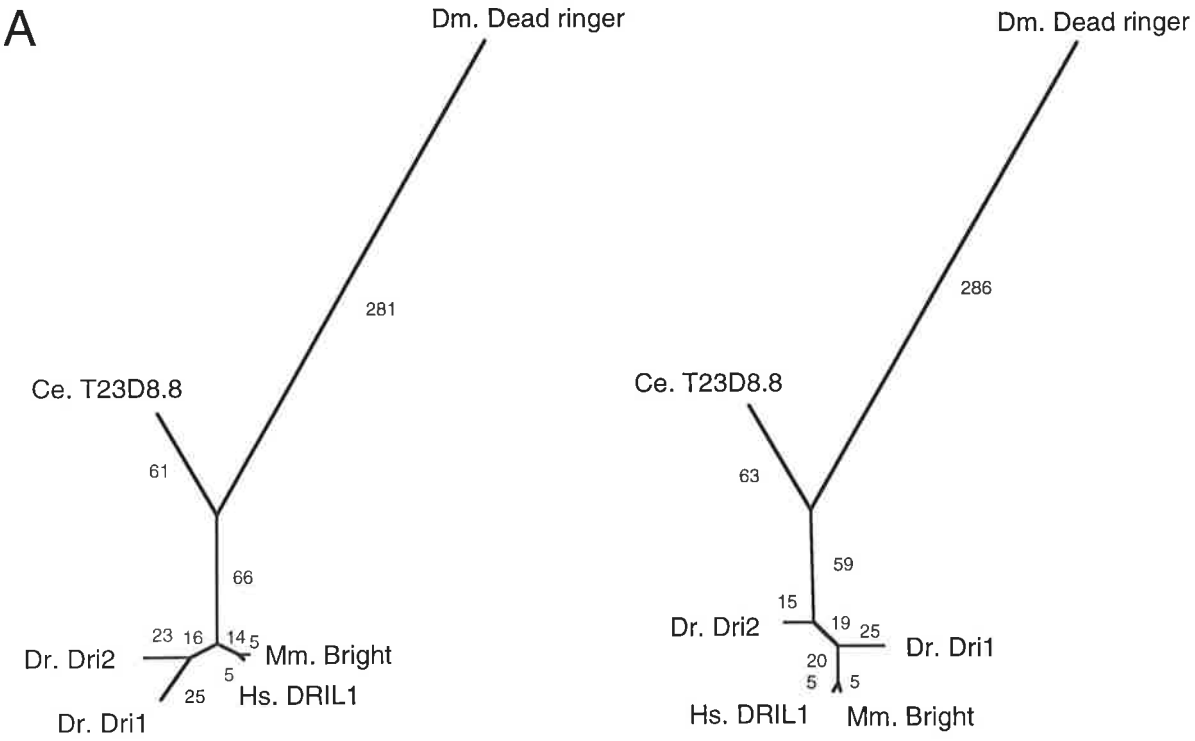
The size of the REKLES domain varies within the eARID group of proteins, ranging from 113 amino acids in the predicted *C. elegans* protein T23D8.8 to 345 amino acids in DRI. The large size of the DRI REKLES is due to an insertion of 237 amino acids relative to the vertebrate domains. This insertion causes the DRI REKLES to appear as a significantly diverged member of the group (figure 3.9A). Discounting the insertion, DRI shows reasonably high levels of conservation with the vertebrate proteins over the entire REKLES domain (table 3.2 and figure 3.9B). The conservation within the vertebrate subgroup is much higher, ranging from 89% identity between Bright and DRIL1 to 47% identity between Dri1 and DRIL1. The last 24 amino acids in the REKLES domain are particularly highly

Figure 3.9 — A. Dendrogram of eARID-containing proteins based on REKLES domain amino acid identity generated using the PAUP (Phylogenetic Analysis Using Parsimony 3.1.1, D. L. Swofford) program. Note that two most parsimonious trees were found. The tree on the left is favoured as it conforms to the structure of the ARID family tree shown in figure3.7 and the phylogram shown in figure3.11. Branch lengths indicate degree of sequence divergence.

B. Alignment of eARID proteins over the REKLES domain generated using the MAP multiple sequence alignment program (Huang 1994) to generate the alignment and the MacBOX program v1.0.8 (a Macintosh implementation of BOXSHADE, M. D. Baron) to generate the shaded figure. Invariant and conserved residues are indicated by white text with a black or dark grey background respectively. Similar residues are indicated by black text with a light grey background. Residues are considered similar for this purpose if they fall into one of the following groups: FYW, ILVM, RK, DE, ST, QN or GA. Additionally, residues not in the groups above are considered similar if a residue in the group, ADEFIKLNPRSVY, is conserved in more than three sequences according to the following scheme: A similar to G; D similar to E or N; E similar to Q; F similar to Y or W; I similar to L,V or M; K similar to H; L similar to V or M; N similar to E or Q; P similar to G, R similar to K or H; S similar to T or A; V similar to I or M; and Y similar to W. The consensus was generated at each position from three or more similar or identical residues. Note that if a consensus is generated for a position on the basis of similar residues with no residue at or above the threshold of three, the consensus residue shown is the residue existing at that position in the first sequence listed.

Sources for the sequences are: Dead ringer (Gregory *et al.*, 1996), Bright (Herrscher *et al*. 1995), DRIL1 (this study, Kortschak *et al*. 1998), Dri1/Dri2 (this study), T23D8.8 (PID g1628238).

A



B

```
T23D8.8    153  ---MRNDLLDDENTLSLQASGLFGTSYGAEQMAILEAHQRN
DRI        463  MPNLVKQEIESRMMEYLQLIQAKKEQGMPPVLGGNHPHQ--
Dri-1      267  LEQL-REKLES----------------EAPERKLLRLSEEQ--
Dri-2      366  LEQL-RERLETGMGAGAVNAGPAVVEGPERKMARLAEEQ--
Bright     451  LEQL-REKLES----------------TEPPEKKMALVADEQ--
DRIL1      446  LEQL-REKLES----------------AEPPEKKMALVADEQ--
consensus    1  leql rekles                  egperkmailaeeQ

T23D8.8    191  LERAQRAVQQQVARQSLGLTACSNGNGGNIHNSGRESTSSN
DRI        501  ----QQHSQQQQQQHHHQQQQQQSQQQHHL-QQQRQRSQSP
Dri-1      290  ----QRVLQHSTHTN-------------NNTI-NTATAREERQ
Dri-2      403  ----QRLLQQAFQQNLLAMASQVNPTNLRM-NNANTREEKQ
Bright     475  ----QRLMQRAVQQSFLAMTAQL-PMNIRI-NS-QASESRQ
DRIL1      470  ----QRLMQRALQQNFLAMAAQL-PMSIRI-NS-QASESRQ
consensus   42       Qrlvqqavqqn lam aqv p nlri nsgna esrq

T23D8.8    232  DSDI--{22}-TKHSDNSKTSMSVSMEINGITYQQGVLFAL
DRI        538  DLSK-{237}-NGTGDNGEPQLIVNLELNGVKYSGVLVAN
Dri-1      316  DLSL-------SISS-GGSASISVSVEVNGIVYSGNLFAQ
Dri-2      440  DLSL-------SISS-NGSASITVSVEVNGIIYSGSLYAQ
Bright     510  DSAV-------SLTSANGSNSISMSVEMNGIVYTGVLFAQ
DRIL1      505  DSAV-------NLTGTNGSNSISMSVEINGIMYTGVLFAQ
consensus   83  D si         sits ngs sisvsvEiNGivYsGvLfAq
```

conserved, with 7 invariant residues (DRI/Dri1, 42% identity, 67% similarity; DRI/Dri2, 50% identity, 75% similarity; DRI/Bright, 50% identity, 83% similarity; DRI/DRIL1, 54% identity, 83% similarity), suggesting that this portion of the domain plays a critical role in the function of the REKLES domain.

Table 3.2 REKLES domain identities and similarities
(upper value is identity, lower value is similarity)

	DRI	Dri1	Dri2	Bright	DRIL1
T23D8.8	19% 35%	20% 31%	23% 39%	29% 45%	26% 41%
DRI		19% 25%	24% 41%	19% 37%	21% 39%
Dri1			53% 72%	50% 70%	47% 69%
Dri2				49% 72%	48% 73%
Bright					89% 95%

As was the case in comparing the *D. rerio* eARIDs, the REKLES domains of *D. rerio* show divergence suggesting possible divergence in specificity of function. Dri1 has a deletion spanning a region conserved in the vertebrate members of the group (identified by the consensus `fLALaaGl-Pm`, invariant residues shown in uppercase, conserved residues in lowercase, - indicating no consensus, figure 3.9), while Dri2 shows changes from the REKLES consensus (`RERLET` in Dri2) and an insertion immediately carboxy-terminal of the REKLES signature.

Comprehensive database searches have failed to identify any non-eARID proteins possessing the REKLES motif. As the REKLES domain is present in all of the eARID proteins but not in any other known protein, the eARID/REKLES motif complex probably represents an evolutionary cassette, coupling the function of sequence-specific DNA-binding with REKLES motif function. The consensus sequence of the REKLES motif does not suggest a function for the domain. However, a region of the Bright protein which includes the

complete REKLES domain has been shown to be required for tetramerisation (Herrscher *et al.* 1995). Thus the REKLES motif is most likely a protein-protein interaction domain mediating tetramerisation, although other functions cannot be ruled out.

3.7 Phylogenetic analysis of the eARID genes

The identification of two *D. rerio* members of the eARID protein group raises the question of the existence of paralogues of the *dril1* and *bright* genes. To address this question I have performed phylogenetic analysis of the eARID-encoding genes using the DNA specifying the conserved protein domains. Maximum parsimony analysis using the eARID or REKLES boxes alone or in conjunction yields a consensus tree in which the two *D. rerio* genes group together to the exclusion of the other members of the group (figure 3.10) suggesting that the *D. rerio* genes separated after the divergence of the teleost line from the other vertebrate lines approximately 400 million years ago (figure 3.11). Southern analysis of the mouse genome has failed to detect a sequence homologous to *bright* (Phil Tucker, pers. comm.). Additionally, the failure of mouse and EST database searches to identify eARID-encoding sequences other than those already known is consistent with a model of eARID gene evolution in which a single eARID-encoding gene was present in the original vertebrate line, with a duplication event occurring in the teleost line generating two *dead ringer* orthologues in fish.

The suggestion of another *D. rerio* gene by the Southern analysis described above potentially complicates this interpretation. It is possible that the signal obtained by reduced stringency hybridisation is due to hybridisation with a non-eARID, ARID encoding gene. Without the sequence of this putative third ARID encoding gene it it not possible to determine its relationship to the rest of the ARID group of genes.

Figure 3.10 — Phylogram of eARID-encoding genes based on sequence identity generated using the PAUP (Phylogenetic Analysis Using Parsimony 3.1.1, D. L. Swofford) program. Branch lengths indicate degree of sequence divergence, percentiles adjacent branch nodes indicate bootstrap values of confidence from 100 pseudo-replicates.

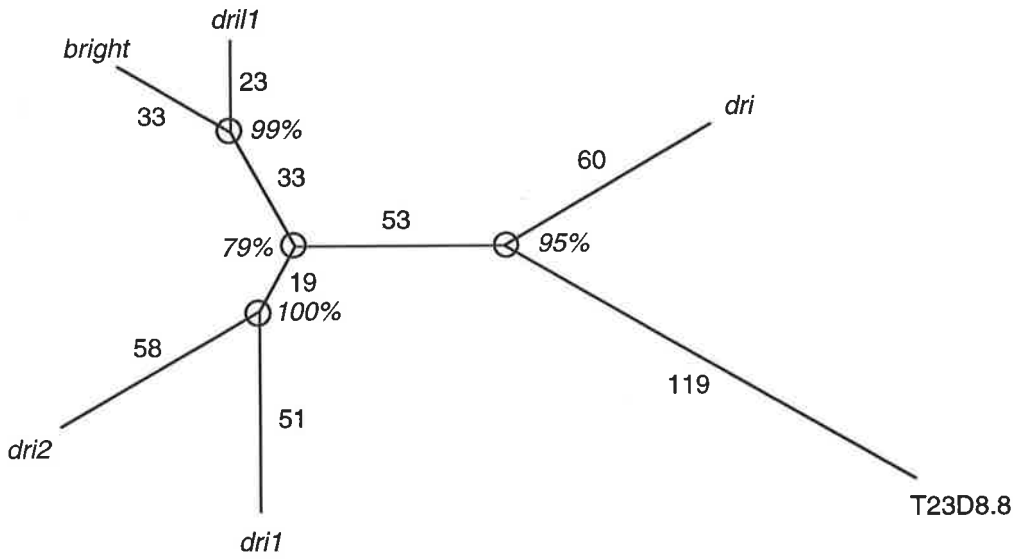
A. Phylogram of eARID encoding DNA

B. Phylogram of REKLES domain-encoding DNA.

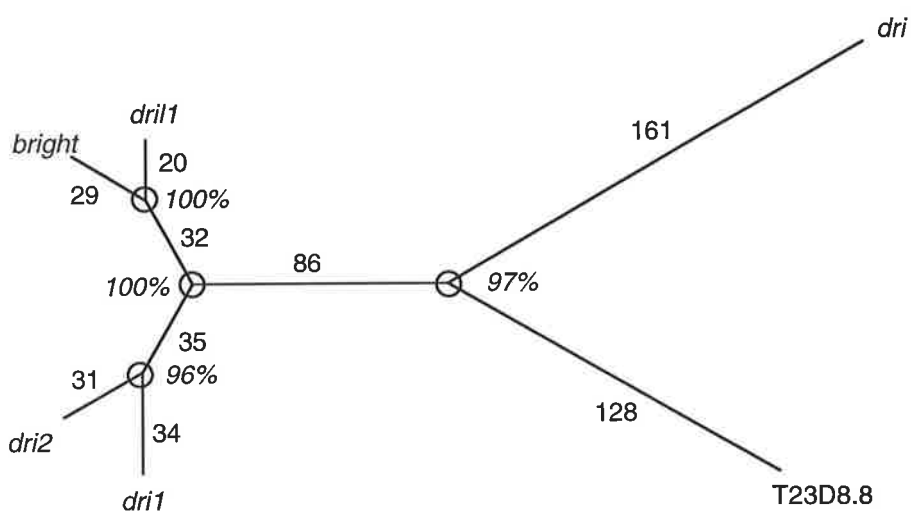
C. Phylogram of eARID and REKLES domain-encoding DNA together.

Sources for the sequences are: *dead ringer* (Gregory *et al*. 1996), *bright* (Herrscher *et al*. 1995), *dril1* (this study, Kortschak *et al*. 1998), *dri1/dri2* (this study), T23D8.8 (worm_chr_I: 8697730-869911). The DNA sequences corresponding to the proteins shown in the alignment in B. was used for phylogenetic analysis. Gaps were introduced into the DNA sequences in positions corresponding to gaps in the protein sequences shown in Figures 3.8 and 3.9.

A


dril1
bright
23
33
99%
33
dri
60
53
79%
95%
19
100%
58
119
51
dri2
T23D8.8
dri1

B


dri
dril1
161
bright
20
29
100%
32
86
100%
97%
35
31
96%
128
dri2
34
T23D8.8
dri1

C

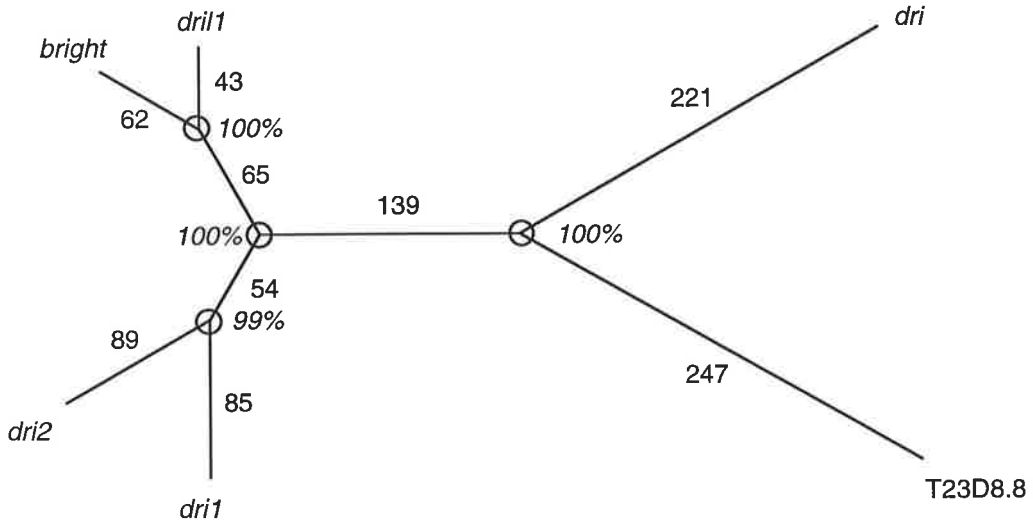
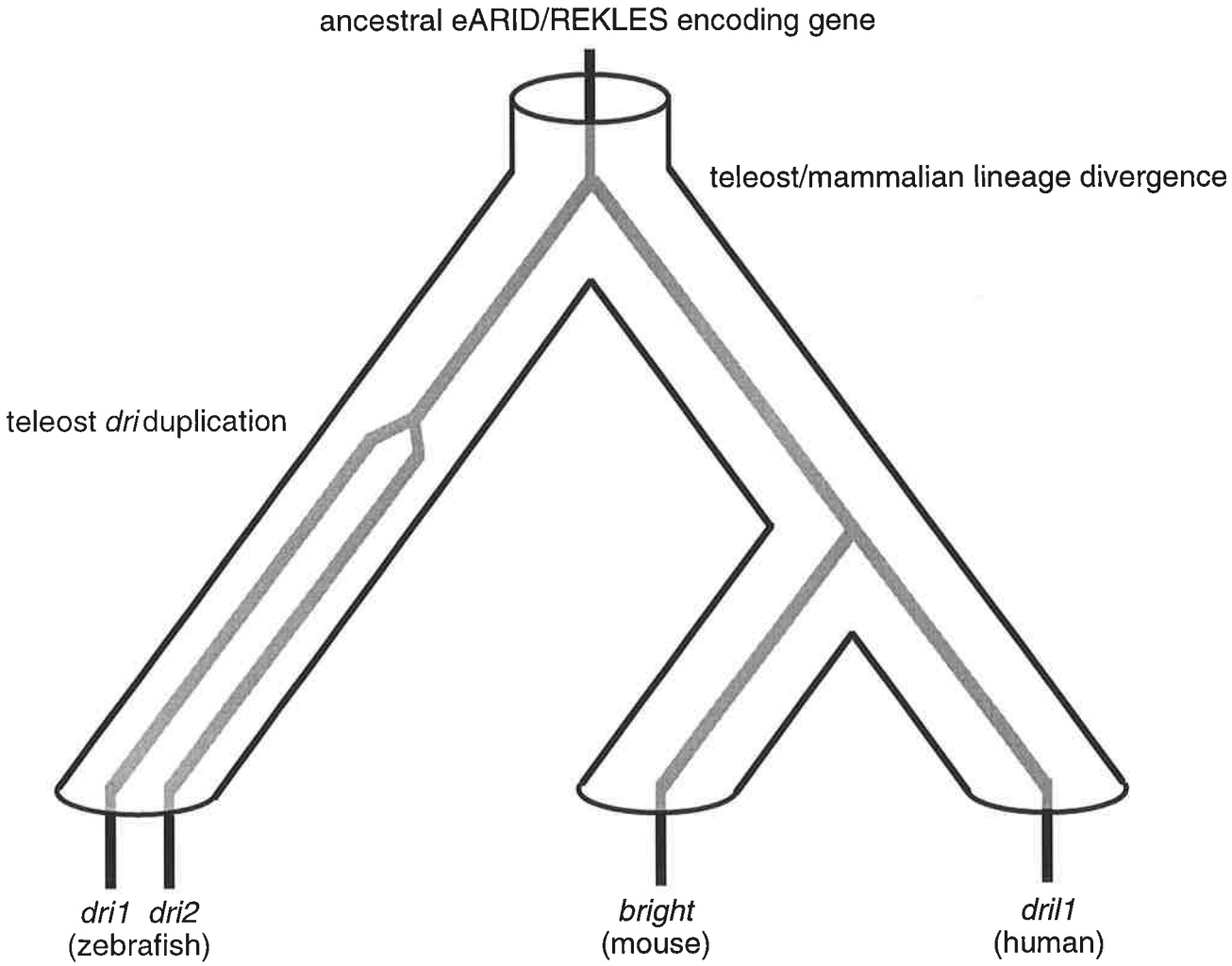

dril1
dri
bright
43
221
62
100%
65
139
100%
100%
54
99%
89
247
dri2
85
T23D8.8
dri1

Figure 3.11 — Branchial diagram showing a proposed phylogeny of the vertebrate eARID genes based in the phylograms shown in Figure 3.11. Pipes indicate the total genome evolution of the organisms shown, while lines indicate dead ringer homologue evolution. The divergence of the ancestral *D. rerio* orthologue after the divergence of the teleost and mammalian lineages is proposed to account for the phylogenies shown in figure 3.10.


ancestral eARID/REKLES encoding gene
teleost/mammalian lineage divergence
teleost *dri* duplication
dri1 *dri2*
(zebrafish)
bright
(mouse)
dril1
(human)

3.8 Discussion

Protein sequence similarity analysis suggests some groupings of ARID family members (figure 3.7). The identification of four eARID proteins in addition to the founding members of the group, Dead ringer and Bright, and phylogenetic analysis presented here significantly strengthen the notion that eARID proteins are an ancient group that serve a crucial role in animal development and function. Although a number of ARID-containing proteins have been shown to bind to, or interact with DNA (Peterson and Herskowitz 1992; Fattaey *et al*. 1993; Treisman *et al*. 1997), none outside the eARID group has been shown to do so in a direct, high-affinity sequence-specific manner. Additionally the core ARID from DRI does not bind DNA in a sequence specific manner (Gregory *et al*. 1996). Thus it appears that an ancestral ARID protein with weak, non-sequence specific DNA-binding capabilities gained the extended regions of the eARID, enabling the protein to bind specific DNA sequences with high affinity. The failure to find eARID-encoding genes in non-metazoan species, including the confirmed absence of an eARID-encoding gene in the completely sequenced *Saccharomyces cerevisiae* genome, indicates that the eARID evolved after the divergence of the metazoan lineage (Cherry *et al.* 1998). The presence of eARID-encoding genes in a variety of widely divergent metazoans suggests that this event occurred at an early stage of metazoan evolution.

The specific aim of identifying a human genetic disorder in order to suggest an underlying feature of the tissues expressing *dead ringer* was not achieved; no mapped human genetic disorder appears likely to correspond to the human *dead ringer* homologue, *DRIL1*. Analysis of *bright*, the mouse homologue, including a planned knockout (P. Tucker, pers. comm.), will undoubtedly aid in the identification of diseases that may be associated with the human locus. The complexity of the expression patterns observed in the zebrafish was reminiscent of those seen in *Drosophila*, making it difficult to draw conclusions about the underlying features of *D. rerio dri* expressing tissues. However, the observation of *dri1* and *dri2* expression in mesodermal and neural tissues in the fish is consistent with the expression

of *dri* seen in *Drosophila* (Gregory *et al*. 1996), strengthening the likelihood of a role of *dri* genes in the development of these tissues.

The existence of two *dri* homologues in *D. rerio* possibly arising from a genome duplication in the teleost line between 100 and 400 million years ago raises some interesting points (Amores *et al.* 1998). The function of the two genes has diverged, reflected in the divergence in both the expression patterns and the sequences of the conserved domains. Similar divergences have been observed in the sequences and expression patterns of the *D. rerio engrailed* and *msx* homeobox genes (Ekker *et al.* 1992; Akimenko *et al.* 1995; Force *et al.* 1998). Whether the ancestral eARID-encoding gene has neofunctionalised or subfunctionalised is not clear from comparison with *Drosophila* as the two organisms are too far diverged. However, comparison of the *D. rerio dri* expression patterns with the expression pattern of the mouse gene *bright* will aid in resolving this as only one eARID-encoding gene appears to be present in this organism, and the homology between tissues and organs in the fish and the mouse is more clearly apparent than between the fish and the fly.

Significantly, a novel domain was identified as a direct result of isolation of the *D. rerio* orthologues of *dri*. Although Herrscher *et al.* (1995) identified a tetramerisation function associated with the region of Bright corresponding to the REKLES domain, they could identify no significant homology with any of the known eARID proteins (P. Tucker, pers. comm.). Indeed, BLAST homology searches using the DRI REKLES domain gave smallest sum probability scores ranging from 0.35 to 0.9991 for the eARID proteins in the Genbank database (Altschul *et al.* 1990). These probability scores indicate a high likelihood of the match being due to chance. Similar searches using the *D. rerio* REKLES domain give scores in the ranges of 0.13 to 0.42 for Dri2 and 3.3×10^{-7} to 8.8×10^{-6} for Dri1. In addition to showing a higher pairwise significance for the REKLES motif, the *D. rerio* sequences broaden the range of proteins carrying the REKLES, strengthening the validity of the motif. Thus the identification of proteins with divergence times intermediate between those already known has been crucial to the identification of the REKLES domain.

A role for the REKLES domain in eARID protein function has been suggested by the observation that deletions in the Bright protein removing regions that include the REKLES domain fail to tetramerise in gel retardation assays (Herrscher *et al*. 1995). The role for the REKLES domain as a protein interaction domain is consistent with the greater level of amino acid sequence divergence between members of the group in the REKLES domain when compared with the level of divergence in the eARID. Although no non-eARID protein has been found that bears the REKLES domain, and the results of Herrscher *et al*. (1995) suggest that the REKLES domain is required for homo-tetramerisation, a more generalised role in which the REKLES domain binds to heterologous proteins cannot be ruled out. Interestingly, the two *D. rerio* Dri proteins show divergence in the REKLES; Dri2 deviates from the REKLES signature, while Dri1 lacks 14 amino acids after the REKLES sequence compared to *Drosophila* DRI, making it more similar to the other vertebrate members of the group in this region. However, the Dri1 REKLES domain also lacks a conserved region present in the other vertebrate proteins. These differences in a putative protein-interaction domain resonate with differences seen in the eARID of the Dri2 protein which may alter its specificity of DNA-binding. The observed differences between the *D. rerio* Dri protein sequences and *D. rerio dri* expression patterns, and possible functional changes resulting from these differences are likely to permit the selective pressure that maintaining the existence of two *dead ringer* genes in this species.

Chapter 4: Functional characterisation of DRI domains

It seemed to me to be a superlative thing — to know the explanation of everything, why it comes to be, why it perishes, why it is. - Socrates

4.1 DRI is a multidomain DNA-binding protein

Comparison of Dead ringer and its orthologues has identified novel motifs, the eARID (Herrscher *et al*. 1995; Gregory *et al*. 1996; Kortschak *et al*. 1998) and REKLES motifs (this study). Although the *in vitro* function of the eARID has been characterised (Herrscher *et al*. 1995; Gregory *et al*. 1996), only circumstantial evidence exists for the *in vitro* function of the REKLES domain (Herrscher *et al*. 1995), and no role for either domain has been demonstrated *in vivo*. Indeed, any *in vivo* role for the ARID has been questioned by the observation that SWI1 lacking the ARID fails to show any loss of SWI1 function (C. Peterson, pers. comm.).

In order to examine the roles of the ARID and the REKLES domain *in vivo* it is necessary to replace the endogenous DRI complement with DRI lacking either ARID or REKLES function. Due to the inability to perform targeted homologous recombination in *Drosophila* this requires a two step approach. Chapter 2 describes the first step towards replacing endogenous DRI with a site-directed mutant transgene, the identification of mutations in the *dri* locus. The second stage involves the generation and introduction of transgenic constructs bearing mutations abolishing specific functions of its protein product.

In the absence of known mutations leading to loss of specific functions, a theoretical analysis was used to determine regions of the protein to target for mutagenesis based on evolutionary conservation and computer based protein structure predictions.

The secondary structure of the DRI eARID, as predicted by the PHD profile fed neural network secondary structure prediction program (Rost and Sander 1993; Rost *et al*. 1994; Rost and Sander 1994), is predominantly α-helical, consisting of 5 α-helices separated by loops of unstructured sequence (figure 4.1). Comparison of predicted eARID secondary structure between the six orthologous proteins showed a high degree of predicted structural

Figure 4.1 — Prediction of structure of the DRI eARID using the PHD program (Rost and Sander 1993; Rost *et al.* 1994; Rost and Sander 1994, Figure 1). The program predicts the three secondary structure types based on the protein sequences of DRI, T23D8.8, Dri1, Dri2, Bright and DRIL1. The structure state of a residue is assigned by choosing the maximal scaled probability value of prH, prE or prL, the probabilities of forming helix, extended β-sheet of unstructured loop respectively. The structure state of the residue is shown as H: helix, E: sheet or blank: unstructured loop or no prediction. The residues of ARID helix 3 deleted in DRIδARID$_{H3}$ are struck through. The reliability index (Rel) giving an indication of the confidence of the prediction is shown below each residue. The accuracies to be expected for residues with values above a particular value of the index are given:

index	0	1	2	3	4	5	6	7	8	9
three state accuracy	72.1%	72.3%	74.8%	77.7%	80.3%	82.9%	85.7%	88.5%	91.1%	94.2%

A subset of the prediction with a reliability index greater than 5 indicating an expected three state accuracy of greater than 82% is shown on the final line, with L indicating loop and "." indicating no reliable prediction.


....,...250...,...260...,...270...,...280...,...290...,...300
AA |ANSASNSSTSSEASNSSQQNNGWSFEEQFKQVRQLYEINDDPKRKEFLDDLFSFMQKRGT|
E HHHHHH EEE HHHHHHHHHHHHH
Rel |999999998998788989999832121664320115434993334775899999996089|
detail:
prH |000000000000100000000000114776553441110003356787899999987410|
prE |000000000000000010000023433000002236552000000000000000000000|
prL |999999988888889989998865452113334311226896633112100000012489|
>82% SUBSET |LLLLLLLLLLLLLLLLLLLLL.....HH......E...LL....HHHHHHHHHHH.LL|
....,...310...,...320...,...330...,...340...,...350...,...360
AA |PINRLPIMAKSVLDLYELYNLVIARGGLVDVINKKLWQEIIKGLHLPSSITSAAFTLRTQ|
HHHHHHHHHHHHHHHHHH EEEHHHHHHHHHH HHHHHHHHHHHHH
Rel |888531799999999999999971992334029899999981998345688999999999|
detail:
prH |001235789999999999999984001132359899999984000666788999999999|
prE |000000000000000000000000003566300000000000000000000000000000|
prL |888654100000000000000001499520024010000001599833211100000000|
>82% SUBSET |LLLL..HHHHHHHHHHHHHHHHH.LL......HHHHHHHHH.LLL..HHHHHHHHHHH|
....,...370...,...380...,...390...,...400...,...410...,...420
AA |YMKYLYPYECEKKNLSTPAELQAAIDGNRREGRRSSYGQYEAMHNQMPMTPISRPSLPGG|
HHHH HHH HHHHHHHH
Rel |787346634123568993899898615754676767777767768799879999999999|
detail:
prH |887521136333210004899898752123100011111111111000100000000000|
prE |000111100321000000000000000000001111000000100000000000000000|
prL |111357762345678995100001247876787777787877878799889999999999|
>82% SUBSET |HHH..LL.....LLLLL.HHHHHHH.LLL.LLLLLLLLLLLLLLLLLLLLLLLLLLLL|

conservation (not shown). Similar analysis of the REKLES domain indicated the presence of a conserved pair of β-sheet strands separated by a loop of unstructured sequence corresponding to the region of highest conservation in the REKLES motif, now named the REKLESβ region (DRI amino acids 788 to 806 QLIVNLELNGVKYSGVLVA, figures 4.2 and 4.3). Comparison of the predicted REKLES domain secondary structures of all eARID proteins revealed only low levels of structural conservation outside this region (not shown).

ARID sequence conservation within the eARID group of proteins is very high, as was discussed in the previous chapter. It was reasoned that residues conserved not only within the eARID group, but across the entire ARID family would be playing a crucial role in ARID function, thus to ensure abolition of eARID function, a region including an invariant residue was chosen. Investigation of the role of a domain by deletion of that domain requires that the remaining protein be expressed and form a stable structure. Thus, a region was selected that was postulated to form a structurally important component, on the basis of the inclusion of an invariant residue and conserved predicted secondary structure, but was small to reduce the likelihood of completely destabilising the protein and causing it to be degraded. For the DRI eARID, the predicted third helix of the core ARID was chosen, that is, amino acids 334 to 342 (KKLWQEIIK), which contains an invariant tryptophan. The REKLES motif is much less highly conserved than the eARID. It was reasoned that the most highly conserved region of the DRI REKLES domain, including the invariant residues of the REKLESβ region (amino acids 792 to 807, NLELNGVKYSGVLVAN, invariant residues are underlined), would make a good target for deletion.

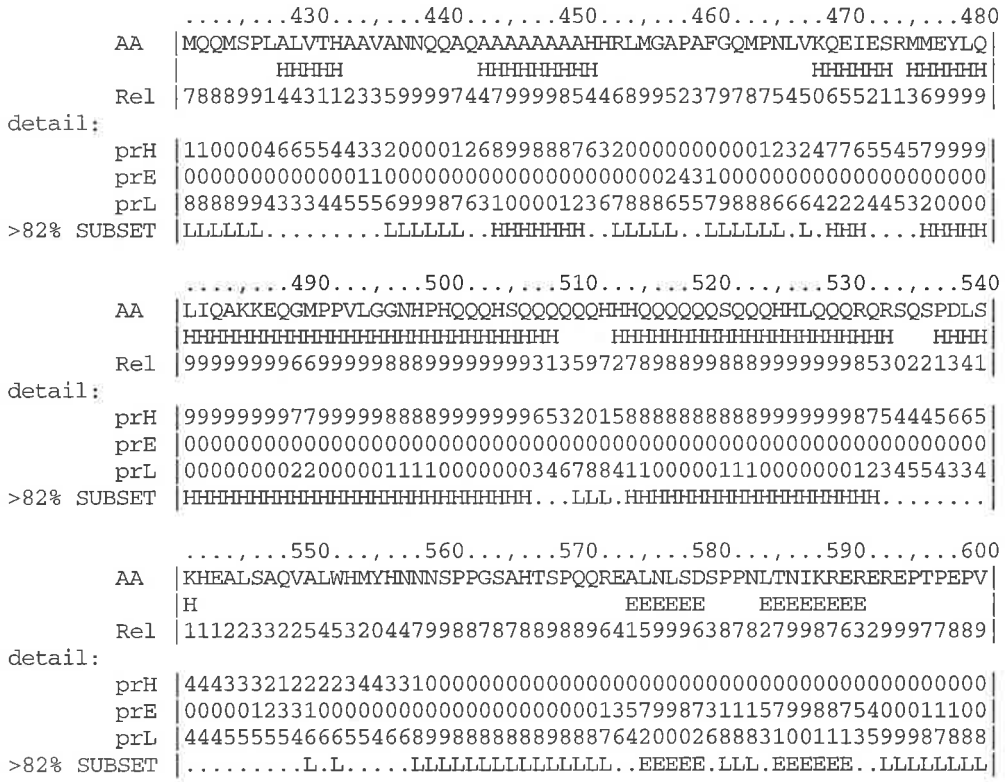
Herrscher *et al*. (1995) have shown that Bright forms tetramers in gel retardation assays and in native protein gel electrophoresis (M. Kaplan, pers. comm.). The region of Bright corresponding to the REKLES domain is necessary for this activity. If a conserved function of the REKLES domain is to mediate tetramerisation then it may be hypothesised that a mutant DRI lacking eARID DNA-binding function but possessing REKLES tetramerisation function would behave antimorphically, as DNA-binding defective monomers would be incorporated into the protein complex, potentially poisoning the function of the

Figure 4.2 — Prediction of structure of the DRI REKLES domain using the PHD program (Rost and Sander 1993; Rost *et al*. 1994; Rost and Sander 1994, Figure 1). The DRI REKLES domain is split showing the amino (A) and carboxy (B) terminal regions, but not the central region with no homology to the other eARID proteins. The structure state of a residue is assigned by choosing the maximal scaled probability value of prH, prE or prL, the probabilities of forming helix, extended β-sheet of unstructured loop respectively. The structure state of the residue is shown as H: helix, E: sheet or blank: unstructured loop or no prediction. The residues of the REKLESβ region (underlined) deleted in DRIδREKLESβ are struck through. The reliability index (Rel) giving an indication of the confidence of the prediction is shown below each residue. The accuracies to be expected for residues with values above a particular value of the index are given:

index	0	1	2	3	4	5	6	7	8	9
three state accuracy	72.1%	72.3%	74.8%	77.7%	80.3%	82.9%	85.7%	88.5%	91.1%	94.2%

A subset of the prediction with a reliability index greater than 5 indicating an expected three state accuracy of greater than 82% is shown on the final line, with L indicating loop and "." indicating no reliable prediction.

A


....,...430...,...440...,...450...,...460...,...470...,...480
AA |MQQMSPLALVTHAAVANNQQAQAAAAAAAAHHRLMGAPAFGQMPNLVKQEIESRMMEYLQ|
HHHHH HHHHHHHHH HHHHHH HHHHHH
Rel |788899144311233599997447999985446899523797875450655211369999|
detail:
prH |11000046655443320000126899888763200000000001232477655457999 9|
prE |00000000000001100000000000000000000024310000000000000000000|
prL |88889943334455569998763100001236788865579888666422244532000 0|
>82% SUBSET |LLLLLL.........LLLLLL..HHHHHHH..LLLLL..LLLLLL.L.HHH...HHHHH|
....,...490...,...500...,...510...,...520...,...530...,...540
AA |LIQAKKEQGMPPVLGGNHPHQQQHSQQQQQQHHHQQQQQQSQQQHHLQQQRQRSQSPDLS|
HHHHHHHHHHHHHHHHHHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHH HHHH
Rel |99999999669999988899999999931359727898899888999999985302213 41|
detail:
prH |99999999779999988889999999653201588888888889999998754445665|
prE |00000000000000000000000000000000000000000000000000000000000|
prL |00000000220000011110000000346788411000001110000001234554334|
>82% SUBSET |HHHHHHHHHHHHHHHHHHHHHHHHHH...LLL.HHHHHHHHHHHHHHHHHH........|
....,...550...,...560...,...570...,...580...,...590...,...600
AA |KHEALSAQVALWHMYHNNNSPPGSAHTSPQQREALNLSDSPPNLTNIKRERERERPTPEPV|
H EEEEEE EEEEEEEE
Rel |11122332254532044799887878898896415999638782799876329997788 9|
detail:
prH |44433321222234433100000000000000000000000000000000000000000000|
prE |00000123310000000000000000000001357998731115799887540001110 0|
prL |44455555466655466899888888898887642000268883100111359998788 8|
>82% SUBSET |.........L.L.....LLLLLLLLLLLLLLL..EEEEE.LLL.EEEEEE..LLLLLLLL|

B

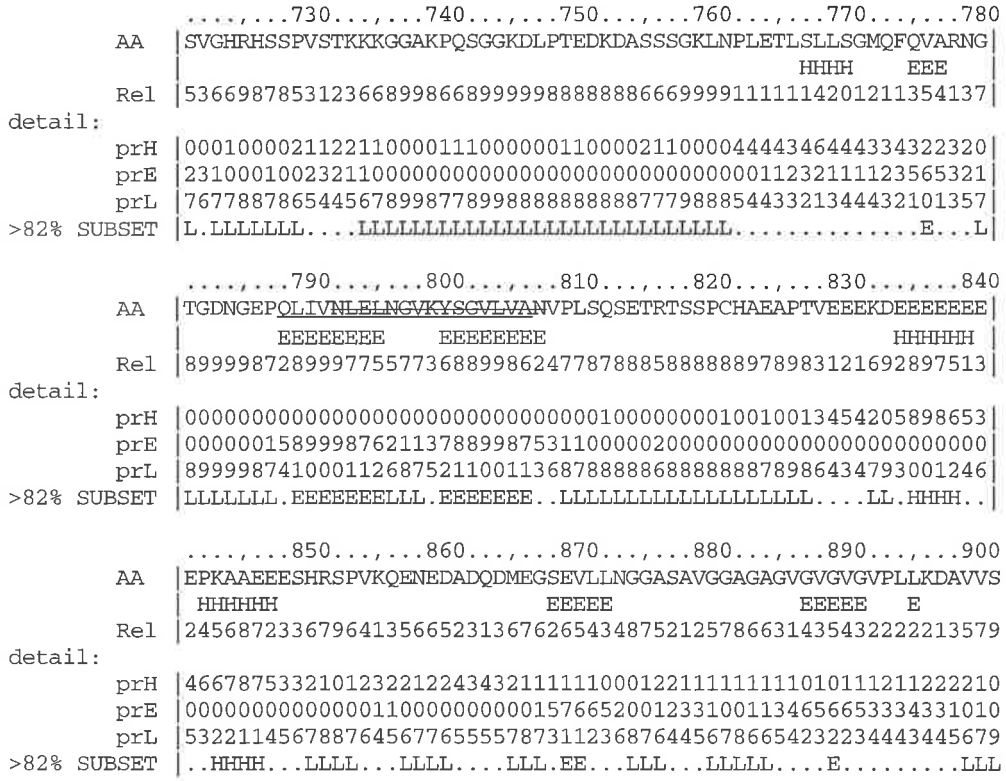

....,...730...,...740...,...750...,...760...,...770...,...780
AA |SVGHRHSSPVSTKKKGGAKPQSGGKDLPTEDKDASSSGKLNPLETLSLLLSGMQFQVARNG|
HHHH EEE
Rel |53669878531236689986689999988888866699991111114201211354137|
detail:
prH |00010000211221100001110000001100002110000444434644334322320|
prE |23100010023211000000000000000000000000000011232111123565321|
prL |76778878654456789987789988888888887779888544332134443210135 7|
>82% SUBSET |L.LLLLLLL....LLLLLLLLLLLLLLLLLLLLLLLLLLLL.............E...L|
....,...790...,...800...,...810...,...820...,...830...,...840
AA |TGDNGEPQLIVNLELNGVKYSGVLVANVPLSQSETRTSSPCHAEAPTVEEEKDEEEEEEE|
EEEEEEEE EEEEEEEE HHHHHH
Rel |89999872899977557736889986247787885888888978983121692897513|
detail:
prH |00000000000000000000000000000010000000010010013454205898653|
prE |00000015899987621137889987531100000200000000000000000000000|
prL |89999874100011268752110011368788888688888887898643479300124 6|
>82% SUBSET |LLLLLLL.EEEEEEELLL.EEEEEEE..LLLLLLLLLLLLLLLLLLL....LL.HHHH..|
....,...850...,...860...,...870...,...880...,...890...,...900
AA |EPKAAEEESHRSPVKQENEDADQDMEGSEVLLNGGASAVGGAGAGVGVGVGVPLLKDAVVS
HHHHHH EEEEE EEEEE E
Rel |24568723367964135665231367626543487521257866314354322222135 79
detail:
prH |46678753321012322122434321111110001221111111110101112112222 10
prE |00000000000000110000000000157665200123310011346566533343310 10
prL |53221145678876456776555578731123687644567866542322344434456 79
>82% SUBSET |..HHHH...LLLL...LLLL....LLL.EE...LLL...LLLLL....E.........LLL

Figure 4.3 — Comparison of outputs for the structure of the REKLES β region by the PHD program (Rost and Sander 1993; Rost *et al*. 1994; Rost and Sander 1994, Figure 1). Six runs of the program were performed using the six eARID protein sequences as inputs to the program and varying the target sequence. The structure state of the residue is shown as H: helix, E: sheet or blank: unstructured loop or no prediction. The reliability index giving an indication of the confidence of the prediction is shown below each residue. The accuracies to be expected for residues with values above a particular value of the index:

index	0	1	2	3	4	5	6	7	8	9
three state accuracy	72.1%	72.3%	74.8%	77.7%	80.3%	82.9%	85.7%	88.5%	91.1%	94.2%

Comparison of REKLESβ secondary structures

```
DRI          NGTGDNGEPQLIVNLELNGVKYSGVLVANVPLS
prediction           EEEEEEEE    EEEEEEEE
Rel index    37899998728999775577368899862 4778

T23D8.8      TKHSDNSKTSMSVSMEINGITYQGVLFALDETV
prediction           EEEEEEEEE  EEEEEEEEEE
Rel index    66999994579999962344885887563 6988

Dri1         SISS-GGSASISVSVEVNGIVYSGNLFAQKSVS
prediction   E   -   EEEEEEEEEE EEEEEEE
Rel index    2599-99916999999732345453313689 99

Dri2         SISS-NGSASITVSVEVNGIIYSGSLYAQKSGS
prediction       -    EEEEEEEEEEEEEEEEEEE
Rel index    3369-99823799999864255689866289 99

Bright       SLTSANGSNSISMSVEMNGIVYTGVLFAQPPPP
prediction           EEEEEEEE  EEEEEEEE
Rel index    22478897169999853231441222117 8999

DRIL1        NLTGTNGSNSISMSVEINGIMYTGVLFAQPPAP
prediction           EEEEEEEE    EEEEE
Rel index    88889997256999985362132432115 8999
```

complex. Thus, a significant test for a generalised model of eARID protein function, based on the work of Herrscher *et al*. will be to test whether DRI lacking eARID function is antimorphic.

4.2 Generation of wild-type and domain-specific mutant *dri* constructs

As no full length *dri* cDNA was isolated during the screen for cDNAs (Gregory 1996), a full length dri cDNA was generated by cloning the 3', *Sty*I/*Sma*I, cDNA fragment from p*dri*2T into the *Sty*I/*Eco*RV sites of p*dri*ϕ4. The resulting clone, p*drif*, releases a 3.7kb *Eco*RI dri cDNA (not shown). The integrity of the junction of p*dri*λ2T and p*dri*ϕ4 DNA at the StyI site was confirmed by sequence analysis (data not shown). The full length *dri* cDNA was cloned into *Eco*RI sites of the p*P*[UAST] and pCaSpeR-hs *Drosophila* transformation vectors to generate p*P*[UAS::*dri*] and p*P*[hs::*dri*]; correct orientation of the cDNA was confirmed by restriction and sequence analysis (data not shown).

To generate domain specific mutants for transformation into *Drosophila*, p*drif* was mutagenised using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). Primer pairs, Δhelix3+/Δhelix3- and ΔREKLESβ+/ΔREKLESβ- were used to generate the *dri* mutant plasmids p*dri*δARID$_{H3}$ and p*dri*δREKLES$_{\beta}$ respectively. p*dri*δARID$_{H3}$ mutants were identified by screening for loss of Δhelix3 tester primer dependent PCR amplification product; p*dri*δREKLES$_{\beta}$ mutants were identified by loss of the *Sca*I site within the REKLES$_{\beta}$ box (not shown). The mutated sites in mutant clones were confirmed by sequencing (figure 4.4A). Mutant *dri* cDNAs were cloned into the *Eco*RI site of p*P*[UAST] to generate p*P*[UAS::*dri*δARID$_{H3}$] and p*P*[UAS::*dri*δREKLES$_{\beta}$]; correct orientation of the cDNA was confirmed by restriction analysis (data not shown).

To analyse the *in vitro* function of the ARID and REKLES domains, protein expression constructs were generated. A fragment generated by PCR using the primers, ARID box 5' and ARID box 3' was cloned into the *Bam*HI and *Eco*RI sites of pGEX1 to generate pGEX::*dri*$_{ARID}$ (with amino acids 258-410, QQNNG. . . QMPMT). Accuracy of PCR, and

Figure 4.4 — Sequence analysis of domain specific deletions generated by QuikChange™ cycle mutagenesis.

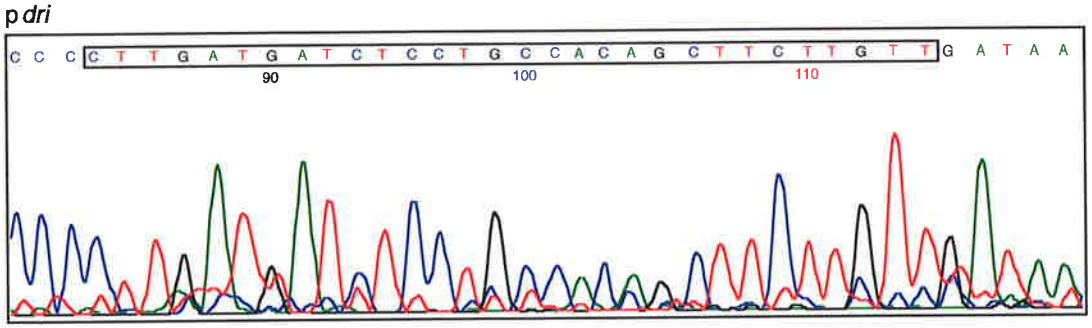
A. Sequence obtained from the clones p*dri*, p*dri*δARID$_{H3}$ and pGEX::*dri*$_{ARIDδH3}$ using the driRT sequencing primer. Note that sequence from driRT-primed sequencing reactions is inverted with respect to the normal 5' to 3' reading of the cDNA.

B. Sequence obtained from the clones p*dri* and p*dri*δREKLESβ using the driD sequencing primer. Sequence from driD-primed sequencing reactions is colinear with the normal 5' to 3' reading of the cDNA.

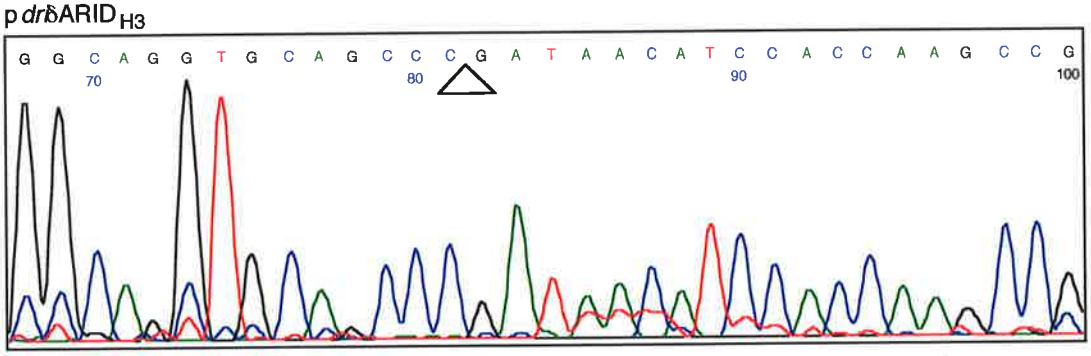
Boxed region in wild-type sequences indicate regions deleted in the mutants. Triangles below mutant sequences indicate mutant sequence junctions.

A

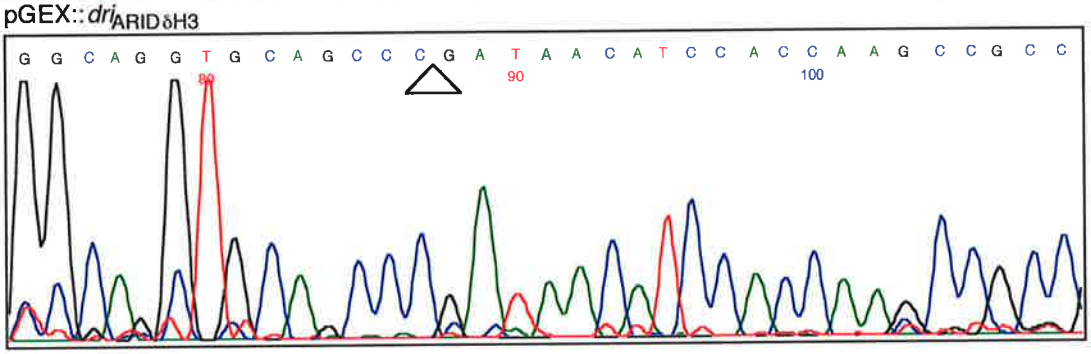
p *dri*



p *dri*δARID$_{H3}$

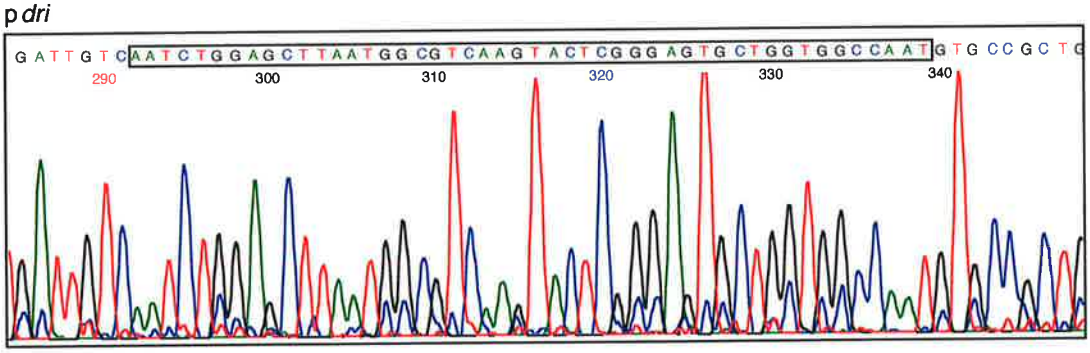


pGEX::*dri*$_{ARID\,\delta H3}$

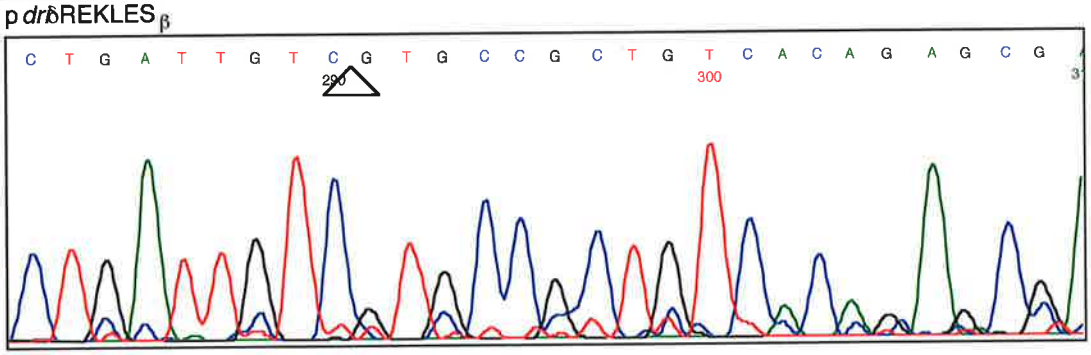


B

p *dri*



p *dri*δREKLES$_{\beta}$



orientation and frame of insertion of the fragment were confirmed by sequence analysis (data not shown). A deletion construct, pGEX::*dri*$_{\text{ARID}\delta\text{H3}}$, lacking the putative eARID helix 3 was generated from pGEX::*dri*$_{\text{ARID}}$ using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) with the primer pair Δhelix3+/Δhelix3-. Mutant clones were identified as described for p*dri*δARID$_{\text{H3}}$ and were confirmed by sequence analysis (figure 4.4B). Similarly, wild-type and mutant REKLES constructs were generated; fragments generated by PCR from the templates, p*drif* and p*dri*δREKLES$_{\beta}$ with *Pfu* polymerase using the primers, REKLES box 5' and REKLES box 3' were cloned into the *Bam*HI and *Eco*RI sites of pGEX1 to generate pGEX::*dri*$_{\text{REKLES}}$ (with amino acids 400-901, YEAMH...DAVVS) and pGEX::*dri*$_{\text{REKLES}\,\delta\beta}$ (as above, but lacking the REKLES$_{\beta}$ region) respectively. Orientation and frame of insertion of the fragments were confirmed by sequence analysis (data not shown).

4.3 Generation of DRI specific antibodies in rats

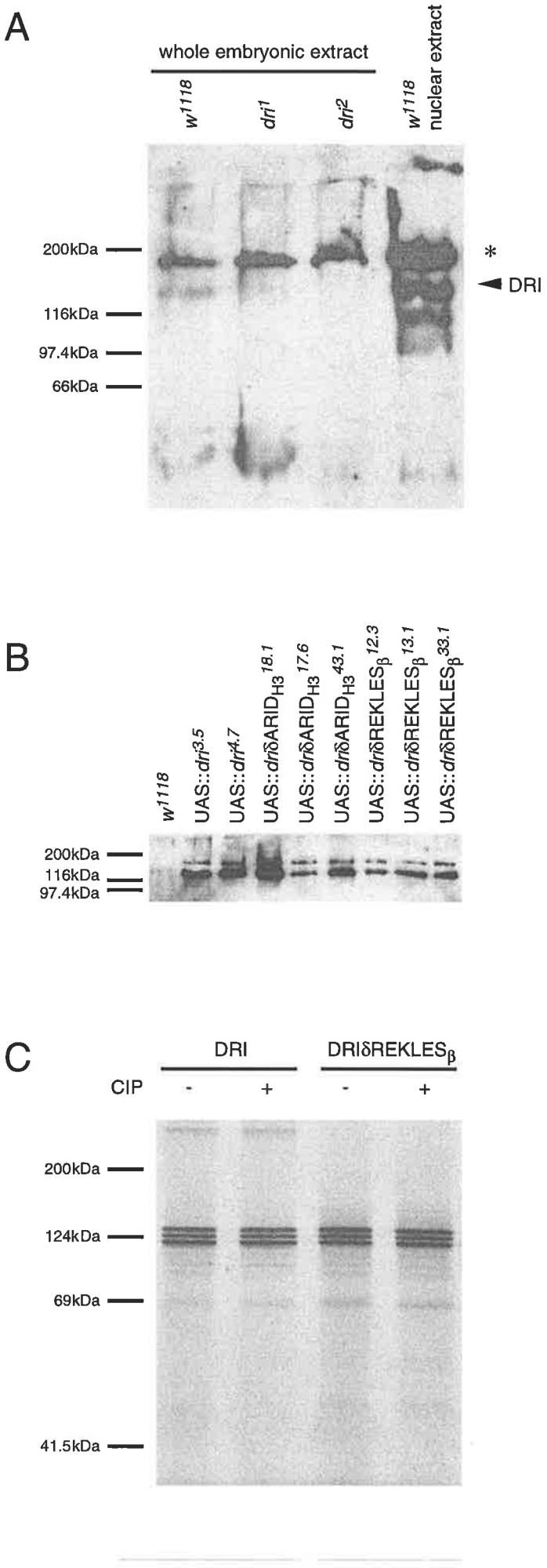
In order to adequately control for the ability to rescue function with wild-type and mutant forms of *dead ringer*, it is necessary to show that the levels of expression from the constructs are comparable. As *in situ* hybridisation does not allow detection of potential protein expression and stability problems, it was necessary to generate antibodies to DRI protein to be used for determining expression levels.

A GST-DRI fusion protein, covering amino acids 128 to 671 of DRI, was expressed from pGEX1ϕ10 (Gregory *et al*. 1996) and used to immunise rats as described in chapter 6. Specificity of the antiserum was shown by western analysis of protein extracts from embryos maternally and zygotically amorphic for *dri* (figure 4.5A). w^{1118} embryos show two bands: at 120kDa and 150kDa. The 120kDa band is lost in *dri* mutant embryonic extracts. The 150kDa band appears to be an artefact of the western analysis as whole mount immunohistochemistry on maternally and zygotically *dri* deficient embryos or *dri* somatic clones using anti-DRI antibody shows no detectable signal (T. Shandala, unpublished observations).

Figure 4.5 — A. Analysis of anti-DRI antibody specificity. Embryonic extracts were prepared from 10µl of w^{1118} embryos or embryos derived from mothers carrying homozygous dri^1 or dri^2 germline clones (provided by T. Shandala). To show nuclear localisation of DRI, embryonic nuclear extracts from w^{1118} embryos (provided by S. S. Robert) were used. DRI was detected by affinity purified anti-DRI antibody. Two bands are present in w^{1118} embryos, one at 150kDa and one at 120kDa. Two additional bands are present in w^{1118} nucelar extracts, at 116kDa and 100kDa. These bands may be the result of degradation of DRI during the nuclear extraction. In embryos from mothers lacking *dri* in the germ line the 120kDa band is not seen indicating that this band corresponds to DRI. The apparent size of DRI is consistent with its predicted size of ~100kDa. The 150kDa band appears to be the artefactual result of western analysis as no anti-DRI immunostaining is observed in whole mount embryos or larval clones carrying the dri^1 or dri^2 mutations (T. Shandala, pers. comm.).

B. Western analysis of transgenic DRI expression from various *dri*-transgenic lines used for *in vivo* analysis of DRI domain function. Extracts from 5 pairs of larval eye-antennal imaginal disks were analysed by SDS-PAGE and transferred to nitrocellulose. DRI was detected by affinity purified anti-DRI antibody. Larvae used for transgenic protein expression were w^{1118} and carried one copy of $GAL4^{GMR.PF}$ and one copy of the *P*[UAS::*dri**] (* indicates wild-type or mutated) construct indicated. w^{1118} larval extract is shown to indicate endogenous levels of DRI in eye-antennal imaginal disks. The absence from w^{1118} larval extracts of the 150kDa band seen in embryonic extracts (A) is repeatable and does not appear to be the result of loading differences (data not shown).

C. Phosphorylation state analysis of *in vitro* transcribed and translated DRI. Radiolabelled wild-type DRI and DRIδREKLESβ protein was analysed by SDS-PAGE after treatment with calf intestinal alkaline phosphatase or no treatment. No difference was observed between treated and untreated DRI or between DRI and DRIδREKLESβ.


A
whole embryonic extract
w1118
dri1
dri2
w1118 nuclear extract
200kDa
116kDa
97.4kDa
66kDa
*
DRI
B
w1118
UAS::dri3.5
UAS::dri4.7
UAS::driδARID_H3 18.1
UAS::driδARID_H3 17.6
UAS::driδARID_H3 43.1
UAS::driδREKLES_β 12.3
UAS::driδREKLES_β 13.1
UAS::driδREKLES_β 33.1
200kDa
116kDa
97.4kDa
C
DRI
DRIδREKLES_β
CIP
-
+
-
+
200kDa
124kDa
69kDa
41.5kDa

4.4 The predicted ARID helix 3 is necessary for NP binding

The DRI eARID, spanning the sequence from amino acid 258 to 410 (QQNNG...QMPMT) has previously been shown to be sufficient for sequence specific DNA binding to trimeric NP oligonucleotide (Gregory *et al.* 1996). The ARID-box mutant constructs described above lack the sequence encoding helix 3 of the ARID (amino acids 334 to 342, KKLWQEIIK). To determine whether helix 3 of the DRI ARID is necessary for NP binding, electrophoretic mobility shift assays were conducted using both the wild-type ARID (GST-DRI_{ARID}) as has been previously described, and the mutant ARID (GST-$DRI_{ARID\delta H3}$) lacking the helix 3 sequence. Using elevated protein levels to detect any potential binding by the mutant, GST-DRI_{ARID} was able to completely retard the available NP_6 while GST-$DRI_{ARID\delta H3}$ was not able to retard the NP_6 DNA (figure 4.6A and B). Thus helix 3 of the DRI ARID is necessary for NP binding *in vitro*.

4.5 The REKLESβ region is not required for self-association

The conservation of the REKLESβ signature in the six eARID proteins was suggestive of a significant role in the function of these proteins. Herrscher *et al.* (1995) showed that a region corresponding to the REKLES domain was required for tetramerisation of the Bright protein in gel retardation assays. To determine whether the REKLESβ region of DRI is necessary for multimerisation a western blot of wild-type REKLES (GST-DRI_{REKLES}) and mutant REKLES (GST-$DRI_{REKLES\delta\beta}$) lacking the predicted beta-region of the REKLES domain, was probed with ^{35}S-methionine-labelled DRI (figure 4.6C). No difference in signal intensity was observed between the wild-type and mutant forms, indicating that the REKLESβ region is not necessary for self-association.

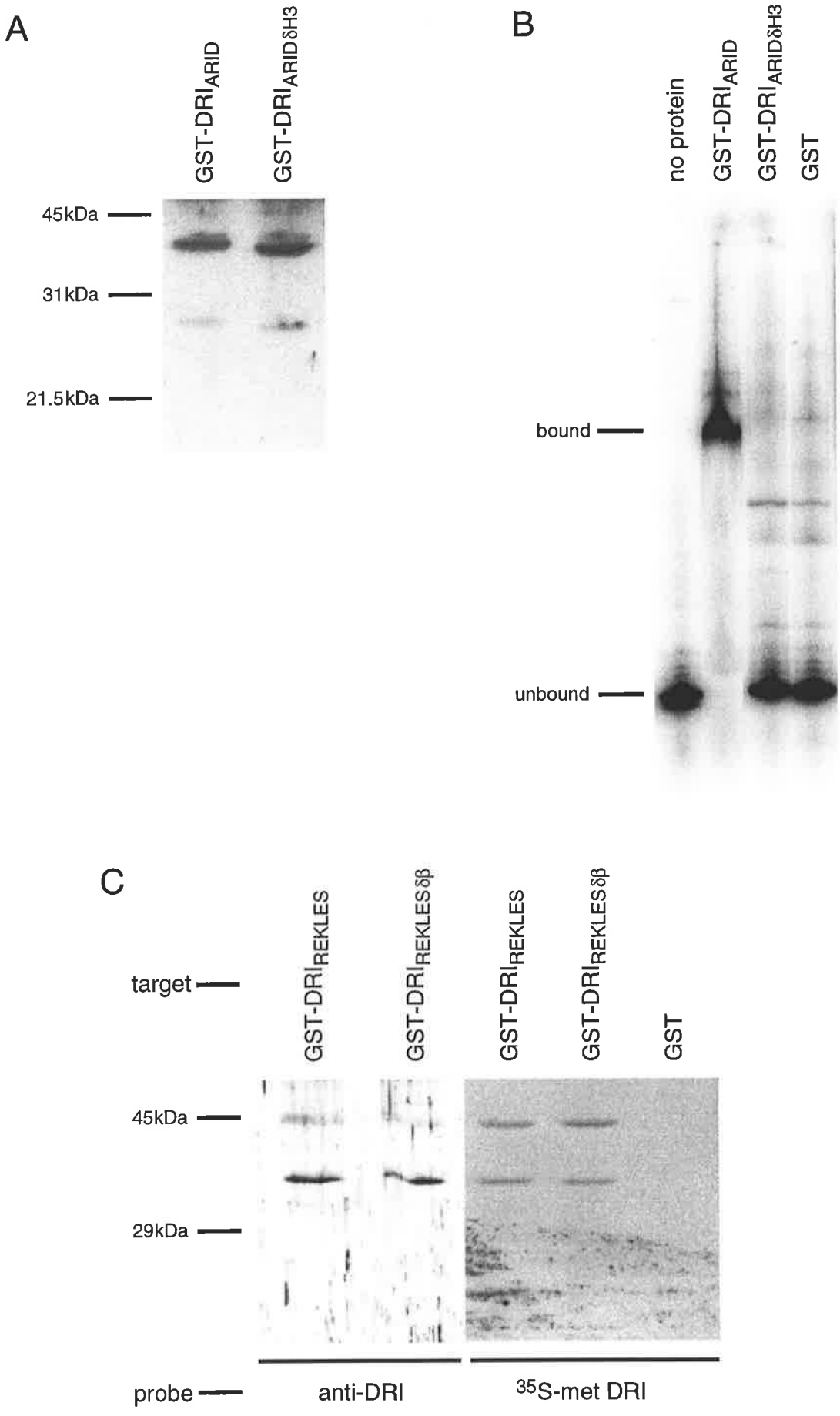
The REKLESβ motif contains a completely conserved tyrosine residue, which may act as a regulatory site on the protein. To investigate the possibility that DRI is phosphorylated to regulate DRI function, *in vitro* translated ^{35}S-methionine-labelled DRI was treated with calf-

Figure 4.6 — Functional analysis of the conserved DRI domains.

A. Western analysis of bacterial GST-DRI_{ARID} and GST-$DRI_{ARID\delta H3}$ expression using anti-DRI antibodies.

B. Gel electrophoretic mobility shift analysis of GST-DRI_{ARID}, GST-$DRI_{ARID\delta H3}$ and GST. The NP_6 oligonucleotide was used as a probe in all lanes and the same amounts of GST-DRI_{ARID} and GST-$DRI_{ARID\delta H3}$ protein were used as those shown in the western analysis in A. No western-based control was possible for GST loading, but equivalent quantities were used as estimated by coomassie stained GST bands generated by polyacrylamide gel electrophoresis (not shown).

C. Western analysis of DRI REKLES domain self-association activity. GST-DRI_{REKLES} and GST-$DRI_{REKLES\,\delta\beta}$ were electrophoresed and transferred to nitrocellulose membranes. Membranes were probed with either anti-DRI antibody or radiolabelled full-length DRI protein produced by *in vitro* transcription and translation. The 29kDa GST protein unfused to DRI and probed with radiolabelled DRI protein is also shown. Note that the GST-DRI_{REKLES} and GST-$DRI_{REKLES\,\delta\beta}$ proteins appear as two bands both of which are less than the size predicted for a GST-DRI_{REKLES} protein. The reduction in apparent size and the presence of two bands are probably the result of degradation of the full length product.


A
GST-DRI$_{ARID}$
GST-DRI$_{ARIDδH3}$
45kDa
31kDa
21.5kDa
B
no protein
GST-DRI$_{ARID}$
GST-DRI$_{ARIDδH3}$
GST
bound
unbound
C
target
GST-DRI$_{REKLES}$
GST-DRI$_{REKLESδβ}$
GST-DRI$_{REKLES}$
GST-DRI$_{REKLESδβ}$
GST
45kDa
29kDa
probe
anti-DRI
^{35}S-met DRI

intestinal alkaline phosphatase and analysed by SDS-PAGE (figure 4.5C). No difference in electrophoretic mobility or number of bands was observed between treated and untreated samples, or between the mutant and wild-type forms.

4.6 Generation of transgenic lines

The p*P*[UAS::*dri*] and p*P*[hs::*dri*] plasmids were independently injected into w^{1118} syncytial embryos in the presence of Δ2-3 helper plasmid. From approximately 300 embryos injected with each construct, four confirmed independent transgenic lines were established carrying the *P*[UAS::*dri*] insertion, but no *P*[hs::*dri*] transformant was generated. The four *P*[UAS::*dri*] insertion lines were mapped genetically to the X chromosome (UAS::$dri^{1.2}$), the second chromosome (UAS::$dri^{3.5}$ and UAS::$dri^{4.19}$) and the third chromosome (UAS::$dri^{4.7}$). To determine whether the failure to generate a *P*[hs::*dri*] insertion line was due to toxicity of leaky expression of *dri*, the line UAS::$dri^{3.5}$ was crossed to a line homozygous for hs::GAL4 on the third chromosome. When raised at 25°C flies heterozygous for UAS::$dri^{3.5}$ on the second and *hs*::GAL4 on the third appeared normal, whereas at 29°C these flies exhibited significantly reduced viability. Heat shock treatment for 30 minutes of third instar larvae carrying the same insertions resulted in complete mortality. The lethality or reduced viability caused by *dri* expression under heat shock control suggested that ubiquitous expression of *dri* would not be a practical approach for attempting rescue of *dri* function *in vivo*. Indeed, the lethality of ectopic *dri* expression has been demonstrated by using a number of *GAL4* expression lines in conjunction with UAS::*dri* transgenes (not shown; A. Lumsden, T. Shandala and R. Saint unpublished observations)

The p*P*[UAS::*dri*δARID$_{H3}$] and p*P*[UAS::*dri*δREKLES$_{\beta}$] plasmids were independently injected into w^{1118} syncytial embryos in the presence of Δ2-3 helper plasmid. From 300 embryos injected with pP[UAS::*dri*δARID$_{H3}$] and 258 embryos injected with p*P*[UAS::*dri*δREKLES$_{\beta}$], 14 independent *P*[UAS::*dri*δARID$_{H3}$] and 9 independent *P*[UAS::*dri*δREKLES$_{\beta}$] transgenic lines were established. The insertions were mapped

genetically to the X chromosome (UAS::*dri*δARID$_{H3}^{18.1}$ and UAS::*dri*δREKLES$_{\beta}^{12.1}$), the second chromosome (UAS::*dri*δARID$_{H3}^{19.2, 21.1, 28.1, 52.1, 17.2}$ and $^{54.1}$ and UAS::*dri*δREKLES$_{\beta}^{11.1, 30.1}$ and $^{29.1}$), and the third chromosome (UAS::*dri*δARID$_{H3}^{17.6,}$ $^{21.3, 36.1, 38.1, 43.1, 47.1}$ and $^{49.1}$ and UAS::*dri*δREKLES$_{\beta}^{12.3, 13.1, 33.1, 36.1}$ and $^{4.1}$). All lines were homozygous viable except UAS::*dri*δARID$_{H3}^{17.2}$ and $^{54.1}$ and UAS::*dri*δREKLES$_{\beta}^{29.1}$ and $^{4.1}$ which were maintained over the appropriate balancer chromosome (CyO or TM6B).

Expression of wild-type and mutant forms of DRI was confirmed by western analysis of eye imaginal disks expressing DRI under the control of the *GAL4*$^{GMR.PF}$ transgene (figure 4.5B).

4.7 The *dri* cDNA rescues *dri* when expressed in the endogenous *dri* pattern

Using the *dri*^{7} *P*-element allele to induced male recombination in the presence of a *P*[*GAL4*] insertion, the original insertion in *dri* was replaced with *P*[*GAL4*] (T. Shandala and R. Saint , unpublished results). The *dri*$^{P[GAL4]}$ insertion lines generated by this *P*-element replacement express GAL4 in patterns similar to the endogenous pattern of *dri* expression and fail to complement the original *dri*^{7} *P*-element allele (T. Shandala, pers. comm.). The *P*[UAS::*dri*$^{3.5}$] insertion was recombined onto an EMS generated *dri* protein null chromosome *dri*^{1}, generating *P*[UAS::*dri*$^{3.5}$] *dri*^{1}. The *dri*^{1} chromosome failed to complement the *P*[*GAL4*] *dri* replacement lines used to rescue *dri* function (table 1). However, when *P*[UAS::*dri*$^{3.5}$] was included, viability increased to 56.8-66.5% depending on the *P*[*GAL4*] used. As the *P*[UAS::*dri*$^{3.5}$] construct driven by the *dri*$^{P[GAL4]}$ GAL4 reporter failed to completely rescue viability, an arbitrary threshold for viability was defined. At a threshold for viability of 70%, no rescue is observed when the *lacZ* enhancer-trap *P*-element allele *dri*^{7} is used (table 1; at a 1% confidence level, failure to rescue by *dri*$^{P[lacZ]}$ is non-significant if the threshold for viability is reduced to 11%). Thus the *dri* cDNA, when expressed indirectly under the control of *dri* enhancers, rescues *dri* loss of function.

Table 1. Rescue of *dri* transheterozygous lethality by GAL4 driven *dri* expression.				
	UAS::$dri^{3.5}$ dri^{1}	CyO	percent of expected viability	probability of N°: viability > 70%
dri^{89} (*GAL4*)	36	137	62.4%	p=0.43
dri^{29F} (*GAL4*)	36	154	56.8%	p=0.15
dri^{139} (*GAL4*)	43	151	66.5%	p=0.70
dri^{7} (*lacZ*)	0	174	0.0%	$p<10^{-12}$

4.8 *dri*δARID$_{H3}$ acts antimorphically

It was reasoned that removal of DNA-binding activity from a multimeric DNA-binding protein might result in a dominant negative effect. To test this proposition, wild-type DRI and DRIδARID$_{H3}$ lacking DNA-binding activity was expressed in the wild-type *dri* pattern using the *GAL4*-expressing dri^{139} allele that was shown to allow rescue of *dri* lethality by UAS::*dri* transgenes. No flies were observed to carry both the UAS::*dri*δARID$_{H3}$ transgene and the dri^{139} *P*[*GAL4*] insertion, indicating that these genes are synthetically lethal (table 2A; an arbitrary threshold for viability of 10% was chosen for statistical analysis). UAS::*dri*δREKLES$_{\beta}$ transgenes were similarly tested, but showed no significant reduction in viability when compared to a strain in which *E. coli lacZ* was driven by the dri^{139} *P*[*GAL4*] insertion. Interestingly, expression of wild-type *dri* in the same system caused a reduction in viability, suggesting that either the level of *dri* expression is important or that dri^{139} *GAL4* expression is not restricted to the endogenous *dri* pattern. Note that the synthetic lethality caused by dri^{139} and UAS::*dri*δARID$_{H3}$ is not a result of a general interaction of UAS::*dri*δARID$_{H3}$ with a *dri P*-element insertion as the *lacZ*-expressing dri^{7} enhancer-trap allele shows no significant reduction in viability (table 2B).

Table 2A. Lethal function of *dri* expression constructs driven by *dri*::GAL4 .				
	dri^{139}	CyO	percent of expected viability	probability of N°: viability > 10%
UAS::*dri*δARID$_{H3}$$^{18.1}$	0	195	0%	p=0.001
UAS::*dri*δARID$_{H3}$$^{17.6}$	0	190	0%	p=0.002
UAS::*dri*δARID$_{H3}$$^{43.1}$	0	186	0%	p=0.002
UAS::*dri*δREKLES$_{\beta}$$^{12.3}$	73	130	71.9%	p=1
UAS::*dri*δREKLES$_{\beta}$$^{13.1}$	86	108	88.7%	p=1
UAS::*dri*δREKLES$_{\beta}$$^{33.1}$	183	170	103.7%	p=1
UAS::*dri*$^{3.5}$	28	121	37.6%	p=1
UAS::*dri*$^{4.7}$	40	154	41.2%	p=1
UAS::*lacZ*	117	139	91.4%	p=1

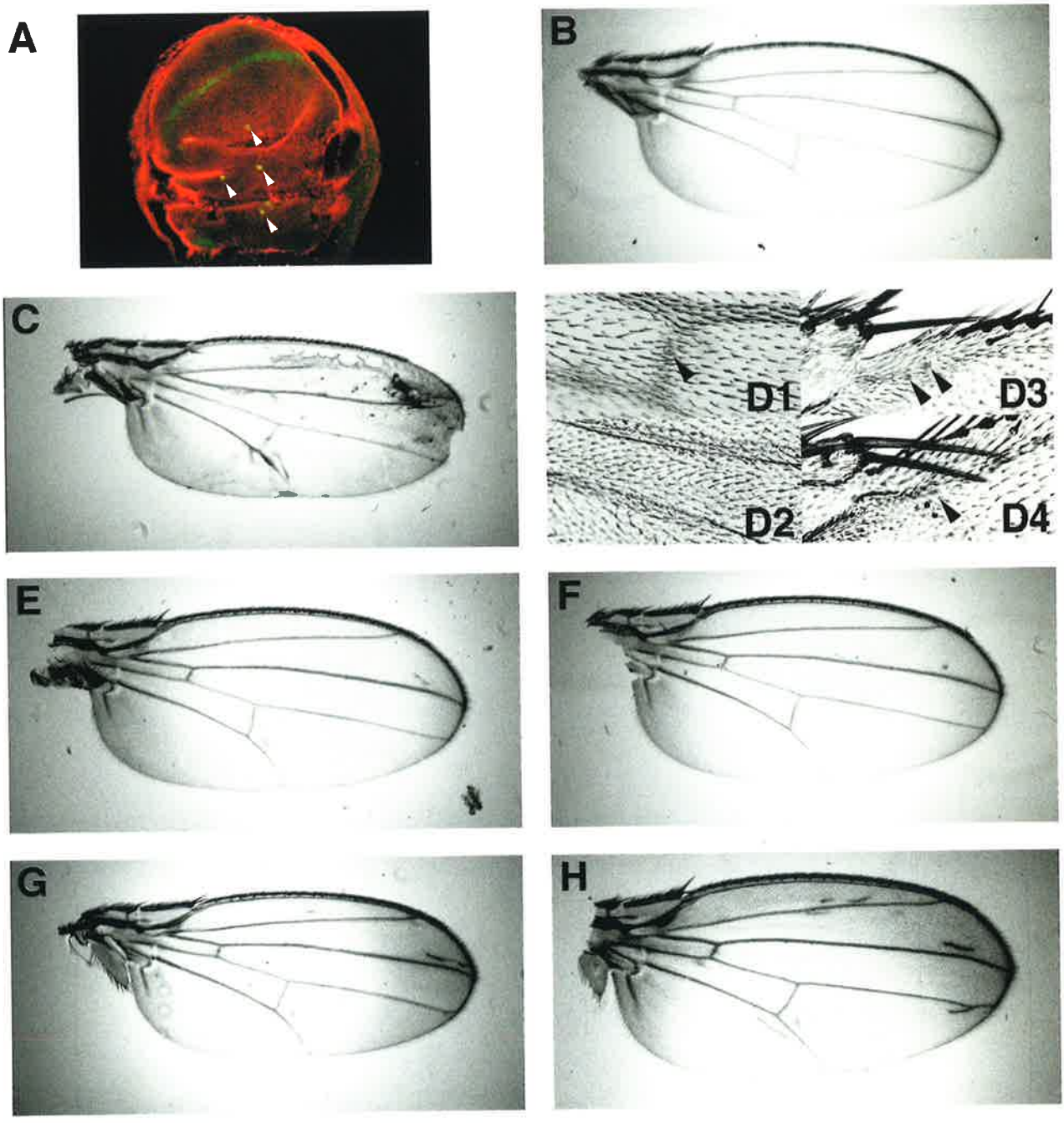
Table 2B. Interaction of *dri* expression constructs with the *dri*^{7} *P*-element insertion.				
	dri^{7}/+	CyO	percent of expected viability	probability of N°: no change in viability
UAS::*dri*δARID$_{H3}$$^{18.1}$	215	193	105.4%	p=0.28
UAS::*dri*δARID$_{H3}$$^{17.6}$	198	184	103.7%	p=0.47
UAS::*dri*δARID$_{H3}$$^{43.1}$	204	199	101.2%	p=0.80
UAS::*dri*δREKLES$_{\beta}$$^{12.3}$	147	151	98.7%	p=0.82
UAS::*dri*δREKLES$_{\beta}$$^{13.1}$	150	167	94.6%	p=0.34
UAS::*dri*δREKLES$_{\beta}$$^{33.1}$	171	196	93.2%	p=0.19
UAS::*dri*$^{3.5}$	210	173	109.7%	p=0.06
UAS::*dri*$^{4.7}$	174	177	99.1%	p=0.87

In order to confirm the antimorphic activity of *dri*δARID$_{H3}$, it was decide to examine the genetic interaction of an amorphic *dri* allele in the heterozygous state with expression of *dri*δARID$_{H3}$ in a non-essential tissue which normally expresses *dri*. Examination of larval *dri* expression patterns revealed that it is expressed in a subset of cells in the wing imaginal disk which express the homeodomain protein Cut (figure 4.7A). These Cut and DRI expressing cells are known to the precursors of a group of sense organs, known as the campaniform sensilla, located on the wing (Huang *et al.* 1991; Blochlinger *et al.* 1993). Additionally, DRI

Figure 4.7 — Effects of mutant and wild-type *dri* expression under $GAL4^{71B}$ control in the wing-blade anlage of the wing imaginal disk.

A. Confocal image of Cut (green) and DRI (red) expression in the larval wing imaginal disk. White arrow heads indicate the location of sense organ precursor cells expressing both Cut and DRI.

B-H. Photomicrographs of wings from females of the following genotypes: B. w^{1118}, C. P[UAS::*dri*δARID$_{H3}{}^{18.1}$]/w^{1118};+;$GAL4^{71B}$/+ (full wing), D1. as in B. and D2. as in C. showing magnification of the region around the anterior cross vein, D3 as in B and D4 as in C. showing magnification of the region around the twin campaniform sensilla of the margin (black arrow heads indicate the presence of campaniform sensilla), E-F. w^{1118};+;P[UAS::*dri*δARID$_{H3}{}^{17.6}$]/$GAL4^{71B}$, G. w^{1118};+;$GAL4^{71B}$ P[UAS::*dri*δARID$_{H3}{}^{17.6}$]/+ and H. w^{1118};dri^{1}/+;$GAL4^{71B}$ P[UAS::*dri*δARID$_{H3}{}^{17.6}$]/+.


A
B
C
D1
D3
D2
D4
E
F
G
H

appears to be expressed at low levels throughout the wing imaginal disk. Thus the wing was chosen as a target tissue for confirmation of *dri*δARID$_{H3}$ antimorphism.

UAS::*dri*δARID$_{H3}$ expression was targeted to the wing using the *GAL4*^{71B} enhancer-trap which expresses *GAL4* in the wing-blade anlage of the wing imaginal disk (Brand and Perrimon 1993). Expression of UAS::*dri*δARID$_{H3}$ under *GAL4*^{71B} control resulted in variable wing vein defects and losses of campaniform sensilla, in regions consistent with the endogenous expression pattern of DRI (figure 4.7B-H). These defects are consistent with effects observed in flies carrying unmarked somatic clones lacking functional *dri* (T. Shandala, pers. comm.). The severity of wing defects correlated with the level of protein expression from the trangenic constructs driven by *GAL4*$^{GMR.PF}$ in the larval eye imaginal disk (figures 4.7CEF and 4.5B). Expression of UAS::*dri* under *GAL4*^{71B}-control was lethal to the flies at 18°C and 25°C (0/130 and 0/163 respectively) suggesting that *GAL4*^{71B} expression is not confined to the wing imaginal disk. Thus it was not possible to examine the effect of *GAL4*^{71B} driven expression of wild-type *dri*.

It was reasoned that reducing the levels of endogenous *dri* would agonise the activity of *dri*δARID$_{H3}$ if the mutant were antimorphic. Thus, to establish whether *dri*δARID$_{H3}$ acts antimorphically, the dose of endogenous *dri* was halved by placing *GAL4*^{71B} and *P*[UAS::*dri*δARID$_{H3}$] in a *dri* heterozygous mutant background. A chromosome carrying both *GAL4*^{71B} and P[UAS::*dri*δARID$_{H3}^{17.6}$], a weakly expressing UAS::*dri*δARID$_{H3}$ transgene (figures 4.5B and 4.7E-F), was generated by recombination to facilitate the coordinate manipulation of the two genes. The wing phenotype generated by the genes in *cis* was slightly stronger than when they were arranged in *trans* (figure 4.7E-G). To test whether *GAL4*^{71B} *P*[UAS::*dri*δARID$_{H3}^{17.6}$] acts antimorphically in the wing, the *GAL4*^{71B} *P*[UAS::*dri*δARID$_{H3}^{17.6}$] chromosome was placed in a *dri*^{1} heterozygous background. Although the distribution of phenotype expression resulting form the *GAL4*^{71B} driven UAS::*dri*δARID$_{H3}$ expression in the *dri*^{1} heterozygous background overlapped that seen in the dri homozygous wild-type background, the severity of the phenotype in *dri*^{1} heterozygotes was greater than in the wild-type when comparing the most extreme cases of

each class (figure 4.7G-H). The variability of phenotypic severity observed to result from *GAL4*^{71B} driven UAS::*dri*δARID$_{H3}$ expression is consistent with the high degree of phenotypic variability seen in embryonic *dri* mutant studies (T. Shandala, pers. comm.)

glass (*gl*) regulates expression of *lacZ* in the enhancer-trap *P*-element insertion in *dri*^{8} (Treisman and Rubin 1996), suggesting that eye development would be a good system for analysis of DRI function in development. However, effects on eye development were not examined in detail as a GAL4-inducible wild-type *dri* construct is lethal at the pupal stage when expressed under *gl* control through the *GAL4*$^{GMR.PF}$ gene while the GAL4 inducible *dri*δARID$_{H3}$ causes no gross morphological defects (data not shown). The apparent paradox of lethality caused by GL/GMR driven *dri* expression is probably a result of ectopic expression as *GAL4*$^{GMR.PF}$ driven expression of *D-cbl* has been shown to cause defects in tissues other than the eye, indicating that activity of the GMR enhancer is not restricted to the eye (G. Hime, pers.comm.).

4.9 *dri* lacking the REKLESβ region is able to rescue *dri* function *in vivo*

In order to examine the *in vivo* role of the REKLESβ region of DRI, rescue experiments similar to those described for wild-type *dri* were performed. The third chromosome insertions, *P*[UAS::*dri*$^{4.7}$], *P*[UAS::*dri*δREKLESβ$^{13.1}$] and *P*[UAS::*dri*δREKLESβ$^{33.1}$] were homozygosed into the *dri*^{1} background, generating the lines *dri*^{1}/CyO;*P*[UAS::*dri*$^{4.7}$], *dri*^{1}/CyO;*P*[UAS::*dri*δREKLESβ$^{13.1}$] and *dri*^{1}/CyO;*P*[UAS::*dri*δREKLESβ$^{33.1}$]. When these lines were crossed to *dri*^{139}, viability was observed to be 56.2-61.5% indicating rescue of the transheterozygous *dri* lethality (table 3A). However, no rescue is observed when the *lacZ* enhancer-trap *P*-element allele *dri*^{7} is used in place of *dri*^{139} (table 3B), indicating that rescue is *dri*::*GAL4*-dependent. These results show that the REKLESβ region of DRI is not essential for DRI function during stages of embryonic and imaginal development rescued by the wild-type transgene.

Table 3A. Rescue of *dri* transheterozygous lethality by GAL4 driven *driδ* REKLES$_\beta$ expression.				
	dri^{139}/*dri*^{1}	*dri*^{1}/CyO	percent of expected viability	probability of N°: viability > 70%
UAS::*driδ* REKLES$_\beta$ $^{13.1}$	46	181	60.8%	p=0.27
UAS::*driδ* REKLES$_\beta$ $^{33.1}$	38	165	56.2%	p=0.12
UAS::*dri*$^{4.7}$	58	225	61.5%	p=0.26

Table 3B. Requirement for GAL4 by *dri* expression constructs for rescue.				
	dri^{7}/*dri*^{1}	*dri*^{1}/CyO	percent of expected viability	probability of N°: viability > 70%
UAS::*driδ* REKLES$_\beta$ $^{13.1}$	0	178	0%	$p<10^{-12}$
UAS::*driδ* REKLES$_\beta$ $^{33.1}$	0	193	0%	$p<10^{-13}$
UAS::*dri*$^{4.7}$	0	180	0%	$p<10^{-12}$

4.10 Discussion

The preceding results chapters have described the identification of *dri* mutants and conserved domains within the eARID group of proteins, and have described an examination of the feasibility of genomic rescue of *dri* mutant lethality . The experiments described in this chapter examine the roles of conserved domains in the DRI eARID protein *in vivo*. Structural prediction analysis of the conserved eARID and of the REKLES domain of DRI was used to identify targets for deletion mutagenesis in order to examine the roles of the conserved domains. Two highly conserved regions within DRI were chosen for mutagenesis, the third ARID helix of the eARID and the REKLES$_\beta$ region of the REKLES domain.

The *in vitro* functions of the conserved domains were examined. Gel mobility shift assays of NP_6 DNA-binding by the DRI eARID showed that the $ARID_{H3}$ region is required for NP_6 DNA-binding. Western analysis of the REKLES domain using radio-labelled DRI as a probe showed that the REKLES domain is capable of mediating self-association, consistent with the finding of Herrscher *et al*. (1995) that the region of Bright corresponding to the REKLES domain confers tetramerisation activity. However, deletion of the REKLES$_\beta$ region

did not abolish self-association, indicating that this region is not required for this activity. The presence of an invariant tyrosine residue within the REKLES$_\beta$ region suggested that the motif may be the target of a kinase. Examination of phosphorylation states of the REKLES domain failed to show differential phosphorylation of the mutant and wild-type REKLES domains in the *in vitro* translated protein. This result does not preclude the phosphorylation of this site *in vivo*, as the *in vitro* translation system used was derived from rabbit reticulocytes and thus may not possess the specific kinase activity required. A role for the REKLES$_\beta$ region has been suggested by preliminary observations that a GRO-DRI interaction, identified by Valentine *et al.* (1998), is abolished with DRI lacking the REKLES$_\beta$ region (A. Courey, pers. comm.).

The observations that the REKLES domain is able to mediate self-association and that the REKLES$_\beta$ region is required for GRO-DRI interaction, but not for self-association, suggests that the REKLES domain is in fact two domains. This assertion is supported by the observation that the *Drosophila* DRI and *C. elegans* T23D8.8 REKLES domains possess insertions amino-terminal of the REKLES$_\beta$ region. On the basis of these observations the two regions of the REKLES domain split by the insertion are now named the REKLES$_\alpha$ (amino-terminal region) and REKLES$_\beta$ (carboxy-terminal region) domains. To test the possibility that the REKLES$_\alpha$ domain does in fact mediate self-association, DRI with deletions spanning the REKLES$_\alpha$ signature (amino acids 463-473, MPNLVKQEIES) should be examined for self-association activity.

The identification of the *P*-element insertions *l(2)02535* and *l(2)05096* as *dri* alleles in chapter 2 was confirmed by the finding that EMS induced mutations failing to complement these insertions show loss of anti-DRI immunoreactivity as assayed by immunohistochemistry (T. Shandala, pers. comm.) and western analysis. This result was confirmed by showing that lethality due to mutations failing to complement the two *P*-element insertions was reduced by expression of the *dri* cDNA in the endogenous *dri* pattern.

A requirement for the ARID$_{H3}$ and REKLES$_\beta$ regions for DRI function *in vivo* was assayed by expression of DRI lacking these regions. As predicted on the basis of other studies

of multimeric DNA-binding proteins (Norton *et al.* 1998), UAS::*dri*δARID$_{H3}$ was found to act antimorphically. This antimorphic activity is presumably a result of the incorporation of DRIδARID$_{H3}$ into a DRI-multimer complex, compromising the DNA-binding activity of the complex. No affects were observed when UAS::*dri*δREKLES$_{\beta}$ was expressed in the endogenous *dri* pattern.

The ability of wild-type and REKLES$_{\beta}$ mutant *dri* transgenes to complement lack of zygotic *dri* activity was examined by expressing the transgenes in the endogenous *dri* pattern. The high level of rescue observed with UAS::*dri*δREKLES$_{\beta}$ transgenes indicates that the REKLES$_{\beta}$ deletion does not result in a significant loss of zygotic *dri* function. However, the reduction in viability observed when wild-type *dri* transgenes were expressed in the endogenous *dri* pattern under the control of the *dri*^{139} *P*[*GAL4*] transgene was not as marked with UAS::*dri*δREKLES$_{\beta}$ transgenes. These results indicate that DRIδREKLES$_{\beta}$ lacks functions present in the wild-type protein, although these functions have only presented themselves in artificial situations thus far. Further analysis of the requirement for the REKLES$_{\beta}$ domain *in vivo* is necessary to determine functional roles for this domain in DRI function.

The antimorphic activity of the UAS::*dri*δARID$_{H3}$ suggests a test for the role of the proposed REKLES$_{\alpha}$ domain. Abolition of a self-association domain would be expected to relieve the antimorphism resulting from loss of DNA-binding. Thus a *dri* mutant transgene lacking both the ARID$_{H3}$ region and the REKLES$_{\alpha}$ domain would be expected to lack antimorphic activity.

Chapter 5: General discussion

Don't bite my finger—look where it's pointing. - Warren S. McCulloch

5.1 Introduction

This thesis describes the characterisation of the *Drosophila dead ringer* (*dri*) gene and conserved domains within the DRI-like, eARID proteins. DRI has been shown to possess a strongly conserved DNA-binding domain shared with a number of other proteins (Gregory *et al.* 1996; Herrscher *et al.* 1995; Kortschak *et al.* 1998) and to act as a transcription factor in conjunction with the Groucho (GRO) corepressor to repress ventral expression of *zerknüllt* (Valentine *et al.* 1998). The previous failure to find conserved domains outside the eARID DNA-binding domain raised the question of how the interactions of DRI with GRO and other possible interactors may be mediated (Gregory *et al.* 1996; Herrscher *et al.* 1995). Evolutionary analysis of DRI orthologues from vertebrate species was used to identify conserved domains outside the eARID. The conserved domains in the DRI protein, including the previously described DNA-binding domain, were characterised functionally *in vitro*. The *in vivo* roles of conserved domains in DRI were partially characterised by replacement of endogenous DRI with specifically mutated *dri* transgenes. Endogenous *dri* function was removed by loss of function mutations in *dri*, identified through the analysis of the genomic structure of *dri* and *P*-element insertions known to be located in the vicinity of *dri* (Kortschak 1993; Gregory *et al.* 1996).

5.2 Structure of the *dri* locus

The primary aims of characterising the *dri* locus were to examine the potential for generation of mutant genomic transgenesis constructs and the identification of *dri* mutations. The large size of the *dri* transcription unit and the uncertainty of the location of *cis*-regulatory elements made specific mutation and reintroduction of genomic construct unfeasible due to

technical constraints of *Drosophila* transgenesis and uncertainties in the distribution of required regulatory elements. However, two non-complementing *P*-element insertions in the vicinity of the *dri* transcription initiation site were identified (dri^7 and dri^8). Expression of the enhancer-trap reporter-gene from the *P*-elements has been shown to be coincident with the endogenous *dri* expression pattern (Gregory 1996). Two homozygous lethal EMS-generated alleles (dri^1 and dri^2) failing to complement the *P*-element insertions show loss of anti-DRI antibody staining as determined by Western analysis, and immunohistochemistry on homozygous embryos and FLP-FRT generated somatic clones (T. Shandala, pers. comm.). Further, the *dri* cDNA under the control of the GAL4 UAS system is able to rescue the transheterozygous lethality of dri^1 over the $dri^{P[GAL4]}$ GAL4 expressing enhancer-trap insertions. Thus, the dri^7 and dri^8 *P*-element insertions can be concluded to disrupt *dri* function.

Genomic characterisation of the *dri* locus revealed that the structure of the *dri* gene is complex. The transcription unit is broken into 12 exons and spans more than 20kb. The complex structure of the transcription unit in conjunction with the observed complexity of the *dri* expression pattern throughout development is suggestive of a complex regulatory system involving a number of factors. One of these factors has been identified previously, namely the Glass (GL) zinc-finger protein (Treisman and Rubin 1996). Enhancer-trap reporter-gene expression from dri^8 was shown to be *gl*-dependent in the eye. The complexity of *dri* regulation is supported by the observation that *P[GAL4]* insertion lines targeted to *dri* by *P*-element replacement using the dri^7 *P*-element insertion show a variety of enhancer-trap reporter-gene expression patterns, overlapping with the endogenous *dri* pattern (T. Shandala, pers. comm.). Possible methods for determining factors involved in the regulation of *dri* expression will be discussed below.

5.3 A novel domain specifically associated with the eARID

A question was raised in the introduction of this thesis of how DRI mediates effector function when bound to DNA. The domain model of transcription factor structure proposed by Ptashne (1988) suggests that DRI-like proteins would possess a second domain, outside the eARID, capable of interacting with other proteins and mediating functions necessary for *trans*-regulation. To test this hypothesis against the counter hypothesis that the eARID is an integrated DNA-binding and effector domain, orthologues of the known *dri*-like genes were identified and the sequences of their protein products compared. In addition to the eARID, a second region of conservation was brought to light, named the REKLES motif. Extensive searches of sequence databases failed to find the REKLES motif, in full or in part, associated with any non-eARID protein, indicating that the REKLES domain is specifically associated with the eARID. The specific association of the REKLES domain with the eARID suggests that DRI-like (eARID) protein function is mediated through activities associated with the REKLES domain, the eARID/REKLES domain complex acting as an integrated DNA-binding and effector system in accordance with the model proposed by Ptashne (1988).

As no significant similarities were identified between the REKLES domains of the known eARID proteins and any other published protein sequences, the function of the REKLES domains could not be inferred on this basis. However, a region of the mouse Bright protein corresponding to the REKLES domain in that protein has been shown to be required for tetramerisation (Herrscher *et al*. 1995), suggesting that the REKLES domain may be responsible for self-association. The division of the DRI REKLES domain into two conserved regions by the insertion of 237 amino acids is suggestive that the REKLES domain may mediate more than one function. Investigations of functions of REKLES domain regions, both in experiments described in chapter 4 of this thesis and work by others, will be discussed below.

5.4 Phylogeny of *dri* genes

The specific association of an effector domain with a DNA-binding domain in such a small set of proteins as seen in the eARID proteins raises questions about how such a domain system arose evolutionarily. Phylogenetic analysis of the known eARID/REKLES domain proteins suggests some partial answers to this question. Thus far no eARID proteins have been identified in phyla outside the metazoan lineage while two ARID proteins have been identified outside the metazoan lineages, the *S. cerevisiae* Swi1 and YM8010.06 proteins, suggesting that the eARID/REKLES system is a specifically metazoan factor.

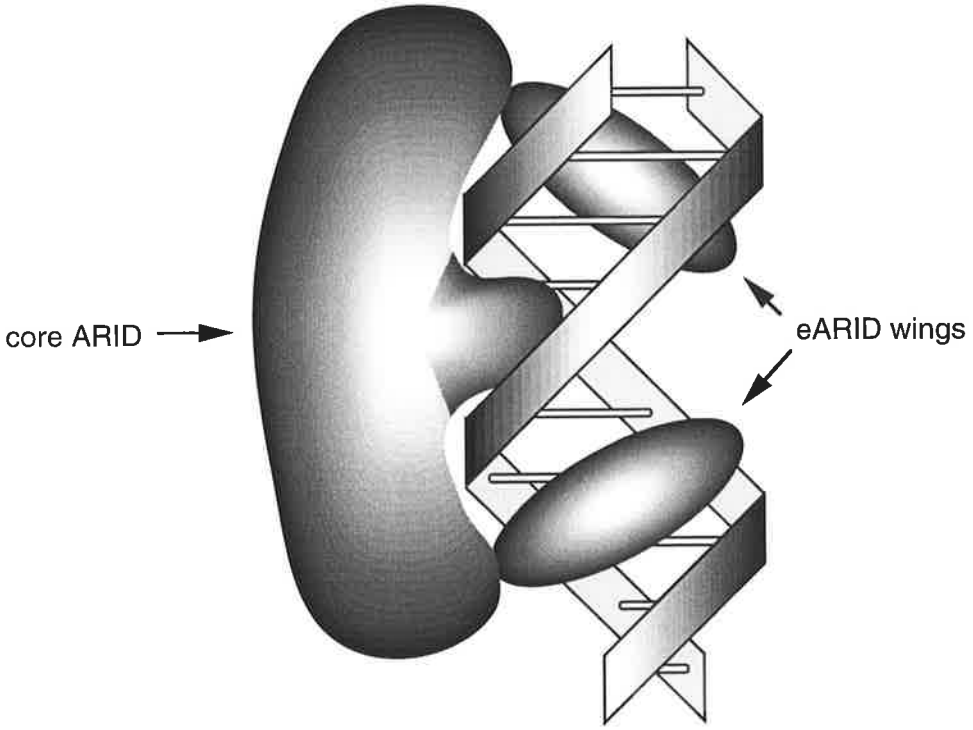
The evolution of the eARID appears to have resulted from the acquisition of flanking regions (wings) conferring site specificity by the ARID, shown to be a non-site specific DNA-binding domain (Huang *et al*. 1996, K. Itakura pers. comm.; Treisman *et al*. 1997, J. Treisman, pers. comm.). If this is the case it is likely that the core of the eARID binds DNA through the minor groove (Herrscher *et al*. 1995), a surface presenting few cues for a protein to identify a specific target, with the wings wrapping around the DNA helix to make contacts with specific groups in the major groove (figure 5.1A). This is consistent with the observations that the SWI/SNF complex binds through the minor groove, possibly via the Swi1 ARID (Quinn *et al*. 1996), and that the DRI core ARID is unable to bind the NP oligonucleotide (Gregory 1996). This proposition can be tested by the replacement of the core ARID of an eARID protein with a heterologous core ARID from a non-eARID protein. The prediction made by the flanking region hypothesis is that the chimeric eARID should bind DNA with similar specificity to that of the original eARID. A further prediction of the hypothesis is that specificity of binding by the eARID is defined by the wings. A test of this would be to examine the binding specificities of the two *D. rerio* Dri proteins as they diverge in the amino-terminal wing. More specific tests would involved generation of specific mutations in the wings, which should modify binding specificity of the eARID. A further test for eARID wing function would be to examine the potential for co-occupancy of the NP site by DRI and EN, as the flanking region hypothesis predicts that DRI binding would preclude EN binding due to obstruction of the major groove by the eARID wings.

Figure 5.1 — Proposed models of eARID-DNA interaction and DRI domain structure.

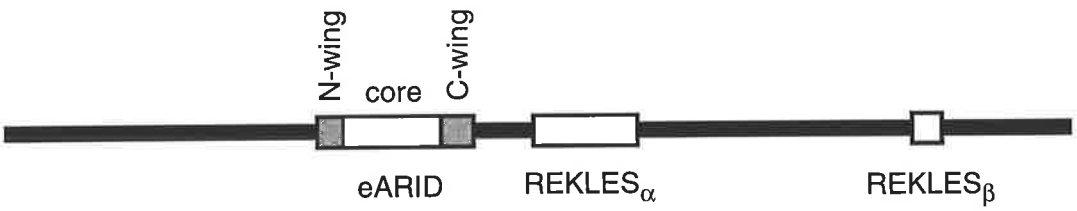
A. Model of eARID DNA-binding conformation based on experiments described in this thesis and evidence from other sources, outlined in the text. The eARID is proposed to bind DNA through the interaction of the core ARID with the minor groove of the DNA double helix in conjunction with interactions of the eARID-specific "wings" with the major groove which provide site-specificity for eARID DNA-binding.

B. Proposed primary structure of domain organisation within the DRI protein, showing the relative positions of the eARID (amino acids 263-399, core 287-369), and the REKLES$_\alpha$ (amino acids 463-541) and REKLES$_\beta$ (amino acids 779-807) domains. Abolition of eARID DNA-binding activity results in antimorphic activity *in vivo*. Deletion of the majority of the REKLES$_\beta$ domain results in context dependent hypomorphism *in vivo*. The *in vivo* activity of the REKLES$_\alpha$ domain has not been examined in *Drosophila*. However, deletion of regions of the REKLES$_\alpha$ domain and the REKLES$_\beta$ domain in the mouse homologue, Bright, abolishes tetramerisation *in vitro* (Herrscher *et al*. 1995). This result, in conjunction with the observation that deletion of the DRI REKLES$_\beta$ domain does not prevent DRI self-association *in vitro*, suggests that the REKLES$_\alpha$ domain mediates this interaction.

A


core ARID
eARID wings

B


N-wing
core
C-wing
eARID
REKLES$_\alpha$
REKLES$_\beta$

The acquisition of the mediatory function provided by the REKLES domain appears to have coincided quite closely with the evolution of the eARID as the eARID and REKLES domains do not exist independently, although this observation should be regarded with caution as only six eARID/REKLES domain proteins are known. Thus a likely scenario is that the evolution of sequence-specific binding enabled the acquisition of a mediatory function. The converse scenario, that the REKLES domain preceded the evolution of the eARID seems unlikely as the possession of transcriptional effector domains, as is proposed to be the function of the REKLES domain, by a non-site specific DNA-binding protein would be likely to be deleterious to the organism. The evolutionary origin of the REKLES domain is less clear than that of the eARID as there is no apparent ancestral form of the REKLES domain.

The use of comparative genomics as illustrated by the identification of a novel domain in the eARID group of proteins suggests a method for identifying regulatory factors controlling the complex *dri* locus. Comparison of sequences in the Drosophila *bw* gene and the vertebrate β-globin locus control region has been used to identify *cis*-regulatory elements mediating interactions with *trans*-acting factors (Martin-Morris and Henikoff 1995; Shelton *et al.* 1997). Comparison of the upstream region of the *dri* locus with the upstream region of *dril1* showed no significant sequence conservation, probably due to the long period of divergence between the two genes. However, identification of *cis*-elements controlling the regulation of *dri* should be feasible by comparison of the *D. melanogaster* and *D. virilis* sequences, as has been done with the *bw* gene (Martin-Morris and Henikoff 1995).

5.5 Functional characterisation of DRI domains *in vivo*

The identification of the conserved DNA-binding eARID in the DRI and Bright transcription factors suggests that this domain is responsible for the targeting of DRI and Bright proteins to the DNA enabling interactions with other proteins on the DNA or the recruitment of other factors. However, the role of the ARID *in vivo* has been questioned by

the observation that the Swi1 protein does not require the ARID for SWI/SNF function (C. Peterson, pers. comm.). To determine whether the ARID is required for DRI function, a mutant form of the protein lacking a highly conserved region of the ARID was generated, namely DRIδARID$_{H3}$. The eARID encoded by *dri*δARID$_{H3}$ showed no detectable DNA-binding activity in electrophoretic mobility shift assays under conditions in which the DRI eARID exhibits DNA-binding activity, indicating that DRIδARID$_{H3}$ lacks ARID-mediated DNA binding activity. As no DNA-binding experiment was performed with full length DRIδARID$_{H3}$ the presence of other DNA-binding domains cannot be ruled out.

As discussed above, sequence comparison of the six known eARID proteins identified a novel conserved motif, the REKLES motif. Coincidence of the REKLES motif in the Bright protein with the Bright tetramerisation domain, identified functionally (Herrscher *et al*. 1995), suggests a role of self-association for the REKLES domain. The proposed self-association activity of the REKLES domain was confirmed by Western analysis of the DRI REKLES domain probed with radio-labelled full length DRI. However, deletion of the REKLES$_\beta$ region, the most conserved region of the REKLES domain, did not abolish self-association, indicating that this region is not required for self-association. However, work by others has suggested that the REKLES$_\beta$ region is required for interaction of DRI with the GRO corepressor (A. Courey, pers. comm.). The observation of REKLES$_\beta$-independent self-association and REKLES$_\beta$-dependent GRO interaction supports the hypothesis stated above that the REKLES domain is actually two domains, separated by a large insertion in *Drosophila* DRI; the amino-terminal REKLES$_\alpha$ domain mediates self-interaction and the REKLES$_\beta$ domain mediates GRO binding (figure 5.1B). Confirmation of this hypothesis will require the generation of *dri*δREKLES$_\alpha$ mutants and analysis of *in vitro* self-association activity.

Comparison of the *D. rerio* proteins indicated differences between Dri1 and Dri2 in the REKLES$_\alpha$ domain. The function of the REKLES$_\alpha$ domain proposed here is to mediate self-association. Thus differences between the two Dri REKLES$_\alpha$ domains may impart specificity to ensure that the paralogues do not cross-associate.

As DRI has been shown to self-associate, it was reasoned that expression of DRI lacking DNA-binding activity might act antimorphically by poisoning the DNA-binding activity of DRI complexes in which it is contained, in the same manner as the vertebrate Id helix-loop-helix protein prevents MyoD DNA-binding (Norton *et al*. 1998). The observation that splice variants of *bright* mRNA lacking the tetramerisation domain act antimorphically (P. Tucker, pers. comm.) suggested that the REKLES domain deletions might also act antimorphically. To test the possibility of antimorphic activity of the deletion constructs, *dri*δARID$_{H3}$ and *dri*δREKLES$_{\beta}$ were expressed in the endogenous *dri* pattern using the *dri*$^{P[GAL4]139}$ enhancer-trap used to drive wild-type *dri* expression to give rescue (described above). Expression of *dri*δARID$_{H3}$ resulted in complete lethality, while *dri*δREKLES$_{\beta}$ expression showed no consistent reduction in viability. The failure to observe results equivelent to those seen with Bright deletions may be due to the fact that the *dri*δREKLES$_{\beta}$ deletion construct does not lack the entire region covered by the deletions in Bright resulting from the variant splicing of exons.

The lethality of *dri*δARID$_{H3}$ when expressed in the endogenous *dri* pattern suggested that the UAS::*dri*δARID$_{H3}$ transgene acts antimorphically. To confirm this antimorphic activity a non-essential model tissue was examined for enhancement of the *dri*δARID$_{H3}$ by reduction of endogenous *dri* levels. The adult wing is not essential for viability and larval tissues giving rise to the wing express *dri*. Thus the wing was chosen as an appropriate tissue for study. Expression of UAS::*dri*δARID$_{H3}$ transgenes in imaginal wing-blade anlag resulted in venation and sense organ defects. The severity of the defects correlated with the levels of DRIδARID$_{H3}$ as assayed by western analysis. Consistent with the hypothesis that UAS::*dri*δARID$_{H3}$ transgenes act antimorphically, reduction in the dose of endogenous dri resulted in an increase in the severity of wing defects.

The *in vivo* expression of *dri*δARID$_{H3}$ in the endogenous *dri* pattern and specifically in the wing imaginal disks indicates that the core ARID is essential for DRI function and that DRI lacking ARID function acts antimorphically.

Expression of *dri*δREKLESβ in the endogenous *dri* pattern was found to be able to rescue lethality caused by *dri* loss of function mutations, nearly to the same extent as the wild-type transgene. However, the lethality caused by ectopic expression of wild-type *dri*, or the expression of dri under the control of *dri* P[GAL4] enhancer-trap insertions was not reiterated with *dri*δREKLESβ in the same systems, indicating that the DRIδREKLESβ product does not possess the complete set of DRI activities. The ability of UAS::*dri*δREKLESβ to rescue *dri* lethality indicates that REKLESβ domain mediated functions are not required during the stages of development rescued by the UAS transgenes, that is, zygotic *dri* activity essential for viability. Thus, REKLESβ domain deletion results in a context-dependent hypomorphism with respect to *dri* function. Currently, REKLESβ domain mediated functions are only revealed in artificial situations involving ectopic expression. However, the possibility that the REKLESβ domain is responsible for mediating the DRI-GRO interaction offers the possibility that replacement of maternal wild-type *dri* with *dri*δREKLESβ will disrupt *zerknüllt* (*zen*) expression, a target of *dri* regulation (Valentine *et al*. 1998).

Groucho-related proteins form a family of transcriptional corepressors (Fisher and Caudy 1998). These proteins function as non-DNA-binding repressors of transcription which act by interacting with specific subsets of DNA-binding transcription factors, including DRI in an interaction which has been proposed to mediate repression of the *zen* gene (Valentine *et al.* 1998). GRO mediated repression in the terminal regions of the Drosophila embryos has been shown to be antagonised by the Torso (TOR) protein tyrosine kinase signalling pathway (Paroush *et al.* 1997). However, the target of TOR kinase activity has not been identified. With the suggestion that the REKLESβ domain is responsible for DRI-GRO interaction and the observation that the REKLESβ domain possesses an invariant tyrosine residue, it is tempting to speculate that the target of TOR kinase activity is the REKLESβ domain of DRI and that TOR-dependent phosphorylation of the invariant tyrosine results in disruption of the DRI-GRO interaction. An interaction of DRI with GRO regulated by phosphorylation would also explain the differential repression of *zen* within the ubiquitous expression domain of *dri*

seen at the stage at which *dri* acts in *zen* regulation (Gregory *et al*. 1996; Valentine *et al*. 1998). The proposition that TOR phosphorylates DRI can be tested by immunoprecipitation of DRI using anti-DRI antiserum and detection of phosphotyrosine with anti-phosphotyrosine antibodies. Confirmation of the specificity of the TOR kinase activity for the REKLES$_\beta$ domain will require the generation of epitope-tagged, wild-type and REKLES$_\beta$ domain deletion constructs to enable the specific isolation of protein lacking the proposed target of TOR activity.

One of the points raised in the introduction was the role that transcription factors may play in chromatin regulation. The self-association activity demonstrated for DRI and Bright suggests that they may act as protein bridges between homologous chromosomes in a similar manner to that observed with Z (Pirrotta 1990). Indeed, given that Bright, and probably DRI, forms a tetramer, and given the likely arrangement of subunits in a tetrameric complex, the only way to allow all four eARID DNA-binding domains to be occupied would be to invoke either DNA looping or associations between *trans*-binding sites (figure 5.2).

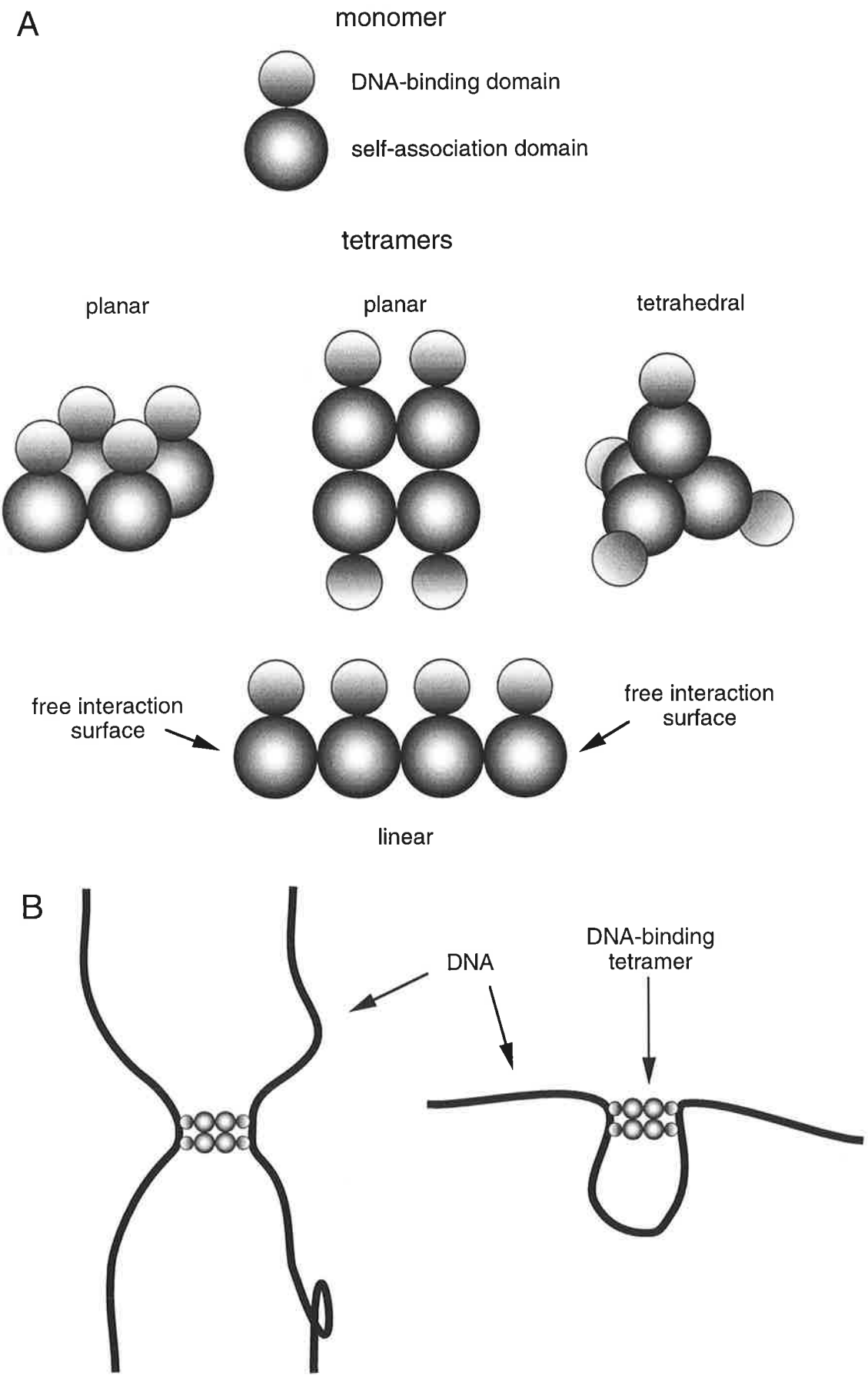
5.6 *dri* mutants as tools for developmental systems analysis

The identification of the antimorphic *dri*δARID$_{H3}$ activity of transgenic expression constructs suggests a potential route to investigation of *dri* function in development. *dri* mutations have highly pleiotropic effects in embryonic development and *dri* has been shown to influence a number of imaginal developmental processes (T. Shandala, R. D. Kortschak and R. Saint, unpublished results; G. Rubin, pers. comm.). The high degree of pleiotropism and the embryonic lethality of *dri* mutations have necessitated the use of random clonal analysis to investigate the function of *dri* in specific tissues and in postembryonic development. By using the antimorphic activity of UAS::*dri*δARID$_{H3}$ expressed under the control of tissue-specific *GAL4*-expressing enhancer-traps, the role of *dri* in these tissues can be examined.

Figure 5.2 — Implications of putative tetrameric quaternary structure of DRI and possible involvement of DRI in chromatin architecture.

A. Arrangements of subunits within a homotetrameric complex. If all four subunits are identical, only planar or tetrahedral tetramers are possible as the terminal subunits in a linear complex possess interaction surfaces that are not occupied by interaction with other subunits. Thus the linear complex would be expected to exists as a polymer unless a cap subunit were able to prevent further polymerisation.

B. Implication of a tetrameric quaternary structure for DRI. If DRI exists as a tetrameric complex, it is not possible for all four DNA-binding domains to be occupied DNA sites in a *cis* configuration unless the linker region between the DNA-binding domain and the self-association domain is flexible enough to allow folding of the protein on itself. This suggests that a tetramer would participate in the association of *trans* DNA molecules or in the looping of DNA.


A
monomer
DNA-binding domain
self-association domain
tetramers
planar
planar
tetrahedral
free interaction surface
free interaction surface
linear
B
DNA
DNA-binding tetramer

5.7 Summary

Gene expression and the maintenance of genetic material in the eukaryotic nucleus are regulated by a large array of interacting factors. At the simplest level, gene expression is controlled by the action of DNA-binding transcription factors that modulate the recruitment of RNA polymerases in a similar manner to that observed in prokaryotic systems. In eukaryotes, however, additional layers of complexity exist which serve to modify the accessibility of the DNA binding-sites to transcription factors. These higher order regulatory systems presumably allow a greater degree of control of gene expression in the larger genomes of eukaryotes and allow the genome to be efficiently packed within the nucleus. The existence of layers of gene regulatory control systems suggests that a hierarchy of gene regulatory systems acts in eukaryotes. However, this notion appears not to be the case as each layer appears to effect the activity of other layers, resulting in an entangled network of interactions leading to the appropriate regulation of gene expression. DNA-binding transcription factors constitute core components of this network, mediating interactions between the DNA, RNA polymerases and the higher order regulatory systems, through their DNA-binding and effector domains.

A number of members of the ARID family of DNA-binding proteins appear to be involved in aspects of chromatin regulation and control of gene expression. Members of a subgroup of the ARID family, the metazoan specific eARID group, have been shown to possess transcription factor activity. Evolutionary analysis of the eARID group of proteins has been used to identify conserved domains outside the DNA-binding eARID which might be involved in interactions with other factors. Two novel motifs were identified, the REKLES$_\alpha$ and REKLES$_\beta$ domains, which are proposed to mediate self-association and GRO corepressor binding respectively. Analysis of the *in vivo* function of the REKLES$_\beta$ domain showed that it is not required for rescue of zygotic *dri* function, but is required for some aspects of *dri* activity in artificial systems. Further work will be required to confirm the molecular activities of the REKLES$_\alpha$ and REKLES$_\beta$ domains and to examine the roles of these activities *in vivo*. Deletion of a highly conserved region of the eARID DNA-binding

domain resulted in the generation of an antimorphic transgene, indicating an essential role for the eARID in DRI function and suggesting that DRI acts as a multimer *in vivo*. The generation of the antimorphic UAS::*dri*δARID$_{H3}$ transgene will provide an invaluable tool for the examination of the developmental role of *dri* during post embryonic development and the development of specific tissues in the embryo.

6. Methods and Materials

Do not infest your mind with beating on the strangeness of this business. - William Shakespeare

6.1 Abbreviations

A_{260}: optical absorbance at the wavelength 260nm
A_{600}: optical absorbance at the wavelength 600nm
AED: after egg deposition
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 pairs
CIP: calf intestinal alkaline phosphatase
DMSO: dimethylsulphoxide
DNA: deoxyribonucleic acid
DNaseI: deoxyribonuclease I
dNTP: deoxyribonucleoside triphosphate
DTE: dithioerythritol
DTT: dithiothreitol
EDTA: ethylenediaminetetraacetic acid
GST: glutathion-S-transferase
hpf: hours post fertilisation
IPTG: isopropyl ß-D-thiogalactopyranoside
kb: kilobase pairs
kDa: kilodaltons
NBT: 4-nitro blue tetrazolium chloride
PAGE: polyacrylamide gel electrophoresis
PEG: polyethylene glycol
PSB: phage storage buffer
rATP: ribo-adenosine triphosphate
RNA: ribonucleic acid
RNaseA: ribonuclease A
SDS: sodium dodecyl sulphate
TEMED: N,N,N',N-tetramethylenediamine
U: unit of enzyme activity
UV: ultraviolet light

6.2 Materials

6.2.1 Chemicals and Reagents

All chemicals were obtained from departmental stocks and were of analytical grade or the highest purity available.

Most chemicals and reagents were obtained from a range of suppliers, the major sources of the more important chemicals and reagents are listed below.

rATP, dNTPs, DTT, ethidium bromide and IPTG: Sigma

Acrylamide, chloroform and DMSO: BDH, Ltd.

Agarose (ultra pure), Ammonium persulphate, bisacrylamide and TEMED: Bio-Rad

BCIG, BCIP, NBT and glycogen: Boehringer Mannheim

Agarose (molecular biology grade): Promega

Urea: Merck

Phenol: Wako pure chemical industries, Ltd.

6.2.2 Enzymes

Enzymes were obtained from the following sources:

Restriction endonucleases: Pharmacia, New England Biolabs and Boehringer Mannheim.

T4 DNA Ligase: Boehringer Mannheim

Calf Intestinal Alkaline Phosphatase: Boehringer Mannheim

Ribonuclease A: Sigma

Proteinase K: Boehringer Mannheim

Lysozyme: Sigma.

6.2.3 Kits

DIG RNA labelling kit:Boehringer Mannheim

Gene Clean kit: Bio 101

HiTrap Protein G Column: Pharmacia

Megaprime DNA radiolabelling kit: Amersham

Qiagen DNA prep kit: Qiagen

QIAprep Spin: Qiagen

QIAquick Gel Extraction kit: Qiagen

QuikChange™ Site-Directed Mutagenesis Kit: Stratagene

Sequenase double stranded DNA sequencing kit: USB

TNT™ Coupled Reticulocyte Lysate System: Promega

6.2.4 Radionucleotides

α-^{32}P-dATP (specific activity, 3000Ci/mmole; concentration, 10mCi/ml): Amersham

6.2.5 Antibiotics

Ampicillin and kanamycin were obtained from Sigma.

6.2.6 Molecular weight standards

Phage λ (c1857 *ind*1 *Sam*7, New England Biolabs) DNA restricted with *Bst*EII and *Sal*I was used as DNA molecular weight standards. Prestained molecular weight markers (Gibco BRL and BioRad) were used as protein molecular weight standards.

6.2.7 Oligonucleotides

Vector complementary oligonucleotides:

P-element primer:	5' CGA CGG GAC CAC CTT ATG 3'
KS primer:	5' CGA GGT CGA CGG TAT CG 3'
SK primer:	5' TCT AGA ACT AGT GGA TC 3'
T3 primer:	5' ATT AAC CCT CAC TAA AG 3'
T7 primer:	5' AAT ACG ACT CAC TAT AG 3'

dead ringer complementary oligonucleotides:

18.7kb:	5' GGA ATT CAT CCC CCT ATT CAC ATT TCA G 3'
check:	5' TTC CCA CTG CCC CAC AAC TC 3'
conexon 3':	5' CGG GAT CCT ATG AGC GCC CCT TGA AAA AC 3'
conexon 5':	5' GGA ATT CGT ATG TCG TTA TGT TGT GC 3'
Dri1 5':	5' AAA GGC CAC AGC GTT GT 3'
dri37:	5' GAC AGC TCT ATG AAA TC 3'
dri47:	5' TAG TCT CGC TCT GTG ACA 3'
dri48:	5' CAA CCA GAT GCC GAT GAC G 3'
dri5' genomic:	5' TTA CTG CTC CAC ACG CTC ACA 3'
dri5' genomic2:	5' GGT CGG ATA ACG GCT TTT CAA 3'
dri5' genomic3:	5' TCT TAA GCG CAC TCT CCT GTT 3'
dri53:	5' AGG CCA AGA AGG AGC AGG 3'
dri54:	5' TTC TCA ATG GCC GAG TCG 3'
driA:	5' GAT CTC CTG CTT GAC CA 3'
driB:	5' TGC TGC GGC GAG GTG TG 3'
driCR:	5' TGT CAT CCT GGT CCA CG 3'
driD:	5' CAC CAC CAC CAT CAG GC 3'
driE2R:	5' GGC ATC GGC GGA CTA CAC AGC 3'
driE3R:	5' GAA CGG ACG ACG AGA TG 3'
driE4:	5' CTC AGG GGC AGC TCG GTC 3'
driE4R:	5' GAC CGA GCT GCC CCT GAG 3'
driE:	5' CAA TGT GCC GCT GTC AC 3'
driF:	5' GAC AAG GAA AGG AAT ACC 3'
driRT:	5' CGA GTG CGA GAA AAA GAA T 3'
driRTR:	5' ATT CTT TTT CTC GCA CTC G 3'
exon 1 5"	5' CGA ATT CTT TCA CCT GGA TAC TTC TAC T 3'
exon 1 3'	5' CGG GAT CCC TTA AGC GCA CTC TCC TGT TT 3'
G1:	5' TCC AGG TCC GGC TGT G 3'
inside:	5' GCT CCA TCC ATT ATT CTG CT 3'

intron 1 3' end: 5' CGG GAT CCT TAC ACC ACC AAA ACT CAT T 3'

intron 5 5' end: 5' GGA ATT CTA CAC CAC CAC TTT CCA ACT C 3'

RT-PCR 5': 5' GGA ATT CGT TAT ATT CGT GAC ATT C 3'

RT-PCR 3': 5' CGG GAT CCT TGT CTC TTC TTT CCA TG 3'

Δhelix3 tester: 5' GCT GTG GCA GGA GAT CAT CAA 3'

ARID box 5' 5' CGG GAT CCG CAG CAG AAT AAT GGA TGG A 3'

ARID box 3' 5' GGA ATT CGT CAT CGG CAT CAT CTG GTT GTG 3'

REKLES box 5' 5' CGG GAT CCC TAC GAG GCC ATG CAC AAC CAG 3'

REKLES box 3' 5' GGA ATT CAT CCT TCC TAG CTG ACC ACG G 3'

Mutagenic oligonucleotides:

Δhelix3+: 5' GCG GCT TGG TGG ATG TTA TCG GGC TGC ACC TGC CCT CCA GC 3'

Δhelix3-: 5' GCT GGA GGG CAG GTG CAG CCC GAT AAC ATC CAC CAA GCC GC 3'

ΔREKLESβ+: 5' GGC GAA CCG CAG CTG ATT GTC GTG CCG CTG TCA CAG AGC GAG 3'

ΔREKLESβ-: 5' CTC GCT CTG TGA CAG CGG CAC GAC AAT CAG CTG CGG TTC GCC 3'

Degenerate oligonucleotides:

ARID 5' outside: 5' CGG GAT CCA ARM GNA ARG ART TYY T 3'

ARID 3' outside: 5' GGA ATT CYK YTC RCA YTC RTA NGG 3'

ARID 5' inside: 5' CGG GAT CCG AYY TNT TYW SNT TYA TG 3'

ARID 3' inside: 5' GGA ATT CGR TAN ARR TAY TTC ATR TA 3'

DRIL1 cDNA complementary oligonucleotides:

HeLaA: 5' CCG TTG GTG CCG TCA GGT TCA 3'

HeLaB: 5' CGG CAC CAA CGG CAG CAA CAG 3'

HeLaC: 5' CAC CCT GCG GAC CCA ATA CAT 3'

HeLaD: 5' CCC CGC TTC TCA CAC TCG TAG 3'

HeLaE: 5' TGT GAT GGG GAT GGC TGA GTC 3'

HeLaF: 5' CCG CCG CCG CCC CCT TTG TTG 3'

HeLaG: 5' CAG CCA TCC CCA TCA CAG TCC 3'

Danio rerio complementary oligonucleotides:

Dr ARID#1 lower: 5' GCC TCC CTT CTC CGT CAC CAG 3'

Dr ARID#1 upper: 5' TCA TGC AGA AAC GAG GGA CAC 3'

Dr ARID#2 lower: 5' CCG CCC CGG ATA CAG CAA CAG 3'

Dr ARID#2 upper: 5' GGG CTG CTC CGC TGG GTG AAG 3'

Dr ARID#3 lower: 5' CTG CAT GTA AGC GAA GAC GTC 3'

Dr ARID#3 upper: 5' GAA CCT GCC CAC CTC CAT CAC 3'

Dr#1 (1014+): 5' GCA TCG TTT ACT CAG GGA ATC 3'

Dr#1 (586+):	5' ACA CCC TGT TCC ACC TTA GTC 3'
Dr#1 (637-):	5' GCT GCA TTG TGG GCG GAA GTC 3'
Dr#1 (814-):	5' CTC CGG CGC TTC AGA CTC CAG 3'
Dr#1.6-7+:	5' CCT GAA CCT GCC CAC CTC CAT 3'
Dr#1.6-7-:	5' AAG GCT TTA CGC TCA CAC TCG 3'
Dr#2 (1317+):	5' GGA CTT GTC TCT CAG CAT TTC 3'
Dr#2 (1362-):	5' ATF CCG TTG ACC TCC ACA GAC 3'
Dr#2 (386-):	5' GTC ATC CAG GTT CCA ATC TCC 3'
Dr#2 (703+):	5' GAA CCT GCC CAC CTC CAT CAC 3'
Dr#2 (2055+):	5' TCT GGC TTT TGC TGT CTA TTA 3'
Dr#2 (-x+):	5' CAA GCT AAA CTG GAG ATG GAG 3'
Dr#2 (85-):	5' CAT CTC ATC AGC AGC GTC CTC 3'
Dr#2 (1724+):	5' GCT AAA AGG AGG GAC CAC ACA 3'
Dr#2 (1724-):	5' TGT GTG GTC CCT CCT TTT AGC 3'
Dr#2.6-7-:	5' CCC TCT CTG CGG TTA CTA TCG 3'
Dr#2 (3296+):	5' CTC CTG AAA CGA ATG ACA CTA 3'
Dr#2.(3296-):	5' GAT ATG TGA GGT CTG AAG TTC 3'
Dr#2.(2719-):	5' AAT CTG ACC CCT TTT CTA ACC 3'
Dr#2.(3616-):	5' GGC AAG ACT ACA CCA ACA TAA 3'
4112.3 (355+):	5' TCG GCA AGA CTA CAC CAA CAT 3'
2111.1 (418+):	5' GCT CTC ATT TCA TCA CGA GCC 3'

6.2.8 Clones and cloning vectors

Cloned DNA sequences:
pGEX*dri*ϕ10 (derived from lambda ϕ10), pBluescript*dri* (derived from bk60), p*dri*ϕ4 (derived from lambda ϕ4) and p*dri*2T (derived from λ2T) were obtained from Stephen L. Gregory (Gregory 1996).

Cloning vectors:
pBluescript KS+ was obtained from Briony Patterson.

6.2.9 Libraries

Drosophila melanogaster library:
The *Drosophila melanogaster* Canton-Special genomic library in EMBL3 SP6/T7 was obtained from Clontech, Palo Alto, California.

Danio rerio libraries:
The total embryonic random primed polyA$^+$ RNA derived *D. rerio* library in λ ZAP was obtained from José Campos-Ortega.
Developmentally staged (gastrula 6-9 hour, neurula 9-16 hour, post-somitogenesis 20-28 and adult), poly A$^+$ RNA derived *D. rerio* cDNA libraries in λ ZAP were obtained from David J. Grunwald.

Homo sapiens libary:
The HeLa derived cDNA library in λ ZAP was obtained from Clontech, Palo Alto, California.

6.2.10 Buffers and solutions

All solutions were prepared with distilled and deionised water and sterilized by autoclaving, except heat labile reagents, which were filter sterilized.

Agarose gel loading buffer:	50% (v/v) glycerol 10mM EDTA 0.2% (w/v) xylene cyanol 0.4% (w/v) bromophenol blue
10x AmpliTaq™ reaction buffer:	as supplied by Perkin-Elmer
10x injection buffer:	50mM KCl 1mM PO_4 pH 6.8
10x larval fixation buffer:	1M HEPES 0.5M EGTA 0.1% Nonidet P40
10x ligation buffer:	660mM Tris-HCl pH 7.9 10mM rATP 100mM $MgCl_2$ 50mM DTE
NE buffers: 1,2,3,4 and *Sal*I:	as supplied by New England Biolabs
nitrocellulose pre-hybe mix:	50% deionized formamide 5x SSC 0.5% blotto (skim milk powder) 100mg/ml sonicated/denatured salmon sperm DNA

NTMTL buffer:	100mM NaCl 100mM Tris-HCl pH 9.5 50mM $MgCl_2$ 0.1% Tween-20 0.024g levamisole
nylon pre-hybe mix:	90% nitrocellulose pre-hybe mix 1% SDS
PBS:	7.5mM Na_2HPO_4 2.5mM NaH_2PO_4 145mM NaCl
PBST:	1x PBS 0.1% Tween-20
phenol/chloroform:	50% phenol 48% chloroform 2% isoamyl alcohol stored under TE in the dark
TAE:	40mM Tris-acetate 20mM sodium acetate 1mM EDTA pH 8.2.
TBE:	100mM Tris-borate pH 8.3 2.5mM EDTA
TBST:	50 mM Tris-HCl pH 7.5 150 mM NaCl 0.05% Tween-20
TE:	10mM Tris-HCl pH 7.4 1 mM EDTA
TEN:	10mM Tris-HCl pH 7.5 100µM EDTA 150mM NaCl
SSC:	150mM NaCl 15mM sodium citrate

STET:	100mM NaCl 10mM Tris-HCl pH 8.0 1mM EDTA 5% Triton X-100
zebrafish prehybe solution:	50% formamide (deionised) 5x SSC 2% Blocking reagent (Boehringer Mannheim) 0.1% Tween-20 0.5% CHAPS (Sigma) 50µg/mL yeast RNA 5mM EDTA 50µg/mL heparin

6.2.11 Bacterial strains

Escherichia coli LE392 was obtained from Stanley S. Robert. *E. coli* DH5α cells were obtained from Julianne Camerotto. *E. coli* XL1-Blue and SOLR strains were obtained from Sharon Orford.

6.2.12 Bacterial media

i. Liquid media:

Ampicillin or kanamycin (50µg/ml) was added from a sterile stock solution after the media had been autoclaved.

L-broth:	NaCl (10g), bactotryptone (10g) and yeast extract (5g). Water was added to 1 litre and the pH was adjusted to 7.0 before autoclaving.
SOC:	NaCl (0.5g), bactotryptone (20g) and yeast extract (5g). Water was added to 1 litre and the pH was adjusted to 7.0 before autoclaving, then 20ml of 1M glucose was added
Tryptone broth:	NaCl (5g) and bactotryptone (10g). Water was added to 1 litre and the pH was adjusted to 7.0 before autoclaving.

ii. Solid media:

L-agar plates: contained L-broth with 1.5% (w/v) bactoagar. Plates were also available with ampicillin (50µg/ml) or kanamycin (50µg/ml).

T-agar plates: contained L-broth with 1.5% (w/v) bactoagar.

6.2.13 Fly strains

Unless otherwise stated, the following fly strains were obtained from the Indiana, Bloomington or Umeå stock centres according to availability. Information about the stocks used is available on FlyBase (http://flybase.bio.indiana.edu/).

w^{1118} (nominated wild-type)

$\frac{l(2)02535}{\text{CyO}}$; *ry* $\left(\frac{dri^{7}}{\text{CyO}} ; ry\right)$

$\frac{l(2)05096}{\text{CyO}}$; *ry* $\left(\frac{dri^{8}}{\text{CyO}} ; ry\right)$

$w^{1118} ; \frac{+}{\text{CyO}} ; \frac{Df(3R)ro^{XB3}}{\text{TM6b}, Hu}$

$w^{1118} ; \frac{Sco}{\text{CyO}}$

P[*GAL4-vg.M*] $\left(w ; GAL4^{vg.PM}\right)$

w ; + ; *P*[w^{+mC}=*GAL4-ninaE.GMR*]12 $\left(w ; + ; GAL4^{GMR.PF}\right)$

yw ; + ; *P*[$w^{+mW.hs}$=GawB]71B $\left(yw ; + ; GAL4^{71B}\right)$

$\frac{P[ry^{+t7.2}=GAL4\text{-}Hsp70.sev]2}{\text{CyO}}$; ry $\left(\frac{GAL4^{sev}}{\text{CyO}} ; \text{ry}\right)$

P[*GAL4-Antp.P1.A*] $\left(GAL4^{Antp.P1}\right)$

Hs-GAL4 $\left(GAL4^{hs}, \text{insertion on the 3}^{rd}\text{ chromosome}\right)$

The following fly strains were obtained from various laboratories:

P. Whitington:

P[*GAL4-ato*] ($GAL4^{ato}$, insertions on X, 2nd and 3rd chromosomes)

L. Jan and J. N. Jan:

$\frac{+}{CyO}$; *P*[*GAL4*]109-88

$\frac{+}{CyO}$; $\frac{P[GAL4]116\text{-}64}{TM6b, Hu}$

P[*GAL4*]109-11 (insertion on 3rd chromosome)

P[*GAL4*]109-29 (insertion on 3rd chromosome)

P[*GAL4*]109-43 (insertion on 3rd chromosome)

P[*GAL4*]109-5 (insertion on 2nd chromosome)

P[*GAL4*]109-82 (insertion on 3rd chromosome)

$\frac{P[GAL4]112\text{-}10}{TM6b, Hu}$

$\frac{P[GAL4]112\text{-}27}{TM6b, Hu}$

P[*GAL4*]112-32 (insertion on 3rd chromosome)

P[*GAL4*]112-64 (insertion on 3rd chromosome)

The following fly strains were obtained from T. Shandala:

w^{1118} ; $\frac{dri^{1}}{CyO}$ (EMS induced)

w^{1118} ; $\frac{dri^{2}}{CyO}$ (EMS induced)

w^{1118} ; $\frac{dri^{1}}{CyO}$; $\frac{Df(3R)ro^{XB3}}{TM6b, Hu}$ (EMS induced)

w^{1118} ; $\frac{dri^{2}}{CyO}$; $\frac{Df(3R)ro^{XB3}}{TM6b, Hu}$ (EMS induced)

w^{1118} ; $\frac{dri^{29F}}{CyO}$ (*P*[w^{+} *GAL4*] insertion)

w^{1118} ; $\frac{dri^{89}}{CyO}$ (*P*[w^{+} *GAL4*] insertion)

w^{1118} ; $\frac{dri^{139}}{CyO}$ (*P*[w^{+} *GAL4*] insertion)

Stocks were routinely cultured and maintained at 18°C or 25°C in vials or half pint plastic bottles containing F1 medium.

6.2.14 Fly media

Fortified (F1) *Drosophila* medium contained 1%(w/v) agar, 18.75% compressed yeast, 10% treacle, 10% cornmeal (polenta) and 2.5% tegosept mix (10% *para*-hydroxybenzoate in ethanol) and 1.5% acid mix (47% propionic acid/4.7% orthophosphoric acid).

6.3 Methods

Generally established molecular biological techniques were carried out according to protocols published previously (Ausubel *et al.* 1994; Sambrook *et al.* 1989) .

6.3.1 Restriction endonuclease digestion

DNA was dissolved in H_2O and 1/10th volume of the appropriate 10x NE buffer was added. For complete digestion 3-5 units of enzyme were added per µg of DNA and incubated at 37°C for at least 1 hour.

6.3.2 Agarose gel electrophoresis

Molten agarose in TAE was poured onto glass slides, with plastic combs to form well slots, and used as horizontal gels. The gels were submerged in TAE, in an electrophoresis tank and DNA samples containing a suitable amount of loading buffer were loaded into the gel slots. 70V-90V were applied until the dye had migrated the required distance. The DNA was visualized, after staining the gel with ethidium bromide, by illumination with short wave UV for non-preparative gels, or long wave UV for preparative gels.

Typical agarose percentages used to separate DNA fragments were:

Agarose concentration (% w/v)	DNA size (kb)
0.8	10-50
1.0-1.2	0.4-15
2.0	0.1-0.5

6.3.3 Isolation of DNA restriction fragments from agarose gels

DNA was isolated from agarose gels by staining the gel with ethidium bromide and excising the band of DNA under long wave UV light. DNA was isolated from the gel slice by one of two methods depending on requirements.

Fragments to be used for radiolabelled probes were extracted by placing the slice in a 0.5ml microfuge tube with a pierced bottom containing a plug of siliconized glass wool. The tube was then placed inside a 1.5ml microfuge tube and centrifuged at 3500g in a microfuge for 10 minutes, to collect the eluate. The eluted DNA was precipitated with 1/10th volume of 5M KOAc pH 5.5 and 2.5 volumes of ethanol and resuspended in the desired volume of water.

Fragments to be cloned were extracted using a Gene Clean or QIAquick Gel Extraction kits following the suppliers' protocols.

6.3.4 Dephosphorylation of vector DNA

After the vector DNA was linearized by restriction enzyme digestion, 2 units of CIP were added to the restriction endonuclease digestion mix and incubated at 37°C for 1 hour.

6.3.5 Ligation of restriction fragments to vector DNA

DNA fragments to be ligated were placed in a mix containing 1/10th volume of 10x ligation buffer and 1U of T4 DNA ligase in a total volume of 10-20µl. The reaction was placed at 18°C overnight.

6.3.6 Transformation procedure for plasmid recombinants

A 500ml flask of L-broth was innoculated with 5ml of an overnight culture of *E. coli* XL1-Blue cells and grown to an A_{600} of 0.5-1. The culture was then chilled in an ice slurry for 15 to 30 minutes and the cells harvested by centrifugation at 4000g for 15 minutes. The cells were then resuspended in 500ml of ice-cold H_2O, pelleted, resuspended in 250ml of ice-cold H_2O, pelleted, resuspended in 10ml of ice-cold 10% glycerol, repelleted and finally resuspended in 1.5ml of ice-cold 10% glycerol. The competent cells were then snap frozen and stored as 45µl aliquots at -80°C. For transformation, cells were thawed at RT, added to a portion of ligation reaction mixture and incubated on ice for at least 30 seconds. Cells were then transferred to an ice-cold 2mm electroporation cuvette and electroporated in a Bio-Rad Gene Pulser at 2500V. The cuvette was immediately washed out with 1ml of SOC, and the suspension incubated at 37°C for 30 minutes. Cells were then pelleted by centrifugation in a bench centrifuge for 5 minutes at 700g. 800µl of the supernatant was removed, and the cells gently resuspended in the remaining SOC. Half of the cell suspension was plated onto L-agar plates supplemented with 50µg/ml ampicillin and incubated at 37°C overnight. If selection for

β-galactosidase activity (blue/white colour selection) was required, 10µl of 10% IPTG and 10µl of 20% BCIG were added prior to plating.

6.3.7 PCR amplification of DNA

All PCR amplifications were performed using a Corbett Research capillary thermal cycler.

i. standard PCR
1µl of 1ng/µl simple template or 1µg/µl complex template (genomic DNA or library) was added to 19µl of PCR mixture containing each primer at 1ng/µl for standard PCR or 50ng/µl for degenerate PCR, each dNTP at 2.5µM, 1x AmpliTaq™ reaction buffer, 1-4mM Mg^{2+} and 0.1U of *Taq* thermostable DNA polymerase. The capillary PCR tube was heat sealed and then themally cycled through 30-40 cycles using the appropriate annealing temperature and extension time after an initial 2 minute denaturation step.

ii. high fidelity PCR
1µl of 1ng/µl template was added to 19µl of PCR mixture containing each primer at 1ng/µl, each dNTP at 2.5µM, 1x *Pfu* reaction buffer X-YmM Mg^{2+} and 1U of *Pfu* thermostable DNA polymerase. The capillary PCR tube was heat sealed and then themally cycled as described for the standard PCR protocol.

6.3.8 Isolation of plasmid DNA

i. rapid small scale preparation

2ml of an overnight culture supplemented with the appropriate antibiotics, and incubated overnight at 37°C, with shaking. Cells were harvested by centrifugation at 3500g in a microcentifuge for 2 minutes. The bacterial pellet was then resuspended in 200µl of STET, followed by addition of 10µl of 10mg/ml lysozyme. The suspension was heated at 100°C for 50 seconds and centrifuged at 15000g in a microcentrifuge for 15 minutes. The pellet was removed with a sterile toothpick. Plasmid DNA was then precipitated from the supernatant with 240µl of isopropanol, followed by two washings in ice cold 80% ethanol, dried and resuspended in 20µl H_2O.

ii. large scale preparation

A single colony was used to innoculate 50ml of L-broth supplemented with the appropriate antibiotics, which was then incubated overnight at 37°C with shaking. Cells were harvested by centrifugation at 5000g for 5 minutes and the bacterial pellet resuspended in 2ml of TES. Cells were lysed by the addition of 2ml of freshly prepared 0.2M NaOH/1%SDS and incubation at RT for 5 minutes after gentle mixing. Bacterial debris was precipitated by addition of 2ml of ice-cold 5M KOAc pH 5.5 and incubation on ice for 10 minutes followed by centrifugation at 15000g for 15 minutes. 5µl of 10mg/ml RNaseA was added to the supernatant, which was then incubated at 37°C for 30 minutes. The mixture was extracted with an equal volume of phenol/chloroform and the aqueous phase separated by centrifugation at 8000g for 2 minutes. DNA was precipitated from the supernatant by addition of 1/10th volume of 5M KOAc pH 5.5 and 2.5 volumes of ethanol, and centrifugation at 27000g for 15 minutes. The DNA pellet was washed in 70% ethanol, dried and resuspended in 200µl H_2O.

iii. preparation of ultrapure DNA
Ultrapure DNA was prepared using the Qiagen Midi-prep DNA preparation kit according to the manufacturer's protocol.

6.3.9 Isolation of λ phage DNA

i. rapid small scale preparation

A single plaque plug was added to 500µl of an overnight culture of phage host cells in TB, with 10µl of 1M $MgSO_4$. The mixture was incubated at 37°C for 30min and then added to 10ml of TB and incubated at 37°C until the culture cleared. 500µl of chloroform was added to the lysate and the lysate was mixed by shaking. 10µl of 10mg/ml RNaseA and 10µl of 10mg/ml DNaseI were added and the lysate incubated at 37°C for 30min and then centrifuged at 12000g for 15min. 500mg of PEG_{6000} and 440mg of NaCl were added to the supernatant which was then incubated at 0°C for at least 2hrs. Phage were then pelleted by centrifugation at 12000g for 15min. The phage pellet was resuspended in 500µl of PSB. The phage suspension was then chlorofom extracted with 500µl of chloroform. If neccesary, a portion of the phage suspension (100µl) was retained as a high titre stock. 20µl of 0.5M EDTA, 10µl of 10% SDS and 5µl of 5mg/ml proteinase K were added and the mix incubated at 65°C for 30min. The mixture was then phenol/chloroform extracted and the DNA precipitated with 1/10 volume 5M KOAc pH 5.5 and 2 volumes of ethanol and then centrifuged at 15000g for

15 minutes in a microcentrifuge. The pellet was washed with cold 70% ethanol, dried and resuspended in 50µl of H_2O.

ii. large scale preparation

A 50-100ml culture of phage host cells in TB supplemented with 10mM $MgSO_4$ was infected with phage at an m.o.i. of 0.01 when the culture reached an OD_{600} of 0.4. The culture was incubated at 37°C until the culture cleared. 1ml of chloroform was added to the lysate and the lysate was mixed by shaking. RNaseA and DNaseI were added to a final concentration of 10µg/ml and the lysate incubated at 37°C for 30min and then centrifuged at 12000g for 15 minutes. PEG_{6000} was added to a final cencentration of 5% and NaCl was added to a final concentration of 0.75M and the lysate incubated at 0°C for at least 2hrs. Phage were then pelleted by centrifugation at 12000g for 15 minutes. The phage pellet was resuspended in 500µl of PSB. The phage suspension was then chlorofom extracted with 500µl of chloroform. 20µl of 0.5M EDTA, 10µl of 10% SDS and 5µl of 5mg/ml proteinase K were added and the mixture incubated at 65°C for 30min. The mixture was then phenol/chloroform extracted and the DNA precipitated with 50µl of 5M KOAc pH 5.5 and 1ml of ethanol and then centrifuged at 15000g for 15 minutes in a microcentrifuge. The pellet was washed with cold 70% ethanol, dried and resuspended in 100-200µl of H_2O.

6.3.10 Isolation of *Drosophila melanogaster* genomic DNA

20 adult female flies were collected and frozen at -20°C. The frozen flies were kept on ice to prevent thawing, and homogenised in 100µl of ice-cold 0.1M Tris-HCl pH 9.0, 0.1M EDTA in a microfuge tube. 100µl of prewarmed 2% SDS was then added and the mixture incubated at 65°C for 60 minutes. After cooling to room temperature, 42 µl of 5M KOAc, pH 5.5, was added and the mixture incubated on ice for 30 minutes. Cell debris was pelleted by centrifugation at 15000g for 10 minutes in a microcentrifuge. The supernatant was then placed on ice for a further 10 minutes and the centrifugation repeated. The supernatant was phenol/chloroform extracted twice, incubated with 1µl of 1mg/ml RNaseA at 37°C for 10 minutes, and then reextracted with phenol/chloroform. The DNA was then precipitated with 100µl of isopropanol and pelleted by centrifugation at 15000g for 10 minutes in a microcentrifuge. The DNA pellet was washed twice with 70% ethanol, dried and resuspended in 20µl H_2O.

6.3.11 Determination of DNA concentration

The concentration of the DNA sample was estimated by comparing it to bands of known concentration on an agarose gel or by determining the A_{260} of the DNA solution and calculating the DNA concentration using the extinction coefficient (ε) of 0.02µl/ng.

6.3.12 Radiolabelling of DNA fragments

DNA restriction fragments were labelled with α-^{32}P-dATP using the Amersham Megaprime kit according to the suppliers protocol, small PCR product were labelled using the specific primers used in the PCR. Radio-labelled DNA was separated from unincorporated nucleotides by spin-column chromatography. Chromatography columns were prepared using a 0.5ml microfuge tube with a pierced at the bottom with a 21 guage needle, placed inside a 1.5ml microfuge tube. A drop of acid-washed glass beads, suspended in TE, was placed in the bottom of the 0.5ml tube and 500µl of a slurry of Pharmacia Sepharose CL6B in TE, added to the top of the glass beads. The tubes were centrifuged at 500g for 3 minutes in a bench centrifuge. The labelling reaction mix loaded onto the column and centrifuged as above. 50µl of H_2O was added to the column and centrifugation repeated to elute remaining labelled DNA.

6.3.13 Hybridisation of radio-labelled DNA probe to nylon or nitrocellulose filters

Filters were prehybridised in the appropriate pre-hybe mix for at least 4 hours at 42°C. Radiolabelled probe was added to the filter/pre-hybe mix. The filters were hybridised for at least 16 hours at 42°C. Probe was removed and the filters washed twice in 2x SSC/0.1% SDS for at least 15 minutes at room temperature and twice in 2x SSC/0.1% SDS for at least 15 minutes at 65°C. Filter bound probe was then detected by either autoradiography at -80°C with an intensifier screen, or by Phosphorimager detection using a Fujix BAS 1000 phosphorimager.

6.3.14 Southern blot transfer

DNA was electrophoresed as described above. After agarose gel electrophoresis the gel was soaked in 0.25M HCl for 7 minutes, rinsed in H_2O, and then soaked in 1.5M NaCl/0.5M NaOH for 7 minutes twice. After rinsing in H_2O, the gel was neutralized in 1.5M NaCl, 1.0M

Tris-HCl pH 8.0 for 15 to 20 minutes before transferring the gel to the transfer apparatus. The transfer apparatus consisted of a sheet of 3MM paper soaked in 20x SSC and placed on a glass plate. The glass plate was suspended over a bath of 20x SSC with the ends of the 3MM paper submerged in the SSC bath to create a wick. A sheet of plastic film was placed over the paper with a window removed over the gel. A sheet of nitrocellulose filter, was placed in direct contact with the gel and subsequently covered with three sheet of 20xSSC saturated 3MM paper and a 5cm thick wad of paper towels. A weighted glass plate was used to provide limited pressure to the paper towels for at least 4 hours. The DNA was then crosslinked to the filter by UV irradiation, using a Stratagene Stratalinker. The filter was then prehybridised and hybridised, and any bound radio-labelled probe detected, as described for Southern analysis.

6.3.15 λ library screening

Genomic phage libraries were plated at a density of 50000 plaques/140mm plate for the first round of screening. Filters were lifted in duplicate from plates by placing a nylon filter (Plaque Screen, DuPont) on the plate until the filter had wet. The phage DNA was denatured by placing the filters, plaque side up, on Whatmann 3MM soaked in 1.5M NaCl/0.5M NaOH for 5 minutes. The filters were then neutralized on 3MM soaked in 1.5M NaCl/1.0M Tris-HCl pH 8.0. After UV crosslinking, the filters were washed in 2x SSC and debris scraped off with a glass spreader. Filters were probed with radio-labeled probe and washed according to 2.3.14. Plaques corresponding to bound probe were picked by pulling plugs of agar from the plate with a pasteur pipette and eluting the phage from the agar in 500µl of PSB. Phage picked from the first round of screening were screen twice again: plating at a density of 1000-2000 plaques per 70mm plate for the second round screen and at a density of 100-200 plaques per 70mm plate for the third round screen.

6.3.15 λ ZAP insert excision

Excising inserts from λ ZAP library clones was performed according to the protocol provided by Stratagene.

6.3.16 Site directed mutagenesis

Site directed muatgenesis was performed according to the protocol provided with the QuikChange™ Site-Directed Mutagenesis Kit by Stratagene, with the exception that

electrocompetent *E. coli* XL1-Blue cell were transformed with 1µl of mutagenesis reaction by electroporation.

6.3.17 DNA sequencing

i. Sequencing reactions
All sequences were obtained from at least one reaction in each direction. Ambiguities between clones were resolved by sequencing both clones in both directions over the ambiguity. DNA was sequenced using the ABI Prism™ Dye Teminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer), essentially as descibed in the manufacturer's protocol with the modication of using half the descibed amount of reaction mix. Reactions were cycled through 25 cycles 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes with a temperature ramp setting of 2 in a Corbett Research capillary thermal cycler. Extension products were purified by ethanol precipitation with 2µL of 3M NaOAc pH 4.6 and 50 µL 95% ethanol and chilling for 10 minutes on ice. The precipitate was pelleted by centrifugation at 15000g for 30 minutes, then washed with 250µL of 70% ethanol and dried in a vacuum centrifuge.

ii. Electrophoresis of reactions
Electrophoresis of sequencing reactions was performed on an ABI 373 or ABI 377 automated sequencing machine.

6.3.18 Expression and extraction of bacterial fusion protein

GST-fusion proteins were expressed and crude lysates extracted essentially as described by (Smith, 1988). Specifically, an overnight culture of cells carrying the expression construct was diluted 100 fold into LB medium containing ampicillin and incubated at 37°C for 1-3 hours. Protein expression was induced by adding IPTG to a final concentration of 1mM and incubating for a further 3 hours. Cells were pelleted by centrifugation at 3500g for 10 minutes, and then resuspended in 1/100 volume of PBS. Cells were lysed by incubation with 50µg/ml lysozyme for 15 minutes, followed by sonication 3 times for 30 seconds on ice ensuring samples remained cold. The insoluble and soluble fractions were seperated by centrifugation at 3500g for 10 minutes. The insuble fraction was solublised in PBS + 1% Triton X-100.

6.3.19 Protein gel electrophoresis and western blotting

All SDS-PAGE of protein samples and subsequent western transfer to nitrocellulose membrane, Coomassie blue staining of gels and Ponceau S or India ink staining of blots was performed exactly as described by (Harlow and Lane 1988).

Nitrocellulose blots were washed thoroughly with PBST and then blocked for 1 hour in PBST 5% Blotto. Primary and secondary antibody incubations were carried out overnight at 4°C and for 45 minutes at room temperature respectively, with the appropriate dilutions of antibody in the described blocking solution. The secondary antibodies were almost always horseradish peroxidase conjugated (Jackson) and therefore detection was either by the Enhanced Chemiluminescence method (Amersham) or by colorimetric detection using nickel enhanced DAB staining (Harlow and Lane 1988).

6.3.20 Generation of anti-DRI antiserum

DRI protein was separated by SDS-PAGE protein was visualised by staining the gel with Coomassie prepared in water (Harlow and Lane 1988), and the appropriate gel slice excised. The slice was homogenised in an equal volume of PBS by passage through gradually narrower needles. The gel slurry was mixed with adjuvant and administered to rats. Approximately 10μg of antigen was given to each rat at each injection. Unfortunately, no pre-immune serum was harvested. As a pre-immune control, serum from rats of the same strain and from the same animal house was used. Serum was tested on nitrocellulose strips containing the DRI antigen, to determine antibody activity. Antiserum was intially purified by Protein-G chromatography using HiTrap Protein-G (Pharmacia) as described in the manufacturer's instructions. Antiserum to be used for western analysis was affinity purified by incubation in PBS with nitrocellulose strips bound with GST-DRI. Strips were washed extensively with PBS and bound antibody was eluted by treatment with 1M glycine-HCl pH2.6 for 5 minutes followed by neutralisation with 1M Tris pH7.

6.3.21 Gel Retardations

Electrophoretic Mobility Shift Assays were performed by incubating 0.1ng of ^{32}P end-labelled double-stranded hexamer of the consensus Engrailed binding site TCAATTAAATGA (NP_6) with 10-100ng of fusion protein in 20μl of a buffer containing 10mM Tris pH7.5, 1mM EDTA, 100mM KCl, 0.1mM dithiothreitol, 5% glycerol, 50μg/ml BSA and 1μg of salmon sperm DNA as non-specific competitor. Specific competitor

oligonucleotide was added as indicated, using 100ng of NP_6. Binding reactions were incubated for 20 minutes at 25°C, followed by electrophoresis on a 4% polyacrylamide 10% glycerol gel in 0.5x TBE buffer.

6.3.22 *In vitro* expression of ^{35}S-labelled proteins by reticulocyte lysate

The TNT™ Coupled Reticulocyte Lysate System (Promega) was used according to the manufacturer's instructions. Unincorperated ^{35}S-methionine was removed by size exclusion chromatography using 500µl of Sephadex G-25 equilibrated in TEN in a 1ml syringe. The was loaded with the reaction and centrifuged at 1500g for 5 minutes, the column was eluted by washing twice with 50µl of TEN, centrifuging at 1500g for 5 minutes each time. To determine ^{35}S incorporation 5µl of the eluate was analysed by SDS-PAGE and autoradiographed.

6.3.23 Direct protein-binding to proteins immobilised on nitrocellulose membranes

Nitrocellulose blots of the desired proteins were prepared by standard western blotting descibed above. The filter was washed thoroughly in TBST and then treated with HBBZ (20mM HEPES-KOH pH 7.6, 5mM KCl, 5mM MgCl2, 1mM $ZnCl_2$, 10mM β-mercaptoethanol) containing 6M guanidine-HCl for 30 minutes at 4°C. To renature the membrane-bound protein, the filter was serially washed with HBBZ containing half the previous concentration of guanidine-HCl for 30 minute periods at 4°C until the guanidine-HCl concentration reached 0.75M. The filter was then washed extensively in HBBZ alone at 4°C, followed by HBBZ/5% Blotto. ^{35}S-methionine–labelled *in vitro* translated ligand from one TNT™ Coupled Reticulocyte Lysate System reaction (Promega) was added to the membrane in approximately 3mL of HBBZ/1% Blotto and allowed to bind at 4°C overnight. The membrane was washed at room temperature 3 times for 10 minutes in PBS/0.2% Triton X-100 and then dried completely. Bound ligand was detected by phosphorimaging using a Fujix BAS 1000 phosphorimager.

6.3.24 Analytical preparation of protein extract from *Drosophila*

Approximately 100 dechorionated embryos or 5-10 pairs of imaginal disks were homogenised and boiled in 100µL of sample buffer.

6.3.25 Immunohistochemical analysis of larval protein expression

i. Fixation of larval tissues

Larval disks were dissected in 1x PBS and kept briefly on ice. Dissected disks were then fixed for 30 minutes at room temperature in 1x larval fixation buffer with gentle mixing. After fixation the disks were washed 3 times briefly in 1x PBT and then washed for 1 hour in 70% ethanol/PBT with gentle mixing. At this point disks may have been stored at -20°C for up to 6 months.

ii. Immunostaining of fixed larval tissues

Fixed disks were washed twice for 5 minutes in ethanol and then rehydated with a series of 5 minute washes with 80% ethanol/PBT, 50% ethanol/PBT and 20% ethanol/PBT followed by 3 brief washes in 1x PBT and one 1 hour wash in 1x PBT. Washed disks were blocked in either 5% Blotto or 5% goat serum for 2 hours at room temperature. Blocked disks were incubated with primary antibody at a dilution of 1/5-1/20 for monoclonal antisera or 1/100-1/500 for polyclonal antisera overnight at 4°C. Unbound antibody was removed by 3 brief washes followed by 4 washes of 20 minutes in 1x PBT. Disks were incubated with secondary and tertiary antibodies either overnight at 4°C or for 2 hours at room temperature and then washed as above. Stained tissue was further disected and mounted in 80% glycerol/PBT

iii. Imaging of fluorescently labelled tissues

Fluorescently labelled disks were imaged using either a Bio-Rad MRC 1000 laser scanning confocal microscope connected to a Nikon microscope with a 40x water-immersion objective or an Olympus AX70 epifluorescence microscope with a 40x dry objective and a Photometrics CH250/A CCD camera controlled by a Photometrics CE200A Camera Electronics Unit and a Macintosh 640AV running Photometrics software.

6.3.26 Production of digoxygenin labelled RNA probes

Clones carrying cDNA inserts to be labelled were digested with a restriction enzyme cleaving only at the 5' end of the cDNA for antisense probe or the 3' end for sense probe. T7 or T3 RNA polymerase was used to incorporate DIG-11-dUTP into the RNA product of an *in vitro* run-off translation reaction using a DIG RNA labelling kit (Boehringer Mannheim) according to the manufacturers protocol.

6.3.27 *In situ* hybridisation to *D. rerio* mRNA

i. fixation of embryos

Zebrafish embryos were fixed (with or without chorions) in PBS+4% formaldehyde overnight at 4°C. Embryos were manually dechorionated in PBT in a dish under sterile conditions. Embryos were rinsed once in methanol and then stored at -20°C for at least one hour in fresh methanol. After treatment with methanol, embryos were rehydrated by serial 5 minute washes of 75% methanol/25% PBS, 50% methanol/50% PBS and 25% methanol/75% PBS follwed by four 5 minute washes in PBS. Embryos were treated with 10µg/ml proteinase K in 1ml of PBT for a period depending on the stage of the embryos: 0-4 hpf, no treatment;. 4-8 hpf, 1 minute treatment; 8-16 hpf, 3 minute tretment; 16-24 hpf, 5 minute treatment; and 48+ hpf, 10min treatment. After treatment with proteinase K, the embryos were fixed for 20 minutes in PBS+4% formaldehyde and then washed four times for 5 minutes in PBT.

ii. hybridisation of probe to embryos

To prepare for hybridisation, embryos were rinsed with 1mL zebrafish prehybe solution and then incubate at least one hour at 70°C in 1ml of fresh zebrafish prehybe solution with gentle agitation. 200-300ng of probe, produced as described above, was added to the prehybridised embryos and mixed gently but completely by inversion, avoiding direct contact of the concentrated probe with the embryos. Probe was allowed to hybridise overnight at 70°C with gently agitation. After hybridisation embryos were rinsed three times with 1 mL of prewarmed prehybridisation solution and then washed twice for 15 minutes in prehybridisation solution, once for 30 minutes in 50% prehybridisation solution+50% 2x SSC,. once for 15 minutes. in 2x SSC+0.1% CHAPS, twice for 30 minutes in 0.2x SSC+0.1% CHAPS with all washes at 70°C and preceeded by one rinse in the solution to be used in the wash. Finally, embryos were washed twice 10 minutes and once for 5 minutes in PBT at room temperature and then rinsed once and washed for 2-3 hours in PBT+1% BSA at 4°C.

iii. antibody detection of bound probe and staining

Hybridised embryos were incubated with PBT+1% BSA at 4°C for at least 1 hour and then with anti-DIG Fab fragments conjugated to alkaline phosphatase diluted 1:4000 in PBT+1% BSA overnight at 4°C with gentle agitation. Embryos were rinsed three times and then washed five times for 1 hour in PBT+0.1% BSA at room temperature with gentle agitation.

Embryos were then rinsed once and washed 3 times for 5 minutes in freshly made NTMTL buffer at room temperature The NTMTL was replaced with 1 mL of NTNTL+(340ng/ml NBT, 175ng/ml BCIP) and the embryos incubated in the dark at room temperature with gentle agitation for variable time depending on the rapidity of staining.

iv. post-fixation and storage

After staining embryos were rinsed once and washed once in NTMTL for 10 minutes and the rinsed once and washed once in PBT for 10 minutes. Embryos were then fixed in PBT+4% formaldehyde for 30 minutes to overnight at 4°C. Fixative solution was replaced with 1 mL 80% glycerol and the embryos placed at 4°C. When the embryos had sunk to the bottom of the glycerol, indicating they had equilibrated with the glycerol, they were stored at -20°C or -80°C until used.

6.3.28 Fluorescent *in situ* hybridisation to human chromosomes

A DNA fragment labelled with biotin-14-dATP by nick-translation was hybridized *in situ* to metaphases from two normal males at a final concentration of 10ng/µl as described by Callen *et al.* (1990) .

6.3.29 P-element mediated transformation of *Drosophila*

i. micro-injection of embryos

High purity DNA for injection was prepared using the Qiagen DNA preparation kit described above. The construct DNA at 700ng/µl and the transposase activity plasmid, pπ25.7*wc*(Δ2-3), at 300ng/µl were combined in injection buffer. w^{1118} embryos, staged between 30 and 45 minutes AED at 18°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 mixture and the embryos were left at 18°C in a humidified chamber to hatch and crawl into a yeast paste encircling them.

ii. screening for transformants

Adults that developed from injected embryos were individually crossed to w^{1118} virgins or males allowing transformed lines to be identified amongst the progeny by the w^{+} eye colour marker. Eye colours obtained varied from pale yellow to strong orange but were consistent within each sex for each independent event.

iii. generating stably transformed lines

Independent transformants were crossed to the doubly balanced stock, w^{1118};+/CyO;*Df(3R)ro*^{XB3}/TM6b, *Hu*. Male transformant flies carrying the CyO and TM6b chromosomes were selected and back crossed to w^{1118} virgins in the next generation. The progeny of this cross were scored to determine whether the P-element insert was segregating

from either the second chromosome by the absence of w^+ Cy progeny, or the third chromosome by the absence of w^+ *Hu* progeny. P-element insertions on the X chromosome were identified by the absence of w^+ male progeny. Any lines not segregating with one of these three chromosomes were assigned to the fourth chromosome and discarded as long as at least three other insertion events 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, maintaining the insertion over a balancer chromosome such as CyO or TM6b.

6.4 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.

References

I hate quotations. Tell me what you know. - Ralph Waldo Emerson

Agulnik, A. I., M. J. Mitchel, J. L. Lerner, D. R. Woods and C. E. Bishop (1994a). A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. *Hum. Mol. Genet.* **3**: 873-878.

Agulnik, A. I., M. J. Mitchell, M. G. Mattei, G. Borsani, P. A. Avner, J. L. Lerner and C. E. Bishop (1994b). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum. Mol. Genet.* **3**: 879-884.

Akimenko, M.-A., S. L. Johnson, M. Westerfield and M. Ekker (1995). Differential induction of four *msx* homeobox gene during fin development and regeneration in zebrafish. *Development* **121**: 347-357.

Altman, C. R., R. L. Chow, R. A. Lang and A. Hemmati-Brivanlou (1997). Lens induction by *Pax-6* in *Xenopus laevis*. *Dev. Biol.* **185**: 119-123.

Altschul, S. F., W. Gisch, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Amores, A., A. Force, M. Ekker, Y.-L. Yan, C. Amemiya, A. Fritz, R. K. Ho, L. Jolly, J. Langeland, V. Prince, Y.-L. Wang, M. Westerfield and J. H. Postlethwait (1998). Genome duplications in vertebrate evolution: evidence from zebrafish *hox* clusters. *Science* **sumitted**:

Amos, C. I., D. Bali, T. J. Thiel, J. P. Anderson, I. Grourley, M. L. Frazier, P. M. Lynch, M. A. Luchtefeld, A. Young, T. J. McGarrity and M. F. Seldin (1997). Fine mapping of a genetic locus for Peutz-Jeghers Syndrome on chromosome 19p. *Cancer Res.* **57**: 3653-3656.

Anderson, K. V. and J. A. Lengyel (1979). Rates of synthesis of major classes of RNA in *Drosophila* embryos. *Dev. Biol.* **70**: 217-231.

Aparicio, O. M., B. L. Billington and D. E. Gottschling (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell* **66**: 1279-1267.

Aramayo, R. and R. L. Metzenberg (1996). Meiotic transvection in fungi. *Cell* **86**: 103-113.

Armstrong, J. A. and B. M. Emerson (1998). Transcription of chromatin: these are complex times. *Curr. Opin. Gen. Dev.* **8**: 165-172.

Ausubel, S. F., R. Brent, R. E. Kingston, D. Moore, J. G. Seidman, J. A. Smith and K. Struhl (1994). *Current protocols in molecular biology*. New York, Wiley.

Ayala, F. J., A. Rzhetsky and F. J. Ayala (1998). Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 606-611.

Azpiazu, N. and M. Frasch (1993). *tinman* and *bagpipe*: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev.* **7**: 1325-1340.

Babu, P. and S. G. Bhat (1980). Effect of *zeste* on *white* complementation. *Development and neurobiology of Drosophila*. New York, Plenum Press. 35-38.

Baxevanis, A. D. and D. Landsman (1995). The HMG-1 box protein family: classification and functional relationships. *Nucleic Acids Res.* **23**: 1604-1613.

Baylies, M. K. and M. Bate (1996). *twist*: a myogenic switch in *Drosophila*. *Science* **272**: 1481-1484.

Benson, M. and V. Pirrotta (1988). The *Drosophila* Zeste protein binds cooperatively to sites in many gene regulatory regions: implications for transvection and gene regulation. *EMBO J.* **7**: 3907-3915.

Bickel, S. and V. Pirrotta (1990). Self-association of the *Drosophila* Zeste protein is responsible for transvection effects. *EMBO J.* **9**: 2959-2967.

Blochlinger, K., L. Y. Jan and Y. N. Yan (1993). Postembryonic patterns of expression of *cut*, a locus regulating sensory organ identity in *Drosophila*. *Development* **117**: 441-450.

Bodmer, R. (1993). The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila*. *Development* **118**: 719-729.

Bodmer, R. and T. V. Venkatesh (1998). Heart development in *Drosophila* and vertebrates: conservation of molecular mechanisms. *Dev. Genet.* **22**: 181-186.

Brand, A. H. and N. Perrimon (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**: 401-415.

Brent, R. (1986). How to run a very small factory. *Cell* **47**: 329-330.

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

Brown, S. A. and R. E. Kingston (1997). Disruption of downstream chromatin directed by a transcriptional activator. *Genes Dev.* **11**: 3116-3121.

Brownell, J. E., J. Zhou, T. Renalli, R. Kobayashi, D. G. Edmondson, S. Y. Roth and C. D. Allis (1996). *Tetrahymena* histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**: 5237-5245.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**: 563-578.

Bucher, P. and E. N. Trifonov (1986). Compilation and analysis of eukaryotic POL II promoter sequences. *Nucleic Acids Res.* **14**: 10009-10026.

Buck, S. W. and D. Shore (1995). Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between *HMR* and telomeres in yeast. *Genes Dev.* **9**: 370-384.

Burns, L. G. and C. L. Peterson (1997a). Protein complexes for remodeling chromatin. *Biochim. Biophys. Acta* **1350**: 159-168.

Burns, L. G. and C. L. Peterson (1997b). The yeast SWI-SNF complex facilitates binding of a transcriptional activator to nucleosomal sites *in vivo*. *Mol. Cell. Biol.* **17**: 4811-4819.

Cairns, B. R., Y. Lorch, Y. Li, M. Zhang, L. Lacomis, H. Erdjument-Bromage, P. Tempst, J. Du, B. Laurent and R. D. Kornberg (1996). RSC, an essential, abundant chromatin-remodelling complex. *Cell* **87**: 1249-1260.

Callen, D. F., E. Baker, H. J. Eyre, J. E. Chernos, J. A. Bell and G. R. Sutherland (1990). Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosomal painting. *Ann. Genet.* **33**: 219-221.

Cao, Y., B. R. Cairns, R. D. Kornberg and B. C. Laurent (1997). Sfh1p, a component of a novel chromatin-remodeling complex, is required for cell cycle progression. *Mol. Cell. Biol.* **17**: 3323-3334.

Chan, C.-S., L. Rastelli and V. Pirrotta (1994). A *Polycomb* response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J.* **13**: 2553-2564.

Cherry, J. M., C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng and D. Botstein (1998). SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* **26**: 73-79.

Cockerill, P. N. and W. T. Gerrard (1986). Chromosomal loop anchorage o the kappa immunoglobulin gene occurs next to the enhancer in the region containing topoisomerase II sites. *Cell* **44**: 273-282.

Cook, P. R. (1997). The transcriptional basis of chromosome pairing. *J. Cell Sci.* **110**: 1033-1040.

Côté, J., J. Quinn, J. L. Workman and C. L. Peterson (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265**: 53-60.

Csink, A. and S. Henikoff (1996). Genetic modifications of heterochromatic associations and nuclear organization in *Drosophila*. *Nature* **381**: 529-531.

Darwin, C. (1876). *The origin of species.* New York, New York University Press.

De Robertis, F., D. Kadosh, S. Henchoz, D. Pauli, G. Reuter, K. Struhl and P. Spierer (1996). The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**: 589-591.

Dernburg, A. F., K. W. Broman, J. C. Fung, W. F. Marshall, J. Philips, D. A. Agard and J. W. Sedat (1996). Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* **85**: 745-759.

Dingwall, A. K., S. J. Beek, C. M. McCallum, J. W. Tamkun, G. V. Kalpana, S. P. Goff and M. P. Scott (1995). The *Drosophila* Snr1 and BRM proteins are related to yeast SWI/SNF proteins and are components o a large protein complex. *Mol. Biol. Cell* **6**: 777-791.

Dreeson, T. D., S. Henikoff and K. Loughney (1991). A pairing-sensitive element that mediates *trans*-inactivation is associated with the *Drosophila brown* gene. *Genes Dev.* **5**: 331-340.

Dunaief, J. L., B. E. Stober, S. Guha, P. Khavari, K. Ålin, J. Lubin, M. Begemann, G. R. Crabtree and S. P. Goff (1994). The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell* **79**: 119-130.

Durrin, L. K., R. K. Mann and M. Grunstein (1992). Nucleosome loss activates *CUP1* and *HIS3* promoters to full induced levels in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**: 1621-1629.

Ekker, M., J. Wegner, M. A. Akimenko and M. Westerfield (1992). Coordinate embryonic expression of three zebrafish *engrailed* genes. *Development* **116**: 1001-1010.

Elfring, L. K., C. Daniel, O. Papoulas, R. Deuring, M. Sarte, S. Mosely, S. J. Beek, W. R. Waldrip, G. Daubresse, A. DePace, J. A. Kennison and J. W. Tamkun (1998). Genetic analysis of *brahma*: the *Drosophila* homolog of the yeast chromatin remodeling factor SWI2/SNF2. *Genetics* **148**: 251-265.

Elfring, L. K., R. Deuring, C. M. McCallum, C. L. Peterson and J. W. Tamkun (1994). Identification and characterization of *Drosophila* relative of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell. Biol.* **14**: 2225-2234.

Engelkamp, D. and V. van Heyningen (1996). Transcription factors in disease. *Curr. Opin. Gen. Dev.* **6**: 334-342.

Engels, W. R. (1989). *P* elements in *Drosophila*. *Mobile DNA*. Washington, D.C., American Society for Microbiology. 437-484.

Fattaey, A. R., K. Helin, M. S. Dembski, N. Dyson, E. Harlow, G. A. Vuocolo, M. G. Hanobik, K. M. Haskell, A. Oliff, D. Defeo-Jones and A. R. E. Jones (1993). Characterization of the retinoblastoma binding proteins RBP1 and RBP2. *Oncogene* **8**: 3149-3156.

Fauvarque, M.-O. and J.-M. Dura (1993). *polyhomeotic* regulatory sequences induce developmental regulator dependent variegation and targeted *P*-element insertions in *Drosophila*. *Genes Dev.* **7**: 1508-1520.

Fisher, A. L. and M. Caudy (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. *Genes Dev.* **12**: 1931-1940.

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

Foley, T. R., T. J. McGarrity and A. B. Abt (1988). Peutz-Jeghers Syndrome: a clinicopathologic survey of the "Harrisburg Family" with a 49-year follow-up. *Gastroenterology* **95**: 1535-1540.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-L. Yan and J. Postlethwait (1998). Preservation of duplicate genes by complementary, degenrative mutation. *Genetics* **sumitted**:

Franke, A., M. DeCamillis, D. Zink, N. Cheng, H. W. Brock and R. Paro (1992). Polycomb and Polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* **11**: 2941-2950.

Fung, J. C., W. F. Marshall, A. Dernburg, D. A. Agard and J. W. Sedat (1998). Homologous chromosome pairing in *Drosophila melanogaster* proceeds through multiple independent initiations. *J. Cell Biol.* **141**: 5-20.

Gasser, S. M., M. Gotta, H. Renauld, T. Laroche and M. Cockell (1998). Nuclear organization and silencing: trafficking of Sir proteins. *Novartis Found. Symp.* **214**: 114-132.

Gasser, S. M. and U. K. Laemmli (1986). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated gene of *D. melanogaster*. *Cell* **46**: 521-530.

Gemkow, M. J., P. J. Verveer and D. J. Arndt-Jovin (1998). Homologous association of the Bithorax-Complex during embryogenesis: consequences for transvection in *Drosophila melanogaster*. *Development* **125**: 4541-4552.

Georgel, P. T., T. Tsukiyama and C. Wu (1997). Role of histone tails in nucleosome remodeling by *Drosophila* NURF. *EMBO J.* **16**: 4717-4726.

Georgiev, P. G. and V. G. Corces (1995). The Su(Hw) protein bound to gypsy sequences in one chromosome can repress enhancer-promoter interactions in the paired gene located in the other homolog. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 5184-5188.

Geyer, P. K., M. M. Green and V. G. Corces (1990). Tissue specific transcriptional enhancers may act in *trans* on the gene located in the homologous chromsome: the molecular basis of transvection in *Drosophila*. *EMBO J.* **9**: 2247-2256.

Gitelman, I. (1997). Twist protein in mouse embryogenesis. *Dev. Biol.* **189**: 205-214.

Gloor, G. B., N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston and W. R. Engels (1991). Targeted gene replacement in drosophila via *P* element-induced gap repair. *Science* **253**: 1110-1117.

Goldberg, M. L., R. A. Colvin and A. F. Mellin (1989). The *Drosophila zeste* locus is nonessential. *Genetics* **123**: 145-155.

Goldsborough, A. S. and T. B. Kornberg (1996). Reduction of transcription by homologue asynapsis in *Drosophila* imaginal discs. *Nature* **381**: 807-810.

Gotta, M. and S. M. Gasser (1996). Nuclear organization and transcriptional silencing in yeast. *Experientia* **52**: 1136-1147.

Gotta, M., T. Laroche, A. Fomenton, L. Maillet, H. Scherthan and S. M. Gasser (1996). The clustering of telomeres and colocalisation with Rap1, Sir3 and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *J. Cell Biol.* **134**: 1349-1363.

Grant, P. A., L. Duggan, J. Côté, S. M. Roberts, J. E. Brownell, R. Candau, R. Ohba, T. Owen-Hughes, C. D. Allis, F. Winston, S. L. Berger and J. L. Workman (1997). Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* **11**: 1640-1650.

Gregory, P. D. and W. Hörz (1998). Life with nucleosomes: chromatin remodelling in gene regulation. *Curr. Opin. Cell Biol.* **10**: 339-345.

Gregory, P. D., A. Schmid, M. Zavari, L. Lin, S. L. Berger and W. Hörz (1998). Absence of Gcn5 HAT activity defines a novel state in the opening of chromatin at the *PHO5* promoter in yeast. *Mol. Cell* **1**: 495-505.

Gregory, S. L. (1996). *Characterisation of the dead ringer gene.* PhD Thesis: University of Adelaide.

Gregory, S. L., R. D. Kortschak, B. Kalionis and R. Saint (1996). Characterization of the *dead ringer* gene identifies a novel, highly conserved family of sequence-specific DNA-binding proteins. *Mol. Cell. Biol.* **16**: 792-799.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature* **389**: 349-352.

Hadjiolov, A. A. (1984). *The nucleolus and ribosome biogenesis.* New York, Springer-Verlag/Wien.

Halder, G., P. Callaerts and W. J. Gehring (1995). Induction of ectopic eyes by targeted expression of the *eyeless* gene in *Drosophila.. Science* **267**: 1788-1792.

Han, M. and M. Grunstein (1988). Nucleosome loss activates yeast downstream promoters *in vivo*. *Cell* **55**: 1137-1145.

Harfe, B. D., A. V. Gomes, C. Kenyon, J. Liu, M. Kause and A. Fire (1998). Analysis of a *Caenorhabditis elegans twist* homolog indentifies conserved and divergent aspects of mesodermal patterning. *Genes Dev.* **12**: 2623-2635.

Harlow, E. and D. Lane (1988). *Antibodies: a laboratory manual.* New York, Cold Spring Harbour Laboratory.

Hassig, C. A., T. C. Fleischer, A. N. Billin, S. L. Schreiber and D. E. Ayer (1997). Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**: 341-347.

Hecht, A., T. Laroche, S. Strahl-Bolsinger, S. M. Gasser and M. Grunstein (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* **80**: 583-592.

Hemminki, A., D. Markie, I. Tomlinson, E. Avizienyte, S. Roth, A. Loukola, G. Bignell, W. Warren, M. Aminoff, P. Höglund, H. Järvinen, P. Kristo, K. Pelin, M. Ridanpää, R. Salovaara, T. Toro, W. Bomber, S. Olschwang, A. S. Olsen, M. R. Stratton, A. de la Chappelle and L. A. Aaltonen (1998). A serine/threonine kinase gene defective in Peutz-Jeghers Syndrome. *Nature* **391**: 184-187.

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

Henikoff, S., J. M. Jackson and P. B. Talbert (1995). Distance and pairing effects on the $brown^{Dominant}$ heterochromatic element in *Drosophila. Genetics* **140**: 1007-1017.

Herrscher, R. F., M. H. Kaplan, D. Lelsz, C. Das, R. Scheuermann and P. W. Tucker (1995). The immunoglobulin heavy-chain matrix associating regions are bound by Bright: a B cell-specific *trans*-activator that describes a new DNA-binding protein family. *Genes Dev.* **9**: 3067-3082.

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

Hirschhorn, J. N., S. A. Brown, C. D. Clark and F. Winston (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* **6**: 2288-2298.

Hochstrasser, M. and J. W. Sedat (1987). Three-dimensional organization of *Drosophila melanogaster* interphase nuclei. II. Chromosome spatial organization and gene regulation. *J. Cell Biol.* **104**: 1471-1483.

Hopwood, N. D., A. Pluck and J. B. Gurdon (1989). A *Xenopus* messenger-RNA related to *Drosophila twist* is expressed in response to induction in the mesoderm and the neural crest. *Cell* **59**: 893-903.

Huang, F., C. Dambly-Chaudière and A. Ghysen (1991). The emergence of sense organs in the wing disc of *Drosophila*. *Development* **111**: 1087-1095.

Huang, T. H., T. Oka, T. Asai, B. W. Merrills, R. H. Gertson, R. H. Whitson and K. Itakura (1996). Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Res.* **24**: 1695-1701.

Huang, X. (1994). On global sequence alignment. *Comput. Appl. Biosci.* **10**: 227-235.

Ito, T., M. Bulger, M. J. Pazin, R. Kobayashi and J. T. Kadonaga (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**: 145-155.

Jenne, D. E., H. Reimann, J.-i. Nezu, W. Friedel, S. Loff, R. Jeschke, O. Müller, W. Back and M. Zimmer (1998). Peutz-Jeghers Syndrome is caused by mutations in a novel serine threonine kinase. *Nature Genetics* **18**: 38-43.

Jeppesen, P., A. Mitchell, B. Turner and P. Perry (1992). Antibodies to defined histone epitopes reveal variation in the chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. *Chromosoma* **101**: 322-332.

Johnson, L. M., P. S. Kayne, E. S. Kahn and M. Grunstein (1990). Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent matin loci in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 6286-6290.

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

Karpen, G. H., J. E. Schaefer and C. D. Laird (1988). A *Drosophila* rRNA gene located in euchromatin is active in transcription and nucleolus formation. *Genes Dev.* **2**: 1745-1763.

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

Kassis, J. A., E. P. Van Sickle and S. M. Sensabaugh (1991). A fragment of *engrailed* regulatory DNA can mediate transvection of the *white* gene in *Drosophila*. *Genetics* **128**: 751-761.

Kornberg, R. D. and Y. Lorch (1995). Interplay between chromatin structure and transcription. *Curr. Opin. Cell Biol.* **7**: 371-375.

Kortschak, R. D. (1993). *Genomic characterisation of the dead ringer gene of Drosophila melanogaster.* Honours: University of Adelaide.

Kortschak, R. D., H. Reimann, M. Zimmer, H. J. Eyre, R. Saint and D. E. Jenne (1998). The human *dead ringer/bright* homolog, DRIL1: cDNA cloning, gene structure and mapping to D19S886, a marker on 19p13.3 which is strictly linked to the Peutz-Jeghers-Syndrome. *Genomics* **51**: 288-292.

Kruger, W., C. L. Peterson, A. Sil, C. Coburn, G. Arents, E. N. Moudrianakis and I. Herskowitz (1995). Amino acid substitutions in the structured domain of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev.* **9**: 2770-2779.

Kuo, M.-H., J. Zhou, P. Jambeck, M. E. A. Churchill and C. D. Allis (1998). Targeted histone acetyltransferase activity of yeast Gcn5p is required for the activation of downstream genes *in vivo*. *Genes Dev.* **12**: 627-639.

Kwon, H., A. N. Imbalzano, P. A. Khavari, R. E. Kingston and M. R. Green (1994). Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* **370**: 477-481.

Laherty, C. D., W.-M. Yang, J.-M. Sun, J. R. Davie, E. Seto and R. Eisenman (1997). Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell* **89**: 349-356.

Landsman, D. and M. Bustin (1993). A signature for the HMG-1 box DNA-binding proteins. *BioEssays* **15**: 539-546.

LaSalle, J. M. and M. Lalande (1996). Homologous asociation of oppositely imprinted chromosomal domains. *Science* **272**: 725-728.

Laurent, B. C., I. Treich and M. Carlson (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev.* **7**: 583-591.

Laurent, B. C., M. A. Tritel and M. Carlson (1990). The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. *Mol. Cell. Biol.* **10**: 5616-5625.

Laurent, B. C., M. A. Tritel and M. Carlson (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 2687-2691.

Laurent, B. C., X. Yang and M. Carlson (1992). An essential *Saccharomyces cerevisiae* gene homologous to *SNF2* encodes a helicase-related protein in a new family. *Mol. Cell. Biol.* **12**: 1893-1902.

Leiserson, W. M., N. M. Bonini and S. Benzer (1994). Transvection in the *eyes absent* gene of *Drosophila*. *Genetics* **138**: 1171-1179.

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

Lewis, E. B. (1954). The theory and application of a new method of detecting chromosomal reaarangements in *Drosophila melanogaster*. *The American Naturalist* **88**: 225-239.

Li, W. and Ö. Wrange (1995). Accessibility of a glucocorticoid response element in a nucleosome depends on its rotational positioning. *Mol. Cell. Biol.* **15**: 4375-4384.

Lindsley, D. L. and G. G. Zimm (1992). *The genome of Drosophila melanogaster.* San Diego, Academic Press.

Lints, T. J., L. M. Parsons, L. Hartley, I. Lyons and R. P. Harvey (1993). Nkx2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic decendents. *Development* **119**: 419-431.

Logie, C. and C. L. Peterson (1997). Catalytic activity of the yeast SWI/SNF complex on reconstituted nucleosome arrays. *EMBO J.* **16**: 6772-6782.

Lonie, A., R. D'Andrea, R. Paro and R. Saint (1994). Molecular characterization of the *Polycomblike* gene of *Drosophila melanogaster*, a *trans*-acting negative regulator of homeotic gene expression. *Development* **120**: 2629-2636.

Loo, S. and J. Rine (1996). Silencers and domains of generalized repression. *Science* **264**: 1768-1771.

Luger, K., A. W. Mäder, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). Crystal structure of the nucleosome core particle at 2.8Å resolution. *Nature* **389**: 251-260.

Luger, K. and T. J. Richmond (1998). The histone tails of the nucleosome. *Curr. Opin. Gen. Dev.* **8**: 140-146.

Lustig, A., C. Liu, C.Zhang and J. P. Hanish (1996). Tethered Sir3p nucleates silencing at telomeres and internal loci in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**: 2483-2495.

Lustig, A. J. (1998). Mechanisms of silencing in *Saccharomyces cerevisiae*. *Curr. Opin. Gen. Dev.* **8**: 233-239.

Lyons, I., L. M. Parsons, L. Hartley, R. Li, J. E. Andrews, L. Robb and R. P. Harvey (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene *Nkx2-5*. *Genes Dev.* **9**: 1654-1666.

Maillet, L., C. Boscheron, M. Gotta, S. Marcand, E. Gilson and S. M. Gasser (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev.* **10**: 1796-1811.

Marcand, S., S. W. Buck, P. Moretti, E. Gilson and D. Shore (1996). Silencing of genes at nontelomeric sites in yeast is controlled by sequestration of silencing factors at telomeres by Rap1 protein. *Genes Dev.* **10**: 1297-1309.

Marshall, W. F., A. F. Dernburg, B. Harmon, D. A. Agard and J. W. Sedat (1996). Specific interactions of chromatin with the nuclear envelope: positional determination within the nucleus in *Drosophila melanogaster*. *Mol. Biol. Cell* **7**: 825-842.

Marshall, W. F., J. C. Fung and J. W. Sedat (1997). Deconstructing the nucleus: global architecture from local interactions. *Curr. Opin. Gen. Dev.* **7**: 259-263.

Martin-Morris, L. E. and S. Henikoff (1995). Conservation of *brown* gene *trans*-inactivation in *Drosophila*. *Genetics* **140**: 193-199.

Mathog, D. (1990). Transvection in the *Ultrabithorax* domain of the Bithorax Complex of *Drosophila melanogaster*.

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

McKee, B. D., L. Habera and J. A. Vrana (1992). Evidence that intergenic spacer repeats of Drosophila melanogaster rRNA genes function as X-Y pairing sites in male meiosis and a general model for achiasmatic pairing. *Genetics* **132**: 529-544.

Messmer, S., A. Frenke and R. Paro (1992). Analysis of the functional role of the *Polycomb* chromo domain in *Drosophila melanogaster*. *Genes Dev.* **6**: 1241-1254.

Miller, O. L., Jr. (1981). The nucleolus, chromsomes, and visualization of genetic activity. *J. Cell Biol.* **91**: 15s-27-.

Mizuguchi, G., T. Tsukiyama, J. Wisniewski and C. Wu (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol. Cell* **1**: 141-150.

Moretti, P., K. Freeman, L. Coodly and D. Shore (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere binding protein. *Genes Dev.* **8**: 2257-2269.

Müller, J. (1995). Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J.* **14**: 1209-1220.

Newman, C. S. and P. A. Krieg (1998). *tinman*-related genes expressed during heart development in *Xenopus*. *Dev. Genet.* **22**: 230-238.

Norton, J. D., R. W. Deed, G. Craggs and F. Sablitzky (1998). Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol.* **8**: 58-65.

O'Hara, P. J., H. Horowitz, G. Eichinger and E. T. Young (1988). The yeast *ADR6* gene encodes homopolymeric amino acid sequences and a potential metal-binding domain. *Nucleic Acids Res.* **16**: 10153-10169.

O'Neill, L. P. and B. M. Turner (1995). Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J.* **14**: 3946-3957.

Olins, D. E., A. L. Olins, A. Herrmann, R. Lin, C. D. Allis and M. Robert-Nicoud (1991). Localization of acetylated histone H4 in the macronucleus of *Euplotes*. *Chromosoma* **100**: 377-385.

Östlund Farrants, A.-K., P. Blomquist, H. Kwon and Ö. Wrange (1997). Glucocorticoid receptor-glucocorticoid response element binding stimulates nucleosome disruption by the SWI/SNF complex. *Mol. Cell. Biol.* **17**: 895-905.

Owen-Hughes, T., R. T. Utley, J. Côté, C. L. Peterson and J. L. Workman (1996). Persistent site-specific remodelling of a nucleosome array by transient action of the SWI/SNF complex. *Science* **273**: 513-516.

Paladino, F., T. Larroche, E. Gilson, A. Axelrod, L. Pillus and S. M. Gasser (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**: 543-555.

Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. *Trends Genet.* **6**: 416-421.

Paroush, Z. e., S. M. Wainwright and D. Ish-Horowicz (1997). Torso signalling regulates patterning in Drosophila by antogonising Groucho-mediated repression. *Development* **124**: 3827-3834.

Peterson, C. L. (1996). Multiple SWItches to turn on chromatin? *Curr. Opin. Gen. Dev.* **6**: 171-175.

Peterson, C. L. and I. Herskowitz (1992). Characterization of the yeast *SWI1*, *SWI2* and *SWI3* genes, which encode a global activator of transcription. *Cell* **68**: 573-583.

Pilat, D., T. Fink, B. Obermaier-Skrobanek, M. Zimmer, H. Wekerle, P. Lichter and D. E. Jenne (1994). The human Met-ase gene (GZMM): structure, sequence and close physical linkage to the serine protease gene cluster on 19p13.3. *Genomics* **24**: 445-450.

Pirrotta, V. (1990). Transvection and long-distance gene regulation. *BioEssays* **12**: 409-414.

Pirrotta, V. (1997). PcG complexes and chromatin silencing. *Curr. Opin. Gen. Dev.* **7**: 249-259.

Pirrotta, V., S. Bickel and C. Mariani (1988). Developmental expression of the *Drosophila zeste* gene and localization of Zeste protein on polytene chromosomes. *Genes Dev.* **2**: 1839-1850.

Pollard, K. and C. L. Paterson (1997). Role for *ADA/GCN5* products in a antagonizing chromatin-mediated transcriptional repression. *Mol. Cell. Biol.* **17**: 6212-6222.

Ptashne, M. (1987). *A genetic switch: gene control and phage λ*. Cambridge, Massachusetts, Blackwell Scientific Publications and Cell Press.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature* **335**: 683-689.

Quinn, J., A. M. Fyrberg, R. W. Ganster, M. C. Schmidt and C. L. Peterson (1996). DNA-binding properties of the yeast SWI/SNF complex. *Nature* **379**: 844-847.

Quiring, R., U. Walldorf, U. Kloter and W. J. Gehring (1994). Homology of the *eyeless* gene of *Drosophila* to the *Small eye* gene in mice and *Aniridia* in humans. *Science* **265**: 785-798.

Roberts, S. and F. Winston (1997). Essential functional interaction of SAGA, a *Saccharomyces cerevisiae* complex of Spt, Ada and Gcn5 proteins, with the Snf/Swi and Srb mediator complexes. *Genetics* **147**: 451-455.

Rost, B. and C. Sander (1993). Prediction of protein structure at better than 70% accuracy. *J. Mol. Biol.* **232**: 584-599.

Rost, B. and C. Sander (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**: 55-72.

Rost, B., C. Sander and R. Schneider (1994). PHD - an automatic mail server for protein secondary structure prediction. *Computer Application in Biological Science* **10**: 53-60.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular cloning: a laboratory manual.* Cold Spring Harbour, Cold Spring Harbor Laboratory Press.

Schloßherr, J., H. Eggert, R. Paro, S. Cremer and R. S. Jack (1994). Gene inactivation in *Drosophila* mediated by the *Polycomb* gene product or by position-effect variegation does not involve major changes in the accessibility of the chromatin fibre. *Mol. Gen. Genet.* **243**: 453-462.

Sharon-Friling, R., J. Richardson, S. Sperbeck, D. Lee, M. Rauchman, R. Maas, A. Swaroop and G. Wistow (1998). Lens-specific gene recruitment of ζ-crystallin through Pax6, NRL-Maf, and brain suppressor sites. *Mol. Cell. Biol.* **18**: 2067-2076.

Shelton, D. A., L. Stegman, R. Hardison, W. Miller, J. H. Bock, J. L. Slightom, M. goodman and D. L. Gumucio (1997). Phylogenetic footprinting of hypersensitive site 3 of the β-globin locus control region. *Blood* **89**: 3457-3469.

Sheng, G., E. Thouvenot, D. Schmucker, D. S. Wilson and C. Desplan (1997). Direct regulation of *rhodopsin 1* by *Pax-6/eyeless* in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Dev.* **11**: 1122-1131.

Shore, D., D. Stillman, A. Brand and K. Nasmyth (1987). Identification of silencer binding proteins from yeast: possible roles in *SIR* control and DNA replication. *EMBO J.* **6**: 461-467.

Sigrist, C. J. A. and V. Pirrotta (1997). Chromatin insulator elements block the silencing of a target gene by the *Drosophila* Polycomb response element (PRE) but allow *trans* interactions between PREs on different chromosomes. *Genetics* **147**: 209-221.

Slatis, H. M. (1955). A reconsideration of the *brown-dominant* position effect. *Genetics* **40**: 246-251.

Sommerville, J. (1986). Nucleolar structure and ribosome biogenesis. *Trends in Biochemical Science* **11**: 438-442.

Stafford, G. A. and R. H. Morse (1997). Chromatin remodeling by transcriptional activation domains in a yeast episome. *J. Biol. Chem.* **272**: 11526-11534.

Strutt, H. and R. Paro (1997). The Polycomb group protein complex of *Drosophila* has differential composition at different target genes. *Mol. Cell. Biol.* **16**: 3621-3633.

Sulston, J., Z. Du, K. Thomas, R. Wilson, L. Hillier, R. Staden, N. Halloran, P. Green, J. Thierry-Mieg, L. Qiu, S. Dear, A. Coulson, M. Craxton, R. Durbin, M. Berks, M. Metzstein, T. Hawkins, R. Ainscough and R. Waterston (1992). The *C. elegans* genome sequencing project: a beginning. *Nature* **356**: 37-41.

Takeuchi, T., Y. Yamazaki, Y. Katoh-Fukui, S. Kondo, J. Motoyama and T. higashinakagawa (1995). Gene trap capture of a novel mouse gene, *jumonji*, required for neural tube formation. *Genes Dev.* **9**: 1211-1222.

Tamkun, J. W., R. Deuring, M. P. Scott, M. Kissinger, A. M. Pattatucci, T. C. Kaufman and J. A. Kennison (1992). Brahma: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF/SWI2. *Cell* **68**: 561-572.

Taniura, H., C. Glass and L. Gerace (1995). A chromatin binding site in the tail domain of nuclear lamins that interacts with core histones. *J. Cell Biol.* **131**: 33-44.

Tartof, K. D. and S. Henikoff (1991). *Trans*-sensing effects from *Drosophila* to humans. *Cell* **85**: 201-203.

Taunton, J., C. A. Hassig and S. L. Schreiber (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**: 408-411.

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.

Thisse, C. and B. Thisse (1992). Dorsoventral development of the *Drosophila* embryo is controlled by a cascade of transcriptional regulators. *Development* **Supplement**: 173-181.

Thompson, J. S., X. Ling and M. Grunstein (1994). Histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* **369**: 245-247.

Treisman, J. E., A. Luk, G. M. Rubin and U. Heberlein (1997). *Eyelid* antagonises *wingless* signalling during *Drosophila* development and has homology to the Bright family of DNA-binding proteins. *Genes Dev.* **11**: 1949-1962.

Treisman, J. E. and G. M. Rubin (1996). Targets of *glass* regulation in the *Drosophila* eye disc. *Mech. Dev.* **56**: 17-24.

Tsukiyama, T., C. Daniel, J. Tamkun and C. Wu (1995). *ISWI*, a member of the *SWI2/SNF2* ATPase family encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* **83**: 1021-1026.

Tsukiyama, T. and C. Wu (1997). Chromatin remodeling and transcription. *Curr. Opin. Gen. Dev.* **7**: 182-191.

Valentine, S. A., G. Chen, T. Shandala, J. Fernandez, S. Mische, R. Saint and A. J. Courey (1998). Dorsal-mediated repression requires the formationof a multiprotein repression complex at the ventral silencer. *Mol. Cell. Biol.* **18**: 6584-6594.

Varga-Weisz, P. D. and P. B. Becker (1998). Chromatin-remodeling factors: machines that regulate? *Curr. Opin. Cell Biol.* **10**: 346-353.

Varga-Weisz, P. D., M. Wilm, E. Bonte, K. Duma, M. Mann and P. B. Becker (1997). Chromatin-remodelling factor CHRAC contain the ATPase ISWI and topoisomerase II. *Nature* **388**: 598-602.

Wang, W., J. Côté, Y. Xue, S. Zhou, P. A. Khavari, S. R. Bigger, C. Muchardt, G. V. Kalpana, S. P. Goff, M. Yaniv, J. L. Workman and G. R. Crabtree (1996a). Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J.* **15**: 5370-5382.

Wang, W., Y. Xue, S. Zhou, A. Kuo, B. R. Cairns and G. R. Crabtree (1996b). Diversity and specialization of mammalian SWI/SNF complexes. *Genes Dev.* **10**: 2117-2130.

Weiler, K. S. and J. R. Broach (1995). Heterochromatin and gene expression in *Drosophila*. *Ann. Rev. Genet.* **29**: 577-605.

Wilson, C. J., D. M. Chao, A. N. Imbalzano, G. R. Schnitzler, R. E. Kingston and R. A. Young (1996). RNA polymerase II holoenzyme contains SWI/SNF transcriptional activators involved in chromatin remodeling. *Cell* **84**: 235-244.

Wilson, R., R. Ainscough, K. Anderson, C. Baynes, M. Berks, J. Bonfield, J. Burton, M. Connell, T. Copsey, J. Cooper, A. Coulson, M. Craxton, S. Dear, Z. Du, R. Durbin, A. Favello, L. Fulton, A. Gardener, P. Green, T. Hawkins, L. Hillier, M. Jier, L. Johnston, M. Jones, J. Kershaw, J. Kirsten, N. Laister, P. Latreille, J. Lightning, C. Lloyd, A. McMurray, B. Mortimore, M. O'Callaghan, J. Parsons, C. Percy, L. Rifken, A. Roopra, D. Saunders, R. Shownkeen, M. Smaldon, A. Smith, E. Sonnhammer, R. Staden, J. Sulston, J. Thierry-Mieg, K. Thomas, M. Vaudin, K. Vaughan, R. Waterson, A. Watson, L. Weinstock, J. Wilkinson-Sproat and P. Wohldman (1994). 2.2Mb of contiguous nucleotide sequence from chromsome III of *C. elegans*. *Nature* **368**: 32-38.

Wobbe, C. R. and K. Struhl (1990). Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol. Cell. Biol.* **10**: 3859-3867.

Wolf, C., C. Thisse, C. Stoetzel, B. Thisse, P. Gerlinger and F. Perrin-Schmitt (1991). The M-Twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the Xenopus X-Twi and the Drosophila twist genes. *Dev. Biol.* **143**: 363-373.

Wong, J., Y.-B. Shi and A. P. Wolffe (1997). Determinants of chromatin disruption and transcriptional regulation instigated by the thyroid hormone receptor: hormone regulated chromatin disruption is not sufficient for transcriptional activation. *EMBO J.* **16**: 3158-3171.

Wright, J. H., D. E. Gottschling and V. A. Zakian (1992). *Saccharomyces* telomeres assume a non-nucleosomal chromatin structure. *Genes Dev.* **6**: 197-210.

Wu, C. (1997). Chromatin remodeling and the control of gene expression. *J. Biol. Chem.* **272**: 28171-28174.

Wu, C.-t. (1993). Transvection, nuclear structure, and chromatin proteins. *J. Cell Biol.* **120**: 587-590.

Wu, C.-T. and M. L. Goldberg (1989). The *Drosophila zeste* gene and transvection. *Trends Genet.* **5**: 189-194.

Wu, C.-t. and M. Howe (1995). A genetic analysis of the *Suppressor 2 of zeste complex* of *Drosophila melanogaster*. *Genetics* **140**: 139-181.

Wu, J., J. Ellison, E. Salido, P. Yen, T. Mohandas and L. J. Shapiro (1994). Isolation and characterization of *XE169*, a novel human gene that escapes X-inactivation. *Hum. Mol. Genet.* **3**: 153-160.

Ye, Q. and H. J. Worman (1996). Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to *Drosophila* HP1. *J. Biol. Chem.* **271**: 14653-14656.

Yin, Z., X.-L. Xu and M. Frasch (1997). Regulation of the Twist target gene *tinman* by modula *cis*-regulatory elements during early mesoderm development. *Development* **124**: 4971-4982.

Yoshinaga, S. K., C. L. Peterson, I. Herskowitz and K. R. Yamamoto (1992). Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. *Science* **258**: 1598-1604.

Zalokar, M. and I. Erk (1976). Division and migration of nuclei during early embryogenesis of *Drosophila melanogaster*. *Journal of Microscopy and Biology of the Cell* **25**: 97-106.

Zhang, Y., R. Iratni, H. Erdjument-Bromage, P. Tempst and D. Reinberg (1997). Histone deacetylase and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**: 357-364.

Zimmer, M., R. L. Medcalf, T. M. Fink, C. Mattman, P. Lichter and D. E. Jenne (1992). Three elastase-like genes co-ordinately expressed in the myelo-monocyte lineage are organized as a single genetic locus on 19pter. *Proc. Natl. Acad. Sci. U.S.A.* **89**:

If we shadows have offended,

Think but this — and all is mended —

That you have but slumber'd here

While these visions did appear.

And this weak and idle theme,

No more yielding but a dream,

Gentles, do not reprehend;

If you pardon, we will mend.

And, as I'm an honest Puck,

If we have unearned luck

Now to 'scape the serpent's tongue,

We will make amends ere long;

Else the Puck a liar call:

So, good night unto you all.

Give me your hands, if we be friends,

And Robin shall restore amends.

- Shakespeare, "A Midsummer Night's Dream"