TATA-binding protein-associated factor 1 (TAF1) is an essential component of the general transcription factor IID (TFIID), which nucleates assembly of the preinitiation complex for transcription by RNA polymerase II. TATA-binding protein and TAF1-TAF2 heterodimers are the only components of TFIID shown to bind specific DNA sequences (the TATA box and initiator, respectively), raising the question of how TFIID localizes to gene promoters that lack binding sites for these proteins. Here we demonstrate that Drosophila TAF1 protein isoforms TAF1-2 and TAF1-4 directly bind DNA independently of TAF2. DNA binding by TAF1 isoforms is mediated by cooperative interactions of two identical AT-hook motifs, one of which is encoded by an alternatively spliced exon. Electrophoretic mobility shift assays revealed that TAF1-2 bound the minor groove of adenine-thymine-rich DNA with a preference for the sequence AAT. Alanine-scanning mutagenesis of the alternatively spliced AT-hook indicated that Lys and Arg residues made essential DNA contacts, whereas Gly and Pro residues within the Arg-Gly-Arg-Pro core sequence were less important for DNA mobility shift assays revealed that TAF1 bound the minor groove of adenine-thymine (A-T) tracts (25, 26). The AT-hook motif contains a central core of Arg-Gly-Arg-Pro residues flanked by Lys and Arg residues (24). AT-hooks were first described in the high mobility group non-histone chromosomal protein HMGA (also known as HMG-I(Y)) and have been shown to bind DNA through contacts with the minor groove of adenine-thymine (A-T) tracts (25, 26). The AT-hook.

Binding of core promoters by the general transcription factor TFIID correlates with transcription activity for most RNA polymerase II genes (1–3). In metazoans, core promoters for RNA polymerase II genes comprise ~70 bp surrounding the transcription start site (TSS) and are made up of ~8-bp elements (4, 5). Core promoter elements include the TATA box located 26–31 bp upstream of the TSS, the initiator element located at the TSS, and the downstream promoter element located 28–33 bp downstream of the TSS. Multiple subunits of TFIID engage in DNA binding at the core promoter. TATA-binding protein binds the TATA box and TATA-binding protein-associated factors (TAFs) bind downstream elements (6, 7). Several lines of evidence point to the importance of TAF1-core promoter DNA interactions. Studies in yeast and mammalian cells indicate that core promoters, not activator-binding sites, render genes TAF1-dependent (8–10); cross-linking studies in Drosophila place TAF1 in close proximity to the TSS of the hsp70 gene (11); studies with purified human TFIID reveal TAF1 as the major TAF species that can be cross-linked to the downstream core element of several promoters (12); and TAF1-TAF2 heterodimers preferentially bind initiator sequences in vitro (8, 13, 14). However, TAF1 domains required for core promoter interactions have not been identified, and direct DNA binding by TAF1 on its own has not been observed.

TFIID is a family of complexes with different TAF components (15, 16). In Drosophila and humans, germline homologs of general TAFs are necessary for transcription during germ cell differentiation (17–21). In addition, changes in the TAF composition of TFIID result from alternative splicing of general TAF pre-mRNAs. For example, the human TAF6 pre-mRNA is alternatively spliced in response to apoptotic signals, and the encoded TAF6 isoform, TAF6δ, modulates the transcription of pro-apoptotic genes (22). Thus, regulated expression of TAF isoforms is an important determinant of gene-specific transcription.

In Drosophila, alternative splicing of the TAF1 pre-mRNA generates four TAF1 mRNA isoforms that differ by the inclusion of two small exons 12a and 13a. The functional importance of TAF1 alternative splicing is suggested by the fact that TAF1 isoform levels differ between Drosophila tissues and are regulated in response to genotoxic stress. TAF1-1 and TAF1-3 encode one predicted AT-hook motif, whereas TAF1-2 and TAF1-4 encode two predicted AT-hooks. The DNA-binding AT-hook motif contains a central core of Arg-Gly-Arg-Pro residues flanked by Lys and Arg residues (24). AT-hooks were first described in the high mobility group non-histone chromosomal protein HMGA (also known as HMG-I(Y)) and have been shown to bind DNA through contacts with the minor groove of adenine-thymine (A-T) tracts (25, 26). The AT-hook.
motif is found in one or more copies in a large number of transcription factors and components of chromatin remodeling complexes (24, 27, 28). AT-hooks may serve as accessory DNA-binding domains that anchor chromatin-associated proteins to particular DNA sequences or participate in cooperative DNA-binding with other proteins. Here we demonstrate direct DNA binding by the two AT-hook containing TAF1-2 and TAF1-4 proteins at the TSS of several Drosophila genes. We find that Lys and Arg residues of the AT-hook are critical for DNA binding, whereas Ala substitutions are tolerated at highly conserved Gly and Pro residues. TAF1 binds the minor groove of DNA with a preference for A-T tracts, especially an AAT sequence element. Direct DNA binding by TAF1 isoforms whose expression is regulated by developmental and stress signals has implications for the mechanism of gene-specific transcription initiation.

EXPERIMENTAL PROCEDURES

Plasmids—TAF1 expression plasmids were constructed by cloning exons 12–14 of TAF1 cDNAs into pET21a (Novagen) cut with NdeI and HindIII. The TAF1-1 plasmid encodes amino acids 1696–2065 of CG17603-PA, the TAF1-2 plasmid encodes amino acids 1696–2098 of CG17603-PC, and the TAF1-4 plasmid encodes amino acids 1696–2129 of CG17603-PC. The TAF1-3 plasmid was generated by joining a cDNA encoding amino acids 1696–1827 of CG17603-PA to a cDNA encoding amino acids 1860–2129 of CG17603-PC. Site-directed mutagenesis was carried out using the QuikChange mutagenesis kit (Stratagene) and was used to generate plasmids expressing TAF1-1 variants mutAT1 (R1793D), mutAT2 (R1816D), the double mutant, mutAT1/2, and each of the 31 alanine-scanning mutants.

Purification of Recombinant TAF1—BL21 Rosetta cells (Novagen) were transformed with pET21a-TAF1 expression plasmids and grown to mid-log phase, and protein expression was induced by addition of isopropyl-/β-D-thiogalactopyranoside to a concentration of 150 μM. Cell lysis was carried out using a cell lytic kit (Sigma). His-tagged TAF1 fusion proteins were purified on nickel-nitrilotriacetic acid columns (Novagen). Eluted proteins were concentrated using Amicon columns (Millipore), and the concentration was checked by Bradford assay (Bio-Rad), equalized by dilution, and dialyzed against Buffer A (50 mM Tris, pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 10% glycerol, 1 mM dithiothreitol) overnight. TAF1-2 protein used in the DNase I footprinting assay was purified as above except that protein eluted from the nickel-nitrilotriacetic acid column was further purified by gel filtration chromatography on a Sephacryl S300 column (Amersham Biosciences) pre-equilibrated with Buffer A containing 500 mM NaCl. Fractions containing TAF1-2 were concentrated and dialyzed against Buffer A.

Electrophoretic Mobility Shift Assay—For EMSAs, complementary oligonucleotides were annealed according to the manufacturer’s protocol (Integrated DNA Technology). dsDNA probes were 5’ end-labeled with T4 polynucleotide kinase and [γ-32P]ATP and analyzed on 20% polyacrylamide gels to verify >95% homogeneity prior to use. Sequences of probe plus
strands (oriented 5’ to 3’) are indicated in Fig. 5A. Protein-DNA complexes were formed by incubation for 30 min at 25 °C in a volume of 20 μl under the following conditions: 10 mM Tris, pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 1 mM dithiothreitol, 100 μg of bovine serum albumin/ml, 5% glycerol, 50 ng of poly(dG-dC)/μl, and 40 fmol of probe. The reactions were fractionated on 6.5% (29:1 acrylamide-bisacrylamide) native gels that included 0.5× TBE (Tris-borate-EDTA). Gels (18 cm × 20 cm × 1 mm) were electrophoresed at 4 °C for 2.5 h at 10 V/cm before exposure to x-ray film or a phosphorimaging plate. Quantitation was performed with ImageQuant software (Molecular Dynamics).

DNA Binding by TAF1 AT-hooks

RESULTS

TAF1 Alternative Exon 12a Encodes an AT-hook Motif—Comparison of conceptually translated Drosophila TAF1 exon 12a to protein motif data bases revealed a 9-amino acid DNA-binding motif called an AT-hook (Fig. 1). Hereafter, we refer to this motif as AT-hook 12a. An identical sequence is encoded by exon 12, and we refer to this motif as AT-hook 12. Thus, TAF1-1 and TAF1-3 proteins contain one AT-hook, whereas TAF1-2 and TAF1-4 proteins contain two AT-hooks separated by a 14-amino-acid spacer. The functional importance of TAF1 AT-hooks is suggested by their sequence conservation between Drosophila species (30). TAF1 AT-hooks are identical in sequence to the first AT-hook in human HMGA1 and are classified by Huth et al. (26) as low affinity type II AT-hooks because they have a central Gly-Arg-Pro peptide with flanking Lys and Arg residues but lack additional carboxyl-terminal residues that contact the sugar-phosphate backbone of DNA. Here we have focused on determining functional differences between TAF1 proteins that contain one versus two AT-hooks and characterizing the DNA binding properties of TAF1 AT-hooks.

TAF1 Isoforms with Two AT-hooks Directly Bind DNA—To determine the extent to which predicted AT-hooks 12 and 12a bind DNA, we performed EMSAs using a 34-bp fragment of the human interferon-β gene enhancer (INF-β) as a probe and His-tagged recombinant proteins encoding exons 12–14 of the four TAF1 isoforms (Figs. 1, 2A, and 5A) (31–33). INF-β, which contains PRDII-1, PRDII, and NRDI enhancer elements, was chosen as a probe because it is bound by HMGA AT-hooks in vivo and in vitro (26, 31–37). TAF1-2 and TAF1-4, the isoforms that contain two AT-hooks, altered the mobility of the INF-β probe, indicating that they bind INF-β DNA (Fig. 2B). In contrast, TAF1-1 and TAF1-3, the isoforms that contain one AT-hook, did not affect the mobility of the INF-β probe, indicating that they do not bind INF-β DNA. These findings suggest that AT-hook 12 is not sufficient for DNA binding, whereas AT-hook 12a in combination with AT-hook 12 is competent for DNA binding. Therefore, alternative splicing of the TAF1 pre-mRNA is a mechanism to control the DNA binding activity of TAF1 proteins.

To determine whether binding of TAF1-2 to INF-β DNA depends on AT-hook 12a or both AT-hooks, we assayed...
DNA Binding of TAF1-2 isoforms whose AT-hook motifs were mutated. TAF1-2 isoforms (mutant AT-hook 12 (mutAT1), mutant AT-hook 12a (mutAT2), and both mutant AT-hooks (mutAT1/2)) were produced in which the AT-hooks were mutated from KRGRGRPRK to KRGDGRPRK (Fig. 2, A and C). In the HMGA NMR structure, the corresponding basic Arg residue makes extensive contact with DNA bases within the minor groove, implying that substitution of this residue with an acidic Asp would disrupt the DNA binding ability of the AT-hook (26). EMSAs showed that DNA binding by mutAT1 or mutAT2 was reduced greater than 15-fold relative to TAF1-2, and DNA binding by mutAT1/2 was reduced below detectable levels (Fig. 2C). These results indicate that both AT-hooks contribute to binding of INF-β DNA by TAF1-2 and presumably TAF1-4. Therefore, inclusion of exon 12a confers a DNA binding activity upon TAF1-2 and TAF1-4 that is not present in TAF1-1 or TAF1-3. Furthermore, the DNA binding affinity of the AT-hooks together was greater than the sum of their individual affinities, suggesting that TAF1 AT-hooks bind DNA in a cooperative manner.

To establish the apparent equilibrium dissociation constant ($K_{D(app)}$) for TAF1-2, we measured the fraction of probe bound in reactions with different concentrations of TAF1-2. Based on maximal binding of the INF-β probe by TAF1-2 at ~85%, we estimate the $K_{D(app)}$ to be 1 μM (data not shown). This value falls within the nM to mM range of $K_{D(app)}$ for other AT-hook proteins and is comparable with the $K_{D(app)}$ of 4–10 μM determined for the HMGA high affinity AT-hook DBD3 bound to a PRDII dodecamer (26, 38).

Mutational Analysis of TAF1-2 AT-hook 12a—To determine which amino acids contribute to DNA binding by AT-hook 12a we performed alanine-scanning mutagenesis of the 33 amino acids encoded by exon 12a. TAF1 provides an ideal model for characterizing an AT-hook, because residues necessary for DNA binding are contained within the polypeptide encoded by alternatively spliced exon 12a. EMSAs were performed with the INF-β probe and each of the 31 mutant proteins resulting from all possible single Ala substitutions (2 of 33 residues are Ala in the wild type protein) (Fig. 3). Experiments performed

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Alanine-scanning mutagenesis of amino acids encoded by TAF1 exon12a identified amino acids important for DNA binding. A, EMSAs using radiolabeled INF-β probe and wild type or mutant TAF1-2 proteins at 0.5 μM are shown. Shifted TAF1-DNA complex is shown for wild type TAF1-2 (lane 1) or Ala substituted TAF1-2 proteins (lanes 2–20). Mutated amino acids are indicated above each lane. Each mutant protein was tested at least twice, and typical results are shown. B, EMSAs using radiolabeled INF-β probe and increasing amounts of wild type or mutant TAF1-2 proteins. Proteins shown in lanes 2, 7, 12, and 17 or 22 and 27 had equal concentrations. DNA binding of R1814A, G1815A, and G1817A was quantitated relative to equal concentrations of TAF1-2 shown in lanes 1–5, and DNA binding of P1818A was quantitated relative to equal concentrations of TAF1-2 shown in lanes 21–25. Dashes denote no protein, and triangles denote increasing amounts of indicated TAF1 proteins in 2-fold increments. Positions of the shifted TAF1-DNA complex and free DNA are indicated on the right. C, the graph depicts a qualitative assessment of the DNA binding affinity of TAF1-2, TAF1-1, or TAF1-2 alanine-scanning mutants for the INF-β probe. Maximal binding was observed with TAF1-2, and no binding was observed with TAF1-1. Ala substitution of amino acids within the region indicated as AT-hook 12a, with the exception of Gly1815, had an intermediate level of DNA binding. Mutated amino acids are indicated below each lane. Ala residues within wild type TAF1-2 are indicated in gray letters.
DNA Binding by TAF1 AT-hooks

---

**FIGURE 4.** A TAF1 protein with two AT-hooks binds dsDNA but not single-stranded DNA and prefers a narrow DNA minor groove. A, EMSAs using radiolabeled single-stranded INF-β probes (lanes 1, 2, 4, and 5) or double-stranded INF-β probe (lanes 3 and 6) and TAF1-2 (lanes 1–3) or mutAT1/2 (lanes 4–6) are shown. A and B refer to the sense and antisense strands, respectively. Positions of the shifted TAF1-DNA complex and free DNA are indicated on the right. B, EMSAs using radiolabeled INF-β probe and TAF1-2 are shown. Shifted TAF1-2-DNA complex was competed with increasing amounts of poly(dG-dC) (dg-dc, lanes 2–5), poly(dl-dc) (dl-dc, lanes 6–9), or poly(dA-dT) (da-dT, lanes 10–13). Dashes denote no protein, and triangles denote increasing amounts of indicated competitor DNA in 2-fold increments. Positions of the shifted TAF1-DNA complex and free DNA are indicated on the right.

---

using different methods of comparison with wild type TAF1-2 DNA binding resulted in qualitatively identical results (Fig. 3C). In experiments, such as that shown in Fig. 3A, EMSAs were performed using TAF1-2 mutant proteins at 0.5 μM, the concentration required for half-maximal binding by wild type TAF1-2 protein. In experiments, such as those shown in Fig. 3B, EMSAs were performed over a range of concentrations, starting with proteins of equal concentration. Evaluation of the fraction of probe shifted by mutant proteins relative to wild type TAF1-2, indicating that these positions do not strongly contribute to binding by TAF1-2. These data show that the functional limits of AT-hook 12a are confined to a 9-amino acid sequence of adjacent Arg side chains. The remaining 23 mutant TAF1-2 proteins that contained individual Ala substitutions at positions Val1806, Val1812, Gly1815, or Asp1822–Glu1838 possessed >90% of the DNA binding activity of wild type TAF1-2, indicating that these positions do not strongly contribute to binding by TAF1-2. These data show that the functional limits of AT-hook 12a are confined to a 9-amino acid region and raise questions about the functional importance of Gly and Pro residues, within the Arg-Gly-Arg-Pro core sequence, which are invariant in the current definition of an AT-hook (24).

**TAF1-2 Binds dsDNA in the Minor Groove—**HMGA AT-hooks bind the minor groove of DNA with a preference for A-T tracts. To determine whether TAF1 AT-hooks exhibit similar specificity, we performed EMSAs with TAF1-2 and single-stranded or double-stranded INF-β probes. As shown in Fig. 4A, TAF1-2 altered the mobility of the double-stranded INF-β probe but neither the top nor bottom single-stranded INF-β probes. As expected, mutAT1/2 failed to bind all three probes. Therefore, DNA binding by the TAF1-2 AT-hooks is specific for dsDNA and is not the result of simple electrostatic interaction between a solvent-exposed basic patch and the acidic DNA backbone of single-stranded DNA.

To determine the specificity of TAF1-2 for structural features of dsDNA, we performed EMSAs with TAF1-2 and the INF-β probe in the presence of increasing concentrations of unlabeled dsDNA competitors: poly(dG-dC), poly(dA-dT), or poly(dl-dc). Poly(dl-dc) effectively competes with A-T-rich sequences for binding by HMGA, presumably because poly(dl-dc) DNA adopts a structure with a narrow minor groove more similar to poly(dA-dT) than poly(dG-dC) (37). As shown in Fig. 4B, poly(dA-dT) and poly(dl-dc) competed with INF-β for binding to TAF1-2 at 4 μg/20 μl, but poly(dG-dC) did not. These observations are consistent with a model in which TAF1-2 AT-hooks recognize the structural characteristics of the minor groove of dsDNA with a preference for the narrow minor groove of A-T-rich DNA.

**TAF1-2 Binds to the TSS Sequence of Several Promoters—**Because TAF1 is a component of TFIIID and is expected to make physiologically relevant interactions with core promoter DNA, we assayed the ability of TAF1-2 to bind DNA sequences derived from core promoters of several *Drosophila* genes. In EMSAs, TAF1-2 protein bound 40-bp probes centered on the *Sdic* (sperm-specific dynein intermediate chain) TSS, the *hsp70* (heat-shock-protein 70bc) TSS, the *β2t* (*β2-tubulin*) TSS, and the *dj* (don juan) TSS, but not the *Cdic* (cytoplasmic dynein intermediate chain) TSS (Fig. 5). *Sdic*, *β2t*, and *dj* are highly expressed in *Drosophila* testes, whereas *Cdic* is ubiquitously expressed (39–41). Because TAF1-2 is the most abundant TAF1 isoform in testes, the observation that TAF1-2 can bind core promoter sequence of the testis-specific *Sdic*, *β2t*, and *dj* genes in vitro raises the possibility that TAF1-2 may contribute to the transcriptional activation of these genes in vivo.

**TAF1-2 Binds DNA That Contains an AAT Sequence—**In the HMGA NMR structure, AT-hooks intimately contact the INF-β PRDII sequence AAT (26). To test whether TAF1-2 also binds AAT, we performed EMSA with TAF1-2 and the INF-β probe or the INF-β mut probe, in which the AAT sequence was changed to AGT (Fig. 5A). This experiment revealed that TAF1-2 weakly bound the INF-β mut probe, suggesting that TAF1-2 AT-hooks directly interact with the AAT sequence (Fig. 5B). The hsp70 TSS probe, to which TAF1-2 strongly bound, contained two AAT sequences (Fig. 5A). Mutation of one of these sequences to AGT, in the hsp70 mut probe, substantially reduced binding by TAF1-2 (Fig. 5B). Attempts to determine a minimal DNA sequence bound by TAF1-2 revealed that binding decreased with the length of the DNA probe, even with probes containing AAT sequences, suggesting that high affinity binding requires additional contacts (data not shown). Taken together these experiments define AAT as a preferred site for binding by TAF1-2 AT-hooks.

To further refine our understanding of TAF1 DNA binding specificity, we used DNase I footprinting to identify TAF1-2
interaction sites on the *Drosophila Sdic* promoter (40). This analysis revealed five regions of protection by TAF1-2 on the 265-bp Sdic probe, which is centered at the TSS (Fig. 6). One protected site was located just upstream of the Sdic TSS and within the region predicted to be bound by TFIID. Each of the five protected regions contained \( \text{/H11022} \) 66\% A-T bp, and taken together the five regions contained 75\% A-T bp. Four of the five protected regions contained the AAT sequence element. These data indicate TAF1-2 AT-hooks bind DNA with a preference for the narrow minor groove of A-T-rich sequences and especially the sequence AAT.

**DISCUSSION**

This work provides evidence that *Drosophila* TAF1 directly interacts with DNA through AT-hook motifs. *Drosophila* appears to be the only organism in which TAF1 contains AT-hooks, which may account for the different DNA binding capabilities of *Drosophila* relative to mammalian TAF1s (7, 8, 13).4 TAF1-2 AT-hooks bound, with variable affinity, TSS sequences from several *Drosophila* genes and mutation of the initiating A-T bp in the hsp70 TSS probe drastically reduced binding (Fig. 5B). Thus, we speculate that evolutionary accretion of *Drosophila* TAF1 AT-hooks may have occurred to recognize A-T-rich promoters. Aerts et al. (42) found differences in the base composition surrounding TSSs in *Drosophila versus* vertebrates. Both *Drosophila* and vertebrates have a greater A-T than G-C content in intergenic regions. But although *Drosophila* have increased A-T content at the TSS, relative to intergenic regions, vertebrates have a greater G-C than A-T content at the TSS. Furthermore, *Drosophila* genes can be divided into classes with small and large increases in A-T content at the TSS, which suggests a functional meaning to the A-T content of *Dro-

4 C. E. Metcalf and D. A. Wassarman, unpublished observation.
DNA Binding by TAF1 AT-hooks

sophila promoters, possibly analogous to CpG islands in vertebrates (43). Hence, TAF1 may add to a growing list of proteins that have acquired AT-hooks in a particular organism or species to regulate DNA-dependent events through A-T-rich elements (44). For example, in Schizosaccharomyces pombe, the amino terminus of the Orc4p subunit of the origin recognition complex contains nine AT-hook motifs, which bind DNA replication origins that are extremely rich in A-T bp (45). AT-hooks in Orc4p and extremely A-T-rich replication origins do not occur in other eukaryotic organisms, including other yeast species such as Saccharomyces cerevisiae. In addition, in plants there are several examples of AT-hook-containing transcription factors that activate transcription through binding A-T-rich enhancer elements (46, 47). These observations imply that other species-specific AT-hook proteins regulate DNA-dependent events controlled by A-T-rich elements.

This work has also extended our understanding of AT-hook-DNA interactions. TAF1 isoforms that contain one versus two AT-hooks greatly differ in their capacity to bind DNA. Neither TAF1 AT-hook 12 nor 12a has DNA binding activity, but together the AT-hooks bind DNA with low micromolar affinity. This finding highlights the importance of intramolecular cooperativity of AT-hooks for DNA binding, as shown for HMGA family proteins (26, 31, 34).

Like HMGA family proteins, TAF1-2 binds the minor groove of A-T-rich dsDNA (26). DNA binding by TAF1-2 was competed by poly(dA-dT) and poly(dI-dC), which have a similar narrow minor groove, but not poly(dG-dC), and TAF1-2 failed to bind single-stranded DNA (Fig. 4). Thus, although TAF1 and HMGA AT-hooks have different sequences and spacing, they appear to bind DNA in a similar manner.

Systematic mutational analysis of the AT-hook region encoded by exon 12a supports the definition of the AT-hook as a 9-amino acid Lys/Arg-rich motif but brings into question the importance of DNA binding of Gly and Pro residues within the Arg-Gly-Arg-Pro core sequence (24). Although 25 of 33 amino acids encoded by exon 12a are conserved among Drosophila species, only the six conserved Lys or Arg amino acids within the predicted AT-hook motif drastically reduced DNA binding (Fig. 3). Similarly, protein hydroxyl radical footprinting analysis of HMGA showed that DBD1, which is identical in sequence to TAF1 AT-hooks, but not surrounding residues contact INF-/H9252 DNA (35). In contrast, residues surrounding DBD2 and DBD3 substantially contribute to DNA binding. Thus, conserved residues surrounding TAF1 AT-hook 12a are likely to serve functions independent of DNA binding.

The strong requirement for Lys and Arg residues flanking the Arg-Gly-Arg-Pro core was unexpected given the frequency of Gly and Pro residues at these positions in predicted AT-hook motifs (24). Also unexpected was the relatively minor requirement for Gly and Pro residues at these positions in predicted AT-hook motifs (24). Also unexpected was the relatively minor requirement for Gly and Pro residues within the Arg-Gly-Arg-Pro core sequence. Proteins containing Ala substitutions at these positions retained substantial DNA binding activity. In the NMR structure of HMGA DBD2 and DBD3 bound to a fragment of the INF-/H9252 enhancer, the two AT-hooks made similar contacts with DNA (26). The Arg-Gly-Arg portion of the Arg-Gly-Arg-Pro core sequence formed a concave surface that inserted into

FIGURE 6. DNase I footprinting demonstrates that TAF1-2 contacts A-T-rich sequences of the Sdic gene. Footprint analysis of TAF1-2 binding to the Sdic promoter is shown. Vertical lines on the right indicate regions of protection by TAF1-2. Lanes 3 and 7 show the DNase I digestion pattern in the absence of TAF1-2, and lanes 4–6 show the pattern in the presence of increasing μM concentrations of TAF1-2. Lane 8 shows the digestion pattern in the absence of DNase I (some of the faint bands result from spillover from lane 7 into lane 8). As sequencing markers, lanes 1 and 2 contain Sdic DNA that was chemically cleaved by Maxam-Gilbert A-C and G-A reactions. The bracket on the left indicates the region of the Sdic promoter used in the Sdic TSS probe, and the arrow denotes the TSS.

DNA Binding by TAF1 AT-hooks

OCTOBER 6, 2006 • VOLUME 281 • NUMBER 40

JOURNAL OF BIOLOGICAL CHEMISTRY 30021
DNA Binding by TAF1 AT-hooks

the minor groove of A-T tracts, the Pro directed the peptide backbone away from the minor groove, and flanking Arg and Lys residues made extensive hydrophobic and polar contacts with the DNA. Our findings indicate that the methyl side group of Ala, within the sequence Arg-Ala-Arg, is tolerated for interactions with the minor groove and that reorientation of flanking sequences by the Pro residue is not necessary for DNA binding.

These data suggest that the AT-hook motif, and more specifically the Gly-Arg-Pro, is more divergent than that described by Aravind and Landsman (24). In support of this proposal, DNA binding by Lens epithelium-derived growth factor 75 (LEDGF/P75) requires three AT-hooks (AA(R/G)RK, SPKRGRPA, and PKPRGPKR, where the underlined sequence corresponds to the Gly-Arg-Pro sequence), the first of which contains a Lys instead of a Pro within the putative invariant Gly-Arg-Pro sequence (48). Furthermore, our searches of the Drosophila data base have identified several proteins with potential AT-hooks that do not contain the Gly-Arg-Pro sequence. For example CG2995, which encodes a chromatin-modifying enzyme, contains five potential AT-hooks within a 168-amino acid region (KRGRGRARK, EKPKGPRKRK, VKRMRRSVR, PKKRGRRKAK, and PKRKRRIK). Thus, the AT-hook may be a more prevalent DNA-binding motif than previously predicted.

As expected based on experiments with other AT-hook proteins, EMSA and footprinting studies revealed that TAF1-2 AT-hooks bind DNA sequences containing A-T tracts (23, 26, 31, 32, 34, 49). The trinucleotide sequence AAT may be a preferred binding site for TAF1-2 because four of the five regions protected by TAF1-2 in the footprinting analysis contained an AAT and mutation of AAT to AGT within the hsp70 TSS or INF-β probes severely reduced binding by TAF1-2 (Figs. 5 and 6). These data are consistent with the prediction, based on NMR analysis of HMGa, that AA(T/A)T is an optimal AT-hook-binding site and other studies showing that AT-hooks bind an A-T tract of 3–5 bp (26, 49).

In summary, experiments presented here suggest that modulation of the TAF1-DNA interaction, by regulated expression of TAF1 isoforms containing one versus two AT-hooks, contributes to gene-specific transcription. Further characterization of TAF1 should shed light on the relationship between TFIIID-promoter affinity and the specificity and efficiency of transcription initiation. Additionally, detailed analysis of the determinants of the TAF1-DNA interaction and comparison with the determinants of the HMGa-DNA interaction suggests the existence of unrecognized AT-hook-containing proteins, elevating the importance of understanding the molecular consequences of AT-hook-DNA interactions in vivo.

Acknowledgments—We thank T. Kokubo and R. Tjian for providing plasmids, D. Brow and W. Ross for advice on experiments, and members of the Wassarman laboratory for many helpful discussions.

REFERENCES
1. Huisinga, K. L., and Pugh, B. F. (2004) Mol. Cell. 13, 573–585
2. Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005) Nature 436, 876–880
3. Chen, B. S., and Hampsey, M. (2002) Curr. Biol. 12, R620–R622
4. Smale, S. T., and Kadonaga, J. T. (2003) Annu. Rev. Biochem. 72, 449–479
5. Ohler, U., Liao, G., Niemann, H., and Rubin, G. M. (2002) Genome Biol. 3, research0087.1–0087.12
6. Matangkasombut, O., Auty, R., and Buratowski, S. (2004) Adv. Prot. Chem. 67, 67–92
7. Verrijzer, C. P., Chen, J.-L., Yokomori, K., and Tjian, R. (1995) Cell 81, 1115–1125
8. Dehm, S. M., Hilton, T. L., Wang, E. H., and Bonham, K. (2004) Mol. Cell. Biol. 24, 2296–2307
9. Shen, W.-C., and Green, M. R. (1997) Cell 90, 615–624
10. Wang, E. H., and Tjian, R. (1994) Science 263, 811–814
11. Wu, C.-H., Madabusi, L., Nishioka, H., Emanuel, P., Sypes, M., Arkhipova, L., and Gilmour, D. S. (2001) Mol. Cell. Biol. 21, 1593–1602
12. Lee, D.-H., Wu, J., Green, R. D., and Ren, B. (2005) Mol. Cell. Biol. 25, 9674–9686
13. Chalkley, G. E., and Verrijzer, C. P. (1999) EMBO J. 18, 4835–4845
14. Verrijzer, C. P., Yokomori, K., Chen, J.-L., and Tjian, R. (1994) Science 264, 933–941
15. Hochheimer, A., and Tjian, R. (2003) Genes Dev. 17, 1309–1320
16. Müller, F., and Tora, L. (2004) EMBO J. 23, 2–8
17. Miller, M., Chen, X., Pringle, M. J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T.-Y., Marino, S., and Fuller, M. T. (2004) Development 131, 5297–5308
18. Miller, M. A., Lin, T.-Y., Wood, C., and Fuller, M. T. (2001) Genes Dev. 15, 1021–1030
19. Falender, A. E., Freiman, R. N., Geles, K. G., Lo, K. C., Hwang, K., Lamb, D. J., Morris, P. L., Tjian, R., and Richards, J. S. (2005) Genes Dev. 19, 794–803
20. Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E., and Tjian, R. (2001) Science 293, 2084–2087
21. Pointud, J. C., Mengus, G., Brancorsini, S., Monaco, L., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2003) J. Cell. Sci. 116, 1847–1858
22. Bell, B., Scheer, E., and Tora, L. (2001) Mol. Cell. 8, 591–600
23. Solomon, M. J., Strauss, F., and Varshavsky, A. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 1276–1280
24. Aravind, L., and Landsman, D. (1998) Nucleic Acids Res. 26, 4413–4421
25. Reeves, R. (2001) Gene (Amst.) 277, 63–81
26. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N., Reeves, J., Gronenborn, A. M., and Clore, G. M. (1997) Nat. Struct. Biol. 4, 657–665
27. Cairns, B. R., Schlichter, A., Erdjument-Bromage, H., Tempst, P., Kornberg, R. D., and Winston, F. (1999) Mol. Cell. 4, 715–723
28. Lours, C., Bardot, O., Godt, D., Laski, F. A., and Couderc, J.-L. (2000) Nucleic Acids Res. 28, 5389–5398
29. Lours, C., Bardot, O., Godt, D., Laski, F. A., and Couderc, J.-L. (2000) Nucleic Acids Res. 28, 3949–4959
30. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 591–600
31. Schauwach, V. (2002) Biol. Linn. Soc. 76, 21–37
32. Yie, J., Liang, S., Merika, M., and Thonas, D. (1997) Mol. Cell. Biol. 17, 3649–3662
33. Schwanbeck, R., Manfoletti, G., and Wisniewski, J. R. (2000) J. Biol. Chem. 275, 1793–1801
34. Faivo, J. V., Thanos, D., and Maniatis, T. (1995) Cell 83, 1101–1111
35. Chalkley, G. E., and Verrijzer, C. P. (1999) J. Biol. Chem. 269, 33042–33048
36. Frank, O., Schwanbeck, R., and Wisniewski, J. R. (1998) J. Biol. Chem. 273, 20015–20020
37. Schwanbeck, R., Gerharz, M., Drung, A., Rogalla, P., Piekielko, A., Blank, C., Bullerdiek, J., and Wisniewski, J. R. (2000) Biochemistry 39, 14419–14425
38. Banks, G. C., Mohr, B., and Reeves, R. (1999) J. Biol. Chem. 274, 16536–16544
39. Dragan, A. I., Liggins, J. R., Crane-Robinson, C., and Privalov, P. L. (2003) J. Mol. Biol. 327, 393–411
40. Blumer, N., Schreiter, K., Hempel, L., Santel, A., Hollman, M., Schafer, M. A., and Renkawitz-Pohl, R. (2002) Mech. Dev. 110, 97–112
41. Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D., and Hartl, D. L. (1998) Nature 396, 572–575
42. Santel, A., Kaufmann, J., Hyland, R., and Renkawitz-Pohl, R. (2000) Nucleic Acids Res. 28, 1439–1446
42. Aerts, S., Thijs, G., Dabrowski, M., Moreau, Y., and De Moor, B. (2004) BMC Genomics 5, 34
43. Caiafa, P., and Zampieri, M. (2005) J. Cell. Biochem. 94, 257–265
44. Koonin, E. V., Aravind, L., and Kondrashov, A. S. (2000) Cell 101, 573–576
45. Chuang, R.-Y., and Kelly, T. J. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2656–2661
46. Christiansen, H., Hansen, A. C., Vijn, I., Pallisgaard, N., Larsen, K., Yang, W. C., Bisseling, T., Marcker, K. A., and Jensen, E. O. (1996) Plant Mol. Biol. 32, 809–821
47. Korfhage, U., Trezzini, G. F., Meier, I., Hahlbrock, K., and Somssich, I. E. (1994) Plant Cell 6, 695–708
48. Turlure, F., Maertens, G., Rahman, S., Cherepanov, P., and Engelman, A. (2006) Nucleic Acids Res. 34, 1663–1675
49. Maher, J. F., and Nathans, D. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 6716–6720