EXPERIMENTAL CELL RESEARCH **236,** 472–481 (1997)
ARTICLE NO. EX973742

The High-Mobility Group Protein T160 Binds to both Linear and Cruciform DNA and Mediates DNA Bending as Determined by Ring Closure

Marisa Gariglio,* Guo-Guang Ying,* Laura Hertel,* Mirella Gaboli,* Roger G. Clerc,† and Santo Landolfo*‡[1]

*Institute of Microbiology and Department of Medical Sciences, Medical School of Novara, University of Torino, Torino, Italy; †Roche Ltd. Research Laboratories, CH-4070 Basel, Switzerland; and ‡CNR, Center of Immunogenetics and Experimental Oncology, 10126 Torino, Italy

The high-mobility group protein T160 was isolated by screening a phage library from a murine pre-B-cell line L1210. South–Western experiments have previously shown that this protein binds to V-(D)-J recombination signal sequences, suggesting that it may be a sequence-specific DNA-binding protein. However, neither gel-shift nor footprinting analyses have been successfully employed with the T160 protein, despite an extensive effort. In this study, the T160 protein or truncated forms made soluble through denaturing and renaturing cycles in urea were successfully used in gel-shift experiments showing that T160 binds to cruciform or linear duplex DNA with no apparent sequence specificity. Furthermore, fragments longer than 100 bp efficiently formed covalently closed circular monomers in the presence of T160 and T4 DNA ligase, indicating that the protein is capable of introducing bends into the duplex. Last, tissue distribution by Western blotting analysis showed that the T160 protein is expressed in various murine tissues in addition to those of lymphoid origin. Considering its broad evolutionary conservation (from plants to mammals) also, these results suggest that the functional role of the T160 protein is not limited to V-(D)-J recombination, but might be involved in basic processes such as DNA replication and repairing, where irregular DNA structures are generated and very likely recognized by HMG domain proteins.

INTRODUCTION

The high-mobility group (HMG) chromosomal proteins are the most abundant and ubiquitous nonhistone proteins found in the nuclei of higher eukaryotes, which contain three HMG families—HMG-1/2, HMG-14/17, and HMG-I(Y) distinguished by their ability to be extracted from chromatin in 0.35 *M* NaCl and solubility in 5% perchloric or trichloroacetic acid, molecular masses, amino acid sequence motifs, and DNA-binding characteristics [1, 2].

The HMG-1/2 proteins have two structural domains (A and B) and a significant homology with each other [3]. Analysis of the RNA polymerase I transcription factor UBF [4, 5] has identified a repeated 85-amino-acid region as a novel DNA-binding motif known as the "HMG box" on account of its homology with the A and B domains. The HMG box motif displays an abundance of aromatic residues, prolines, and basic amino acids. The HMG-1/2 family is divided into two subfamilies according to the number of HMG domains present in the protein, their specificity of sequence recognition, and their evolutionary relatedness [6–8]. The first subfamily comprises the HMG-1/2 proteins, the yeast ARS-binding protein ABF-2 [9], UBF [10], and the mitochondrial transcription factor mTF-1 [11]. It is the most abundant and present in all cell tissues. Its proteins have multiple HMG domains and preferentially bind to non-B-DNA conformations, such as B-Z junctions, stem-loops, cruciforms, four-way junctions, and cisplatin-modified DNA with no obvious, or low, sequence specificity [12, 13]. The second subfamily includes the lymphoid enhancer-binding factor LEF-1 (also termed TCF-1α) [14–16] and TCF-1 [17], the sex-determining factor SRY [18] and related Sox proteins [19], and the fungal regulatory proteins Mat-Mac [20], Ste 11, and Rox 1 [21]. These proteins have a single HMG box domain, interact with specific nucleotide sequences (A/TA/TCAAAG), and have a restricted cell-type distribution [22, 23].

HMG-1 box proteins have been implicated in a wide variety of cellular functions. These include a role in nucleosome assembly and disassembly [24, 25] and in both facilitating and interfering with the transcription process [26, 27]. Nightingale *et al.* (1996) [28] have demonstrated that HMG-1, like histone B4, can interact with nucleosomal DNA with very similar consequences to those that follow assembly of histone H1

M. Gariglio and G. G. Ying contributed equally to this work.

[1] To whom reprint requests should be addressed. Fax: +39-11-6636436. E-mail: virmol@inrete.it.

0014-4827/97 $25.00

into the nucleosome. LEF-1 or TCF1, members of a family of T-lymphocyte enhancer-binding factors, seem to be involved in the regulation of the T-cell receptor enhancer α [14, 16]. Thus, although HMG box proteins have been studied extensively, a clear functional and structural role remains to be established.

In previous studies, we have used the promoter of an IFN-activatable gene, the 202 gene, to characterize the nuclear factors responsible for gene activation at the transcriptional level in L1210, a murine pre-B-leukemia-cell line [29]. The structure of the 202 promoter has been determined by transfection experiments in which a reporter gene was linked to the 202 5′ flanking region. An IFN-stimulated response element (ISRE), designated GA box because of its high content of CGAAA motifs, appears to be necessary and sufficient for IFN-α induction [30, 31].

In this study, we describe the cloning of a HMG box-containing protein which binds to an ISRE trimer from the 202 gene. Nucleotide sequencing revealed that this was identical to the mouse HMG1-related DNA-binding protein T160, isolated by using the V-(D)-J recombination signal probes [32]. This protein was specific for 12-bp spacer signals and failed to bind a sequence with a base change in the third position of the heptamer. However, DNA binding by the 86-kDa T160 protein could only be detected with Southwestern blot analysis; the *in vitro*-translated material could not be demonstrated to bind DNA according to electrophoretic mobility shift and DNA footprinting analysis [33]. The human homolog was isolated by screening a mature B-cell cDNA library with a platinated (non-joining signal) DNA probe and found to be ubiquitously expressed at the mRNA level [34]. Last, the T160 yeast homolog designated Pob3 for its capability to interact with the DNA polymerase 1 has been knocked out and the phenotype was lethal [35].

Our results demonstrate that: (i) the T160 HMG box domain is responsible for DNA binding according to electrophoretic mobility shifts; (ii) T160 binds to cruciform DNA as well as to linear double-stranded DNA with no sequence specificity; (iii) it bends DNA and mediates ligase-catalyzed cyclization of DNA fragments longer than 100 bp; and finally (iv) the T160 protein is expressed not only in lymphoid tissues, but in various mouse tissues of different histological origin, containing actively proliferating cells.

MATERIALS AND METHODS

Mice. Pathogen-free, 6- to 8-week-old DBA\2 female mice were purchased from Charles River Laboratories (Calco, Italy). All animals were maintained in our animal laboratory only for the duration of the experiments.

Cell culture and reagents. The L1210 cell line was grown in RPMI 1640 medium supplemented with 10% fetal bovine serum. Recombinant hybrid human IFN-α A/D (sp act 4×10^8 units/mg of protein) used as the source of murine IFN-α was kindly provided by Dr. G. Garotta, Hoffmann-La Roche.

Oligonucleotides. Oligonucleotides were synthesized on an Applied Biosystem 308A DNA synthesizer and purified in denaturing 18% (w/v) polyacrylamide/7 *M* urea gels. From 5′ to 3′ the sequences are: (1) CCCTATAACCCCTGCATTGAATTCCAGTCTGATAA. (2) GTAGTCGTGATAGGTGCAGGGGTTATAGGG. (3) AACAGTAGCTCTTATTCGAGCTCGCGCCCTATCACGACTA. (4) TTTATCAGACTGGAATTCAAGCGCGAGCTCGAATAAGAGCTACTGT. (5) GTAGTCGTGATAGGGCGCGAGCTCGAATAAGAGCTACTGT. (6) TTTATCAGACTGGAATTCAATGCAGGGGTTATAGGG. Complete and incomplete cruciform DNAs and control linear duplexes were made as described previously by Bianchi (1988) [36]. DNA molecules were labeled with T4 polynucleotide kinase before annealing. The crude labeled preparations of complete or incomplete cruciform DNA were then loaded on a 10% polyacrylamide gel in TBE (89 m*M* Tris base, 89 m*M* boric acid, 2 m*M* EDTA) and electrophoresed at 20 V/cm. The band of pure material was identified by autoradiography, cut out, and eluted.

Library screening. A λgt11 cDNA expression library constructed from poly(A)$^+$ RNA prepared from L1210 treated with IFN-α for 30 min was screened as described previously with some modifications [37, 38]. Briefly, 35,000 plaques in a 150-mm petri dish were induced by isopropyl-β-D-thiogalactopyranoside (IPTG) soaked in nitrocellulose filters. After lifting, the filters were incubated in BLOTTO (5% low-fat milk powder, 50 m*M* Tris, pH 7.5, 50 m*M* NaCl, 1 m*M* EDTA, and 1 m*M* dithiothreitol (DTT)) for 60 min at room temperature (RT) and washed twice (1–5 min for each wash) with TNE-50 (10 m*M* Tris, pH 7.5, 50 m*M* NaCl, 1 m*M* EDTA, and 1 m*M* DTT). They were then incubated with ^{32}P-labeled GA trimer (1×10^6 cpm/ml) in TNE-50 supplemented with 7 μg/ml of denatured sonicated calf thymus DNA (Boehringer) for 60 min at RT, washed four times in TNE-50 (5–10 min for each wash), and autoradiographed. By screening 7×10^5 phages one single positive clone was isolated through repeated rounds of screening. The insert of this clone, called 18.2.1, was 1.4 kb in length. The 5′ end of clone 18.2.1 was used as a probe to obtain additional clones with overlapping sequences in the same library.

Sequence determination and analysis. The full-length cDNA was subcloned in pUC18, and the DNA sequence was determined for both strands by the dideoxy chain-termination method, using Sequenase (United States Biochemical) and synthetic primers.

Preparation of DNA fragments. The 40- and 105-bp fragments were generated by digesting the pBend2 plasmid [39] with *Spe*I and *Sal*I restriction enzymes, while the 62- and 84-bp fragments were obtained by digesting the same plasmid with *Spe*I and *Xba*I restriction enzymes, respectively [40]. All fragments had 4-bp 5′ overhangs and were end-labeled by treatment with the Klenow fragment of DNA polymerase I and an equimolar mixture of [α-^{32}P]dCTP, [α-^{32}P]dTTP, dATP, and dGTP. The labeled probes were subsequently separated on a 12% nondenaturing polyacrylamide gel, eluted from the gel, desalted, concentrated, and used directly for the binding and cyclization assays.

The GA trimer was obtained from the plasmid pGD4-3, kindly provided by Professor P. Lengyel. The sequence of the GA box from 5′ to 3′ is TCCAAAGCCAGGGAAATTGAAAGCTATGAACGAAACTGGGAG. The *Hin*dIII–*Bam*HI 167-bp fragment was end-labeled with Klenow in the presence of [α-^{32}P]dCTP, [α-^{32}P]dTTP, [α-^{32}P]dGTP, and dCTP, gel purified, and used to screen the library.

Construction of T160 expression plasmids and purification of T160 protein, T160 C-terminus, and T160 N-terminus peptides. The T160 C-terminus coding region of the T160 cDNA (amino acids 467–645) was obtained by using *Bsm*I and *Sma*I unique restriction sites, filled in with the Klenow fragment of *Escherichia coli* DNA polymerase, and ligated as blunt end in the *Sma*I site of the pGEX 4T2 expression vector (Pharmacia) to regenerate the proper reading frame. In frame fusion with the coding region of the GST gene was confirmed by DNA sequence analysis. The GST–T160 fusion protein was expressed in

E. coli AD202 [41] and affinity-purified on glutathione–Sepharose 4B beads according to the method of Smith and Johnson (1988) [42].

Fusion protein expression was induced by the addition of 0.1 m*M* IPTG for 3 h. The bacterial cells were harvested by centrifugation and resuspended in cold lysis buffer (0.5 mg/ml lysozyme, 25 m*M* Tris–HCl, pH 7.9, 150 m*M* NaCl, 1 m*M* EDTA, 1 m*M* DTT, 10% glycerol, 0.2% Triton X-100 containing 2 m*M* PMSF, 50 m*M* pepstatin A, and 50 m*M* leupeptin as protease inhibitors). Cells were lysed by sonication and centrifuged at 10,000*g* for 20 min. Fusion proteins were purified from the cleared lysate by glutathione–Sepharose affinity chromatography. Cleavage of the fusion protein while still attached to the beads was achieved by the addition of thrombin (10 cleavage units/mg fusion protein) with incubation overnight at RT. The recombinant proteins were then eluted from the column using 25 m*M* Tris–HCl, pH 7.9, 150 m*M* NaCl.

The DNA fragment comprising the coding region of full-length T160 (amino acids 60–708) was obtained from the plasmid pUCT160 by using *Nco*I and *Bsu*36I unique restriction sites, filled in, and ligated in the *Eco*RV site of plasmid pET30a (Novagen). The gene construct encoding the T160 N-terminus protein (amino acids 60–467) was generated by inserting the *Nco*I–*Bsm*I DNA fragment into the *Eco*RV site of plasmid pET30a. The integrity of all these fusion proteins was verified by sequencing. The recombinant proteins were produced in the bacterial strain BL21 (DE3), after induction by IPTG (0.1 m*M*) [43]. After 3 h of induction, cells were collected by centrifugation (7000*g* for 10 min at 4°C). The pellets were resuspended in 1/15 culture volume of 0.5 mg/ml lysozyme, 25 m*M* Tris–HCl, pH 7.9, 150 m*M* NaCl, 1 m*M* EDTA, 1 m*M* DTT, 10% glycerol, 0.1% NP-40 plus protease inhibitors and lysed by mild sonication. Cell lysates were centrifuged (10,000*g* for 20 min at 4°C), and the pellets, containing most of the overproduced T160, were resuspended in buffer L (25 m*M* Tris–HCl, pH 7.9, 150 m*M* NaCl, 0.1% NP-40) with 2 *M* urea, sonicated briefly, and centrifuged as above. These washed pellets were then dissolved in buffer L with 8 *M* urea. The T160 N-terminus protein was already pure to near homogeneity; solubilized protein was separated from insoluble proteins by centrifugation and extensively dialyzed against buffer L. This procedure yields highly purified soluble T160 N-terminus.

T160 full-length protein in buffer L with 8 *M* urea was then passed over a nickel chelate column (Pharmacia) charged with 50 m*M* $NiSO_4$ (recombinant proteins bind to the nickel column due to the presence of six His residues added to the NH_2-terminal portion of the protein by the expression vector). The column was washed with buffer A (25 m*M* Tris–HCl, pH 7.9, 500 m*M* NaCl, 8 *M* urea) and bound proteins were eluted from the column with buffer A plus 0.5 *M* imidazole. Final purification was achieved by chromatography on a gel permeation column S-300 (Pharmacia) with buffer A as eluent. Renaturation was then performed by stepwise dialysis against buffer L.

Assays for DNA binding. DNA-binding reactions were carried out in a final volume of 20 μl by using approximately 30 fmol of Klenow-labeled DNA probe in the presence of 10 m*M* Tris–HCl, pH 7.5, 50 m*M* NaCl, 1 m*M* EDTA, and 1 m*M* DTT at RT for 20 min. The amount of protein used is specified in the figure legends. Protein–DNA complexes were analyzed by electrophoresis through 6% nondenaturing polyacrylamide or 0.8% agarose gels in 0.5× TBE buffer (89 m*M* Tris–borate, 89 m*M* boric acid, 2 m*M* EDTA). For the cruciform DNA binding, gel mobility-shift experiments were performed by incubating varying amounts of T160 protein with 5 fmol of either the labeled cruciform DNA or the labeled linear DNA for 20 min at RT in 10 μl of binding buffer including 10 m*M* Hepes, pH 7.9, 10 m*M* $MgCl_2$, 1 m*M* EDTA, 0.5 m*M* DTT, 200 mg/ml BSA, 5% Ficoll plus 200 m*M* NaCl for stringent conditions or 5 m*M* KCl and 1 m*M* spermidine for nonstringent condition. Gels were dried on 3MM paper (Whatman) and autoradiographed. For quantitation of radioactivity in the protein–DNA complex and in the free DNA probe, the dried gels were exposed directly on a PhosphorImager (Bio-Rad).

DNA circularization assay. The cyclization reactions were performed as described previously by Putman *et al.* (1994) [44]. DNA-binding reactions were carried out as described above. After 20 min incubation at RT, 1 U of T4 polynucleotide ligase (BRL) was added. Reactions with ligase were terminated after 4 min by immediate addition of an equal volume of 0.5% sodium dodecyl sulfate (SDS)–50 m*M* EDTA followed by heat treatment at 65°C for 10 min to denature the T160 and ligase. Reactions mixtures were then phenol–chloroform extracted, ethanol precipitated, washed with 70% ethanol, and resuspended in 10 m*M* Tris–HCl–1 m*M* EDTA (pH 8.0). They were next treated with exonuclease III (50 U, BRL) and incubated at 37°C for 30 min in a 20-μl volume. The enzyme was then inactivated at 65°C for 10 min before addition of marker dyes (0.025% each of xylene cyanol and bromophenol blue in 30% vol/vol glycerol). The samples were electrophoresed at RT in 1× TBE buffer at approximately 10 V/cm for ~3 h on 5% nondenaturing polyacrylamide gels. The gels were then vacuum-dried and autoradiographed.

Antibody production and immunoblot analysis. Antisera against T160 were raised by injecting rabbits with the purified GST–T160 N-terminus fusion protein. The sera obtained after bleeding at 1 week after the fourth immunization were precipitated with ammonium sulfate at 45% saturation. The precipitate was then resuspended in PBS and further purified on a protein A affinity column (Pharmacia) according to the specifications of the supplier.

For Western blotting, adult mouse tissues were lysed in a buffer containing 125 m*M* Tris–HCl, pH 6.8, 3% SDS, 10 m*M* DTT, 10% glycerol, briefly sonicated, and insoluble material was removed by centrifugation. Proteins were electrophoresed through 8.5% SDS–PAGE and electrotransferred to PVDF membrane (Amersham). Immunoblots were blocked with 5% nonfat milk for 2 h at RT and then incubated with anti-T160 antibody (diluted 1:3000) recognizing the N-terminus. Goat anti-rabbit IgG–horseradish peroxidase conjugate was used as second antibody at a dilution of 1:3000 and detected by ECL (Amersham). Monoclonal anti-actin antibody (Boehringer) was used as an internal control.

RESULTS

Cloning of the T160 Protein Binding to the 202 Gene ISRE

To isolate proteins that bind to the GA box, a cDNA library was prepared from poly(A)$^+$ RNA obtained from L1210, a murine leukemia pre-B-cell line highly sensitive to the IFN-α antiproliferative activity. Recombinant phages expressing cDNA-encoding proteins were screened as described by Singh *et al.* (1989) [38] by using a double-stranded trimer probe of the 202 GA box. Of the 7×10^5 cDNA clones screened in the presence of calf thymus DNA, one clone containing a 1.4-kb insert, designated 18.2.1, remained positive throughout subsequent rounds of screening. This insert was then used as a probe to rescreen the library, and cDNA clones with overlapping inserts were obtained and sequenced. A 2650-nt full-length cDNA containing the entire coding information for the GA box-binding protein was obtained. A computer search of the GenBank–EMBL database revealed complete identity with the T160 gene, encoding a high-mobility group protein of 708 aa and an expected molecular mass of 86 kDa (Fig. 1A). In the carboxyl terminus–proximal region, significant homology was noted with

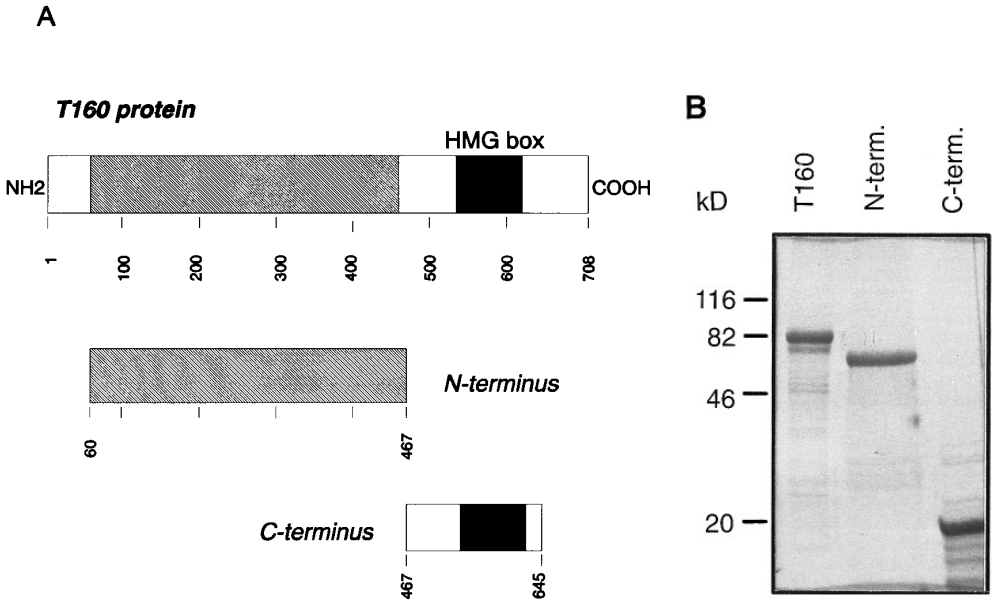

FIG. 1. Schematic representation of the T160 protein. (A) Both T160 and the N-terminus peptide have their amino terminus at position 60. The structures of the truncated proteins, with their amino-terminal and carboxy-terminal amino acid positions denoting the deletion break point, are shown. (B) Coomassie blue-stained SDS–polyacrylamide gel (12%) of the T160 and truncated recombinant proteins. Equal amounts (approximately 2 μg) of each purified protein were electrophoresed. Molecular weight marker positions are indicated in kDa.

the HMG1 protein. Hydropathy profiles performed for the HMG box in the T160 and HMG-1 (box B) proteins demonstrated that the two patterns are very similar and contain two hydrophobic peaks in the hydrophilic regions (data not shown), suggesting that the HMG box could be the DNA-binding domain of the T160 protein. Although T160 and its homologs in different species have been cloned in various laboratories [32, 45–49], the difficulties in keeping this protein in solution due to its physical nature have so far impeded a detailed analysis of its binding characteristics and biological role, despite an extensive effort. The T160 protein produced in bacteria aggregates readily and soluble preparations are only possible in high urea concentrations. These conditions, however, are unfavorable for binding assays. Therefore, to study the DNA-binding properties of T160 in solution, we expressed both the full-length protein and the N-terminal domain using the pET30a plasmid system (Figs. 1A and 1B). Both proteins, purified under denaturing conditions in the presence of high urea concentrations, turned out to be soluble and suitable for band-shift assays when renatured by gradually decreasing urea concentrations. The T160 C-terminus portion, designated C-terminus, expressed as a GST fusion protein, was soluble and suitable for gel shift assays without any denaturation step (Figs. 1A and 1B).

The T160 Protein Binds to the GA Trimer in Solution as Detected by EMSA

Both full-length T160 and truncated polypeptides were examined for DNA binding capability in an electrophoretic mobility-shift assay (EMSA) using the labeled GA trimer as a probe. The T160 protein generated a retarded complex, which retained 98% of the probe at 0.25 μM protein concentration (Fig. 2A, lane 3). At 0.05 μM only a weak band was evident (Fig. 2A, lane 2). Deletion of as many as 467 amino acids from the amino terminus of T160 (C-terminus peptide) did not interfere with DNA binding, as shown in Fig. 2B (lanes 1–4). By contrast, carboxy-terminal truncation of T160 to amino acid 467 (N-terminus peptide) abrogated DNA binding (Fig. 2B, lanes 5–8). These data conclusively indicate that the homology region of T160 with the other HMG box proteins is necessary for DNA binding in solution.

The T160 Protein Binds to the Synthetic Cruciform DNA

Bianchi has described preferential binding of the eukaryotic HMG-1 protein to a synthetic, cruciform DNA [50]. To see whether T160 also binds to the cruciform DNA and to precisely define its binding to four-way junction DNA structures, we have constructed a synthetic cruciform identical to that used by Bianchi for the HMG-1 protein [50]. Figure 3 shows the gel retardation patterns of cruciform DNA upon addition of increasing concentrations of T160 C-terminus under low- or high-salt conditions. At a low-salt concentration (5 mM KCl), the C-terminus (5 nM) generates an initial retarded complex (C1) (Fig. 3, lane 4). When the protein concentration is increased to 0.05 μM, a second, more

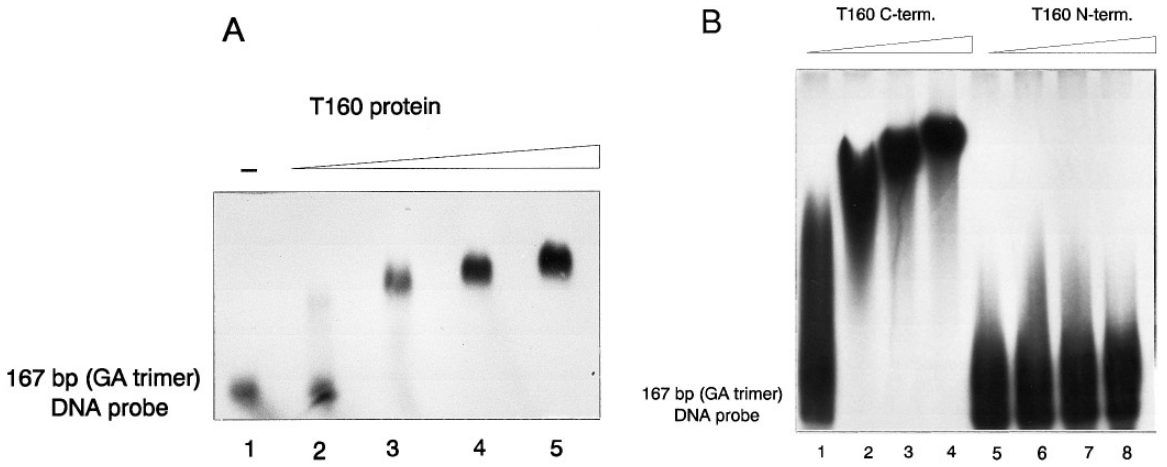

FIG. 2. DNA binding of recombinant T160 and truncated proteins. (A) End-labeled GA trimer was incubated with increasing amounts of T160 protein, such that lane 2 contains 0.05 μM; lane 3, 0.25 μM; lane 4, 0.75 μM; and lane 5, 1.5 μM. The protein–DNA complexes were separated by electrophoresis through a 0.8% agarose gel in 0.5× TBE and visualized by autoradiography. The autoradiogram was exposed for 4 h at −70°C. (B) End-labeled GA trimer was incubated with increasing amounts of either T160 C-terminus (lanes 1–4) or N-terminus peptide (lanes 5–8). The protein–DNA complexes were separated by electrophoresis through a 6% acrylamide gel in 0.5× TBE and visualized by autoradiography. The autoradiogram was exposed for 4 h at −70°C. Lanes 1 and 5 contain 0.2 μM concentrations of the indicated protein; lanes 2 and 6, 1 μM; lanes 3 and 7, 3 μM; lanes 4 and 8, 6 μM.

retarded complex (C2) is observed (Fig. 3, lane 3). At a high-salt concentration (200 m*M* NaCl), the interaction of the C-terminus fragment with cruciform DNA slightly differs. At 5 n*M* protein concentration, no retarded complexes are present (Fig. 3, lane 6). C1 appears at 0.05 μM (Fig. 3, lane 7) and C2 at 0.5 μM (Fig. 3, lane 8), with a significant increase at 5 μM concentration (Fig. 3, lane 9). These results demonstrate that the T160 C-terminus binds to the cruciform DNA at protein concentrations lower than that required for binding linear DNA of appropriate length (Fig. 2B).

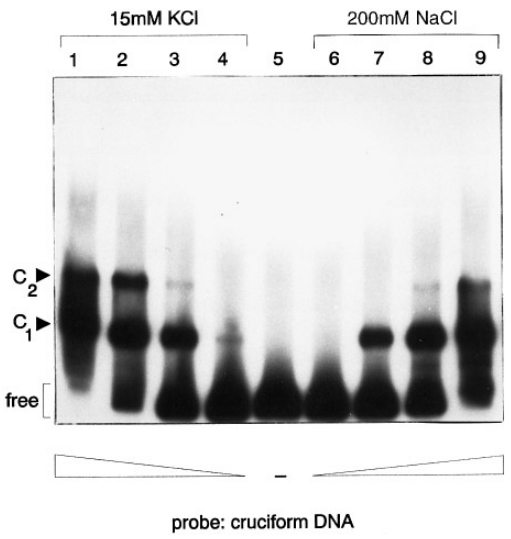

FIG. 3. Binding assays for T160 to complete crucifom DNA. Increasing amounts of T160 C-terminus protein were incubated, as described under Materials and Methods, with 5′ end-labeled cruciform DNA (5 fmol) under both nonstringent (15 m*M* KCl) and stringent (200 m*M* NaCl) conditions. Electrophoresis was performed on a 6% polyacrylamide gel run in the cold. The gel was dried and exposed to film overnight. Lanes 4 and 6 contain 0.005 μM T160 C-terminus; lanes 3 and 7, 0.05 μM; lanes 2 and 8, 0.5 μM; and lanes 1 and 9, 5 μM. The positions of the two DNA–protein complexes C_1 and C_2 are indicated by arrowheads.

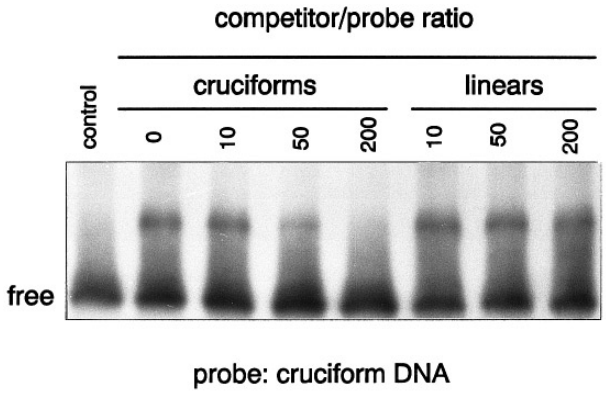

FIG. 4. Competition for binding between cruciform DNA and linear control duplexes. Each reaction mixture in 10 μl (under the stringent condition) contains 5 fmol of labeled cruciform DNA and the indicated amount of unlabeled cruciform DNA or both unlabeled control linear duplexes. After addition of the T160 C-terminus (0.5 μM) and 20 min incubation at room temperature, the samples were electrophoresed on a 6% polyacrylamide gel. The gel was then dried and exposed to film overnight at −70°C. The faster migrating band is free junction DNA (free), the slower is the T160–junction complex.

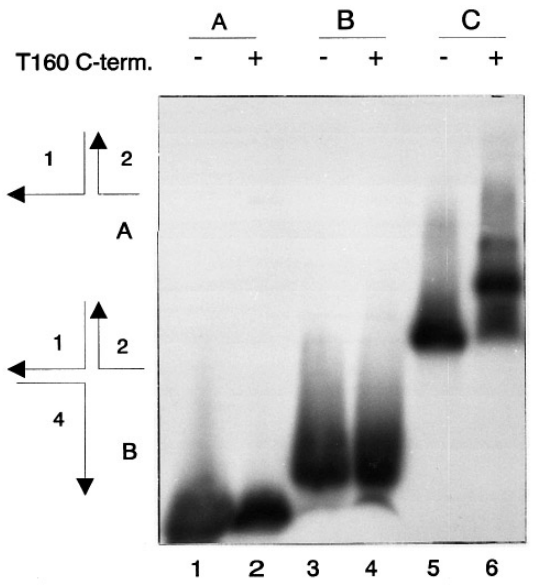

FIG. 5. Binding of the T160 C-terminus protein to incomplete cruciform molecules. Incomplete cruciform molecules were generated by mixing labeled oligonucleotide 1 with unlabeled oligonucleotide 2 (molecule A) or with oligonucleotides 2 and 4 (molecule B). Complete cruciform molecules (C) were generated by mixing together oligonucleotides 1 to 4. The annealed molecules were purified by gel electrophoresis. Incomplete and complete molecules (5 fmol) were then incubated with the T160 C-terminus protein (0.5 μM) under the stringent conditions and electrophoresed on a 6% polyacrylamide gel.

T160 Binds to the Synthetic Cruciform DNA, but Not to Linear Duplex Molecules of Identical Sequence or to Imperfect Cruciform Molecules

To rule out the possibility that T160 recognizes a specific sequence on the synthetic molecule, two additional oligonucleotides were used (5 and 6) to construct linear double-stranded (ds) DNAs (1-6; 5-3), each containing the same sequence as one half of the cruciform DNA [36]. The labeled cruciform DNA was then incubated with a fixed concentration of the T160 C-terminus in the presence of increasing amounts of cold cruciform DNA or linear duplex DNA. Figure 4 shows that the binding activity can be competed off by cold cruciform DNA, but not by a mixture of the two duplex molecules, which are equivalent to the linearized arms of the synthetic cruciform. These findings demonstrate that the T160-binding activity is structure-specific rather than sequence-specific.

To further define this structure specificity, incomplete cruciform molecules were constructed and tested in band-shift assays. Oligonucleotide 1 was labeled at its 5′ end and annealed with oligonucleotide 2, or with oligonucleotides 2 and 4, generating the molecules indicated as A and B, respectively, in Fig. 5. Molecule A is partially ss and partially ds, and is equivalent to the transition from paired to unpaired segments in ss DNA. Molecule B contains two segments of duplex DNA, which can independently rotate around the center of the molecule. Figure 5 shows that the T160 C-terminus does not bind either to molecule A (lane 2) or to molecule B (lane 4). By comparison, the complete cruciform was bound to more than 75% in the same experiments (lane 6).

T160 Binding Has a Strong Dependence on Probe Length

It has been demonstrated that the binding of HMG-box containing proteins, such as UBF, is highly dependent on the length of the DNA fragment [44]. A fixed amount of the end-labeled DNA fragments from 40 to 105 nt in length, obtained from pBend2 plasmid (see Materials and Methods), was therefore tested against increasing concentrations of the T160 C-terminus. Although significant binding was observed with fragments of lengths equal to or greater than 40 bp, strong binding to an 84-bp probe occurred at a T160 concentration (Fig. 6B, lane 2) about three times lower (1 μM) than the 3 μM necessary for binding to a 62-bp element (Fig. 6A, lanes 6–8). With the 105-bp probe binding seems to be even stronger, since almost all of the probe is retained in the retarded complex at a protein concentration of 3 μM (Fig. 6B, lane 7). Significant binding was observed with a 40-bp oligonucleotide (Fig. 6A, lanes 3 and 4), whereas with 25-bp probes no binding was observed, even at the highest concentration used (6 μM) (data not shown). The more stable interaction between T160 and longer DNA fragments may be due to more extensive protein–DNA contacts on the longer probes.

T160 C-terminus Bends Linear DNA Fragments to Facilitate Their Ligation into Circles

HMG box proteins that bend DNA can bring the two ends of a DNA fragment into proximity and facilitate their ligation into circles under dilute conditions [44, 51, 52]. The ability of T160 to bend DNA was examined by studying the cyclization of 84- to 167-bp DNA fragments. Figures 8A and 8B show the ligation products of the 105- and 167 (GA trimer)-bp fragments with blunt ends. In the absence of both the T160 C-terminus and the ligase, the 105-bp probe (Fig. 7A, lanes 1 and 4) and the GA trimer 167-bp probe (Fig. 7B, lane 1) migrate as monomers. In the presence of T4 DNA ligase alone, both fragments formed a set of linear dimers, trimers, and higher molecular weight multimers, but were not converted into a covalently closed circular monomer. Treatment of these multimers with exonuclease III (Exo III), which digests linear duplex DNA molecules, caused the complete disappearance of the multimerized probe (Fig. 7A, lane 6). However, in the presence of both the C-terminus and the ligase, the 105-bp probe

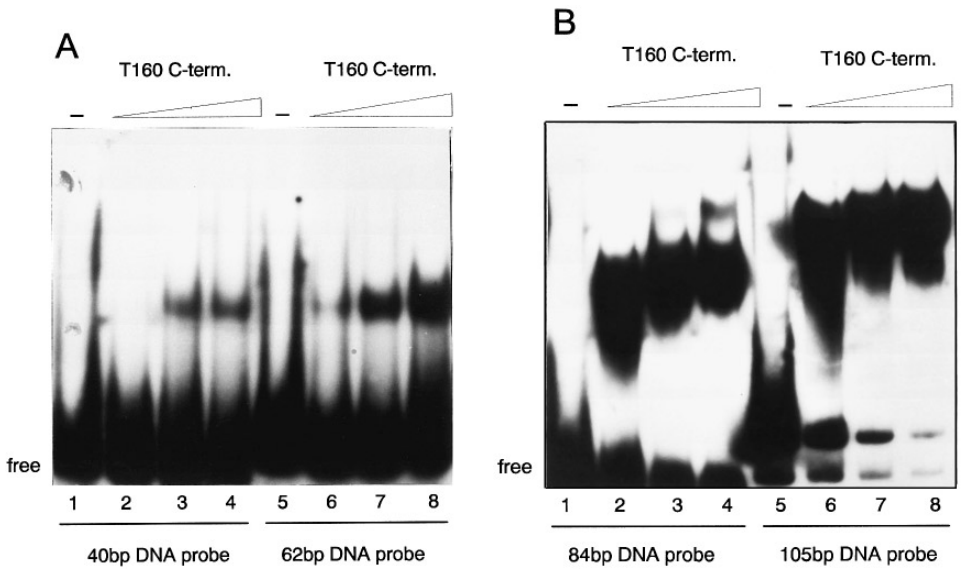

FIG. 6. T160 binding is dependent on DNA length. Increasing amounts of T160 C-terminus were incubated with the indicated end-labeled DNA fragments for 20 min at room temperature. The binding reactions were then loaded onto a native 6% polyacrylamide gel run in the cold. Following electrophoresis, the gel was dried and exposed to film for 4 h at −70°C. In both A and B, lanes 2 and 6 contain 1 μM T160 C-terminus; lanes 3 and 7, 3 μM; lanes 4 and 8, 6 μM.

(Fig. 7A, lane 2) and the GA trimer 167-bp probe (Fig. 7B, lane 2) generated closed-circular monomers that are resistant to the action of Exo III (Figs. 7A, lane 3, and 7B, lane 3, respectively). These results are consistent with the novel DNA being covalently closed. Moreover, the mobility of these ligation products was unchanged by eukaryotic topoisomerases I and II (data not shown), suggesting that they were relaxed, covalently closed circular molecules. By contrast, the 84-bp probe did not notably circularize (data not shown), suggesting that, although able to bind to it, the T160 C-terminus cannot bend an 84-bp DNA fragment to such an extension, allowing the ligase to ligate it. The rapid cyclization in the presence of T160 suggested the ability of the protein to induce DNA distortion, particularly bending of the fragment.

Immunoblotting Analysis of the T160 Expression in Mouse Tissues

The HMB-1/2 subfamily proteins are present to different extents in all mammalian tissues and cell types. They are abundant in rapidly dividing cells and tend to decline in concentration after differentiation [6]. By contrast, the LEF/SRY subfamily proteins have a restricted cell type distribution and their expression is usually temporaly regulated (e.g., SRY during embryonic life) [1].

To gain further insight into the expression pattern of T160, rabbit antibodies highly selective for the T160 N-terminus were generated and used to prove immunoblots of protein extracts from different tissues derived from 6-week-old DBA/2 mice. The T160 protein, with an apparent molecular mass of 86 kDa, was highly expressed in spleen, thymus, gut, brain, testis, and ovary, probably because of their high proliferation, and at low level (detectable only after a very long exposure) in muscle, lung, and heart (Fig. 8). By contrast, all the cell lines tested so far displayed a high level of the T160 protein, regardless their histological origin (data not shown).

DISCUSSION

This report describes the cloning of a DNA-binding protein, T160, from a λgt11 library with a probe containing a trimer of the 202 gene ISRE, designated the GA box. Nucleotide sequencing revealed that T160 contains an HMG box of about 80 aa at the C-terminus which is responsible for binding to DNA. T160 was considered to bind specifically to a heptamer located in the V-(D)-J joining signals [32]. This conclusion was based on previous results showing that the T160–β-galactosidase fusion protein interacted with the 12-bp RSS probe more strongly than with the RSS mutated in the heptamer motif. Owing to the difficulty of keeping the T160 protein in solution, these studies were conducted only with the fusion protein immobilized on a nitrocellulose filter (Southwestern blotting technique), and gel-shift assays or footprinting analyses were unsuccessful [32]. Using either the entire protein, made soluble

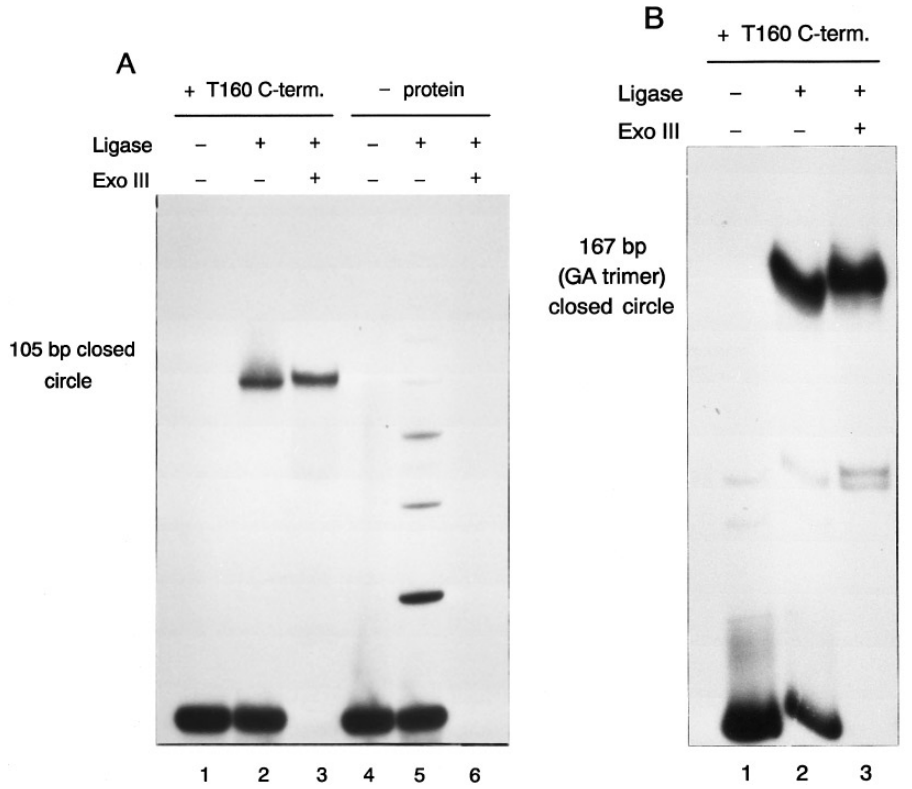

FIG. 7. T160 bends linear DNA upon binding to allow circularization by DNA ligase. Approximately 30 fmol of the end-labeled 105-bp (A) and 167-bp GA trimer (B) fragments were ligated for 4 min in the absence of added protein or in the presence of T160 C-terminus (4 μM). After phenol extraction and precipitation, the DNA was loaded onto a native 5% polyacrylamide gel. Following electrophoresis, the gel was dried and exposed to film overnight at −70°C. In lanes 3 and 6 the DNA was treated with exonuclease (Exo) III to degrade linear molecules. The ligation product formed in the presence of T160 (A and B, lanes 2 and 3) is resistant to exonuclease, indicating that it is circular.

through denaturating and renaturing cycles in urea, or its HMG box-containing domain (C-terminus) we demonstrate, for the first time, in gel-shift assays that T160 binds to cruciform DNA as well as to linear duplex DNA in a nonspecific, sequence-independent manner. T160 binding to cruciform DNA structures is selective, as it does not bind to structures resembling stem and loops of ssDNA, and it binds inefficiently to imperfect cruciform molecules consisting of only three strands. Taken as a whole, our results demonstrate that T160 recognizes non-B-DNA conformations in a nonspecific, sequence-independent manner and shares DNA-binding properties with the HMG-1/2 subfamily [1, 53]. In agreement with our observations, the human T160 counterpart, designated SSRP [34], has also been shown to recognize cisplatin-modified DNA, the motif being probably the local unwinding and bending toward the major groove that occurs upon formation of intrastrand *cis*-$(Pt(NH_3)_2)^{2+}$ d(GpG) and d(ApG) cross-links [13, 45]. On the other hand, *in vitro* binding studies using the *Drosophila* T160 counterpart, DssRP, produced in bacteria and immobilized on a membrane filter, showed that this protein does not bind to structured DNA, but rather to single-stranded DNA and RNA [48].

HMG box-containing proteins recognize structured nucleic acids, such as cruciforms. Canonical HMG-1/2 proteins preferentially bind to non-B-DNA conformations, such as B-Z junctions, stem-loops, cruciforms, four-way junctions, and cisplatin-modified DNA without any sequence-specificity [3]. Other HMG-1 box-containing proteins, such as LEF-1 [15], SRY [54], mtTF1 [11], and ABF2 [9] bind and recognize both specific sequences and non-B-DNA structures. T160, although structurally similar to LEF-1 [16] in that it contains only one HMG-1 box and an extended N-terminus peptide, binds to cruciform and linear DNA with binding properties closer to those of the canonical HMG-1 protein.

Like several other HMG box-containing proteins, T160 does not appear to require interaction with a specific consensus sequence for nucleic acid binding. This was first suggested by the lack of sequence similarity between sequences previously used to clone the T160 protein and its homologs in different species [34, 49, 55]. Another finding is that T160 is also capable of

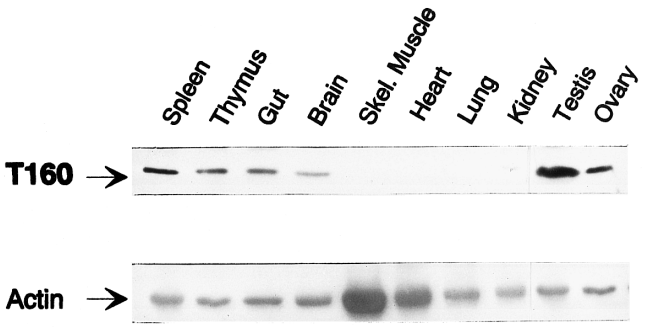

FIG. 8. Expression of T160 protein in mouse tissues as detected by Western immunoblot analysis. Protein lysates (50 μg) from various mouse organs were size fractionated by SDS–PAGE, transferred to PVDF membrane, and probed for the presence of T160 protein with T160 anti-N-terminus antibodies (top). Immunodetection of actin (bottom) using a monoclonal anti-actin antibody was performed as an internal control.

inducing DNA bends, as revealed by the ring closure of short DNA fragments (from 105 to 167 bp) in the presence of T4 DNA ligase. This ring closure was not observed with 84-bp probes, implying a limit to the amount of DNA distortion induced by T160. The question then is, what does T160 recognize at a binding site? Our gel mobility-shift data with those obtained in DNA circularization assays suggest that T160, like other proteins of the HMG-1/2 subfamily, is extremely flexible in its ability to bind DNA. This assumption is further reinforced by our DNase footprinting data showing the lack of a discrete sequence protected by enzymatic digestion on binding to T160 protein (data not shown).

In conclusion, in this report we demonstrate that: (i) the T160 HMG box domain is responsible for DNA binding in solution; (ii) T160 binds to linear as well as to cruciform DNA (four-way junctions) folded to generate an X-shape by the pairwise coaxial stacking of helical arms [56]; (iii) it bends DNA as determined by the cyclization method; and finally (iv) it is highly expressed in various mouse tissues containing actively proliferating cells. Moreover, stable expression of a cDNA encoding antisense T160 RNA in NIH3T3 cells resulted in the permanent down-regulation of the endogenous T160 protein and specifically affected the replication of a mouse DNA virus, the cytomegalovirus [57]. These results, together with the finding that T160 is evolutionary highly conserved (from plants to mammals) and the knockout of the yeast homolog, designated Pob3 for its capability to interact with DNA polymerase 1 [35], resulted in a lethal phenotype, suggest that the T160 protein is very likely involved in more general processes such as DNA replication and/or repairing in addition to V-(D)-J recombination.

We are grateful to Dr. Marco Bianchi for his suggestions, criticisms, and careful revision of the manuscript. We thank Rudolf Grosschedl for providing the LEF-1 cDNAs and B. Reimund for expert experimental contributions. This study was supported by grants from the Italian National Research Council (C.N.R.) (P.F. "A.C.R.O." and Biotecnologíe) and from the Associazione Italiana per la Ricerca sul Cancro (A.I.R.C.).

REFERENCES

1. Baxevanis, A. D., and Landsmann, D. (1995) *Nucleic Acids Res.* **23,** 1604–1613.
2. Johns, E. W. (1982) The HMG Chromosomal Proteins, Academic Press, London.
3. Landsman, D., and Bustin, M. (1993) *BioEssays* **15,** 539–546.
4. Jantzen, H. M., Admon, A., Bell, S. P., and Tjian, R. (1990) *Nature* **344,** 830–836.
5. Jantzen, H. M., Chow, A. M., King, D. S., and Tjian, R. (1990) *Genes Dev.* **6,** 1950–1963.
6. Bustin, M., Lehn, D. A., and Landsman, D. (1990) *Biochim. Biophys. Acta* **1049,** 231–243.
7. Griess, E. A., Rensing, S. A., Grasser, K. D., Maier, U. G., and Feix, G. (1993) *J. Mol. Evol.* **37,** 204–210.
8. Laudet, V., Stehelin, D., and Clevers, H. (1993) *Nucleic Acids Res.* **21,** 2493–2501.
9. Diffley, J. F. X., and Stilmann, B. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7864–7868.
10. Copenhaver, G. P., Putman, C. D., Denton, M. L., and Pikaard, C. S. (1994) *Nucleic Acids Res.* **22,** 2651–2657.
11. Parisi, M. A., and Clayton, D. A. (1991) *Science* **252,** 965–969.
12. Bianchi, M. E., Falciola, L., Ferrari, S., and Lilley, D. M. J. (1992) *EMBO J.* **11,** 1055–1063.
13. Pil, P. M., and Lippard, S. J. (1992) *Science* **256,** 234–237.
14. Giese, K., Amsterdam, A., and Grosschedl, R. (1991) *Genes Dev.* **5,** 2567–2578.
15. Giese, K., Cox, J., and Grosschedl, R. (1992) *Cell* **69,** 185–195.
16. Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991) *Genes Dev.* **5,** 880–894.
17. van de Wetering, M., Oosterwegwl, M., Dooijes, D., and Clevers, H. (1991) *EMBO J.* **10,** 123–132.
18. Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell, B. R., and Goodfellow, P. N. (1992) *Science* **255,** 453–456.
19. Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P., and Ashoworth, A. (1992) *Nucleic Acids Res.* **55,** 2887–2893.
20. Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988) *EMBO J.* **7,** 1537–1547.
21. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y., and Yamamoto, M. (1991) *Genes Dev.* **5,** 1990–1999.
22. Grosschedl, R., Giese, K., and Pagel, J. (1994) *Trends Genet.* **10,** 94–100.
23. Ner, S. S. (1992) *Curr. Biol.* **2,** 208–210.
24. Kuhn, A., and Grummt, I. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7340–7344.
25. Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M. E., and Grummt, I. (1994) *EMBO J.* **13,** 416–424.
26. Shykind, B. M., Kim, J., and Sharp, P. A. (1995) *Genes Dev.* **9,** 1354–1365.
27. Ge, H., and Roeder, R. G. (1994) *J. Biol. Chem.* **269,** 17136–17140.

28. Nightingale, K., Dimitrov, S., Reeves, R., and Wolffe, A. P. (1996) *EMBO J.* **15,** 548–561.
29. Gariglio, M., Gaboli, M., Mana, C., Ying, G. G., Gribaudo, G., Cavallo, R., and Landolfo, S. (1994) *Eur. J. Biochem.* **221,** 731–739.
30. Ito, N., Gribaudo, G., Toniato, E., Thakur, A., Yagi, Y., Barbosa, J., Kamarck, M., Ruddle, F. H., and Lengyel, P. (1989) *in* UCLA Symposium on Growth Inhibitory and Cytotoxic Polypeptides (Sporn, M. *et al.,* Eds.), pp. 169–178, A. R. Liss, New York.
31. Gariglio, M., Foresta, P., Ying, G. G., Gaboli, M., Lembo, D., and Landolfo, S. (1996) *J. Cell. Biochem.* **60,** 83–94.
32. Shirakata, M., Hüppi, K., Usuda, S., Okazaki, K., and Sakano, H. (1991) *Mol. Cell. Biol.* **11,** 4528–4536.
33. Lewis, S. M. (1994) *Adv. Immunol.* **357,** 282–283.
34. Bruhn, S. L., Pil, P. M., Essigmann, J. M., Housman, D. E., and Lippard, S. J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2307–2311.
35. Wittmeyer, J., and Formosa, T. (1996) *Saccharomyces* Genome Database.
36. Bianchi, M. E. (1988) *EMBO J.* **7,** 843–849.
37. Singh, H., LeBowitz, J. H., Baldwin, A. S., and Sharp, P. A. (1988) *Cell* **52,** 415–423.
38. Singh, H., Clerc, R. G., and LeBowitz, J. H. (1989) *BioTechniques* **7,** 252–261.
39. Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989) *Gene* **85,** 15–23.
40. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1992) *EMBO J.* **11,** 4497–4506.
41. Nakano, H., Yamazaki, T., Ikeda, M., Masai, H., Miyatake, S., and Saito, T. (1994) *Nucleic Acids. Res.* **22,** 543–544.
42. Smith, D. B., and Johnson, K. S. (1988) *Gene* **67,** 31–40.
43. Studier, F. W., Rosenberg, H. A., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
44. Putman, C. D., Copenhaver, G. P., Denton, M. L., and Pikaard, C. S. (1994) *Mol. Cell. Biol.* **14,** 6476–6488.
45. Bruhn, S. L., Housman, D. E., and Lippard, S. J. (1993) *Nucleic Acids Res.* **21,** 1643–1646.
46. Bunker, C. A., and Kingston, R. E. (1995) *Nucleic Acids Res.* **23,** 269–276.
47. Hotze, M., Lurz, G., and Schoder, J. (1995) *Gene,* **161,** 295–296.
48. Hsu, T., King, D. L., LaBonne, C., and Kafatos, F. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6488–6492.
49. Wang, L., Precht, P., Balakir, R., and Horton, W. E. (1993) *Nucleic Acids Res.* **21,** 1493.
50. Bianchi, M. E., Beltrame, M., and Paonessa, G. (1989) *Science* **243,** 1056–1059.
51. Bianchi, M. E. (1995) *in* DNA–Protein Structural Intercations (Lilley, D. M. J., Ed.), pp. 177–200, Oxford Univ. Press, Oxford.
52. Pil, P. M., Chow, C. S., and Lippard, S. J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9465–9469.
53. Crothers, D. M. (1993) *Curr. Biol.* **3,** 675–676.
54. King, C. Y., and Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11990–11994.
55. Yamaguchi-Shinozaki, K., and Shinozaki, K. (1992) *Nucleic Acids Res.* **20,** 6737.
56. Lilley, D. M. (1992) *Nature* **357,** 282–283.
57. Gariglio, M., Foresta, P., Sacchi, C., Lembo, D., Hertel, L., and Landolfo, S. (1997) *J. Gen. Virol.* **78,** 665–670.

Received July 4, 1997
Revised version received July 24, 1997