Specific Recognition of AT-Rich DNA Sequences by the Mammalian High Mobility Group Protein AT-hook 2: A SELEX Study†

Tengjiao Cui and Fenfei Leng*

Department of Chemistry and Biochemistry, Florida International University, Miami, Florida 33199

Received June 28, 2007; Revised Manuscript Received September 13, 2007

ABSTRACT: The mammalian high mobility group protein AT-hook 2 (HMGA2) is a transcriptional factor involved in cell differentiation and transformation. Disruption of its normal expression pattern is directly linked to oncogenesis and obesity. HMGA2 contains three "AT-hook" DNA binding domains, which specifically bind to the minor groove of AT-rich sequences. Using a PCR-based systematic evolution of ligands by exponential enrichment (SELEX) procedure, we have identified two consensus sequences for HMGA2, 5′-ATATTCGCGAWWATT-3′ and 5′-ATATTGCGCAWWATT-3′, where W represents A or T. These two consensus sequences have a unique and interesting feature: the first five base pairs are AT-rich, the middle four base pairs are GC-rich, and the last six base pairs are AT-rich. Our results showed that all three of these segments are critical for high-affinity binding of HMGA2 to DNA. For example, if one of the AT-rich sequences is mutated to a non-AT-rich sequence, the DNA binding affinity of HMGA2 is reduced at least 100-fold. Intriguingly, if the GC-segment is replaced by an AT-rich segment, the binding affinity of HMGA2 is reduced approximately 5-fold. Identification of the consensus sequences for HMGA2 represents an important step toward finding its binding sites within the genome.

The mammalian high mobility group protein AT-hook 2 (HMGA2[1]) is a transcriptional factor regulating mesenchymal cell development and differentiation (*1*−*4*). It was first discovered as a nuclear protein in proliferating fibroblasts and embryos and also in transformed thyroid cells infected by oncogenic viruses (*5*, *6*). The aberrant expression of HMGA2 has been attributed to the formation of a variety of tumors, including benign tumors, such as lipomas (*7*, *8*) and uterine leiomyomas (*9*−*11*), and malignant tumors, such as breast cancer (*12*), lung cancer (*13*−*15*), and leukemia (*12*, *16*). The expression level is correlated with the degrees of malignancy and metastatic potential of the transformed cells (*17*), suggesting that HMGA2 can be used as a biomarker for diagnosing the neoplastic transformation and the metastatic potential of many cancers (*13*, *18*). HMGA2 is only expressed in proliferating, undifferentiated mesenchymal cells and is undetectable in normal fully differentiated adult cells (*1*, *19*). Disruption of its normal expression patterns causes deregulations of cell growth and differentiation. For example, *hmga2* knock-out mice developed the pygmy phenotype (*1*). These mutant mice were severely deficient in fat cells and other mesenchymal tissues. Breaking of the *hmga2* gene caused a dramatic reduction in obesity of leptin-deficient mice (Lep^{ob}/Lep^{ob}) in a gene-dosage-dependent manner (*2*). These results suggest that HMGA2 plays an important role in fat cell proliferation and is a potential target for treatment of obesity (*2*).

HMGA2 is a member of the HMGA family, which also includes another two proteins, HMGA1a and -1b. HMGA1a and -1b are splice variants of the same gene, the HMGA1 gene (*20*). HMGA2 is the product of a separate gene, the HMGA2 gene (*21*). One unique feature of HMGA proteins is that all have three "AT-hook" DNA binding domains. These DNA binding domains contain a consensus sequence, PRGRP, flanked on each side by one or two positively charged amino acids (arginine or lysine). The AT-hook DNA binding domain, in the absence of DNA, is "unstructured" (*22*, *23*). However, when it binds to the minor groove of AT-rich sequences, this domain adopts a defined conformation (*23*). The central core, RGR, deeply penetrates into the minor groove of AT base pairs with two arginine residues, forming extensive electrostatic and hydrophobic contacts with the floor of the minor groove (*23*). Two prolines on each side of the RGR core direct the protein away from the floor of the minor groove and position the positively charged arginine or lysine near the negatively charged phosphate backbone, thereby making further contacts. This disordered-to-ordered conformational change significantly increases HMGA proteins' adaptability and allows them to participate in a variety of nuclear activities, such as gene transcription, DNA replication, chromatin remodeling, and DNA repair (*4*, *24*).

Since their discovery, HMGA proteins were shown to preferentially bind to AT-rich DNA sequences in the promoter regions. The early qualitative footprinting studies showed that they could bind to any run of five to six AT base pairs with similar DNA binding affinity (*25*). These studies also suggested that HMGA proteins only recognize

† F.L. acknowledges NIH Grant S06 GM008205 and a grant from the Chinese National Science Foundation (Grant Number 20373026).

* To whom correspondence should be addressed. Tel: 305-348-3277. Fax: 305-348-3772. E-mail: lengf@fiu.edu.

[1] Abbreviations: HMGA2, The mammalian high mobility group protein AT-hook 2; SELEX, systematic evolution of ligands by exponential enrichment; EMSA, electrophoretic mobility shift assay; IPTG, isopropyl-β-D-thiogalactopyranoside.

10.1021/bi701269s CCC: $37.00
Published on Web 10/23/2007

the minor groove configuration of AT base pairs rather than specific sequences. However, more recent studies showed that HMGA proteins should have sequence specificity. For example, the high-affinity binding of HMGA proteins requires two or three appropriately spaced AT-rich sequences as a single multivalent binding site (*26*). The DNA binding affinity of this multivalent binding is much higher than that of the single valent binding (*26*). More significantly, each HMGA protein always simultaneously binds to two or three runs of AT base pairs in the regulatory transcription regions, such as the human interferon β enhancer (*27*, *28*), the promoter regions of interleukin-2 gene (*29*) and interleukin-2 receptor α-chain gene (*30*), and the promoter region of interleukin-15 gene (*31*). These results suggest that HMGA proteins bind to specific DNA sequences as transcriptional factors. More evidence for sequence-specific DNA binding of HMGA proteins comes from the NMR structural studies (*23*). The structure of an HMGA1a−DNA complex showed that the AT-hook DNA binding domain binds to the 5′-AAATT-3′ sequence in a fixed orientation with the N-terminal arginine of the core RGR sequence located near the 3′-end of the sequence and the C-terminal arginine located near the 5′-end of the sequence. This binding orientation resulted from hydrophobic interactions of the arginine side chains of the RGR core with the adenine bases of the DNA sequence and from hydrogen bonding between the AT-hook DNA binding domains and the bases of the binding site. These results suggest that HMGA proteins do not randomly bind to any AT-rich DNA sequences. In this study, we used an in vitro systematic evolution of ligands by exponential enrichment (SELEX) experiment to further investigate the DNA binding specificity of HMGA2. Our results showed that HMGA2 specifically recognizes a type of 15 bp AT-rich DNA sequence: the first five base pairs are AT-rich, the middle four or five base pairs are GC-rich, and the last five or six base pairs are AT-rich. Interestingly, all three segments are critical for high-affinity binding of HMGA2 to DNA.

MATERIALS AND METHODS

Preparation of [^{3}H]HMGA2 (^{3}H-Labeled HMGA2). *Escherichia coli* strain BLR(DE3), which has plasmid pMGM1, was grown in 200 mL of a supplemented M9 medium containing 1× M9 salts, 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 μg/mL thiamine, 40 μg/mL L-leucine, 150 μg/mL L-proline, 50 μg/mL each of the other 18 L-amino acids, and 50 μg/mL kanamycin. The cell growth was monitored by measuring OD_{595}. At $OD_{595} \approx 1$, cells were harvested by centrifugation at 4000 rpm and 4 °C and resuspended in 100 mL of the supplemented M9 medium in the absence of L-lysine. After 40 min incubation at 37 °C, another 100 mL of the supplemented M9 medium (prewarmed to 37 °C) containing 10 mCi of [^{3}H]-L-lysine (10 μg/mL) was added into the cell culture. When the OD_{595} reached ~0.8, HMGA2 expression was induced by addition of 1 mM of IPTG into the cell culture. After an additional 3 h incubation, the cells were harvested by centrifugation at 4000 rpm for 25 min at 4 °C. [^{3}H]HMGA2 was purified as described previously (*32*). The specific activity of [^{3}H]-HMGA2 was determined to be 375 cpm/pmol.

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) Experiment. SELEX experiments started with a 61 bp double-stranded DNA library that contains 15 random nucleotides. This DNA library was constructed through amplification of a 61 base synthetic oligodeoxynucleotide FL-250, 5′-CATGGTACCTCTAGAGGCTCGAG-$(N)_{15}$GCTAGCTGGCATGCAAGCTTCAC-3′(N_{15} represents the 15 random nucleotides and underlined sequences are cleavage sites for *Kpn* I and *Hin*d III, respectively), using two primers, FL-248 (5′-CATGGTACCTCTAGAGGC-3′) and FL-249 (5′-GTGAAGCTTGCATGCCAG-3′), which correspond to the first 18 bases (top strand) and the last 18 bases (bottom strand). For the first round of selection, 80 nM of DNA (9.6×10^{-13} mol of DNA molecules) was incubated with increasing amount of HMGA2 in 12 μL of 1× DNA binding buffer containing 20 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, 50 mM NaCl, and 5% glycerol. After 30 min incubation at 22 °C, the DNA samples were loaded on a 12% native polyacrylamide gel to separate the bound and free species. Band shifts were detected by SYBR Gold staining. The bound DNA was excised and eluted in an elution buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 100 mM NaCl and purified by phenol extraction and ethanol precipitation. The purified DNA was dissolved in water and PCR-amplified for the next round of SELEX selection. Twenty cycles of PCR amplification were performed for 1 min each at 95, 55, and 72 °C with an additional step of 3 min at 72 °C for the final cycle of the PCR reaction. The PCR products were purified by phenol extraction and ethanol precipitation. Ten rounds of SELEX selection were performed. After the eighth round, the concentration of NaCl was increased to 200 mM, and 30 μM (bp) of poly(dG-dC)$_2$ was added to the binding reactions as a competitor to promote selection toward a high-affinity and specificity pool of DNA fragments. After the 10th round of selection, the PCR-amplified products were cloned into *Hin*d III−*Kpn* I sites of pUC18, which were used to transform *E. coli* strain DH5α. The plasmid DNA of each transformed colony was purified and sequenced. DNA sequences, corresponding to the 15 random nucleotides, were analyzed using the programs ClustalX 1.81 (*33*) and MEME 3.5.3 (*34*).

Electrophoretic Mobility Shift Assay (EMSA). EMSA experiments were used to determine the apparent DNA binding constant of HMGA2. DNA oligomers containing AT-rich DNA sequences were labeled with ^{32}P at the 5′ termini by T4 polynucleotide kinase in the presence of γ-[^{32}P]-ATP. The protein−DNA complexes were formed by addition of appropriate amounts of the protein to a solution containing 1 nM of ^{32}P-labeled DNA in the 1× DNA-binding buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 1.5 μM (bp) poly(dG-dC)$_2$, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, and 5% glycerol. After equilibration for 60 min at 22 °C, the samples were loaded on a 12% native polyacryamide gel in 0.5× TBE buffer (0.045 M Tris−borate (pH 8.3) and 1 mM EDTA) to separate free and bound DNA. The gels were subsequently dried and visualized by autoradiography or quantitated using a Fuji FLA 3000 image analyzer. The radioactivity of the free and bound DNA was determined and used to calculate the binding ratio (R), which is equal to the ratio of the radioactivity of the bound DNA divided by the sum of the radioactivity of the bound and free DNA. The apparent DNA binding constant (K_{app}) was obtained by nonlinear-least-

squares fitting the following equation using the program Scientist

$$R = \frac{(a + x + 1/K_{app}) - \sqrt{(a + x + 1/K_{app})^2 - 4ax}}{2a} \quad (1)$$

where a and x represent the total DNA and the total protein concentration, respectively.

EMSA experiments were also used to determine binding stoichiometries of HMGA2 binding to DNA using ^{32}P-labeled ologonucleotides and [^{3}H]HMGA2. In these experiments, 200 nM of ^{32}P-labeled AT-rich DNA oligomers with a specific activity of ~700 cpm/pmol were mixed with 0.5 μM of [^{3}H]HMGA2 in 50 μL in the 1× DNA binding buffer as described above. After incubation at 22 °C for 1 h, the samples were loaded on a 12% native polyacryamide gel in 0.5× TBE buffer to separate free and bound DNA. The gels were subsequently stained by SYBR Gold. DNA bands corresponding to the HMGA2−DNA complexes were excised and electroeluted into a dialysis bag (8000 MWCO, BioDesign, NY) in 1× TBE containing 0.1% SDS. The eluents in the dialysis bag were mixed with 15 mL of Aquasol 2 (Perkin-Elmer, MA) and counted in ^{32}P and ^{3}H channels in a liquid scintillation spectrometer to determine the molar ratio of HMGA2 to DNA oligomers. Alternatively, a method previously described by Carey (*35*) was also used to determine the binding stoichiometries. After EMSA experiments, the DNA bands were located by autoradiography and excised. Gel slices containing the HMGA2−DNA complex were solubilized by oxidation in 1 mL of 21% H_2O_2/17% $HClO_4$ in a tightly sealed scintillation vial at 65 °C for 24 h. Vials were cooled to 22 °C, mixed with 15 mL of Aquasol 2, and then stored at 4 °C for 48 h in dark. Radioactivities of ^{3}H and ^{32}P were counted in ^{3}H and ^{32}P channels using the method as described above. Background was determined by counting the control gel slices excised from the unused regions.

RESULTS AND DISCUSSION

Determining the Binding Stoichiometries of HMGA2 Binding to AT-Rich DNA Oligonucleotides. Mouse HMGA2 is a small basic protein containing 108 amino acid residues. One unique feature of HMGA2 is the asymmetric charge distribution of its primary structure (Figure 1 of ref *32*). The positive charges are mainly concentrated in the three AT hooks and the negative charges at the C-terminus. In solution, HMGA2 may self-associate into homodimers or homo-oligomers through electrostatic interactions and may bind to AT-rich DNA as a homodimer (*36*, *37*). Thanos and colleagues showed that HMGA1a, another member of the HMGA family with similar physical properties, interacts with itself in a glutathione *S*-transferase (GST)-pull down experiment (*36*). Our preliminary results showed that free HMGA2 in solution may be a homodimer (data not shown). We, therefore, decided to determine whether HMGA2 binds to AT-rich DNA as a monomer or as a homodimer.

As demonstrated in our previous publication (*38*), each HMGA2 (monomer) binds to 15 AT base pairs in aqueous buffer solution. In this study, we used two DNA oligomers, FL-AT15 (5′-CATGGTACCTTCAGAGGCTCGAGAAA-AAAAAAAAAAAAGCTGACTGGCATGCAAGCTG-3′) and FL-AT30 (5′-CATGGTACCTTCAGAGGCTCGAG-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTGACT-GGCATGCAAGCTG-3′), to determine the stoichiometries of HMGA2 binding to AT-rich DNA sequences. FL-AT15 contains a 15 bp A-track (one HMGA2-binding site) and FL-AT30 contains a 30 bp A-track (two HMGA2-binding sites). As an initial step, the apparent DNA binding constants of HMGA2 (K_{app}) were determined using an electrophoretic mobility shift assay (EMSA). Figure S1A (Supporting Information) shows the results of HMGA2 titrating into a solution containing FL-AT15. In this EMSA experiment, only one shift band was observed (the bands at the top, Figure S1A). The apparent DNA binding constant of HMGA2 binding to FL-AT15 was estimated to be 2.2×10^6 M^{-1} according to the method as described under Materials and Methods. Since we used 200 mM of NaCl and 1.5 μM (bp) of poly(dG-dC)$_2$ (as a competitor) to reduce the nonspecific binding between HMGA2 and the oligomers (see Figure S1 legend for details), the relatively low value of K_{app} may represent specific binding of HMGA2 to the AT-rich region of FL-AT15. Figure S1B shows results of HMGA2 titrating into a solution containing FL-AT30 in which two shift bands were generated. Because FL-AT30 contains two HMGA2-binding sites, it is reasonable to assume that these two shift bands represent HMGA2 occupying one and two sites on FL-AT30, respectively. Interestingly, the apparent binding constant of HMGA2 binding to the first site of FL-AT30 was estimated to be 2.0×10^7 M^{-1}, 10-fold higher than K_{app} of FL-AT15. These results indicate that HMGA2 may cooperatively bind to DNA containing multiple AT-rich sites. An alternative explanation would be that FL-AT30 contains 15 HMGA2 binding sites, which should result in an order of magnitude higher apparent binding constant for FL-AT30.

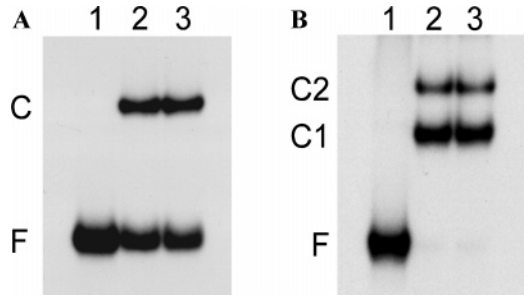

FIGURE 1: Determination of DNA-binding stoichiometries of the HMGA2−DNA complexes by the double-label experiment. EMSA experiments using ^{32}P-labeled FL-AT15 (A) or FL-AT30 (B) and [^{3}H]HMGA2 were performed as detailed under Materials and Methods. The autoradiograms of the ^{32}P-labeled DNA oligomers shown were used to excise the DNA bands for scintillation counting. Lane 1 is the free DNA. Lanes 2 and 3 contained ^{32}P-labeled DNA oligomers and [^{3}H]HMGA2. F is the free DNA; C represents the HMGA2−FL-AT15 complex; C1 and C2 represent the first and second shifted bands of HMGA2−FL-AT30 complexes, respectively.

The binding stoichiometries of HMGA2−DNA complexes were measured directly in EMSA experiments using [^{3}H]-HMGA2 and ^{32}P-labeled FL-AT15 or FL-AT30 (the double-label experiments). In these experiments, high concentrations of HMGA2 and DNA oligomers were used for stoichiometric binding (Figure 1). In addition, 200 mM of NaCl (physiologically relevant salt concentration) and 30 μM (bp) of poly(dG-dC)$_2$ (as a competitor for nonspecific binding) were added to the binding reactions. The shifted bands were

Table 1: DNA Binding Stiochiometries of HMGA2 Binding to Two AT-Rich DNA Oligomers Determined by the Double-Label Experiment

	HMGA2−FL-AT15 complex	HMGA2−FL-AT30 complex	
		first shift	second shift
method I[a]	0.92 ± 0.17	0.98 ± 0.05	1.97 ± 0.23
method II[b]	0.93 ± 0.05	0.92 ± 0.05	1.88 ± 0.13

[a] HMGA2−DNA complexes were excised from the polyacrylamide gels after EMSA experiments as described under Materials and Methods. The ^{32}P-labeled DNA and [^{3}H]HMGA2 were eluted from the gel in 1× TBE containing 0.1% SDS at 50 mA for 4 h. The binding stoichiometries were calculated from the radioactivities of ^{32}P and ^{3}H, determined by scintillation counting. [b] The gel slices containing [^{3}H]HMGA2 and ^{32}P-labeled DNA oligomers were dissolved in 1 mL of solution containing 21% H_2O_2 and 17% $HClO_4$ at 65 °C for 24 h. The binding stoichiometries were calculated from the radioactivities of ^{32}P and ^{3}H, determined by scintillation counting.

excised; the radioactivities of ^{3}H and ^{32}P in the protein−DNA complexes were determined by scintillation counting. As described under Material and Methods, two methods were used to elute [^{3}H]HMGA2 and ^{32}P-labeled oligomers into solutions. We either electroeluted [^{3}H]HMGA2 and the ^{32}P-oligomers into 1×TBE containing 0.1% SDS or dissolved the gel slices into a solution containing 21% H_2O_2 and 17% $HClO_4$ by oxidization. No difference was found between these two methods. Our results are summarized in Table 1. For the HMGA2−FL-AT15 complex, the shifted band gave a 1:1 molar ratio of HMGA2 per FL-AT15 molecule. For the HMGA2−FL-AT30 complex, the first shifted band yielded a 1:1 molar ratio of HMGA2 per FL-AT30 molecule and the second shifted band produced a 2:1 molar ratio of HMGA2 per FL-AT30 molecule. These results indicate that HMGA2 binds to AT-rich sites as a monomer under physiologically relevant salt conditions. Similar results were also obtained by using DNA oligomers containing one or two HMGA2 DNA binding sites determined by SELEX experiments (data not shown; see below for the SELEX experiments).

In Vitro Selection of DNA Oligomers That Bind to HMGA2 with High Affinity. We next used SELEX (*39*−*42*) experiments to determine the consensus DNA binding sequences for HMGA2 (Figure 2A). In these experiments, we utilized a library of DNA oligonucleotides containing a central randomized region of 15 bp, flanked by two 23 bp regions with defined sequences in each end (total 61 bp; Figure 2A). Fifteen base pair randomized sequences were chosen because each HMGA2 molecule binds to 15 AT base pairs (*38*). For the first round of the SELEX experiments, 9.6×10^{-13} mol of DNA oligonucleotides ($\sim 5.8 \times 10^{11}$ molecules) were used to ensure a completely random population for selection (approximately 500 copies of each of the 4^{15} possible

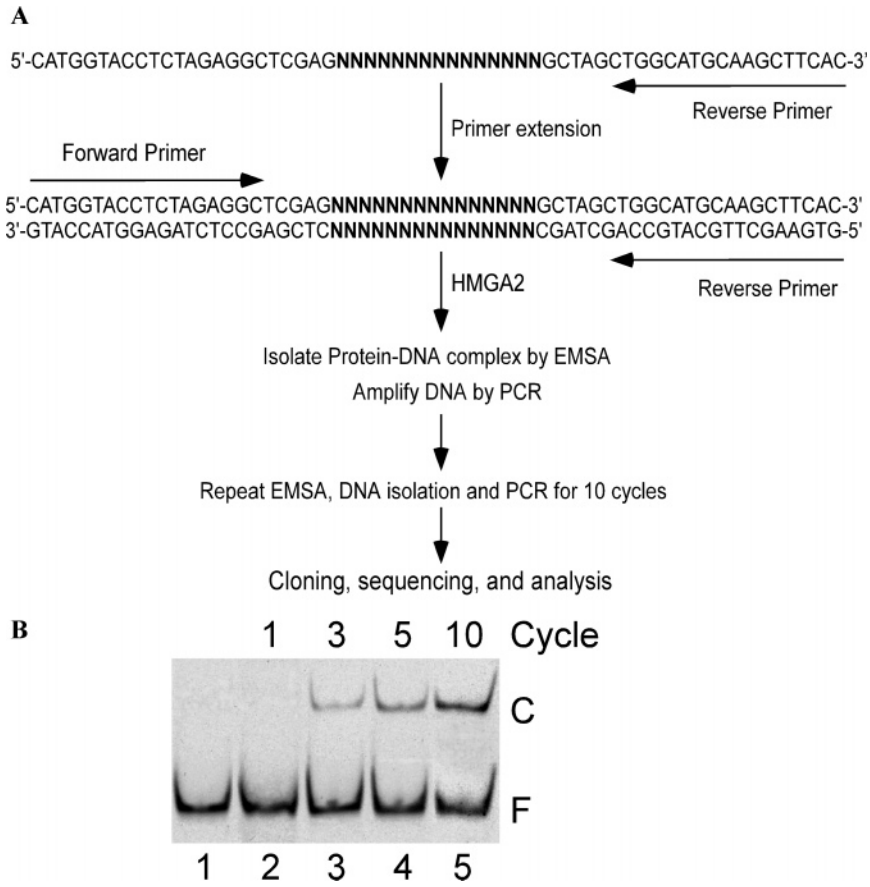

FIGURE 2: The SELEX experiments. (A) Strategy for SELEX analysis of HMGA2 DNA binding sites. (B) Representative EMSA analysis used to monitor the progress of enrichment for HMGA2 binding sites. About 80 nM of the DNA samples isolated after the first (lane 2), third (lane 3), fifth (lane 4), and tenth (lane 5) rounds of the selection were incubated with 200 nM of HMGA2 in an EMSA buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, and 5% glycerol. EMSA experiments were performed as described under Materials and Methods. Following electrophoresis, the PAGE gels were stained with SYBR Gold, destained, and photographed under UV light. Lane 1 is the free DNA library before the selection. The HMGA2−DNA complex and the free DNA were designated as C and F, respectively.

sequences on average). This library of oligonucleotides was mixed with increasing amounts of HMGA2, incubated at 22 °C for 60 min, and loaded on a 12% polyacrylamide gel to separate free and bound DNA. FL-AT15 was used as a control to define the position to which the HMGA2−DNA complex was migrated. The gel containing the HMGA2−DNA complexes was excised; DNA was eluted from the gel and amplified by PCR according to methods as described under Materials and Methods. The PCR products were purified and subjected to the next round of selection. After the eighth round, the concentration of NaCl was increased to 200 mM, and 30 μM (bp) of poly(dG-dC)$_2$ was added to the binding reactions to increase the binding specificity. The progress of the enrichment of DNA molecules recognized by HMGA2 was assessed by an EMSA experiment in which no HMGA2−DNA complex of the original DNA library and the DNA pool of the first round of selection was visible in the presence of 200 nM of HMGA2 (lanes 1 and 2, Figure 2B). In contrast, the DNA pools of rounds 3, 5, and 10 gave approximately 7.1%, 17.5%, and 35.2% of HMGA2−DNA complexes under the same conditions (lanes 3−5, Figure 2B).

After the 10th round of selection, the purified PCR products were cloned into *Hin*d III−*Kpn* I sites of pUC18 and sequenced. We repeated the SELEX experiment twice and sequenced 102 individual clones in which 71 sequences are unique. Figures 3A and S2 (Supporting Information) show our sequencing results. To our surprise, all sequences have a common unique feature: the first five base pairs are AT-rich, the middle four or five base pairs are GC-rich, and the last five or six base pairs are AT-rich. The 71 unique sequences were further analyzed by two multiple-sequence-alignment programs, ClustalX 1.81 (*33*) and MEME 3.5.3 (*34*), and two consensus sequences, 5′-ATATTCGCGAW-WATT-3′ and 5′-ATATTGCGCAWWATT-3′, where W represents A or T, were identified. These two consensus sequences are identical to those obtained from manual alignment of the SELEX sequences (Table 2). Another striking feature of these sequences is that, although each clone is unique, certain sequence motifs appeared multiple times. For example, a 5 bp AT-rich sequence, 5′-ATATT-3′, occurred 58 times; two GC-rich sequences, 5′-CGCG-3′ and 5′-GCGC-3′, appeared 22 and 13 times, respectively (Figures 3A and S2). This finding further suggests that we have identified HMGA2 binding sites with high affinity. We also sequenced 20 clones from the original library to demonstrate that there is no inherent bias in favor of our SELEX experiments. These sequences are shown in Figure 3B. As expected, none of these sequences contains the consensus motifs, such as 5′-ATATT-3′ and 5′-CGCG-3′. A simple addition analysis showed that this pool of sequences has more G+C base pairs (173 G+C base pairs) than A+T base pairs (129 A+T base pairs), further suggesting that this pool of DNA oligonucleotides was not produced in favor of our selection. Although more sequence data are required for a meaningful statistical analysis to demonstrate the unbiased nature of the starting library, we believe that our DNA library was generated randomly. First, the middle 15 bp of the DNA oligonucleotide FL-250, which was used to generate the starting DNA library (Figure 2A), were synthesized randomly. In addition, we included, on average, approximately 500 copies of each of the 4^{15} possible sequences in our first

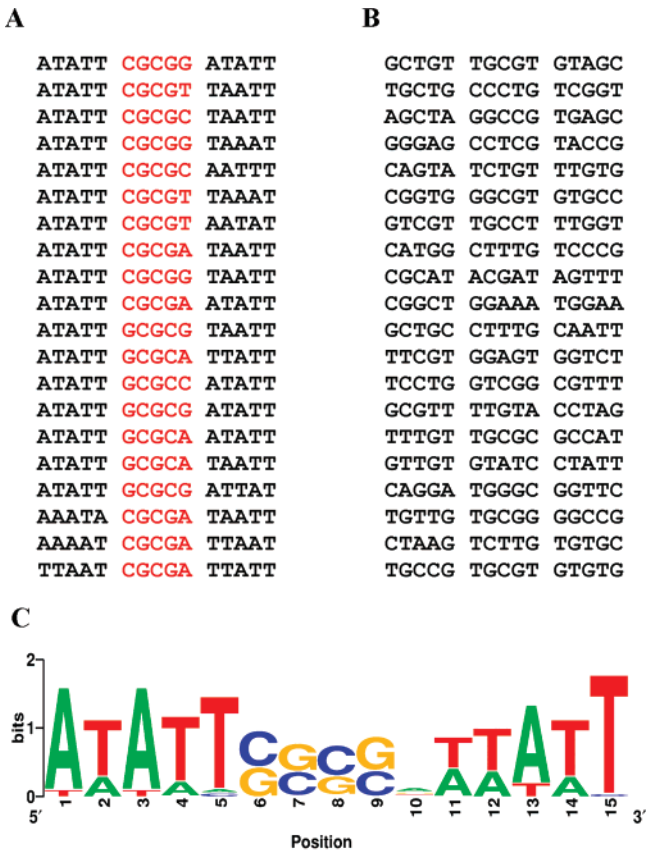

FIGURE 3: Sequence analysis of the SELEX experiments for HMGA2. (A) HMGA2 binding sequences identified after 10 rounds of the SELEX experiments. Twenty sequences are shown here. The other 51 sequences are shown in Figure S2. (B) DNA sequences from the randomized oligonucleotide library before selection. (C) Sequence logo of the 71 SELEX sequences shown in Figures 3A and S2. Sequence conservation, measured in bits of information (2 bits is the highest), is illustrated by the height of stacking of the four letters for each position in the binding sites. The relative heights are proportional to their frequencies shown in the 71 SELEX sequences. The sequence logo was generated by WebLogo (available at www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi).

round of selection to increase the randomness of the DNA library. As mentioned above, we repeated our SELEX experiment twice and used different batches of FL-250 to produce the starting DNA library. Sequences obtained from these two independent SELEX experiments have the same feature: the first five base pairs are AT-rich, the middle four or five base pairs are GC-rich, and the last five or six base pairs are AT-rich. All this evidence suggests that our starting DNA library should contain all possible sequences for the SELEX experiments. We also performed quantitative analyses to show that we have obtained HMGA2 binding sequences with high binding affinity (see below for detail). For example, if one of the AT-rich sequences is mutated to a non-AT-rich sequence, the DNA binding affinity of HMGA2 is reduced at least 100-fold. A single A to G point mutation at the 10th position of FL-SELEX1 (Tables 2 and 3) significantly lowered the DNA binding constant of HMGA2 (Wilson, W. D., and Leng, F., unpublished collaboration results). Addition of one or two AT base pairs to either end of the SELEX sequences does not significantly change the DNA binding affinity of HMGA2 (data not shown). These results suggest that we have likely obtained the "optimal" sequences for HMGA2 under our experimental

Table 2: Occurrence of Nucleotides (%) in the Consensus Sequences Obtained from the SELEX Experiments

position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	**96.2**	23.6	**95.3**	19.3	2.8	1.4	1.4	0	0.9	**42.9**	**49.1**	**36.8**	**83.5**	28.8	0.5
C	0	0	0	0	2.8	**54.7**	**44.3**	**61.3**	**40.1**	15.1	0	0	0	0	1.4
G	0	0	0	1.4	0	**42.9**	**52.8**	**35.9**	**59.0**	25.5	1.4	0	0	0	0
T	3.8	**76.4**	4.7	**79.2**	**94.3**	1.0	1.4	2.8	0	16.5	**49.5**	**63.2**	16.5	**71.2**	**98.1**
consensus sequence 1[b]	A	T	A	T	T	C	G	C	G	A	W	W	A	T	T
consensus sequence 2[b]	A	T	A	T	T	G	C	G	C	A	W	W	A	T	T

[a] Numbers in this table represent the percentage of the 71 unique clones from the SELEX experiments with the indicated bases at that position. Bold numbers correspond to the consensus nucleotides at each position. [b] Two consensus sequences are derived from the SELEX analysis, where W = A or T.

Table 3: DNA Binding Constants for HMGA2 Binding to Different DNA Oligomers

DNA oligomers	sequence (top strand)[a]	K_{app} (M^{-1}) [b]
FL-SELEX1	5′-ATATTCGCGATTATT-3′	$7.45 \pm 0.77 \times 10^6$
FL-304	5′-TATATATATATATAT-3′	$2.16 \pm 0.14 \times 10^6$
FL-AT15	5′-AAAAAAAAAAAAAAA-3′	$2.19 \pm 0.24 \times 10^6$
FL-308	5′-ATATTTATATTTATT-3′	$1.44 \pm 0.27 \times 10^6$
FL-306	5′-ATATTAAAAATTATT-3′	$1.46 \pm 0.40 \times 10^6$
FL-315	5′-ATATTCGCGACTGTC-3′	$\ll 1.0 \times 10^5$
FL-317	5′-GTGTCCGCGATTATT-3′	$\ll 1.0 \times 10^5$
FL-GC	5′-GGGGGGGGGGGGGGG-3′	0

[a] All DNA oligomers are 59 bp double-strand DNA oligonucleotides with the sequence 5′-CATGGTACCTTCAGAGGCTCGAGN_{15}GCTGACTGGCATGCAAGCTG-3′, where N_{15} represents the 15 bp DNA sequence listed in the table. Only top strands are shown. [b] The apparent DNA-binding constants of HMGA2 were measured as described under Materials and Methods. The values are the average of at least three independent determinations.

conditions. Our SELEX results are summarized in Figure 3C as a sequence logo.

Quantitative Analysis of HMGA2−DNA Interactions by EMSA. We previously demonstrated that HMGA2 binds with very high affinity to poly(dA-dT)$_2$ and poly(dA)poly(dT) under low-salt buffer conditions, and the DNA-binding constant of HMGA2 is strongly dependent on the salt concentration (*38*). At physiologically relevant salt conditions, i.e., 200 mM NaCl, it binds to AT-rich DNA with a binding constant from 10^6 to 10^7 M^{-1} (*38*). In the present study, we employed quantitative EMSA to measure the DNA binding constants of HMGA2 to different DNA oligomers in a buffer containing 200 mM NaCl. We also included 1.5 μM (bp) of poly(dG-dC)$_2$ (about 25-fold over 1 nM of ^{32}P-labeled oligonucleotides) as a competitor to reduce nonspecific binding. In this case, the DNA-binding constant should reflect specific recognition of different DNA sequences by HMGA2. Figure 4A shows a typical EMSA experiment in which HMGA2 was titrated into 1 nM of ^{32}P-labeled FL-SELEX1, a consensus sequence obtained from our SELEX experiments. The binding constants were obtained by fitting the binding data to eq 1 as described under Materials and Methods. Our binding curves and binding constants are summarized in Figure 4B and Table 3, respectively. As expected, HMGA2 binds to FL-SELEX1 with the highest affinity. Intriguingly, when the middle GC-rich sequence was mutated to an AT-rich sequence (either 5′-TATA-3′ or 5′-AAAA-3′), the DNA binding affinity was reduced 5-fold. These results indicate that the middle GC-rich sequence is required for high-affinity binding. Our results also demonstrated that one 5 or 6 bp AT-rich sequence is not sufficient

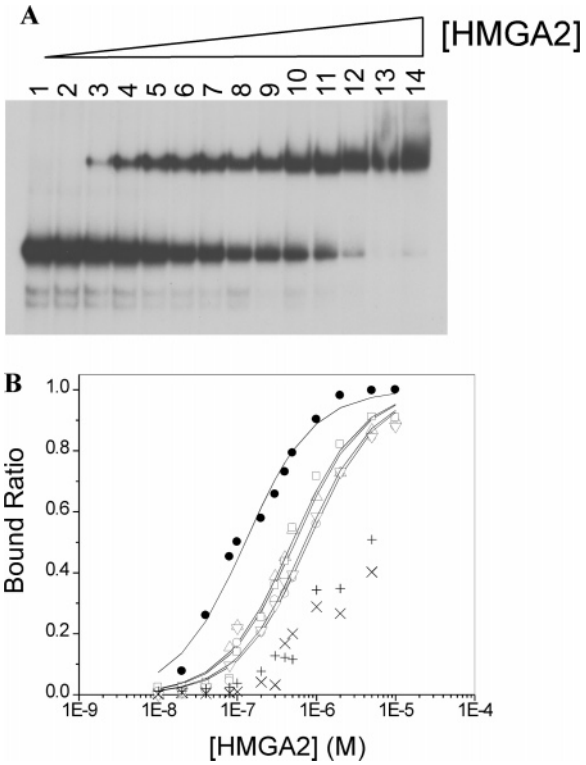

FIGURE 4: Quantitative EMSA experiments to determine the DNA binding constants of HMGA2 binding to different DNA oligomers. (A) Binding of HMGA2 to the DNA oligomer FL-SELEX1. ^{32}P-labeled FL-SELEX1 (1 nM) was incubated with increasing concentrations of HMGA2 in 50 μL of 1× EMSA binding buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 0.5 mM $MgCl_2$, 5% glycerol, and 1.5 μM (bp) poly(dG-dC)$_2$. EMSA experiments were performed as described under Materials and Methods. The autoradiogram of the ^{32}P-labeled FL-SELEX1 was shown. The radioactivities were quantified with a PhosphorImager. Lane 1 is the free FL-SELEX1. In addition to FL-SELEX1, lanes 2−14 also contain 10, 20, 40, 80, 100, 200, 300, 400, 500, 1000, 2000, 5000, and 10 000 nM of HMGA2, respectively. (B) Quantification analysis of the binding data from the EMSA experiments. The bound ratio of DNA was plotted against the protein concentration. The curves are generated by fitting the data to eq 1 as described under Materials and Methods. Symbols: ●, FL-SELEX1; △, FL 300; □, FL 304; ○, FL 306; ▽, FL 308; ×, FL 315; +, FL317.

for high-affinity binding of HMGA2 to DNA. Mutation of either AT-rich sequence of FL-SELEX1 to a non-AT-rich sequence significantly decreases the apparent DNA binding constant. These results suggest that all three segments in FL-SELEX1 are critical to high-affinity binding. Consistent with our previous results (*38*), HMGA2 also tightly binds to 15 bp A-track and alternate AT sequence, although the binding

affinity is 3-fold lower than that of FL-SELEX1; it does not bind to GC-rich sequences.

Biological Implication. The identification of the consensus sequences for HMGA2 is an important step toward characterization of its biological functions in vivo. As demonstrated previously, HMGA2 is a transcriptional factor involved in adipocytic proliferation and differentiation during embryogenesis (*1*, *19*, *43*). However, the mechanism by which HMGA2 regulates fat cell proliferation and differentiation is still unknown. Knowing the DNA binding specificity of HMGA2, it is possible to locate its binding sites within the genome and map its regulatory network in the cell. Furthermore, the discovery of the preferred binding sites of HMGA2 may provide essential information for designing new anticancer or antiobesity drugs. It has been shown that HMGA2 is an oncoprotein associated with many benign and malignant tumors (*44*); it is also involved in obesity (*2*). These studies suggest that HMGA2 is a potential target for treatment of cancers and obesity. We showed here that HMGA2 strongly and specifically binds to the minor groove of the two AT sites in the SELEX sequences, presumably through interactions with three of the AT-hook DNA binding motifs, suggesting that the minor groove binding agents are good inhibitors of HMGA2−DNA interactions. Indeed, we have recently demonstrated that netropsin, a well-known minor groove binder, competitively inhibited HMGA2 binding to the SELEX sequences (Wilson, W. D., and Leng, F., unpublished collaboration results). It is possible to design new minor groove binders to alleviate tumor development and progression through inhibition of the important HMGA2−DNA interactions.

CONCLUSION

Our results, presented in this paper, represent a systematic and unbiased analysis of HMGA2 DNA binding sites under physiologically relevant conditions [200 mM NaCl and 30 μM (bp) of poly(dG-dC)$_2$]. Using a SELEX procedure, we have demonstrated that HMGA2 preferentially binds to a type of 15 bp AT-rich DNA sequence: the first 5 bp are AT-rich, the middle 4 or 5 bp are GC-rich, and the last 5 or 6 bp are AT-rich. Alignment of 71 unique SELEX sequences revealed two consensus sequences for HMGA2: 5′-ATATTCGCGAWWATT-3′ or 5′-ATATTGCGCAWWATT-3′, where W = A or T. Our quantitative EMSA assays showed that the three segments in the consensus sequences (two AT-rich segments and one GC-rich segment) are required for high-affinity binding; mutations of these sequences significantly reduced the DNA-binding affinity of HMGA2. These results indicate that HMGA2 does not randomly recognize any AT-rich sequences. In contrast, it binds to specific AT-rich DNA sequences. We also showed that HMGA2 binds to AT-rich sites as a monomer under physiologically relevant salt conditions.

ACKNOWLEDGMENT

Some of this work was submitted to Florida International University in partial fulfillment of the requirements of the Ph.D. degree (T.C.). The authors would like to thank Drs. Bruce Dunlap (Florida International University) and W. David Wilson (Georgia State University) for reading the paper before submission.

SUPPORTING INFORMATION AVAILABLE

EMSA experiments of HMGA2 binding to FL-AT15 and FL-AT30, and HMGA2 binding sequences identified after 10 cycles of the SELEX experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Zhou, X., Benson, K. F., Ashar, H. R., and Chada, K. (1995) Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C, *Nature 376*, 771−774.
2. Anand, A., and Chada, K. (2000) In vivo modulation of Hmgic reduces obesity, *Nat. Genet. 24*, 377−380.
3. Wolffe, A. P. (1994) Architectural transcription factors, *Science 264*, 1100−1101.
4. Reeves, R. (2003) HMGA proteins: Flexibility finds a nuclear niche, *Biochem. Cell Biol. 81*, 185−195.
5. Goodwin, G. H., Cockerill, P. N., Kellam, S., and Wright, C. A. (1985) Fractionation by high-performance liquid chromatography of the low- molecular-mass high-mobility-group (HMG) chromosomal proteins present in proliferating rat cells and an investigation of the HMG proteins present in virus transformed cells, *Eur. J. Biochem. 149*, 47−51.
6. Giancotti, V., Berlingieri, M. T., DiFiore, P. P., Fusco, A., Vecchio, G., and Crane-Robinson, C. (1985) Changes in nuclear proteins on transformation of rat epithelial thyroid cells by a murine sarcoma retrovirus, *Cancer Res. 45*, 6051−6057.
7. Ashar, H. R., Fejzo, M. S., Tkachenko, A., Zhou, X., Fletcher, J. A., Weremowicz, S., Morton, C. C., and Chada, K. (1995) Disruption of the architectural factor HMGI-C: DNA-binding AT hook motifs fused in lipomas to distinct transcriptional regulatory domains, *Cell 82*, 57−65.
8. Schoenmakers, E. F., Wanschura, S., Mols, R., Bullerdiek, J., Van den, B. H., and Van de Ven, W. J. (1995) Recurrent rearrangements in the high mobility group protein gene, HMGI-C, in benign mesenchymal tumours, *Nat. Genet. 10*, 436−444.
9. Kools, P. F., Wanschura, S., Schoenmakers, E. F., Geurts, J. M., Mols, R., Kazmierczak, B., Bullerdiek, J., Van den, B. H., and Van de Ven, W. J. (1995) Identification of the chromosome 12 translocation breakpoint region of a pleomorphic salivary gland adenoma with t(1;12)(p22;q15) as the sole cytogenetic abnormality, *Cancer Genet. Cytogenet. 79*, 1−7.
10. Schoenmakers, E. F., Huysmans, C., and Van, d. V. (1999) Allelic knockout of novel splice variants of human recombination repair gene RAD51B in t(12;14) uterine leiomyomas, *Cancer Res. 59*, 19−23.
11. Klotzbucher, M., Wasserfall, A., and Fuhrmann, U. (1999) Misexpression of wild-type and truncated isoforms of the high-mobility group I proteins HMGI-C and HMGI(Y) in uterine leiomyomas, *Am. J. Pathol. 155*, 1535−1542.
12. Rommel, B., Rogalla, P., Jox, A., Kalle, C. V., Kazmierczak, B., Wolf, J., and Bullerdiek, J. (1997) HMGI-C, a member of the high mobility group family of proteins, is expressed in hematopoietic stem cells and in leukemic cells, *Leuk. Lymphoma 26*, 603−607.
13. Rogalla, P., Drechsler, K., Schroder-Babo, W., Eberhardt, K., and Bullerdiek, J. (1998) HMGIC expression patterns in non-small lung cancer and surrounding tissue, *Anticancer Res. 18*, 3327−3330.
14. Sarhadi, V., Wikman, H., Salmenkivi, K., Kuosma, E., Sioris, T., Salo, J., Karjalainen, A., Knuutila, S., and Anttila, S. (2006) Increased expression of high mobility group A proteins in lung cancer, *J. Pathol*.
15. Meyer, B., Loeschke, S., Schultze, A., Weigel, T., Sandkamp, M., Goldmann, T., Vollmer, E., and Bullerdiek, J. (2007) HMGA2 overexpression in non-small cell lung cancer, *Mol. Carcinog*.
16. Kottickal, L. V., Sarada, B., Ashar, H., Chada, K., and Nagarajan, L. (1998) Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal in human leukemia, *Biochem. Biophys. Res. Commun. 242*, 452−456.
17. Tallini, G., and Dal Cin, P. (1999) HMGI(Y) and HMGI-C dysregulation: A common occurrence in human tumors, *Adv. Anat. Pathol. 6*, 237−246.
18. Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B., and Goodwin, G. H. (1991) Comparison of multiple forms of the high mobility group I proteins in rodent and human

cells. Identification of the human high mobility group I-C protein, *Eur. J. Biochem. 198*, 211−216.
19. Gattas, G. J., Quade, B. J., Nowak, R. A., and Morton, C. C. (1999) HMGIC expression in human adult and fetal tissues and in uterine leiomyomata, *Genes Chromosomes. Cancer 25*, 316−322.
20. Friedmann, M., Holth, L. T., Zoghbi, H. Y., and Reeves, R. (1993) Organization, inducible-expression and chromosome localization of the human HMG-I(Y) nonhistone protein gene, *Nucleic Acids Res. 21*, 4259−4267.
21. Chau, K. Y., Patel, U. A., Lee, K. L., Lam, H. Y., and Crane-Robinson, C. (1995) The gene for the human architectural transcription factor HMGI-C consists of five exons each coding for a distinct functional element, *Nucleic Acids Res. 23*, 4262−4266.
22. Lehn, D. A., Elton, T. S., Johnson, K. R., and Reeves, R. (1988) A conformational study of the sequence specific binding of HMG-I (Y) with the bovine interleukin-2 cDNA, *Biochem. Int. 16*, 963−971.
23. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans, J. N., Reeves, R., Gronenborn, A. M., and Clore, G. M. (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif, *Nat. Struct. Biol. 4*, 657−665.
24. Goodwin, G. (1998) The high mobility group protein, HMGI-C, *Int. J. Biochem. Cell Biol. 30*, 761−766.
25. Solomon, M. J., Strauss, F., and Varshavsky, A. (1986) A mammalian high mobility group protein recognizes any stretch of six A.T base pairs in duplex DNA, *Proc. Natl. Acad. Sci. U.S.A. 83*, 1276−1280.
26. Maher, J. F., and Nathans, D. (1996) Multivalent DNA-binding properties of the HMG-1 proteins, *Proc. Natl. Acad. Sci. U.S.A. 93*, 6716−6720.
27. Thanos, D., and Maniatis, T. (1992) The high mobility group protein HMG I(Y) is required for NF-kappa B-dependent virus induction of the human IFN-beta gene, *Cell 71*, 777−789.
28. Du, W., Thanos, D., and Maniatis, T. (1993) Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements, *Cell 74*, 887−898.
29. Baldassarre, G., Fedele, M., Battista, S., Vecchione, A., Klein-Szanto, A. J., Santoro, M., Waldmann, T. A., Azimi, N., Croce, C. M., and Fusco, A. (2001) Onset of natural killer cell lymphomas in transgenic mice carrying a truncated HMGI-C gene by the chronic stimulation of the IL-2 and IL-15 pathway, *Proc. Natl. Acad. Sci. U.S.A. 98*, 7970−7975.
30. John, S., Reeves, R. B., Lin, J. X., Child, R., Leiden, J. M., Thompson, C. B., and Leonard, W. J. (1995) Regulation of cell-type-specific interleukin-2 receptor alpha-chain gene expression: Potential role of physical interactions between Elf-1, HMG-I(Y), and NF-kappa B family proteins, *Mol. Cell Biol. 15*, 1786−1796.
31. Reeves, R., and Beckerbauer, L. (2001) HMGI/Y proteins: Flexible regulators of transcription and chromatin structure, *Biochim. Biophys. Acta 1519*, 13−29.
32. Cui, T., Joynt, S., Morillo, V., Baez, M., Hua, Z., Wang, X., and Leng, F. (2007) Large scale preparation of the mammalian high mobility group protein A2 for biophysical studies, *Protein Pept. Lett. 14*, 87−91.
33. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res. 25*, 4876−4882.
34. Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) MEME: Discovering and analyzing DNA and protein sequence motifs, *Nucleic Acids Res. 34*, W369−W373.
35. Carey, J. (1988) Gel retardation at low pH resolves trp repressor-DNA complexes for quantitative study, *Proc. Natl. Acad. Sci. U.S.A. 85*, 975−979.
36. Yie, J., Merika, M., Munshi, N., Chen, G., and Thanos, D. (1999) The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome, *EMBO J. 18*, 3074−3089.
37. Padmanabhan, S., Elias-Arnanz, M., Carpio, E., Aparicio, P., and Murillo, F. J. (2001) Domain architecture of a high mobility group A-type bacterial transcriptional factor, *J. Biol. Chem. 276*, 41566−41575.
38. Cui, T., Wei, S., Brew, K., and Leng, F. (2005) Energetics of binding the mammalian high mobility group protein HMGA2 to poly(dA-dT)2 and poly(dA)-poly(dT), *J. Mol. Biol. 352*, 629−645.
39. Robertson, D. L., and Joyce, G. F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA, *Nature 344*, 467−468.
40. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science 249*, 505−510.
41. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature 346*, 818−822.
42. Cui, Y., Wang, Q., Stormo, G. D., and Calvo, J. M. (1995) A consensus sequence for binding of Lrp to DNA, *J. Bacteriol. 177*, 4872−4880.
43. Li, O., Vasudevan, D., Davey, C. A., and Droge, P. (2006) High-level expression of DNA architectural factor HMGA2 and its association with nucleosomes in human embryonic stem cells, *Genesis 44*, 523−529.
44. Young, A. R., and Narita, M. (2007) Oncogenic HMGA2: Short or small, *Genes Dev. 21*, 1005−1009.

BI701269S