Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID)

Junji Iwahara and Robert T.Clubb¹

Department of Chemistry and Biochemistry, Molecular Biology Institute and the UCLA-DOE Laboratory of Structural Biology and Genetics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1570, USA

¹Corresponding author
e-mail: rclubb@mbi.ucla.edu

The Dead ringer protein from *Drosophila melanogaster* is a transcriptional regulatory protein required for early embryonic development. It is the founding member of a large family of DNA binding proteins that interact with DNA through a highly conserved domain called the AT-rich interaction domain (ARID). The solution structure of the Dead ringer ARID (residues Gly262–Gly398) was determined using NMR spectroscopy. The ARID forms a unique globular structure consisting of eight α -helices and a short two-stranded anti-parallel β -sheet. Amino acid sequence homology indicates that ARID DNA binding proteins are partitioned into three structural classes: (i) minimal ARID proteins that consist of a core domain formed by six α -helices; (ii) ARID proteins that supplement the core domain with an N-terminal α -helix; and (iii) extended-ARID proteins, which contain the core domain and additional α -helices at their N- and C-termini. Studies of the Dead ringer–DNA complex suggest that the major groove of DNA is recognized by a helix–turn–helix (HTH) motif and the adjacent minor grooves are contacted by a β -hairpin and C-terminal α -helix. Primary homology suggests that all ARID-containing proteins contact DNA through the HTH and hairpin structures, but only extended-ARID proteins supplement this binding surface with a terminal helix.

Keywords: AT-rich interaction domain/Dead ringer/DNA binding domain

Introduction

The Dead ringer protein from *Drosophila melanogaster* is an essential transcriptional regulatory protein that interacts with DNA through a highly conserved domain called the AT-rich interaction domain (ARID) (Herrscher *et al.*, 1995; Gregory *et al.*, 1996). The ARID motif is unrelated to any of the common domains used to recognize DNA (homeodomains, zinc fingers, etc.) and presumably interacts with the minor groove (Herrscher *et al.*, 1995; Gregory *et al.*, 1996; Valentine *et al.*, 1998). ARID-containing proteins are present in a variety of eukaryotic organisms and have been shown to participate in several biologically significant processes. Members of the family include, among others, the yeast SWI1 protein, from the SWI–SNF complex involved in global transcriptional activation

(O'Hara *et al.*, 1988); Bright, a B-cell-specific *trans*-activator of IgH transcription (Herrscher *et al.*, 1995); RBP1 and RBP2, retinoblastoma binding factors (Fattaey *et al.*, 1993); PLU-1, a protein that is upregulated in breast cancer cells (Lu *et al.*, 1999); and the mammalian proteins Jumonji (Motoyama *et al.*, 1997), SMCx (Agulnik *et al.*, 1994b), SMCy (Agulnik *et al.*, 1994a), Mrf-1 and Mrf-2 (Huang *et al.*, 1996). Dead ringer is itself highly conserved, with specific orthologs in mammalian genomes (mouse and human), suggesting that it plays a fundamental role in embryonic development.

The Dead ringer protein plays an essential role in early anterior–posterior patterning and muscle development in the *Drosophila* embryo (Shandala *et al.*, 1999). Its mutation in embryonic cells leads to segmentation and head defects and ectopic cephalic furrow formation. These phenotypes appear to be caused by Dead ringer-dependent changes in the expression levels of several key developmental genes including *engrailed*, *wingless*, *even-skipped*, *argos* and *buttonhead*. The Dead ringer protein is a sequence-specific DNA binding protein (Gregory *et al.*, 1996) and mutational studies suggest that it can act as either an activator or a repressor of transcription (Shandala *et al.*, 1999). For example, in mutant *dri* embryos the mutant head phenotype and the presence of ectopic cephalic furrows have been correlated with the reduced expression of *argos* and the derepression of *buttonhead*, respectively.

Dead ringer may also participate in dorsal–ventral axis formation in the early embryo. Repression of the dorsal ectoderm-determining gene *zerknüllt* (*zen*) requires an upstream silencer called the ventral repression region (VRR) (Doyle *et al.*, 1989; Ip *et al.*, 1991). VRR-mediated repression of heterologous genes requires conversion of Dorsal into a repressor by the Cut and Dead ringer proteins (Figure 1) (Ip *et al.*, 1991), which both interact with the conserved nucleotide sequence TATTGAT within sites AT2 and AT3 (Valentine *et al.*, 1998). Dead ringer appears to convert Dorsal to a repressor by cooperatively recruiting the global co-repressor Groucho to the VRR (Valentine *et al.*, 1998). The function of the Cut protein in Groucho recruitment is unknown; however, it may repress transcription through a different mechanism, since human homologs of Cut have been shown to repress transcription actively through an alanine-rich C-terminal region that is also present in the *Drosophila* Cut protein (Dufort and Nepveu, 1994; Mailly *et al.*, 1996). The mechanism of transcriptional silencing by the *zen* VRR is complex and additional proteins and DNA sequences are probably involved. For example, mutation of the AT1 site also causes a derepressed phenotype. However, this site does not contain the conserved binding sequence for the Cut and Dead ringer proteins and it presumably interacts with an as yet unidentified protein (Jiang *et al.*, 1993). In addition, binding sites for the Dorsal switch protein 1 (DSP1) (Lehming *et al.*,

zen VRR

Fig. 1. Schematic of the VRR from the *zen* gene. Dorsal protein binding sites are shown as hatched rectangles (dl1–dl3). Sites AT1–AT3 contain semi-conserved AT-rich sequences that have been shown to interact with proteins (dark squares). The Dead ringer and Cut proteins bind to sites AT2 and AT3. The cognate protein for site AT1 has not been identified. Binding sites for the Dorsal switch protein 1 (NRE) and the NTF-1/Elf-1 protein (DRE) are represented by open and closed diamonds, respectively. Circles correspond to GC-rich sequences that interact with an as yet unidentified protein. The displayed region is located –1.17 to –1.35 kb from the start of transcription.

1994) and the NTF-1/Elf-1 protein (Huang *et al.*, 1995) have been mapped to the VRR.

Here we present the solution structure of the extended-ARID from the Dead ringer protein. Surprisingly, we find that the ARID of Dead ringer adopts a heretofore unseen three-dimensional structure that is markedly different from the previously reported structure of the related protein Mrf-2 (Yuan *et al.*, 1998). Our results suggest that Dead ringer recognizes the major groove of DNA with residues located in a helix–turn–helix (HTH) motif and contacts the adjacent minor grooves with a β -hairpin and C-terminal α -helix. Amino acid sequence conservation suggests that all ARID-containing proteins will contact DNA through the HTH and hairpin structures, but only extended-ARID proteins will supplement this binding surface with a terminal helix.

Results

Structure of the extended ARID from Dead ringer

The solution structure of residues Gly262–Gly398 of the DNA binding domain from the Dead ringer protein (DRI-DBD) was determined using heteronuclear NMR spectroscopy. A total of 2140 restraints were obtained from an analysis of the NMR data and include: 1662 NOEs, 305 dihedral angle restraints, 57 $^3J_{\text{HN}\alpha}$ coupling constant restraints and 116 hydrogen bond distance restraints (two restraints per hydrogen bond, added during the final stages of refinement). Distance geometry and simulated annealing calculations were employed to generate an ensemble of 20 conformers consistent with the NMR data (Figure 2A and B). These structures exhibit good covalent geometry and have no NOE, scalar coupling or dihedral angle violations greater than 0.5 Å, 2 Hz or 5°, respectively. Complete restraint and structural statistics are presented in Table I. The structure of the DRI-DBD is well ordered from residues Ser264 to Leu344 and Thr351 to Asn388 with a root mean square (r.m.s.) deviation between the atomic coordinates of the backbone atoms and all heavy atoms of these residues to the average coordinates of 0.61 ± 0.16 and 1.10 ± 0.15 Å, respectively. Residues at the polypeptide termini and in an internal loop (residues His345–Ile350) are largely unstructured and dynamic as judged by the small magnitude of their ^1H - ^{15}N heteronuclear NOEs (data not shown).

The DNA binding domain of Dead ringer consists of eight α -helices and a short β -hairpin (Figure 2C). The secondary structural elements are arranged in a H1-H2-

B1-B2-H3-H4-H5-loop-H6-H7-H8 topology (H, α -helix; B, β -strand). These elements are organized into two distinct sub-domains that rest against each other to form two lobes of an otherwise globular structure. The smaller of the two sub-domains (called region I) is formed by residues in helices H1 (Phe265–Tyr276), H2 (Pro282–Met295) and a short two-stranded anti-parallel sheet (strands B1, Ile307–Met308; B2, Ser311–Val312). In region I, helices H1 and H2 form a v-shape and are positioned anti-parallel to each other at a $\sim 20^\circ$ angle. The helices are packed against the β -sheet and an ordered extended strand (residues Gln296–Pro306) that connects H2 to strand B1. Both the sheet and preceding strand are anchored to the first two helices by hydrophobic contacts to residues in helix H1, while several hydrophobic contacts from residues in helix H2 and the ordered strand anchor region I to the remainder of the polypeptide. The second lobe of the structure (region II) consists of a bundle of six α -helices. Helix H3 (Leu315–Ala324) leads away from region I and is circled by helices H4 (Leu328–Lys334), H5 (Trp337–Gly343), H6 (Ala353–Tyr364) and H7 (Tyr366–Lys373) which pack against H3 at angles of $\sim 55^\circ$, $\sim 130^\circ$, $\sim 115^\circ$ and $\sim 50^\circ$, respectively. The final α -helix, H8 (residues Pro378–Asn388), is positioned on the outside of the helical bundle and is parallel and adjacent to helix H4. The junction point between the two sub-domains is extensive and buries ~ 1230 Å² of solvent-accessible surface area. The interface is comprised of residues from helix H3, H6 and H7 from region II, which contact residues in region I that are located in helix H2, the adjacent extended strand, and the C-terminal end of helix H1. The hydrophobic core of the DRI-DBD is comprised of amino acids from both sub-domains and is continuous (Figure 2B).

Determination of the surface on the Dead ringer protein used to bind DNA

In order to gain insights into the molecular basis of DNA binding by the Dead ringer protein we studied its complex with DNA using NMR spectroscopy. A 1:1 DRI-DBD–DNA complex was formed using ^{15}N , ^{13}C -labeled DRI-DBD protein and a DNA 15mer that contains its cognate DNA sequence. The NMR spectra of the DRI-DBD–DNA complex were well dispersed enabling the application of double- and triple-resonance NMR techniques to assign unambiguously the backbone atoms of 128 of the 133 non-proline residues within the protein (Figure 3A). Although the spectra were adequate for backbone assignments they

Fig. 2. The NMR solution structure of the DRI-DBD. (A) Cross-eyed stereo picture of the ensemble of 20 structures of the Dead ringer DBD. The backbone atoms (C, C α and N) of residues Ser264 to Asn388 are shown. The ensemble was obtained by superimposing the backbone atoms of Ser264–Leu344 and Thr361–Asn388. The structurally disordered loop between helices H5 and H6 (His345–Ile350) is colored green. (B) Cross-eyed stereo picture of the ensemble of structures showing the backbone and all ordered side chain residues within the hydrophobic core. The displayed side chains include: Val272, Leu275, Phe287, Leu288, Leu291, Phe292, Met295, Pro301, Ile302, Leu305, Leu315, Tyr316, Leu318, Tyr319, Val322, Leu328, Val329, Val331, Ile332, Trp337, Ile340, Ile341, Ala353, Leu357, Tyr364, Leu365, Tyr361, Tyr366, Leu365, Tyr368, Leu375, Leu381, Ala384 and Ile385. The side chains are colored purple and the backbone atoms are colored as in (A). (C) Ribbon drawing of the energy-minimized average coordinates of the DRI-DBD. All secondary structural elements are labeled. Color code: helices, purple β -sheet, red; coil, black.

were not good enough for a structure determination of the complex. A large number of resonances exhibited broadening, presumably as a result of conformation exchange that is intermediate on the chemical shift time-scale. The DNA binding surface was mapped by comparing the backbone chemical shifts of the DRI-DBD protein in

the DNA-free state with those of the DRI-DBD–DNA complex. In this analysis it is anticipated that residues proximal to DNA will exhibit large changes in their chemical shifts as a result of DNA binding, since the chemical shift of a nucleus depends on its magnetic environment. Figure 3B shows the energy-minimized

Table I. Structural statistics^a

	<SA>	(SA) <i>r</i>
R.m.s. deviations from NOE interproton distance restraints (Å) ^b		
All (1662)	0.034 ± 0.003	0.032
Inter-residue sequential ($ i-j = 1$) (427)	0.026 ± 0.005	0.035
Inter-residue short range ($1 < i-j < 5$) (404)	0.044 ± 0.005	0.050
Inter-residue long range ($ i-j \geq 5$) (359)	0.045 ± 0.005	0.031
Intra-residue (472)	0.014 ± 0.003	0.017
R.m.s. deviations from hydrogen bonding restraints (Å) (116)	0.030 ± 0.007	0.013
R.m.s. deviations from experimental dihedral angle restraints (°) (305) ^c	0.058 ± 0.055	0.001
R.m.s. deviations from experimental $^3J_{\text{HN}\alpha}$ coupling constants (Hz) (57)	0.843 ± 0.031	0.704
Deviations from idealized covalent geometry		
Bonds (Å)	0.004 ± 0.0001	0.005
Angles (°)	0.528 ± 0.018	0.666
Impropers (°)	0.406 ± 0.017	0.536
Measures of structure quality		
PROCHECK-NMR ^d		
Most favored regions (%)	85.2 ± 2.3	80.9
Additional allowed regions (%)	12.4 ± 2.0	15.7
Generously allowed regions (%)	1.6 ± 1.1	2.6
Disallowed regions (%)	0.8 ± 0.5	0.9
Coordinate precision (Å) ^e		
Protein backbone	0.61 ± 0.16	
Heavy atoms	1.10 ± 0.15	

^aThe notation of the NMR structures is as follows: <SA> are the final 20 simulated annealing structures; (SA)*r* is the restrained regularized mean structure obtained by averaging the coordinates of the individual SA structures best fitted to one another with respect to residues Ser264–Leu344 and Thr351–Asn388, and restrained regularization of this mean structure. The number of terms for each restraint is given in parentheses. Residues at the C-terminus of the DRI-DBD (Arg390–Gly398) and two residues at the N-terminus that originate from the expression vector (Gly-Ser) displayed no long-range NOE cross-peaks in the data and were omitted from the final simulated annealing calculations.

^bNone of the structures exhibited distance violations >0.5 Å, dihedral angle violations >5° or $^3J_{\text{HN}\alpha}$ coupling constant violations >2 Hz.

^cThe experimental dihedral angle restraints comprised 105 ϕ , 92 ψ , 61 χ_1 and 27 χ_2 angle restraints.

^dLaskowski *et al.* (1996).

^eThe coordinate precision is defined as the average atomic r.m.s. deviation between the 20 individual SA structures and the mean coordinates. The reported values are for residues Ser264–Leu344 and Thr351–Asn388. The backbone value refers to the N, C and CO atoms.

structure of the DRI-DBD with the largest chemical shift changes mapped onto the structure (backbone atoms are colored green if the ^1H or ^{15}N nuclei exhibit absolute shift changes >0.35 or 1.0 p.p.m., respectively). The largest chemical shift changes occur in residues that are located in three regions of the protein: (i) helix H6 and the preceding H5/H6 loop; (ii) residues in and immediately adjacent to the two-stranded β -sheet; and (iii) residues at the C-terminus. Inspection of the structure reveals that these residues all cluster on a single face of the protein, suggesting that they constitute the DNA binding surface. Figure 3C shows the electrostatic surface potential of the DRI-DBD from a similar vantage point to that shown in Figure 3B. The surface of this face of the protein contains several basic residues that form two positively charged patches. In contrast to these basic surfaces that roughly coincide with the binding surface determined by NMR, the opposite face of the protein is primarily negatively charged (Figure 3D) and not affected by DNA binding (rear face of Figure 3B).

Changes in the major groove alter binding affinity

It has been suggested that ARIDs represent a family of minor-groove DNA binding proteins. Evidence for this is based on the observation that the DNA binding affinity of the Dead ringer protein is substantially reduced in the presence of distamycin (Gregory *et al.*, 1996) and Dead ringer does not UV cross-link to 5-iodouracil-substituted DNA (Valentine *et al.*, 1998). Distamycin sensitivity has also been observed in the ARID-containing protein Bright

(Herrscher *et al.*, 1995) and in the SWI–SNF complex, which contains the ARID protein SWI1 (Quinn *et al.*, 1996). In order to gain a greater understanding of the mechanism of DNA binding by the ARID family, we performed a series of binding experiments with wild-type and mutated DNA fragments to determine whether major-groove contacts contribute to DNA binding affinity (Figure 4). Initially, we studied a DRI-DBD–DNA complex formed with a DNA 15mer (d-CGAATATTG-ATTGGG/d-CCCAATCAATATTTCG) that contains the DRI-DBD cognate binding site within the VRR (AT2 and AT3 sites). The results of a gel retardation binding assay shown in Figure 4B indicate that the DRI-DBD binds to this DNA fragment in a sequence-specific manner. Challenge with non-specific DNA does not disrupt the DRI-DBD–DNA complex (Figure 4B, lanes 1 and 2), while challenge with unlabeled DNA containing the consensus binding site disrupts the complex completely (lanes 3 and 4). We next investigated whether major-groove contacts are important for binding affinity. Two mutated oligonucleotides were tested for their ability to disrupt the specific DRI-DBD–DNA complex. These nucleotides replace several thymine bases within the consensus site with uracil, which effectively replaces a methyl group in the major groove with a hydrogen atom. Mutant 1 removes two adjacent methyls at the center of the conserved heptamer at base pairs 3 and 4 (see Figure 4A for numbering). When used to challenge the DRI-DBD–DNA complex, mutant 1 does not disrupt the complex as well as the wild-type sequence (compare lanes 5 and 3), suggesting

Fig. 3. NMR studies of the DRI-DBD–DNA complex. (A) Overlay of the ^{15}N – ^1H HSQC spectra of the 1:1 DRI-DBD–DNA complex (cross-peaks in red) and the free DRI-DBD (cross-peaks in black). Nearly all of the backbone amide groups could be assigned in the 26 kDa DRI-DBD–DNA complex with the exception of Thr351, Ser352 and the N-terminal dipeptide encoded by the expression vector. (B) Ribbon drawing of the DRI-DBD with residues that exhibit large chemical shift changes as a result of DNA binding colored green. A residue was considered to be significantly affected by DNA binding if the absolute chemical shift difference between its free and complex amide ^{15}N or ^1H nuclei was ≥ 1.0 or 0.35 p.p.m., respectively. The backbone nuclei of Thr351 and Ser352 are colored yellow. These residues are either broadened beyond detection or exceed the chemical shift threshold described above. (C) Electrostatic surface plot of the DRI-DBD. The view is identical to (B) and demonstrates that the surface of the DRI-DBD protein most affected by DNA binding is also positively charged. Basic and acidic residues are colored blue and red, respectively. (D) Identical to (C) except that the protein has been rotated by 180°.

that it has reduced affinity for the DRI-DBD. In contrast, removal of the methyl group at position 7 within the consensus sequence has little effect on binding, since this mutant disrupts the DRI-DBD–DNA complex as well as the wild-type sequence (lanes 7 and 8). These results show that the DRI-DBD used in our NMR studies binds DNA sequence specifically and that changes within the major groove alter binding affinity. It appears likely that one or both of the central methyl groups are contacted directly by the protein in the protein–DNA complex and that their removal reduces binding affinity. However, it is conceivable that the conformational properties of the mutated duplex may indirectly reduce its ability to bind protein. For example, it is possible that the DNA binding site is distorted in the complex and removal of the central

methyl groups results in a duplex that can not as readily adopt this distorted conformation.

Discussion

ARID-containing proteins interact with DNA through an HTH motif

Analysis of the DRI-DBD structure reveals that it is unique. A search of the protein data bank for structural homologs using the program DALI (Holm and Sander, 1996) found no proteins that were structurally homologous to the full-length DRI-DBD. The best matches were homologous to at most 36% of the residues in the DRI-DBD fold (46 residues) and had Z-scores < 4.1 . Interestingly, the DALI analysis did reveal that helices

Fig. 4. Gel retardation studies of the DRI-DBD-DNA complex. (A) DNA sequences assayed for their ability to disrupt the DRI-DBD-DNA complex. The sequences enclosed in the box correspond to either the wild-type AT2/AT3 site that contains the cognate binding site for the DRI-DBD or two mutants of this site that replace either one or two thymine bases with uracil. (B) Gel retardation assay of DRI-DBD DNA binding. Each lane measures the formation of the 32 P-labeled DRI-DBD-DNA complex in the presence of varying amounts of cold competitor DNA. In these reactions the protein was added last to a binding reaction mixture that contained unlabeled competitor DNA that was either 10 or 30 times more concentrated than the radiolabeled DNA probe. Lanes 1 and 2 show the effects of adding either a 10- or 30-fold excess of a 15mer that does not contain the specific DRI-DBD binding site. Lanes 3 and 4, 5 and 6 and 7 and 8 show the effects of adding a 10- or 30-fold excess of a 15mer that contains the wild-type AT2/AT3 site, mutant sequence 1 and mutant sequence 2, respectively.

H5 and H6 in the DRI-DBD are structurally homologous to the HTH DNA binding motif. The HTH is a ubiquitous DNA binding motif that was originally identified in several prokaryotic repressor proteins. More recently, a systematic comparison has shown that all HTH-containing DNA binding proteins actually consist of at least three α -helices, the two helices of the HTH motif and a third α -helix that stabilizes this unit (Suzuki and Brenner, 1995; Wintjens and Rooman, 1996). For clarity, we refer to these three helices as an expanded HTH motif. The DALI search revealed that the DRI-DBD helices H3, H5 and H6 adopt a structure that is similar to the expanded HTH motif. Figure 5 shows a comparison of the DRI-DBD protein with three other structurally homologous HTH-containing proteins identified in the DALI search. The structure of the DRI-DBD represents an atypical HTH protein. To the best of our knowledge, it is the only HTH protein of known structure that has an α -helix inserted between the three helices that comprise the expanded HTH unit. As such, the ARID proteins represent another new sub-class of this binding motif. The turn in the putative HTH unit

Fig. 5. Comparison of the DRI-DBD with three homologous HTH-containing DNA binding proteins. The structures of the DRI-DBD, histone H5 (1st-A) (Ramakrishnan *et al.*, 1993), the Mu transposase I β DBD (2ezk) (Schumacher *et al.*, 1997) and the first repeat of the Myb proto-oncogene DBD (1mbe) (Ogata *et al.*, 1995) are displayed. Color code: red, helices of the HTH; gold, the third stabilizing helix; cyan, the turn in the HTH.

of the DRI-DBD is also somewhat unusual, in that it is longer than most (nine residues) and structurally disordered (Figure 2). Although unusual, the features of this turn are not unprecedented and structural studies of other HTH units have revealed large variations in the lengths of the connecting turn and the turn has also been shown to be structurally disordered in several NMR-derived structures (Donaldson *et al.*, 1996; Furui *et al.*, 1998). Members of the ARID family are highly conserved and a sequence alignment indicates that they will all interact with DNA through the HTH motif.

ARID-containing proteins are structurally diverse

The structure of DRI-DBD provides insights into the highly conserved family of ARID-containing DNA binding proteins. Based on primary sequence homology, more than 30 eukaryotic proteins interact with DNA through the ARID structure. Figure 6A displays a structure-based sequence alignment of the ARID family. An inspection of this alignment suggests that the ARID family is structurally diverse with members possessing six to eight α -helices. All ARID proteins contain six conserved helices that correspond to helices H2-H7 in the structure of the DRI-DBD (colored blue in Figure 6A). This core domain contains the putative HTH DNA binding motif and the β -hairpin structure that presumably interact with DNA. Interestingly, approximately half of the ARID proteins supplement the core region with an N-terminal α -helix as seen in the DRI-DBD structure (helix H1 in the DRI-DBD). As shown in Figure 6B, several residues in the

Fig. 6. (A) Sequence alignment of the Dead ringer protein with other members of the ARID family of DNA binding proteins. The sequences are divided into two classes. The sequences of the upper set are expected to contain helix H1 and the those of the bottom set are not expected to contain this helix. The secondary structural elements of the DRI-DBD are indicated above the sequence. Residues are shaded blue if they participate in the hydrophobic core used to construct the central core domain (helices H2–H7). Residues are colored green or brown if they participate in interactions that stabilize the formation of helices H1 or H8, respectively. Amino acids in the putative recognition helix that presumably interact with DNA are marked with an asterisk. The abbreviations are as follows: d-DRI-DBD, *Drosophila* Dead ringer; BRIGHT, Bright protein from *Mus musculus*; h-DRIL, human Dead ringer-like 1; h-BDP, human Bright and Dead ringer gene product homologous protein; HYP-1 (CA), hypothetical Bright protein homolog from *Caenorhabditis elegans*; EYELID, *Drosophila* Eyelid; h-BP120, human brain protein; h-SMCY, human SMCY; m-SMCX, *Mus musculus* SMCX; h-SMCX, human SMCX; e-SMCY, *Equus caballus* SMCY protein; e-SMCX, *Equus caballus* SMCX; m-SMCY, *Mus musculus* SMCY; HYP-2 (CA), hypothetical protein from chromosome 2 in *Caenorhabditis elegans*; h-PLU-1, human Plu-1 protein; h-RBP2, human retinoblastoma binding protein 2; h-RBP1, human retinoblastoma binding protein 1; h-RBP2b, human retinoblastoma binding protein 2 from pre-B cells; HYP. (AT), hypothetical protein from *Arabidopsis thaliana*; HYP-3 (CA), hypothetical protein from *Caenorhabditis elegans*; HYP. (SP), hypothetical protein from *Schizosaccharomyces pombe*; h-MRF1, human modulator recognition factor 1; h-MRF2, human modulator recognition factor 2; m-JUM, *Mus musculus* Jumoni; h-JUM, human Jumoni homolog; SWI1, SWI1 regulatory protein from *Saccharomyces cerevisiae*. **(B)** Expanded view of the packing interactions used to stabilize the first helix in the extended-ARID family. Members of the family that contain all or a subset of these hydrophobic residues are expected to possess a helix analogous to helix H1 in the DRI-DBD. Helices H1 and H2 are positioned at the top and bottom of this figure, respectively.

DRI-DBD structure are instrumental in anchoring the first helix to the body of the protein. In particular, the side chain of Val272 from helix H1 contacts the side chains

of Phe292 in helix H2 and Leu305 on the extended strand preceding the β -sheet. Helix H1 is also anchored by the side chain of Leu275 which packs against the side chains

Fig. 7. The ARIDs from Mrf-2 and Dead ringer have different three-dimensional structures. **(A)** Ribbon drawings of the structures of the DRI-DBD and Mrf-2 (Yuan *et al.*, 1998). The helices have been colored to show residues that share primary sequence homology. Color code: red, structurally conserved core domain (helices H3–H5, DRI-DBD; H2–H4, Mrf-2); dark purple (helix H2 of DRI-DBD; H1 of Mrf-2); magenta, second helix of HTH unit (helix H6, DRI-DBD; H5, Mrf-2); green (helix H7 of DRI-DBD; H6 of Mrf-2); black, additional helices in the extended-ARID of the DRI-DBD that do not share primary sequence homology with Mrf-2 (helix H1 and H8 of DRI-DBD). **(B)** Panels from the 3D ^{13}C -edited NOESY–HSQC spectrum of the DRI-DBD. The panels are labeled a–d and display NOE cross-peaks originating from hydrogen atoms of Leu275δ, Ile278δ, Leu328δ and Val329γ, respectively. The majority of cross-peaks could be confirmed by the presence of the appropriate symmetrically related reflection. The spectrum was recorded in water allowing the observation of NOEs to exchangeable backbone amide protons. All cross-peaks are labeled and the carbon chemical shift of each group is indicated.

of Leu288 (helix H2) and Leu315 (helix H3). Inspection of the sequence alignment indicates that many of these residues are conserved in other ARID-containing proteins (residues shaded green in Figure 6A). This strongly suggests that they possess an N-terminal helix analogous to helix H1. Although nearly half of the ARID proteins contain an N-terminal α -helix, very few proteins appear to possess a C-terminal α -helix analogous to helix H8 in the DRI-DBD structure. An inspection of the sequence alignment reveals that only five proteins contain conserved hydrophobic residues in helices H4 and H8 that are presumably required for the formation of the C-terminal helix (residues shaded brown in Figure 6A). In summary, the ARID DNA binding proteins can be partitioned into three structural classes: (i) minimal ARID proteins that consist of a core domain formed by six α -helices (H2–H7 in the DRI-DBD); (ii) ARID proteins that supplement the core domain with an N-terminal α -helix; and (iii) ARID domains that contain the core domain and additional α -helices at their N- and C-termini. The latter structural class we refer to as extended-ARID proteins, so as to maintain consistency with the previous classification scheme based on primary sequence homology. The structural classification of ARID-containing proteins is likely to be of functional importance, since only the extended-ARID proteins possess the final α -helix (H8 in the DRI-DBD structure), which presumably constitutes part of the DNA binding surface.

The fold of the DRI-DBD differs markedly from the recently determined structure of Mrf-2 (Yuan *et al.*, 1998). This is unexpected, since Mrf-2 shares 30% sequence identity (over 83 amino acids) with the DRI-DBD. The structures are compared in Figure 7A, with regions that have conserved primary sequences colored the same. Inspection reveals that only the central three helices (H3–

H5, DRI-DBD numbering) are structurally conserved (colored red). Although the flanking regions share primary sequence homology, their conformations differ; for example: (i) the orientation of helices H2 (dark purple) in the DRI-DBD is reversed relative to Mrf-2; (ii) helix H6 (magenta) in the DRI-DBD packs against the preceding helix in an HTH configuration; in contrast, the analogous helices in Mrf-2 are nearly anti-parallel; and (iii) helix H7 (green) in the DRI-DBD wraps around the body of the structure, while in Mrf-2 the analogous helix packs against the N-terminal helix. The NMR data of the DRI-DBD are incompatible with the three-dimensional structure of Mrf-2 (Figure 7B). For example, we observe NOEs between residues near the end of helix H1 and residues in the β -sheet (panel a: L275δ to Asp314 α , Leu315δ and Val312γ; panel b: Ile278δ to Asp314 α , β). These contacts are not possible in the Mrf-2 structure, since this region of the protein is near the C-terminus of the polypeptide. The data also position helix H4 near helices H6 (Figure 7B, panel c: L328δ to Tyr361δ, M362 β , γ and Tyr366δ) and H8 (panel d: Val329γ to Ser376 β , Leu281 β , δ and Ala384 β), but this configuration is not observed in the Mrf-2 structure. The origins of the structural differences between these proteins need to be addressed. The data suggest that ARID proteins are structurally diverse, with α -helices at the N- and C-termini playing instrumental roles in defining the three-dimensional structure of the central core domain.

A model of the ARID–DNA complex

A plausible model that describes the relative orientation of the DRI-DBD on its DNA binding site can be constructed from our data (Figure 8). In the model the DRI-DBD is positioned so as to insert helix H6 into the major groove, with additional contacts to the DNA in the adjacent

Fig. 8. Model of the Dead ringer–DNA complex. A ribbon drawing of the DRI-DBD protein is displayed, with the amino acids that exhibit large chemical shift changes as a result of DNA binding colored red. The protein is docked to a ribbon drawing of B-form DNA. The van der Waals surfaces of methyl groups mutated in the gel retardation experiments are displayed (methyls at position 3–4 and 7 are displayed). The nucleotide sequence of the DNA is shown below the figure with the Dead ringer binding site in bold.

minor grooves by residues in helix H8 and the β -hairpin structure. Several lines of evidence support this model. First, the NMR data are consistent with three regions of the protein contacting the DNA (the β -sheet and helices H6 and H8). Secondly, binding affinity measurements suggest that the center of the DNA binding site is contacted by the protein in the major groove (Figure 4B). Thirdly, helices H5 and H6 are structurally homologous to the ubiquitous HTH motif, which has consistently been shown to interact with DNA via the major groove (Harrison, 1991; Pabo and Sauer, 1992; Luisi, 1995). Finally, the DNA binding affinity of the related protein Bright is sensitive to distamycin, suggesting that minor-groove interactions between the protein and DNA are important for binding affinity (Herrscher *et al.*, 1995). In the model, two minor grooves of the binding site are proposed to interact with the protein. It is important to note that the model predicts that residues near helix H6 are proximal to the major groove, which is inconsistent with the finding that Dead ringer does not cross-link to 5-iodouracil-substituted DNA (Valentine *et al.*, 1998) and the model does not define the precise orientation of the DRI-DBD on the duplex or the specific intermolecular contacts in the nucleoprotein complex.

Biological implications

The structure of the DRI-DBD provides insights into the DNA binding specificity of the ARID class of proteins. The extended-ARID proteins Bright and Dead ringer bind sequence specifically to very similar AT-rich sites. The core recognition sequence of Bright is PuATa/tAA (Herrscher *et al.*, 1995), while binding site selection studies performed with the Dead ringer protein indicate that it binds to the consensus PuATTAA (Gregory *et al.*,

1996). The similar specificity of these proteins is in agreement with our model of the protein–DNA complex. In the model, sequence specificity is determined by interactions from the recognition helix of the HTH motif. Inspection of the sequence alignment presented in Figure 6A indicates that all of the potential DNA contact residues in the recognition helix are conserved between Bright and Dead ringer (marked with an asterisk above the sequence). A sequence comparison with other members of the ARID family reveals a large amount of variation in the potential DNA contact residues, suggesting that other members of this large family will recognize different nucleotide sequences.

Binding site selection experiments indicate that Dead ringer has the same sequence specificity as the Engrailed homeodomain: they both bind to the sequence ATTA (Kalionis and O'Farrell, 1993; Gregory *et al.*, 1996). Although Dead ringer and Engrailed both contain HTH motifs, a comparison of their recognition helices reveals no conserved residues that might interact with DNA. This result is surprising, since one would expect residues in the contact helix to be conserved if a similar DNA sequence is recognized. There are two possible explanations for this discrepancy. First it is possible that our binding model is incorrect and the HTH unit does not interact with the central major groove. Since the NMR data do not directly define the position of the DRI-DBD on the duplex, this possibility can not be excluded. A second plausible explanation is that the HTH motif in the DRI-DBD complex is positioned in a different orientation relative to the ATTA sequence as compared with its position in the three-dimensional structures of homeodomain protein–DNA complexes. A different orientation relative to the ATTA site would enable the non-homologous amino acid side chains in the DRI-DBD HTH motif to form a distinct set of intermolecular contacts with the ATTA subsite and with nucleotides that immediately flank this sequence. This explanation is consistent with the fact that two of the strongest physiological binding sites for the DRI-DBD (AT2 and AT3) do not contain the ATTA sequence, but rather the sequence TATTGAT. Regardless of the actual mode of binding and whether or not the ARID–DNA model is correct, our results reveal that two HTH-containing proteins can recognize the same ATTA DNA sequence in different ways.

Materials and methods

Preparation of Dead ringer DBD

A DNA fragment encoding the Dead ringer DBD (residues 262–398) was amplified by PCR from the full-length *dead ringer* gene (a gift from Dr Albert Courey) and inserted into the pGEX-4T-1 vector (Pharmacia) between the *Bam*HI and *Xho*I restriction sites. *Escherichia coli* strain BL21 was then transformed with the resultant plasmid and cultured at 37°C in M9 medium. To make ^{15}N -, or ^{15}N - and ^{13}C -labeled protein, $^{15}\text{NH}_4\text{Cl}$ or $[^{13}\text{C}_6]\text{glucose}$ was used as the sole nitrogen or carbon source in the growth medium. Expression of the GST–ARID fusion protein was induced with 0.2 mM isopropyl- β -D-thiogalactopyranoside when the absorbance of the culture reached ~0.8–0.9 optical density units at 600 nm. The cells were then harvested 3 h after induction and resuspended in 50 mM Tris–HCl pH 7.5, 300 mM NaCl, 2 mM dithiothreitol (DTT), 0.8 mM phenylmethylsulfonyl fluoride, 5% glycerol (20 ml/l of culture). The GST fusion protein was purified as previously described (Iwahara and Clubb, 1999). Typically, ~7 mg of purified Dead ringer protein were obtained per liter of M9 medium. The final purified

protein consisted of residues 262–398 of the Dead ringer protein and an N-terminal Gly-Ser dipeptide derived from the expression vector.

NMR analysis of Dead ringer DBD

Sample conditions for NMR of the DRI-DBD were 1.5 mM protein, 20 mM Tris-HCl pH 6.7, 100 mM NaCl, 1.5 mM ZnCl₂, 2 mM DTT, 0.01% NaN₃ and 5% ²H₂O. ¹H, ¹³C, ¹⁵N resonances were assigned with double- and triple-resonance experiments as described elsewhere (Iwahara and Clubb, 1999). NOE data for the structure calculations were obtained from 2D homonuclear NOESY, 3D ¹⁵N-edited NOESY-HSQC (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989) and 3D ¹³C-edited NOESY-HSQC (Muhandiram *et al.*, 1993) spectra with mixing times of 80 ms. The 2D homonuclear NOESY spectrum was used only for well-resolved NOE cross-peaks, including those from the hydroxyl proton of Tyr319 at 10.81 p.p.m. NOE data for both aliphatic and aromatic groups were obtained in a single 3D ¹³C-edited NOESY-HSQC spectrum measured with ¹³C-WURST inversion pulses (Kupce and Freeman, 1995). ³J_{HN-Hα} values were measured from a water-flip back 3D HNHA spectrum (Vuister and Bax, 1993; Kuboniwa *et al.*, 1994). χ^1 angles were analyzed using 3D HNHB (Archer *et al.*, 1991), 3D ¹⁵N-edited TOCSY-HSQC (Marion *et al.*, 1989), 3D ¹⁵N-edited ROESY-HSQC, ¹³C_{aromatic} ¹⁵N- or ¹³C' spin echo difference HSQC (Hu *et al.*, 1997), and ¹³C'- or ¹⁵N-¹³C_γ spin echo difference CT-HSQC spectra (Grzesiek *et al.*, 1993; Bax *et al.*, 1994). The χ^2 angles of leucine and isoleucine residues were determined from a long-range ¹³C-¹³C correlation spectrum and an analysis of the NOE data (Powers *et al.*, 1993).

The 1:1 DRI-DBD-DNA complex consisted of ¹⁵N- or ¹⁵N- and ¹³C-labeled DRI-DBD and a 15 bp DNA fragment (d-CGAATATTG-ATTGGG/d-CCAATCAATATTCG). The sample conditions for NMR measurements were 1.2 mM DRI-DBD-DNA complex, 20 mM Tris-HCl pH 6.7, 2 mM DTT, 0.01% NaN₃ and 5% ²H₂O. NMR spectra were acquired at 37°C on Bruker DRX-500 and DRX-600 spectrometers. For backbone resonance assignments, the following spectra were used: 2D ¹H-¹³C HSQC, 2D ¹H-¹⁵N HSQC, 3D double ¹⁵N-edited HMQC-NOESY-HSQC, 3D ¹⁵N-edited NOESY-HSQC, 3D HNHA, 3D HCACO, 3D HNCO, 3D HNCA and 3D HN(CO)CA. Detailed descriptions of these experiments, along with their original references, have been reviewed elsewhere (Clare and Gronenborn, 1994; Cavanagh *et al.*, 1996). All NMR data were processed with the program NMRPipe (Delaglio *et al.*, 1995) and analyzed using NMRView software (Johnson and Blevins, 1994) with in-house Tcl/Tk scripts.

Structure calculations

Structures of the free DRI-DBD were calculated using the program X-PLOR (Brünger, 1993) modified to include terms for ³J_{HNα} couplings (Garrett, 1994) and conformational database refinement (Kuszewski *et al.*, 1996). A total of nine residues at the C-terminus of the DRI-DBD (Arg390-Gly398) and two residues at the N-terminus that originate from the expression vector (Gly-Ser) displayed no long-range NOE cross-peaks in the data and were omitted from the final simulated annealing calculations. Distance restraints were grouped into four distance ranges: 1.8–3.0 Å, 1.8–4.0 Å (1.8–4.2 Å for distances involving ¹⁵N-bound protons), 1.8–5.0 Å (1.8–5.4 Å for distances involving ¹⁵N-bound protons) and 1.8–6.0 Å. To account for the increased apparent intensities of methyl resonances, 0.5 Å was added to the upper distance limits of NOE distances involving methyl protons. Distances involving methyl protons, aromatic ring protons and non-stereospecifically assigned methylene protons were represented as a $(\sum r_i^{-6})^{-1/6}$ sum. Hydrogen bond restraints were employed in areas of regular secondary structure and were introduced at the final stages of refinement. Two distance restraints were used for each hydrogen bond ($r_{\text{NH-O}} < 2.5$ Å and $r_{\text{N-O}} < 3.5$ Å). The structures were calculated using a hybrid distance geometry-simulated annealing (DGSA) protocol (Nilges *et al.*, 1988), followed by an additional simulated annealing step on each of the initial DGSA structures. The final simulated annealing protocol is identical to what has been described previously (Connolly *et al.*, 1998) with the following exceptions: (i) the initial phase comprised 15 ps of dynamics (5000 integration time steps of 3 fs each) at 3000 K; (ii) there were 116 cycles of cooling (25 K per cycle), each for 0.258 ps (129 integration times of 2 fs each); and (iii) a total of 350 cycles of Powell minimization were performed using the final values for the various force constants. Figures were prepared using the program MOLMOL (Koradi *et al.*, 1996).

Gel mobility DNA binding assay

The conditions for the binding reactions were as follows: 0.3 μM labeled DNA, 0.7 μM DRI-DBD, 3 or 9 μM cold competitor DNA, 20 mM

Tris-HCl pH 7.5, 100 mM NaCl, 5% glycerol. The 5'-terminus of the 15 bp DNA fragment was labeled with ³²P using T4 polynucleotide kinase and [γ -³²P]ATP. The labeled DNA had the sequence dCCTGTATTGATGTGG/dCCACATCAATACAGG and contains the Dead ringer binding sequence within the AT2 and AT3 sites (underlined). In all binding reactions the protein was added last. As cold competitors, four different 15 bp oligonucleotides were used. These fragments were: (i) the same DNA as the ³²P-labeled DNA; (ii) dCCTGTAUUGATGTGG/dCCACATCAATACAGG (mutant 1); (iii) dCCTGTATTGATGTGG/dCCACATCAATACAGG (mutant 2); and (iv) dCGAAGACGTG-TTGGG/dCCCAACACGTCTTCG which does not contain the cognate binding site. The electrophoresis was performed at 4°C with TBE buffer using a 15% polyacrylamide-TBE gel (Sambrook *et al.*, 1989).

Acknowledgements

We thank R. Peterson for NMR technical support, Dr A. Courey for his comments on this manuscript and Dr R. Saint for providing a copy of his manuscript before publication. We also wish to thank U. Ilangoan, J. Wojciak and K. Connolly for useful discussions. This work was supported by a grant from the US Department of Energy (DE-FC-03-87ER60615). The Protein Data Bank accession number is 1C20.

References

- Agulnik, A.I., Mitchell, M.J., Lerner, J.L., Woods, D.R. and Bishop, C.E. (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., Mitchell, M.J., Mattei, M.G., Borsani, G., Avner, P.A., Lerner, J.L. and Bishop, C.E. (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.
- Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) An alternative 3D-Nmr technique for correlating backbone N-15 with side chain H-β resonances in larger proteins. *J. Magn. Reson.*, **95**, 636–641.
- Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. and Zhu, G. (1994) Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.*, **239**, 79–105.
- Brünger, A.T. (1993) *X-PLOR (Version 3.1): A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G.I. and Skelton, N.J. (1996) *Protein NMR Spectroscopy*. Academic Press, San Diego, CA.
- Clare, G.M. and Gronenborn, A.M. (1994) Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.*, **239**, 349–363.
- Connolly, K.M., Wojciak, J.M. and Clubb, R.T. (1998) Site-specific DNA binding using a variation of the double stranded RNA binding motif. *Nature Struct. Biol.*, **5**, 546–550.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
- Donaldson, L.W., Petersen, J.M., Graves, B.J. and McIntosh, L.P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix–turn–helix DNA binding motif. *EMBO J.*, **15**, 125–134.
- Doyle, H.J., Kraut, R. and Levine, M. (1989) Spatial regulation of *zerknüllt*: a dorsal-ventral patterning gene in *Drosophila*. *Genes Dev.*, **3**, 1518–1533.
- Dufort, D. and Nepveu, A. (1994) The human Cut homeodomain protein represses transcription from the c-myc promoter. *Mol. Cell. Biol.*, **14**, 4251–4257.
- Fattaey, A.R. *et al.* (1993) Characterization of the retinoblastoma binding proteins RBP1 and RBP2. *Oncogene*, **8**, 3149–3156.
- Fesik, S.W. and Zuiderweg, E.R.P. (1988) Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.*, **78**, 588–593.
- Furui, J., Uegaki, K., Yamazaki, T., Shirakawa, M., Swindells, M.B., Harada, H., Taniguchi, T. and Kyogoku, Y. (1998) Solution structure of the IRF-2 DNA binding domain: a novel subgroup of the winged helix–turn–helix family. *Structure*, **6**, 491–500.
- Garrett, D.S. (1994) The impact of direct refinement against three-bond HN-Cα coupling constants on protein determination by NMR. *J. Magn. Reson. Ser. B*, **104**, 99–103.

- Gregory, S.L., Kortschak, R.D., Kalionis, B. and Saint, R. (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.
- Grzesiek, S., Vuister, G.W. and Bax, A. (1993) A simple and sensitive experiment for measurement of Jcc couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. *J. Biomol. NMR*, **3**, 487–493.
- Harrison, S.C. (1991) A structural taxonomy of DNA binding domains. *Nature*, **353**, 715–719.
- Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R. and Tucker, P.W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific transactivator that describes a new DNA binding protein family. *Genes Dev.*, **9**, 3067–3082.
- Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science*, **273**, 595–603.
- Hu, J.S., Grzesiek, S. and Bax, A. (1997) Two-dimensional NMR methods for determining (χ 1) angles of aromatic residues in proteins from three-bond J (C'C γ) and J (NC γ) couplings. *J. Am. Chem. Soc.*, **119**, 1803–1804.
- Huang, J.D., Dubnicoff, T., Liaw, G.J., Bai, Y., Valentine, S.A., Shirokawa, J.M., Lengyel, J.A. and Courey, A.J. (1995) Binding sites for transcription factor NTF-1/Elf-1 contribute to the ventral repression of decapentaplegic. *Genes Dev.*, **9**, 3177–3189.
- Huang, T.H., Oka, T., Asai, T., Okada, T., Merrills, B.W., Gertson, P.N., Whitson, R.H. and Itakura, K. (1996) Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Res.*, **24**, 1695–1701.
- Ip, Y.T., Kraut, R., Levine, M. and Rushlow, C.A. (1991) The dorsal morphogen is a sequence-specific DNA binding protein that interacts with a long-range repression element in *Drosophila*. *Cell*, **64**, 439–446.
- Iwahara, J. and Clubb, R.T. (1999) ¹H, ¹³C and ¹⁵N resonance assignments of the AT-rich interaction domain from the Dead ringer protein. *J. Biomol. NMR*, in press.
- Jiang, J., Cai, H., Zhou, Q. and Levine, M. (1993) Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. *EMBO J.*, **12**, 3201–3209.
- Johnson, B.A. and Blevins, R.A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
- Kalionis, B. and O'Farrell, P.H. (1993) A universal target sequence is bound *in vitro* by diverse homeodomains. *Mech. Dev.*, **43**, 57–70.
- Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
- Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) Measurement of H-N-H- α J couplings in calcium-free calmodulin using new 2d and 3d water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.
- Kupce, E. and Freeman, R. (1995) Stretched adiabatic pulses for broadband spin inversion. *J. Magn. Reson.*, **A117**, 246–256.
- Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.*, **5**, 1067–1080.
- Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
- Lehming, N., Thanos, D., Brickman, J.M., Ma, J., Maniatis, T. and Ptashne, M. (1994) An HMG-like protein that can switch a transcriptional activator to a repressor. *Nature*, **371**, 175–179.
- Lu, P.J. *et al.* (1999) A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J. Biol. Chem.*, **274**, 15633–15645.
- Luisi, B. (1995) DNA-protein interaction at high resolution. In Lilley, D.M. (ed.), *DNA-Protein: Structural Interactions*. Oxford University Press, Oxford, UK, pp. 1–48.
- Mailly, F., Bérubé, G., Harada, R., Mao, P.L., Phillips, S. and Nepveu, A. (1996) The human Cut homeodomain protein can repress gene expression by two distinct mechanisms: active repression and competition for binding site occupancy. *Mol. Cell. Biol.*, **16**, 5346–5357.
- Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Overcoming the overlap problem in the assignment of ¹H NMR spectra of larger proteins by use of three-dimensional heteronuclear ¹H-¹⁵N Hartmann-Hahn multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1 β . *Biochemistry*, **28**, 6150–6156.
- Motoyama, J., Kitajima, K., Kojima, M., Kondo, S. and Takeuchi, T. (1997) Organogenesis of the liver, thymus and spleen is affected in jumonji mutant mice. *Mech. Dev.*, **66**, 27–37.
- Muhandiram, D.R., Farrow, N.A., Xu, G.-Y., Smallcombe, S.H. and Kay, L.E. (1993) A gradient ¹³C NOESY-HSQC experiment for recording ¹³C-labeled proteins dissolved in H₂O. *J. Magn. Reson.*, **B102**, 317–321.
- Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamic simulated annealing. *FEBS Lett.*, **229**, 129–136.
- O'Hara, P.J., Horowitz, H., Eichinger, G. and Young, E.T. (1988) The yeast *ADR6* gene encodes homopolymeric amino acid sequences and a potential metal binding domain. *Nucleic Acids Res.*, **16**, 10153–10169.
- Ogata, K. *et al.* (1995) Comparison of the free and DNA-complexed forms of the DNA binding domain from c-Myb. *Nature Struct. Biol.*, **2**, 309–320.
- Pabo, C.O. and Sauer, R.T. (1992) Transcription factors—structural families and principles of DNA recognition. *Annu. Rev. Biochem.*, **61**, 1053.
- Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1993) The high-resolution, three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry*, **32**, 6744–6762.
- Quinn, J., Fyrberg, A.M., Ganster, R.W., Schmidt, M.C. and Peterson, C.L. (1996) DNA binding properties of the yeast SWI/SNF complex. *Nature*, **379**, 844–847.
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219–223.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schumacher, S., Clubb, R.T., Cai, M., Mizuuchi, K., Clore, G.M. and Gronenborn, A.M. (1997) Solution structure of the Mu end DNA binding 1 β subdomain of phage Mu transposase: modular DNA recognition by two tethered domains. *EMBO J.*, **16**, 7532–7541.
- Shandala, T., Kortshak, R.D., Gregory, S. and Saint, R. (1999) The *Drosophila dead ringer* gene is required for early embryonic patterning through regulation of *argos* and *buttonhead* expression. *Development*, in press.
- Suzuki, M. and Brenner, S.E. (1995) Classification of multi-helical DNA binding domains and application to predict the DBD structures of sigma factor, LysR, OmpR/PhoB, CENP-B, RapI and XylS/Ada/AraC [published erratum appears in *FEBS Lett.* (1996) **381**, 270]. *FEBS Lett.*, **372**, 215–221.
- Valentine, S.A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R. and Courey, A.J. (1998) Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. *Mol. Cell. Biol.*, **18**, 6584–6594.
- Vuister, G.W. and Bax, A. (1993) Quantitative J correlation—a new approach for measuring homonuclear 3-bond J (H (N)H (α) coupling constants in N-15-enriched proteins. *J. Am. Chem. Soc.*, **115**, 7772–7777.
- Wintjens, R. and Rooman, M. (1996) Structural classification of HTH DNA binding domains and protein-DNA interaction modes. *J. Mol. Biol.*, **262**, 294–313.
- Yuan, Y.C., Whitson, R.H., Liu, Q., Itakura, K. and Chen, Y. (1998) A novel DNA binding motif shares structural homology to DNA replication and repair nucleases and polymerases. *Nature Struct. Biol.*, **5**, 959–964.

Received July 19, 1999; revised and accepted September 14, 1999