MOLECULAR AND CELLULAR BIOLOGY, Mar. 1996, p. 792–799 Vol. 16, No. 3
0270-7306/96/$04.00+0

Characterization of the *dead ringer* Gene Identifies a Novel, Highly Conserved Family of Sequence-Specific DNA-Binding Proteins

STEPHEN L. GREGORY, R. DANIEL KORTSCHAK, BILL KALIONIS,† AND ROBERT SAINT*

Department of Genetics, University of Adelaide, Adelaide, South Australia 5005, Australia

Received 17 July 1995/Returned for modification 5 September 1995/Accepted 29 November 1995

We report the identification of a new family of DNA-binding proteins from our characterization of the *dead ringer* (*dri*) gene of *Drosophila melanogaster*. We show that *dri* encodes a nuclear protein that contains a sequence-specific DNA-binding domain that bears no similarity to known DNA-binding domains. A number of proteins were found to contain sequences homologous to this domain. Other proteins containing the conserved motif include yeast SWI1, two human retinoblastoma binding proteins, and other mammalian regulatory proteins. A mouse B-cell-specific regulator exhibits 75% identity with DRI over the 137-amino-acid DNA-binding domains of these proteins, indicating a high degree of conservation of this domain. Gel retardation and optimal binding site screens revealed that the in vitro sequence specificity of DRI is strikingly similar to that of many homeodomain proteins, although the sequence and predicted secondary structure do not resemble a homeodomain. The early general expression of *dri* and the similarity of DRI and homeodomain in vitro DNA-binding specificity compound the problem of understanding the in vivo specificity of action of these proteins. Maternally derived *dri* product is found throughout the embryo until germ band extension, when *dri* is expressed in a developmentally regulated set of tissues, including salivary gland ducts, parts of the gut, and a subset of neural cells. The discovery of this new, conserved DNA-binding domain offers an explanation for the regulatory activity of several important members of this class and predicts significant regulatory roles for the others.

The large number of developmentally significant transcription factors found in a wide variety of metazoans can be grouped into a surprisingly small number of gene families, each encoding proteins with a related DNA-binding or dimerization motif (8, 14). Protein motifs such as the homeodomain, zinc finger, bZIP, POU, Ets, and Forkhead/HNF-3 domains have been widely conserved during evolution and are used in a variety of regulatory roles.

Within the homeodomain group of regulators (see reference 12 for a review), DNA-binding domains encoded by functionally different genes exhibit striking similarities in DNA-binding specificity. For example, the *Ultrabithorax* and *engrailed* genes encode homeodomain proteins (UBX and EN, respectively) that bind in a sequence-specific fashion to the same DNA sequences in vitro (1, 7). One of the ways in which the specificity of homeodomain protein action can be conferred in vivo, despite this similarity in binding properties, is by association with other factors. For example, the affinity of UBX for particular binding sites is modified by cooperative interactions with the EXD homeodomain protein (4, 32). From these recent reports and from the genetic analysis of the in vivo role of the homeodomain (11, 18), it is clear that our understanding of gene regulation during development is far from complete, even at the level of identification of the factors involved.

The similarity of the in vitro DNA-binding characteristics of homeodomain proteins was particularly evident in a screen that identified a large number of homeodomain proteins by their ability to bind an oligonucleotide containing a consensus EN binding site (16). One of the genes identified in that screen, originally termed bk60 but here renamed *dead ringer* (*dri*), is the subject of this report. We describe the detailed characterization of this gene, its developmental expression pattern, and the binding specificity of its DNA-binding domain. Surprisingly, *dri* was found to lack any homology to the homeobox but instead encodes a new, highly conserved DNA-binding domain. Thus, the paradox of the similarity of homeodomain binding specificities is further complicated by the discovery of an entirely new DNA-binding domain with the same specificity. We also show that sequences closely related to the sequence of this new DNA-binding domain are present in regulatory proteins from a wide variety of eukaryotes, including the SWI1 protein of *Saccharomyces cerevisiae* and mouse and human regulatory proteins. The limited number but extensive usage of a few highly conserved domain types lends special significance to the discovery of a new family of sequence-specific DNA-binding proteins.

MATERIALS AND METHODS

Isolation and sequencing of cDNA and genomic clones. The initial *dri* cDNA clone (bk60) was isolated as described previously (16). Subsequent overlapping clones were isolated by screening a λgt11 *Drosophila* 3- to 18-h embryo cDNA library (Clontech) with regions of bk60 as the probes. Positive clones were restriction mapped and sequenced by standard methods.

Sequence comparisons were performed at the facilities of the Australian National Genomic Information Service (ANGIS). A related cDNA from *Danio rerio* was obtained by degenerate PCR on an embryonic cDNA library, with primers covering residues DLFSFM (5′ end) and TSAAFT (3′ end) (see Fig. 2).

The bk60 cDNA clone was used to isolate genomic clones from a Canton-S genomic library in λEMBL3 SP6/T7 (Clontech) spanning 30 kb and extending past the 3′ end of the gene. A further clone was isolated by probing with an *Xho*I fragment from the 5′ end of the cloned region, extending the walk to 43 kb, and covering the full extent of the cloned cDNAs. The clones were restriction mapped with *Sac*I, *Sal*I, *Xho*I, and *Xba*I. The location of exons was determined by Southern and PCR analyses. The positions of all intron-exon boundaries were confirmed by sequence analysis.

Expression of DRI protein and DNA-binding experiments. Proteins were expressed and lysates were made in *Escherichia coli* DH5α essentially as described before (27). The glutathione *S*-transferase (GST)-DRI$_{258–410}$ construct contained *dri* sequences encoding amino acids 258 to 410 cloned into pGEX1. The GST-EN construct contained residues 409 to 552 (including the homeodo-

* Corresponding author. Mailing address: Dept. of Genetics, University of Adelaide, Adelaide, SA 5005, Australia. Phone: 61-8-303 4043. Fax: 61-8-303 4399. Electronic mail address: rsaint@genetics.adelaide.edu.au.

† Present address: Department of Obstetrics and Gynaecology, Flinders Medical Centre, Bedford Park, South Australia 5042, Australia.

main at 453 to 513) from the *engrailed* gene cloned into pGEX1. The Mal-DRI constructs contained *dri* sequences encoding amino acids 256 to 410, 287 to 369, or 307 to 369 cloned into pMal-c2. Protein production was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to ensure that equivalent amounts of each protein were used in subsequent assays. No breakdown products were apparent in any of the induced protein samples.

Electrophoretic mobility shift assays were performed by incubating 0.1 ng of ^{32}P-end-labeled double-stranded trimer of the consensus Engrailed binding site TCAATTAAATGA (NP_3) with approximately 10 ng of fusion protein in 20 μl of a buffer containing 10 mM Tris (pH 7.5), 1 mM EDTA, 100 mM KCl, 0.1 mM dithiothreitol, 5% glycerol, 50 μg of bovine serum albumin per ml, and 100 ng of herring sperm DNA as a nonspecific competitor. Specific competitor oligonucleotide was added as indicated, using 1, 10, or 100 ng of NP_3 or TTA_9. Incubation was for 20 min at 25°C, followed by electrophoresis on a 5% polyacrylamide–10% glycerol gel in 0.5× TBE (Tris-borate-EDTA) buffer.

To purify the fusion protein, glutathione-agarose purification was performed as described before (27), and with the fusion protein still bound to the beads, binding reactions to labeled NP_3 were carried out as described above. Unbound probe was removed by pelleting the beads, removing the liquid, and then resuspending the beads in buffer. After two washes, the amount of probe still bound was detected by scintillation counting. The results obtained with glutathione-purified proteins in this method were consistent with those shown in Fig. 3. Following selection of $DRI_{258–410}$ target sequences (see below), a trimer of a consensus site was synthesized: CCGATTAATCCC. Labeled double-stranded trimer (4 ng) was incubated with approximately 1 ng of each Mal-DRI fusion protein and electrophoresed as above to test the shorter constructs and to demonstrate affinity for the selected consensus. A mutated consensus trimer of CCGATTGATCCC was added as a competitor (as above) to test the preference for a possible alternative site.

Selection and amplification of $DRI_{258–410}$ target sequences. $DRI_{258–410}$ target sequences were isolated by the method of Wilson et al. (34), using the randomer CGGGATCCGTGACTGAGGN$_{20}$TTGATGCCGAGGATCCCG, with amplification primers made to the first (top strand) and last (bottom strand) 18 bases. Binding was carried out as described above, with the addition of herring sperm DNA (2 μg/ml) as a non-specific competitor. Amplification was carried out with an annealing temperature of 50°C for 20 cycles. Following eight cycles of selection and amplification, the products were digested with *Bam*HI, cloned into pBluescript (Stratagene), and sequenced. Of the oligonucleotides selected, only one sequence was found twice; all others were unique. We isolated three sequences that could not be aligned with the consensus or with each other, indicating a low background and high diversity in the final pool of selected oligonucleotides. The variant sequences were CCGTATCGTTCGATCTCTTA, GCGT ATGGGATGGAGGAGAG, and GGTTGTCTCCGGGCATGATG and were not included in Fig. 4 because their alignment, and thus their contribution to the consensus, would be purely arbitrary in the absence of any clear relationship to the other sequences obtained.

Production of DRI antiserum. DRI protein was expressed from a pGEX1 construct to produce amino acids 128 to 671 as a GST fusion protein. A crude bacterial lysate was prepared as described by Smith and Johnson (27) from which DRI-GST fusion protein was isolated by SDS-PAGE, Coomassie blue staining, and band isolation. This band was homogenized and injected into rats, with two boosting immunizations at fortnightly intervals. Sera from the rats were tested by Western (immunoblotting) analysis for specificity and activity and then tested by embryo immunostaining.

Embryo in situ hybridization and immunostaining. In situ hybridization with a digoxigenin-11-dUTP-labeled probe was carried out as described by Patel and Goodman (21), using a probe containing *dri* sequence from positions 500 to 2900. DRI protein expression was determined by using polyclonal anti-DRI antibodies and immunostaining as described by Foe (10). The name *dead ringer* was given to this gene to describe the ringed expression pattern, the high level of conservation of the DNA-binding motif, and its similarity to homeodomain binding specificity.

RESULTS

***dead ringer* encodes a protein with a novel, highly conserved domain.** Overlapping *dri* cDNA clones were isolated from an 18-h embryonic cDNA library by hybridization to a bk60 cDNA probe (16), and their sequences were determined. Compilation of these sequences generated a 3.7-kb cDNA. Genomic mapping revealed nine exons spaced over 23 kb (Fig. 1). The cDNA contig revealed a complete open reading frame encoding a protein of 901 amino acids. An in-frame ATG exists 24 bp upstream of the indicated start of translation (Fig. 1); however, there is a very poor fit to the consensus *Drosophila* start of translation at that position (3), so we would expect most translation to begin where indicated in the figure. Like many eukaryotic developmental regulators, the open reading frame contained several di- and trinucleotide repeat regions (26, 33, 37). No previously identified DNA-binding motif was found to be encoded by this gene.

Comparison of the *dri* sequence with those available in sequence databases revealed an 83-amino-acid region with striking similarity to 10 other proteins (Fig. 2): the human retinoblastoma binding proteins RBP1 and RBP2 (9), DNA-binding proteins MRF1 and MRF2 (33a), a protein encoded by the X-linked but not X-inactivated human XE169 gene (36), the product of the budding yeast *S. cerevisiae* regulatory gene *SWI1* (20, 22), Jumonji, a mouse protein required for neural tube formation (31), and two uncharacterized cDNA clones: R20183 from a human placental library (WashU-Merck EST Project; unpublished data) and C08B11.3 from the Yeast Chromosome III sequencing project (35). This core conserved region begins with the charged residues at 287 (FLDDLFS) in DRI and ends at 369 (KYLYPYE). No other regions of DRI showed significant similarity to other proteins in database searches. Currently, the protein most closely related to DRI is Bright, a B-cell-restricted murine protein identified by its ability to bind to immunoglobulin regulatory sequences (15). Bright shows 84% identity to DRI over the most conserved 83-amino-acid region. The region of similarity in Bright, however, extends further, spanning 132 amino acids with 75% identity to DRI (Fig. 2). The high level of conservation of this domain is comparable to that seen between fly and mouse homeodomain homologs (25). We anticipated that the motif would be similarly conserved among vertebrates, and this was confirmed by using a zebrafish cDNA library to amplify a sequence (zf10) that encoded a domain that was almost identical to the Bright domain (Fig. 2). Southern blot analysis showed the presence of hybridizing bands in human, mouse, rat, bovine, and yeast genomic DNA (results not shown), confirming the wide conservation of this domain. All closely related proteins known to contain this novel motif that have been tested have been shown to interact with DNA, some with characterized binding sites. None has a previously defined DNA-binding motif (9, 15, 20, 31, 33a, 36).

The DRI conserved domain forms a sequence-specific DNA-binding domain. To assess whether the conserved domain was necessary to confer the DNA-binding ability observed for DRI, a region of *dri* encoding 153 amino acids (residues 258 to 410), including the residues conserved between DRI and Bright (residues 263 to 398), was cloned into a pGEX expression plasmid, and a soluble GST-$DRI_{258–410}$ fusion protein was produced. Electrophoretic mobility shift assays with a trimer of a consensus EN homeodomain binding site (NP_3) revealed that this GST-$DRI_{258–410}$ fusion protein in a bacterial lysate bound specifically to the EN binding site (Fig. 3, tracks 1 to 4). Small amounts of purified GST-$DRI_{258–410}$ on a glutathione-agarose matrix could also specifically retain labeled NP_3 (result not shown). Several truncations of the conserved domain fused to pMAL were also made to determine how much of the conserved region was required for binding. A construct containing the region conserved with Bright (Fig. 2, residues 256 to 410) was capable of binding; shorter constructs (287 to 369 or 307 to 369) showed no binding activity (result not shown).

Several characteristics of the DRI DNA-binding domain were found to be similar to the homeodomain. For example, GST-$DRI_{258–410}$ was found to bind more poorly to sites diverged from the consensus site, with preferences similar to those of homeodomain proteins. Like most homeodomains, the DRI DNA-binding domain bound the NP site (7) more strongly than it bound a double-stranded oligomer of dTTA:dAAT, which contains only the core recognition bases (ATTA) of the homeodomain site (Fig. 3). Competition with a 1,000-fold excess of a TTA_9 oligonucleotide did not prevent binding

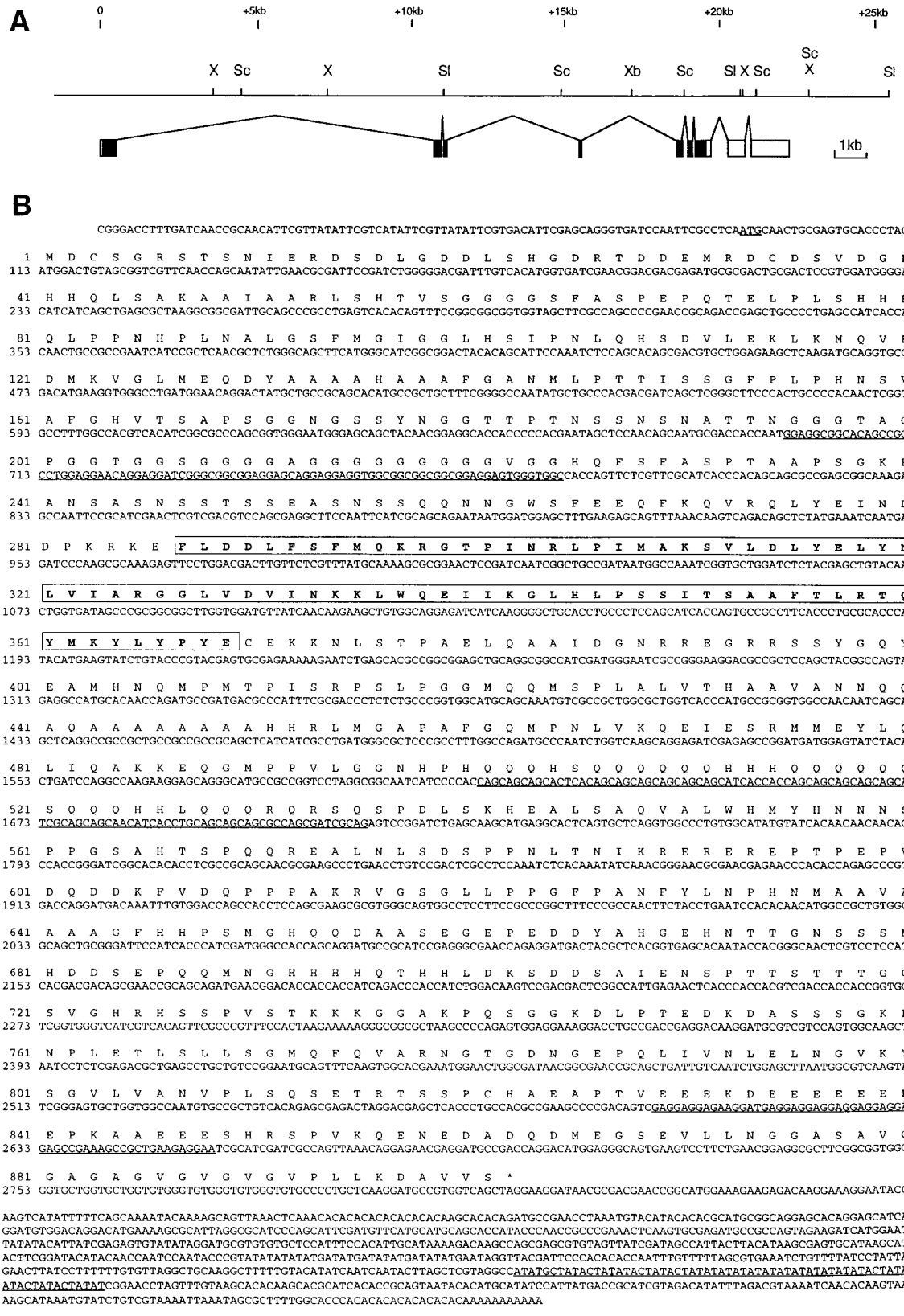
A
0
+5kb
+10kb
+15kb
+20kb
+25kb
X Sc X Sl Sc Xb Sc Sl X Sc Sc X Sl
1kb
B
CGGGACCTTTGATCAACCGCAACATTCGTTATATTCGTCATATTCGTTATATTCGTGACATTCGAGCAGGGTGATCCAATTCGCCTCAATGCAACTGCGAGTGCACCCTACT
1 M D C S G R S T S N I E R D S D L G D D L S H G D R T D D E M R D C D S V D G E
113 ATGGACTGTAGCGGTCGTTCAACCAGCAATATTGAACGCGATTCCGATCTGGGGGACGATTTGTCACATGGTGATCGAACGGACGACGAGATGCGCGACTGCGACTCCGTGGATGGGGAG
41 H H Q L S A K A A I A A R L S H T V S G G G G S F A S P E P Q T E L P L S H H H
233 CATCATCAGCTGAGCGCTAAGGCGGCGATTGCAGCCCGCCTGAGTCACACAGTTTCCGGCGGCGGTGGTAGCTTCGCCAGCCCCGAACCGCAGACCGAGCTGCCCCTGAGCCATCACCAT
81 Q L P P N H P L N A L G S F M G I G G L H S I P N L Q H S D V L E K L K M Q V R
353 CAACTGCCGCCGAATCATCCGCTCAACGCTCTGGGCAGCTTCATGGGCATCGGCGGACTACACAGCATTCCAAATCTCCAGCACAGCGACGTGCTGGAGAAGCTCAAGATGCAGGTGCGC
121 D M K V G L M E Q D Y A A A A H A A A F G A N M L P T T I S S G F P L P H N S V
473 GACATGAAGGTGGGCCTGATGGAACAGGACTATGCTGCCGCAGCACATGCCGCTGCTTTCGGGGCCAATATGCTGCCCACGACGATCAGCTCGGGCTTCCCACTGCCCCACAACTCGGTG
161 A F G H V T S A P S G G N G S S Y N G G T T P T N S S N S N A T T N G G G T A G
593 GCCTTTGGCCACGTCACATCGGCGCCCAGCGGTGGGAATGGGAGCAGCTACAACGGAGGCACCACCCCCACGAATAGCTCCAACAGCAATGCGACCACCAATGGAGGCGGCACAGCCGGA
201 P G G T G G S G G G G A G G G G G G G G G V G G H Q F S F A S P T A A P S G K E
713 CCTGGAGGAACAGGAGGATCGGGCGGCGGAGGAGCAGGAGGAGGTGGCGGCGGCGGCGGAGGAGTGGGTGGCCACCAGTTCTCGTTCGCATCACCCACAGCAGCGCCGAGCGGCAAAGAA
241 A N S A S N S S T S S E A S N S S Q Q N N G W S F E E Q F K Q V R Q L Y E I N D
833 GCCAATTCCGCATCGAACTCGTCGACGTCCAGCGAGGCTTCCAATTCATCGCAGCAGAATAATGGATGGAGCTTTGAAGAGCAGTTTAAACAAGTCAGACAGCTCTATGAAATCAATGAT
281 D P K R K E F L D D L F S F M Q K R G T P I N R L P I M A K S V L D L Y E L Y N
953 GATCCCAAGCGCAAAGAGTTCCTGGACGACTTGTTCTCGTTTATGCAAAAGCGCGGAACTCCGATCAATCGGCTGCCGATAATGGCCAAATCGGTGCTGGATCTCTACGAGCTGTACAAT
321 L V I A R G G L V D V I N K K L W Q E I I K G L H L P S S I T S A A F T L R T Q
1073 CTGGTGATAGCCCGCGGCGGCTTGGTGGATGTTATCAACAAGAAGCTGTGGCAGGAGATCATCAAGGGGCTGCACCTGCCCTCCAGCATCACCAGTGCCGCCTTCACCCTGCGCACCCAA
361 Y M K Y L Y P Y E C E K K N L S T P A E L Q A A I D G N R R E G R R S S Y G Q Y
1193 TACATGAAGTATCTGTACCCGTACGAGTGCGAGAAAAAGAATCTGAGCACGCCGGCGGAGCTGCAGGCGGCCATCGATGGGAATCGCCGGGAAGGACGCCGCTCCAGCTACGGCCAGTAC
401 E A M H N Q M P M T P I S R P S L P G G M Q Q M S P L A L V T H A A V A N N Q Q
1313 GAGGCCATGCACAACCAGATGCCGATGACGCCCATTTCGCGACCCTCTCTGCCCGGTGGCATGCAGCAAATGTCGCCGCTGGCGCTGGTCACCCATGCCGCGGTGGCCAACAATCAGCAG
441 A Q A A A A A A A A H H R L M G A P A F G Q M P N L V K Q E I E S R M M E Y L Q
1433 GCTCAGGCCGCCGCTGCCGCCGCCGCAGCTCATCATCGCCTGATGGGCGCTCCCGCCTTTGGCCAGATGCCCAATCTGGTCAAGCAGGAGATCGAGAGCCGGATGATGGAGTATCTACAG
481 L I Q A K K E Q G M P P V L G G N H P H Q Q Q H S Q Q Q Q Q Q Q H H H Q Q Q Q Q Q
1553 CTGATCCAGGCCAAGAAGGAGCAGGGCATGCCGCCGGTCCTAGGCGGCAATCATCCCCACCAGCAGCAGCACTCACAGCAGCAGCAGCAGCAGCATCACCACCAGCAGCAGCAGCAGCAG
521 S Q Q Q H H L Q Q Q R Q R S Q S P D L S K H E A L S A Q V A L W H M Y H N N N S
1673 TCGCAGCAGCAACATCACCTGCAGCAGCAGCGCCAGCGATCGCAGAGTCCGGATCTGAGCAAGCATGAGGCACTCAGTGCTCAGGTGGCCCTGTGGCATATGTATCACAACAACAACAGC
561 P P G S A H T S P Q Q R E A L N L S D S P P N L T N I K R E R E R E P T P E P V
1793 CCACCGGGATCGGCACACACCTCGCCGCAGCAACGCGAAGCCCTGAACCTGTCCGACTCGCCTCCAAATCTCACAAATATCAAACGGGAACGCGAACGAGAACCCACACCAGAGCCCGTG
601 D Q D D K F V D Q P P P A K R V G S G L L P P G F P A N F Y L N P H N M A A V A
1913 GACCAGGATGACAAATTTGTGGACCAGCCACCTCCAGCGAAGCGCGTGGGCAGTGGCCTCCTTCCGCCCGGCTTTCCCGCCAACTTCTACCTGAATCCACACAACATGGCCGCTGTGGCA
641 A A A G F H H P S M G H Q Q D A A S E G E P E D D Y A H G E H N T T G N S S S M
2033 GCAGCTGCGGGATTCCATCACCCATCGATGGGCCACCAGCAGGATGCCGCATCCGAGGGCGAACCAGAGGATGACTACGCTCACGGTGAGCACAATACCACGGGCAACTCGTCCTCCATG
681 H D D S E P Q Q M N G H H H H Q T H H L D K S D D S A I E N S P T T S T T T G G
2153 CACGACGACAGCGAACCGCAGCAGATGAACGGACACCACCACCATCAGACCCACCATCTGGACAAGTCCGACGACTCGGCCATTGAGAACTCACCCACCACGTCGACCACCACCGGTGGG
721 S V G H R H S S P V S T K K K G G A K P Q S G G K D L P T E D K D A S S S G K L
2273 TCGGTGGGTCATCGTCACAGTTCGCCCGTTTCCACTAAGAAAAAGGGCGGCGCTAAGCCCCAGAGTGGAGGAAAGGACCTGCCGACCGAGGACAAGGATGCGTCGTCCAGTGGCAAGCTC
761 N P L E T L S L L S G M Q F Q V A R N G T G D N G E P Q L I V N L E L N G V K Y
2393 AATCCTCTCGAGACGCTGAGCCTGCTGTCCGGAATGCAGTTTCAAGTGGCACGAAATGGAACTGGCGATAACGGCGAACCGCAGCTGATTGTCAATCTGGAGCTTAATGGCGTCAAGTAC
801 S G V L V A N V P L S Q S E T R T S S P C H A E A P T V E E E K D E E E E E E E
2513 TCGGGAGTGCTGGTGGCCAATGTGCCGCTGTCACAGAGCGAGACTAGGACGAGCTCACCCTGCCACGCCGAAGCCCCGACAGTCGAGGAGGAGAAGGATGAGGAGGAGGAGGAGGAGGAG
841 E P K A A E E E S H R S P V K Q E N E D A D Q D M E G S E V L L N G G A S A V G
2633 GAGCCGAAAGCCGCTGAAGAGGAATCGCATCGATCGCCAGTTAAACAGGAGAACGAGGATGCCGACCAGGACATGGAGGGCAGTGAAGTCCTTCTGAACGGAGGCGCTTCGGCGGTGGGT
881 G A G A G V G V G V G V P L L K D A V V S *
2753 GGTGCTGGTGCTGGTGTGGGTGTGGGTGTGGGTGTGCCCCTGCTCAAGGATGCCGTGGTCAGCTAGGAAGGATAACGCGACGAACCGGCATGGAAAGAAGAGACAAGGAAAGGAATACCA
AAGTCATATTTTTCAGCAAAATACAAAAGCAGTTAAACTCAAACACACACACACACACACAAGCACACAGATGCCGAACCTAAATGTACATACACACGCATGCGGCAGGAGCACAGGAGCATCAA
GGATGTGGACAGGACATGAAAAGCGCATTAGGCGCATCCCAGCATTCGATGTTCATGCATGCAGCACCATACCCAACCGCCCGAAACTCAAGTGCGAGATGCCGCCAGTAGAAGATCATGGAATA
TATATACATTATCGAGAGTGTATATAGGATGCGTGTGTGCTCCATTTCCACATTGCATAAAAGACAAGCCAGCGAGCGTGTAGTTATCGATAGCCATTACTTACATAAGCGAGTGCATAAGCATC
ACTTCGGATACATACAACCAATCCAATACCCGTATATATATATGATATGATATATGATATATGAATAGGTTACGATTCCCACACACCAATTTGTTTTTAGCGTGAAATCTGTTTTATCCTATTAC
GAACTTATCCTTTTTTGTGTTAGGCTGCAAGGCTTTTTGTACATATCAATCAATACTTAGCTCGTAGGCCATATGCTATACTATATACTATACTATATATATATATATATATATACTATAT
ATACTATACTATATCGGAACCTAGTTTGTAAGCACACAAGCACGCATCACACCGCAGTAATACACATGCATATCCATTATGACCGCATCGTAGACATATTTAGACGTAAAATCAACACAAGTAAT
AAGCATAAATGTATCTGTCGTAAAATTAAATAGCGCTTTTGGCACCCACACACACACACACACAAAAAAAAAAA

FIG. 1. Genomic structure and sequence of the *dead ringer* gene. (A) Connected boxes show the splicing of *dri* cDNA, with shading indicating the open reading frame translated left to right. Numbering starts from the beginning of the 5′-most cDNA end. Restriction sites in the genomic DNA: Sc, *Sac*I; Sl, *Sal*I; X, *Xho*I; and Xb, *Xba*I. (B) Contiguous nucleotide sequence from overlapping cDNA clones including the complete *dead ringer* open reading frame and 3′ untranslated region, with the derived amino acid sequence written in the standard one-letter code. Regions containing di- and trinucleotide repeats (GGN, CAG, GAG, and TA) and the alternative start site are underlined. A putative polyadenylation signal (ATTAAA) is double underlined. The boxed amino acids indicate the most widely conserved motif.

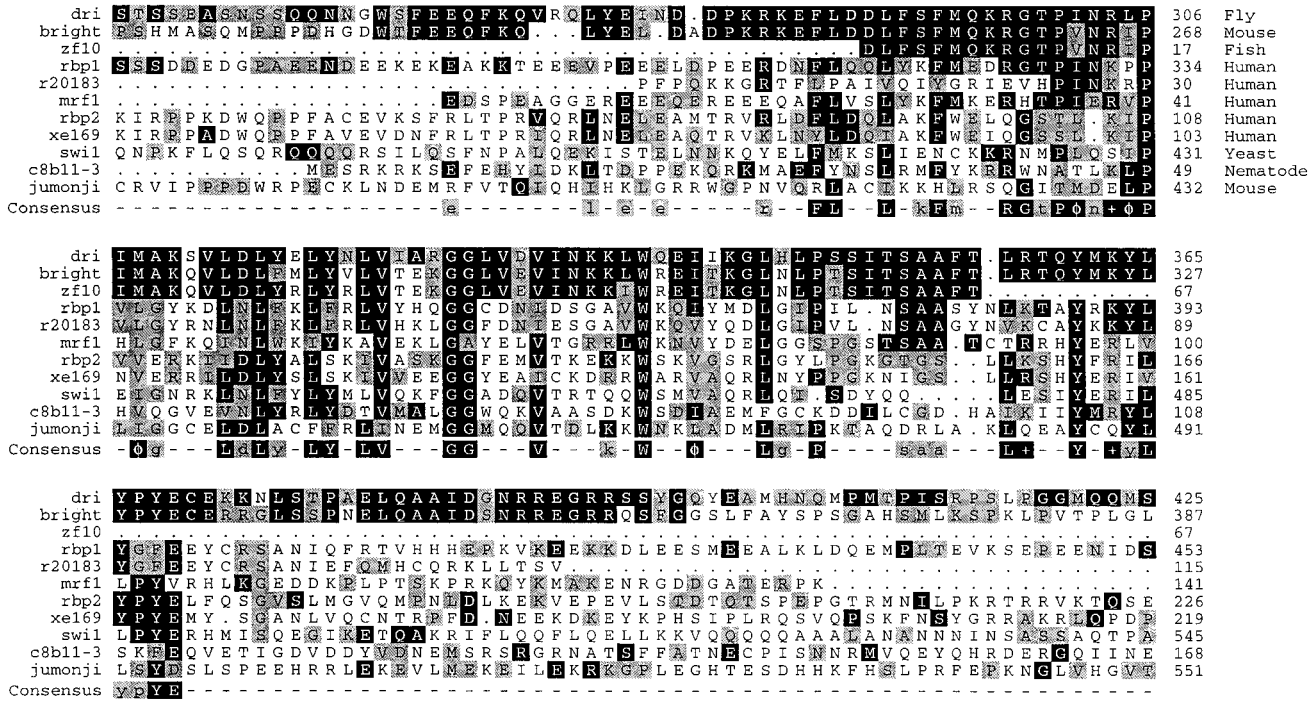

FIG. 2. Alignment of the conserved region in DRI with related proteins. Proteins are listed in order of similarity to the *dead ringer* product. White letters on black indicate residues identical to those in DRI; grey shading indicates residues that are related to those in DRI. Residues are considered related if they both fall into one of the following groups: M, L, I, and V; D, E, N, and Q; R, K, and H; Y, F, and W; and G, S, A, T, and P. The consensus at each position is formed from five or more identical residues or nine or more related residues. The consensus is black and capitalized if seven or more of the proteins have residues identical to the consensus (highly conserved positions). The most widely conserved 83-residue motif corresponds to the section containing residues marked black in the consensus (287 to 369 in DRI). The symbols used are the standard one-letter amino acid code, with + for positively charged and ϕ for hydrophobic positions in the consensus.

to the NP_3 sites, indicating that GST-$DRI_{258–410}$ shows a strong preference for the EN consensus homeodomain binding site (Fig. 3). The reverse experiment showed that TTA_9 binding was only just detectable with elevated protein levels and was effectively competed away by a 10-fold excess of NP_3 (result not shown).

The consensus binding site for the DRI DNA-binding domain resembles an EN homeodomain binding site. While binding to the NP and TTA sequences clearly demonstrates the homeodomain-like properties of the $DRI_{258–410}$ polypeptide, the discovery of the NP site as a binding site was fortuitous (16). The NP site need not have been representative of sites with the highest affinity for the DRI conserved domain. To test this possibility, we determined the preferred binding site for DRI by selection and amplification of random oligomers essentially as described by Wilson et al. (34). Briefly, GST-$DRI_{258–410}$ fusion protein was bound to glutathione-agarose beads and incubated with a pool of oligomers containing a stretch of 20 random bases. Unbound oligomers were washed off, and the remaining pool was amplified by PCR. This process was repeated seven times, and the remaining oligomers were cloned and sequenced. The selected sequences were aligned by inspection to locate the most widely conserved bases within the degenerate region. This process revealed a strongly selected consensus with striking similarity to the EN binding site (Fig. 4). Only 3 of the 45 sequences showed no similarity to this consensus (or to each other—see Materials and Methods). The core bases of many homeodomain binding sites (ATTA) were shown to be highly selected for, with significant selection for the two flanking bases, generating a DRI consensus site of PuATTAA. We could not align any bases other than these to indicate strong selection for a dual or any widely divergent alternative site. However, one of the primers fortuitously contained a site similar to the consensus (NNTTGA), and because this sequence is also found in the NP site (AATTGA), it was possible that many of the oligomers were selected by binding to this site rather than the consensus obtained from the degenerate section. To test this, electrophoretic mobility shift assays were performed with the consensus site and the AATTGA alternative as competitors. Figure 5 shows that 10-fold more AATTGA trimer (track 6) than consensus trimer (track 2) is required to compete away 90% of the protein, suggesting a considerably lower although detectable affinity for this alternative site. Five of the 42 selected sequences contained nothing closer to the other 37 than this PuATTGA site, using four primer bases. This site is clearly bound by the $DRI_{258–410}$ peptide; however, the other 37 contained PuATTAA, chosen from random nucleotides, and this site clearly demonstrated a higher affinity for the protein than AATTGA, strongly suggesting that PuATTAA is the preferred sequence. The selection of oligomers with this consensus site might have been aided by the presence of a potential lower-affinity site in the primer, but this second site is not required (Fig. 5, track 1), nor did it prevent selection of a consensus from the random bases. This consensus must to some extent reflect the in vitro selection conditions and is, of course, based on the assumption that each oligomer is selected by its similarity to some ideal site. Nonetheless, these results are consistent with the DNA-binding experiments described above and those reported previously (16), indicating that the EN consensus sequence is among the highest-affinity binding sites for the DNA-binding domain of DRI. The DNA-binding specificity of the conserved domain of the

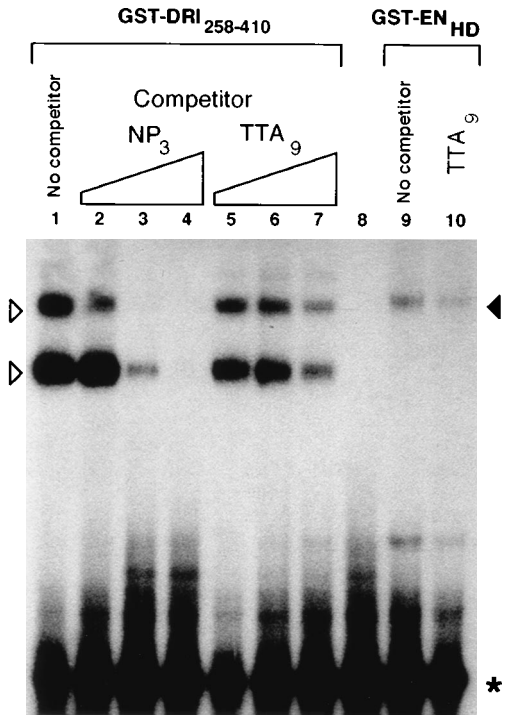

FIG. 3. The conserved domain from DRI specifically binds an Engrailed DNA-binding site. Electrophoretic mobility shift assays show binding of bacterially expressed GST fusion proteins to NP_3, a double-stranded synthetic oligonucleotide trimer of the consensus Engrailed binding site (TCAATTAAATGA). All tracks contain a 1,000-fold excess of nonspecific competitor DNA. Free oligonucleotide is marked with an asterisk. The assays show competition for binding of GST-$DRI_{258–410}$ protein with labeled NP_3 and increasing amounts of unlabeled specific competitor, either NP_3 at a 10-fold (lane 2), 100-fold (lane 3), or 1,000-fold (lane 4) excess or the variant site TTA_9 [(TTATTATTA)×3] at a 10-fold (lane 5), 100-fold (lane 6), or 1,000-fold (lane 7) excess. A 1,000-fold excess of NP_3 competes away binding (track 4), whereas the same concentration of TTA_9 (track 7) has little effect. DRI-bound oligomer migrates at two positions, indicated by the open arrowheads, which suggests the binding of one or two protein molecules per trimer binding site. The preference of the Engrailed homeodomain (GST-EN_{HD}) for the consensus site is shown in tracks 9 and 10, in which a 1,000-fold excess of TTA_9 (track 10) does not eliminate binding to the NP site (solid arrowhead). Track 8 shows that GST alone does not bind to the NP_3 oligonucleotide.

mouse Bright has also recently been identified and contains a core recognition site of PuATa/tAA (15).

***dead ringer* encodes a nuclear protein and is developmentally regulated.** *dri* expression was examined by whole-mount in situ hybridization with *dri* cDNA clones and immunochemical staining with a rat polyclonal antibody raised against a bacterially expressed pGEX-DRI fusion protein. Both approaches revealed a dynamic pattern of expression during embryogenesis. Maternal mRNA was found to be distributed throughout the embryo during the syncytial cleavage divisions (results not shown), while at cellularization, mRNA was found in broad bands at the termini and in a central band (Fig. 6A). At germ band extension, mRNA was found predominantly in the mesoderm (Fig. 6B). DRI protein was found to be nuclearly localized whenever present (Fig. 6 and results not shown). It was found evenly distributed among syncytial nuclei (Fig. 6C). The only instance in which mRNA and protein distribution differed was in late blastoderm embryos. At this stage, the striped appearance of mRNA distribution (Fig. 6A) contrasted with ubiquitous distribution of protein (Fig. 6C), presumably reflecting the persistence of maternal protein after the degradation of maternal mRNA. At germ band extension, protein distribution again reflected mRNA localization, both appearing primarily in the mesoderm (compare Fig. 6B and D). Germ band-retracted embryos exhibited organ-specific expression (Fig. 6E to H), including expression in the pharyngeal muscles, discrete rows of cells in the hindgut epithelium, the amnioserosa (Fig. 6E), the ring gland, a ring of cells at the midgut-hindgut junction, and several distinct cells in the posterior region of each brain lobe (Fig. 6F). Expression was also ob-

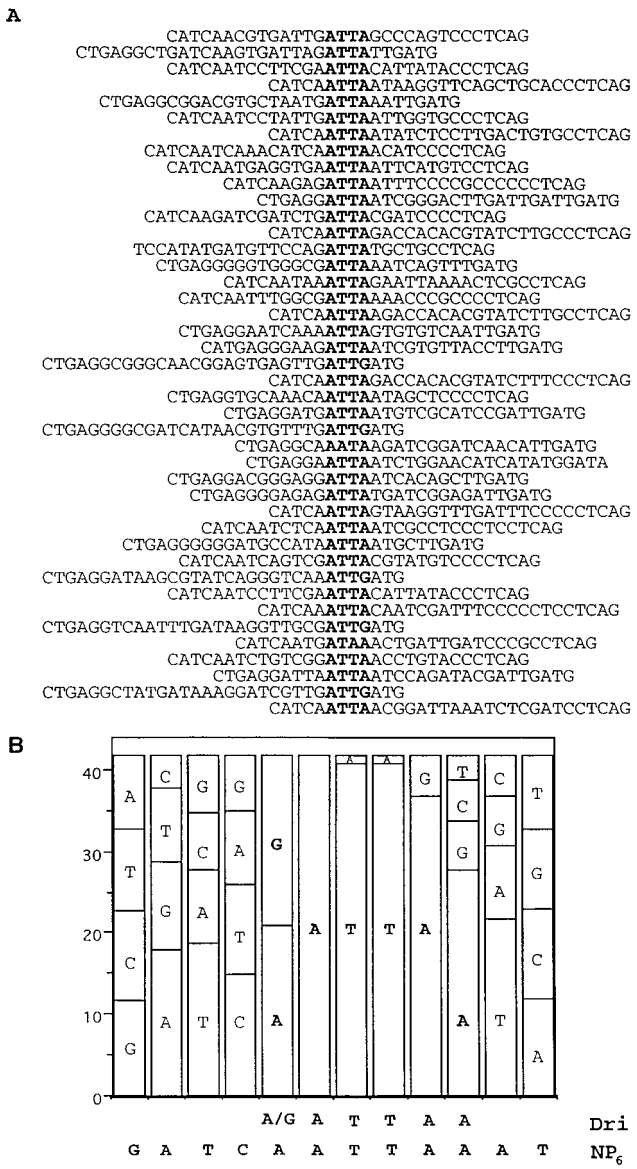

FIG. 4. DNA-binding site consensus for the $DRI_{258–410}$ peptide. (A) Aligned sequences of random oligonucleotides selected by GST-$DRI_{258–410}$ protein binding. The random section of 20 bases plus 6 flanking bases of the primer are shown in each case. These 42 sequences were aligned, giving a consensus of PuATTAA, where Pu stands for either purine base. (B) Consensus diagram to indicate the frequency of each base at each position across the aligned region. Positions at which there has been strong selection have the predominant base(s) marked in boldface. The consensus binding site for the DRI DNA-binding domain is indicated below the figure, aligned with the NP Engrailed binding site, known to specifically bind DRI. The homeodomain "core" binding site ATTA is marked in boldface. Three sequences showed no clear relationship to each other or to the 42 shown above (see Materials and Methods) and are not included, as alignment would be purely arbitrary.

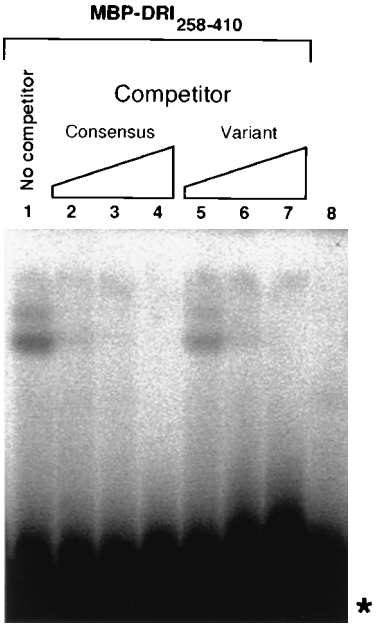

FIG. 5. The conserved domain from DRI specifically binds the identified consensus sequence. Electrophoretic mobility shift assays (EMSAs) show the binding of a bacterially expressed DRI-maltose-binding protein (MBP) fusion to a labeled trimer containing the identified consensus $(CCGATTAATCCC)_3$. All tracks contain a 1,000-fold excess of nonspecific competitor DNA. Free oligonucleotide is marked with an asterisk. EMSAs show competition for binding of MBP-$DRI_{256–410}$ protein with labeled consensus trimer and increasing amounts of unlabeled specific competitor, either the consensus trimer in a 10-fold (track 2) 100-fold (track 3), or 1,000-fold (track 4) excess or the variant site trimer $(CCGATTGATCCC)_3$ in a 10-fold (track 5), 100-fold (track 6), or 1,000-fold (track 7) excess. A 100-fold excess of variant competitor is required to compete to the same level as a 10-fold excess of consensus site (compare tracks 2 and 6), indicating that, although the variant exhibits affinity, there is a clear preference for the consensus site. Track 8 shows that MBP alone does not contribute to the retardation of the labeled oligonucleotide.

served in the cells of the salivary gland duct but not in cells of the salivary gland itself, in a ring of cells at the foregut-midgut junction (Fig. 6G), and in a segmentally repeated pattern in the central nervous system (Fig. 6H).

DISCUSSION

The studies of the *Drosophila dead ringer* (*dri*) gene reported here have led to the identification of a new, highly conserved sequence-specific DNA-binding domain. *dri* has been shown to encode a 901-amino-acid protein containing a domain that confers in vitro sequence-specific DNA-binding activity. Gel retardation experiments showed that a DRI 153-amino-acid peptide but not an internal 83-residue peptide was sufficient for sequence-specific binding. Whole-mount embryos stained with polyclonal antibodies specific for the DRI protein showed that DRI is a nuclear protein, consistent with its in vitro DNA-binding properties.

The in vitro specificity of the DRI DNA-binding domain was shown to be strikingly similar to that of homeodomains that bind the NP Engrailed (EN) consensus binding site. The DRI DNA-binding domain efficiently binds the NP sequence and behaves like an NP-binding homeodomain in competition experiments. Optimal binding site selection experiments showed that the NP consensus EN binding sequence constitutes one of the highest-affinity in vitro binding sites for the DRI DNA-binding domain. Surprisingly, the DRI DNA-binding domain does not exhibit sequence similarity to homeodomains. Computer compilation of the predicted secondary structure of the DRI DNA-binding domain (24) suggested the presence of four helices separated by highly conserved loops, but these also bore no obvious resemblance to the helix-turn-helix structure of homeodomains.

Sequence and database analyses revealed that this DNA-binding domain, although not a homeodomain, is highly conserved. The most highly conserved sequence occurs in the DNA-binding domain of the mouse B-cell-specific regulator Bright (15), which exhibits 75% identity to DRI over a 137-amino-acid region. Striking sequence similarity was also found between the DNA-binding domain and a number of proteins encoded by the yeast, nematode, mouse, and human genomes, defining a new family of genes with a conserved motif. These homologies do not extend as far as the DRI-Bright homology, the region of identity being restricted to an 83-amino-acid region within the 137-amino-acid region of conservation between DRI and Bright. Curiously, the smaller, more widely conserved segments of DRI (this study) and Bright (15) did not exhibit DNA-binding activity in the assays used. The discovery of a DNA-binding domain with the conserved 83-amino-acid motif in the human mrf1 protein (33a) suggests that the flanking sequences can vary widely without compromising the DNA-binding function, perhaps reflecting structural rather than primary sequence conservation in this flanking region.

The gene family identified by conservation of sequences within the DRI DNA-binding domain is an important one, as it includes not only the *Drosophila dri* and mouse Bright genes but also the yeast *SWI1* gene, which encodes a member of the SWI regulatory complex (22); the X-linked but not X-inactivated human gene XE169 (36); and genes that encode the human retinoblastoma binding proteins RBP1 and RBP2 (9), the human DNA-binding proteins MRF1 and MRF2 (33a), the mouse gene jumonji, which is required for neural tube development (31), and several as yet uncharacterized cDNAs from a variety of organisms.

The DRI sequence conservation between *S. cerevisiae*, flies, nematodes, fish, and mammals is very likely to reflect conservation of its DNA-binding function. This has particular significance for the yeast protein SWI1, since the SWI complex interacts with chromatin to enhance the effect of transcriptional activators (6). Neither SWI1 nor any other of the characterized members of this complex have yet been shown to possess sequence-specific DNA-binding activity (22). We predict that SWI1 contributes to the DNA binding of the complex through its DRI homology domain. The recent isolation of *Drosophila* and human protein complexes containing homologs of the SWI complex (23) suggests a possible role for this family of proteins in widespread transcriptional control in eukaryotes. We also predict that the proposed specific DNA-binding activities of RBP1 and RBP2 (9) are conferred by the DRI homology domain of those proteins. It is likely that the conserved domain in these proteins confers or modifies the DNA-binding specificity of the complexes with which they associate.

The homology with the mouse Bright and human MRF proteins that bind to matrix attachment regions (MARs) (15, 33a) may have significance for the mechanism of *dri* action. MARs are often associated with enhancer elements (5, 17, 30), and it has been noted that potential homeodomain binding sites are frequently observed near certain MARs (2). The possibility that this new family of DNA-binding proteins binds in vivo to MAR-associated sites remains to be tested. It should be noted that retinoblastoma protein has been shown to localize to the nuclear matrix during G_1 phase (19), although the involvement of the DRI-homologous region of the two retinoblastoma binding proteins in this process has not been tested.

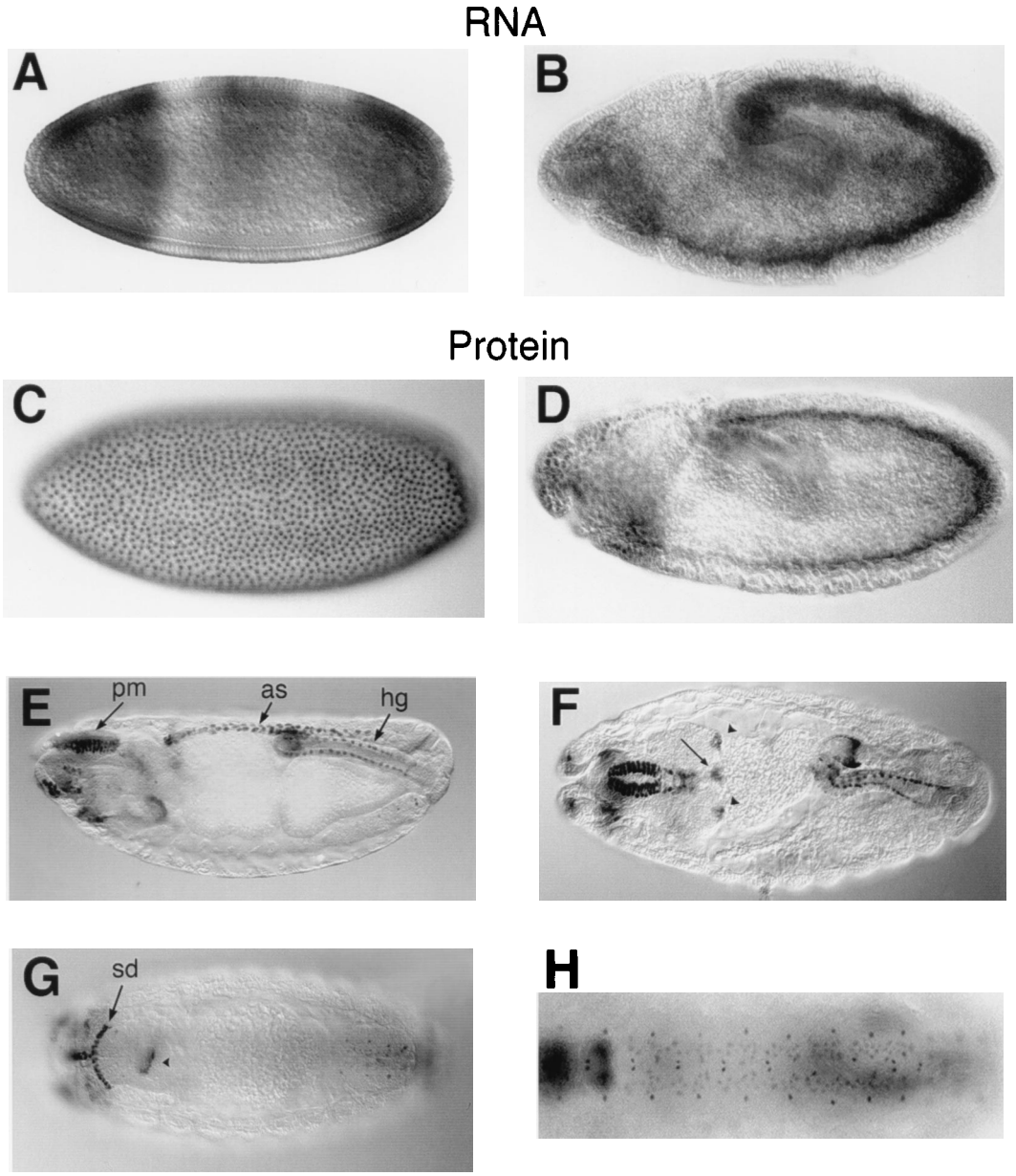

FIG. 6. Expression of *dead ringer* during early development. (A and B) In situ hybridization to *dead ringer* mRNA with a digoxigenin-labeled DNA probe. At 2 to 3 h after fertilization, *dead ringer* is expressed in three bands, one covering each terminus and one broad dorsal stripe across the trunk (A). At 3 to 5 h after fertilization, *dead ringer* mRNA is localized predominantly to the mesoderm (B). (C to E) Immunostaining with polyclonal antibodies against DRI protein. (C) DRI protein is nuclearly localized and found ubiquitously until after gastrulation. (D) The expression of DRI protein after germ band extension is identical to that of its message. (E) In stage 15 to 16 embryos, DRI is expressed in the pharyngeal muscles (pm), amnioserosa (as), and distinct rows of cells in the hindgut (hg). (F) Expression is also seen in the ring gland (arrow), clypeolabrum, a ring of cells around the midgut-hindgut junction, and several cells in each lobe of the brain (arrowheads). (G and H) From a ventral view, DRI expression can be seen in the salivary ducts (sd) and rings of cells at the foregut-midgut junction (arrowhead) (G) and in a repeated pattern through the central nervous system (H).

dri is expressed in a complex developmental pattern that suggests a variety of possible functions during development. The ubiquitous early distribution of maternal *dri* mRNA and protein is followed by mesoderm-specific expression during germ band extension and then zygotic expression in a range of different tissue types, including salivary gland ducts, pharyngeal muscles, discrete rows of cells in the hindgut epithelium, amnioserosa, ring gland, rings of cells at the midgut-hindgut and foregut-midgut junctions, the posterior region of the brain lobes, and some cells of the central nervous system. We anticipate that DRI is a transcription factor required for the specification of these tissues and are in the process of generating *dead ringer* mutants to test this hypothesis.

The in vitro DNA-binding properties of DRI may provide a further clue to *dri* function. The similarity in the DNA-binding properties of the DRI DNA-binding domain to those of a number of homeodomains is interesting for two reasons. The first is that it compounds the question of how different homeobox genes carry out radically different biological functions when they encode proteins with similar in vitro DNA-binding

specificities by adding a member of an entirely new class of proteins to the group that bind such sequences. The second is that it raises the possibility that DRI acts by cooperative or competitive interactions with homeodomain proteins. Interactions between sequence-specific DNA-binding proteins and homeodomain proteins have been described for humans, *D. melanogaster*, and *S. cerevisiae* (13, 28, 29). The idea that DRI may be cooperating with one or other homeodomain proteins is made more plausible by an intriguing observation made about the yeast homolog SWI1. The SWI complex interacts with chromatin and enhances the activity of specific transcription factors (6). Furthermore, it can interact with homeodomain proteins, since the *Drosophila* homeodomain protein encoded by the fushi tarazu gene expressed in *S. cerevisiae* loses its ability to transactivate in SWI1$^-$ yeast cells (22). While the expression pattern does not point to any particular homeodomain protein as a potential target for DRI interaction or competition, the possibility of such an interaction provides a focus for future studies of *dead ringer.*

ACKNOWLEDGMENTS

We thank R. Whitson, K. Itakura, M. Kaplan, and P. Tucker for communicating unpublished work and P. Wigley and the reviewers for helpful comments on the manuscript.

S.L.G. and R.D.K. were supported by Australian Research Council Postgraduate Research Awards. B.K. was supported by a National Health and Medical Research Council C. J. Martin Fellowship. This work was supported by Australian Research Council grants to R.S. and B.K.

REFERENCES

1. **Beachy, P. A., M. A. Krasnow, E. R. Gavis, and D. S. Hogness.** 1988. An Ultrabithorax protein binds sequences near its own and the Antennapedia P1 promoters. Cell **55:**1069–1081.
2. **Boulikas, T.** 1993. Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human β-globin gene complex. J. Cell. Biochem. **52:**23–36.
3. **Cavener, D. R.** 1987. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Res. **15:** 1353–1361.
4. **Chan, S.-K., L. Jaffe, M. Capovilla, J. Botas, and R. S. Mann.** 1994. The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell **78:**603–615.
5. **Cockerill, P. N., and W. T. Garrard.** 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell **44:**273–282.
6. **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.
7. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1988. The sequence specificity of homeodomain-DNA interaction. Cell **54:**1081–1090.
8. **Faisst, S., and S. Meyer.** 1992. Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res. **20:**3–26.
9. **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 R. E. Jones.** 1993. Characterization of the retinoblastoma binding proteins RBP1 and RBP2. Oncogene **8:**3149–3156.
10. **Foe, V. E.** 1989. Mitotic domains reveal early commitment of cells in *Drosophila* embryos. Development **107:**1–22.
11. **Furukubo-Tokunaga, K., M. Muller, M. Affolter, L. Pick, U. Kloter, and W. J. Gehring.** 1992. In vivo analysis of the helix-turn-helix motif of the fushi tarazu homeo domain of *Drosophila melanogaster*. Genes Dev. **6:**1082–1096.
12. **Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting, and K. Wüthrich.** 1994. Homeodomain-DNA recognition. Cell **78:**211–223.
13. **Grueneberg, D. A., S. Natesan, C. Alexandre, and M. Z. Gilman.** 1992. Human and *Drosophila* homeodomain proteins that enhance the DNA-binding activity of serum response factor. Science **257:**1089–1095.
14. **Harrison, S. C.** 1991. A structural taxonomy of DNA-binding domains. Nature (London) **353:**715–719.
15. **Herrscher, R. F., M. H. Kaplan, D. L. Lelsz, C. Das, R. Scheuermann, and P. W. Tucker.** The immunoglobulin heavy chain MARs are bound by Bright: a B-cell specific transactivator that describes a new DNA binding protein family. Genes Dev., in press.
16. **Kalionis, B., and P. H. O'Farrell.** 1993. A universal target sequence is bound in vitro by diverse homeodomains. Mech. Dev. **43:**57–70.
17. **Klehr, D., K. Maass, and J. Bode.** 1991. Scaffold-attached regions from the human interferon beta domain can be used to enhance the stable expression of genes under the control of various promoters. Biochemistry **30:** 1264–1270.
18. **Lockett, T. J., P. A. B. Moretti, J. Lake, and R. Saint.** 1993. rough genes with Deformed homeobox substitutions exhibit rough biological specificity during *Drosophila* eye development. Mech. Dev. **41:**57–68.
19. **Mancini, M. A., B. Shan, J. A. Nickerson, S. Penman, and W.-H. Lee.** 1994. The retinoblastoma gene product is a cell cycle-dependent, nuclear matrix-associated protein. Proc. Natl. Acad. Sci. USA **91:**418–422.
20. **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.
21. **Patel, N. H., and C. S. Goodman.** 1992. Whole mount in situ hybridization with a non-radioactive probe, p. 377–381. *In* C. Kessler (ed.), Non-radioactive labeling and detection of biomolecules. Springer-Verlag, Berlin.
22. **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.
23. **Peterson, C. L., and J. W. Tamkun.** 1995. The SWI-SNF complex:a chromatin remodeling machine? Trends Biol. Sci. **20:**143–146.
24. **Rost, B., and C. Sander.** 1994. Combining evolutionary information and neural networks to predict protein secondary structure. Proteins **19:**55–72.
25. **Scott, M. P., J. W. Tamkun, and G. Hartzell III.** 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta **989:**25–48.
26. **Shaw, D. R., H. Richter, R. Giorda, T. Ohmachi, and H. L. Ennis.** 1989. Nucleotide sequences of Dictyostelium discoideum developmentally regulated cDNAs rich in (AAC) imply proteins that contain clusters of asparagine, glutamine, or threonine. Mol. Gen. Genet. **218:**453–459.
27. **Smith, D., and K. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67:**31–40.
28. **Smith, D. L., and A. D. Johnson.** 1992. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an alpha 2 dimer. Cell **68:**133–142.
29. **Stern, S., and W. Herr.** 1991. The herpes simplex virus trans-activator VP16 recognizes the Oct-1 homeo domain: evidence for a homeo domain recognition subdomain. Genes Dev. **5:**2555–2566.
30. **Stief, A., D. M. Winter, W. H. Stratling, and A. E. Sippel.** 1989. A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature (London) **341:**343–345.
31. **Takeuchi, T., Y. Yamazaki, Y. Katoh-Fukui, R. Tsuchiya, 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.
32. **van Dijk, M. A., and C. Murre.** 1994. extradenticle raises the DNA binding specificity of homeotic selector gene products. Cell **78:**617–624.
33. **Wharton, K. A., B. Yedbovnick, V. G. Finnerty, and S. Artavanis-Tsakonas.** 1985. opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in *Drosophila melanogaster*. Cell **40:**55–62.

33a. **Whitson, R. H., T. H. Huang, B. W. Merrills, T. Asai, and K. Itakura.** Submitted for publication.

34. **Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan.** 1993. Cooperative dimerization of paired class homeo domains on DNA. Genes Dev. **7:** 2120–2134.
35. **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, A. Fraser, L. Fulton, A. Gardner, P. Green, T. Hawkins, L. Hillier, M. Jier, L. Johnston, M. Jones, J. Kershaw, J. Kirsten, N. Laisster, P. Latreille, J. Lightning, C. Lloyd, B. Mortimore, M. O'Callaghan, J. Parsons, C. Percy, L. Rifken, A. Roopra, D. Saunders, R. Shownkeen, M. Sims, N. Smaldon, A. Smith, M. Smith, E. Sonnhammer, R. Staden, J. Sulston, J. Thierry-Mieg, K. Thomas, M. Vaudin, K. Vaughan, R. Waterson, A. Watson, L. Weinstock, J. Wilkinson-Sprout, and P. Wohldman.** 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. Nature (London) **368:**32–38.
36. **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. Human Mol. Genet. **3:**153–160.
37. **Yedvobnick, B., D. Smoller, P. Young, and D. Mills.** 1988. Molecular analysis of the neurogenic locus mastermind of *Drosophila melanogaster*. Genetics **118:**483–497.