The HMG-I(Y) A-T-hook Peptide Motif Confers DNA-binding Specificity to a Structured Chimeric Protein*

(Received for publication, January 27, 1999, and in revised form, March 22, 1999)

Geoffrey C. Banks‡, Brandt Mohrs§, and Raymond Reeves‡¶

From the Departments of ‡Biochemistry/Biophysics and §Genetics and Cell Biology, Washington State University, Pullman, Washington 99164-4660

Chromosomal translocations involving genes coding for members of the HMG-I(Y) family of “high mobility group” non-histone chromatin proteins (HMG-I, HMG-Y, and HMG-IC) have been observed in numerous types of human tumors. Many of these gene rearrangements result in the creation of chimeric proteins in which the DNA-binding domains of the HMG-I(Y) proteins, the so-called A-T-hook motifs, have been fused to heterologous peptide sequences. Although little is known about either the structure or biophysical properties of these naturally occurring fusion proteins, the suggestion has been made that such chimeras have probably assumed an altered in vivo DNA-binding specificity due to the presence of the A-T-hook motifs. To investigate this possibility, we performed in vitro “domain-swap” experiments using a model protein fusion system in which a single A-T-hook peptide was exchanged for a corresponding length peptide in the well characterized “B-box” DNA-binding domain of the HMG-I non-histone chromatin protein. Here we report that chimeric A-T-hook/B-box hybrids exhibit in vitro DNA-binding characteristics resembling those of wild type HMG-I(Y) protein, rather than the HMG-1 protein. These results strongly suggest that the chimeric fusion proteins produced in human tumors as a result of HMG-I(Y) gene chromosomal translocations also retain A-T-hook-imparted DNA-binding properties in vivo.

The mammalian HMG-I(Y) family of “high mobility group” (HMG) chromatin proteins are founding members of a new assemblage of nuclear proteins referred to as “architectural transcription factors,” whose role in gene regulation is the recognition and modulation of both DNA and chromatin structure (recently reviewed in Ref. 1). The human HMG-I(Y) protein family consists of: HMG-I (11.7 kDa), HMG-Y (10.6 kDa), and HMG-I(C) (12 kDa). HMG-I and HMG-Y are isoform proteins (identical in sequence except for an internal 11-amino acid deletion in the latter) derived from an alternatively spliced mRNA transcript coded for by the HMG-I(Y) gene at chromosomal locus 6p21 (1–5). The related HMGI-C protein is coded for by a separate gene located at chromosomal locus 12q14–15 (6). Members of the HMG-I(Y) family share many biochemical and biophysical properties, including possessing three independent “A-T-hook” DNA-binding motifs, which exhibit a marked preference toward A-T-rich B-form DNA (2, 3, 7–9). Therefore, for convenience, we will collectively refer to members of this family as simply HMG-I(Y), unless otherwise noted. Various lines of evidence indicate that the individual A-T-hook motifs of the HMG-I(Y) proteins recognize the structure associated with the narrow minor groove of A-T-rich target DNA rather than its nucleotide sequence (1, 8, 10, 11). The physical basis for this DNA-structural recognition became clear with NMR structural studies, which indicated that in solution the HMG-I(Y) protein is unstructured but that upon substrate association the A-T-hook undergoes a disorder-to-order structural transition necessary for specific A-T-DNA binding (9, 12). During this transition, the arginine side chains of the palindromic “core” sequence, Pro-Arg-Gly-Arg-Pro (PRGRP; Fig. 1) assume a specific planar, crescent-shaped structure that confers selectivity for the narrow minor groove of A-T-rich sequences. This core peptide motif is the most highly conserved peptide sequence found in the HMG-I(Y) family of proteins. Correspondingly, the A-T-hook motif is evolutionarily conserved as a DNA-binding unit in many other proteins and transcription factors found in organisms ranging from bacteria (13) to humans (1).

Chromosomal translocations of the genes coding for the HMG-I(Y) proteins are among the most common rearrangements observed in human neoplasms (reviewed in Ref. 14). In the majority of cases, these chromosomal translocations result in the creation of chimeric genes coding for hybrid proteins in which the A-T-hook motifs derived from one of the HMG-I(Y) gene loci are fused “in frame” to the N-terminal end of ectopic peptide sequences derived from a number of other gene loci. Such rearrangements are quite prevalent in benign mesenchymal human tumors including lipomas, leiomyomas, fibroadenoma, pleomorphic adenomas, aggressive angiomyxomas, and pulmonary hamartomas (14). Table I lists some examples of aberrant chimeric proteins containing HMG-I(Y)-like A-T-hook motifs. However, little information is available about these fusion protein’s role, if any, in the process of tumorigenesis. Nevertheless, it has been suggested that one possible in vivo role for these ectopic fusion sequences in tumor cells is to interfere with the normal gene regulatory function of the HMG-I(Y) protein (15). Another suggestion is that a likely consequence of the attachment of the A-T-hook motifs to other proteins or peptides is to change the DNA-binding characteristics of the hybrids so that they now “mimic” a new substrate binding sites within the nucleus (16–18). In either case, the intrinsic assumption is that the A-T-hook peptides present in…

* This work was supported by National Institutes of Health Grant GM-46352 (to R.R.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
¶ To whom correspondence should be addressed. Tel.: 509-335-1948; Fax: 509-335-9668; E-mail: reeves@wsu.edu.
† The abbreviations used are: HMG, high mobility group nonhistone chromatin proteins; aa, amino acid(s); BLT, DNA from the 3’-untranslated tail region of the bovine interleukin-2 cDNA; CD, circular dichroism; CP, nucleosome core particle; EMSA, electrophoretic mobility shift analysis; GST, glutathione S-transferase; A-T-hook, the highly conserved DNA-binding domain of the HMG-I(Y) family of nonhistone proteins; IF, intrinsic fluorescence; NMR, nuclear magnetic resonance; PCR, polymerase chain reaction; WT, wild type; DTT, dithiothreitol; bp, base pair(s).
chimeric proteins retain their normal A-T-DNA substrate binding specificity. In lieu of direct experimental support in favor of such an assumption, it is just as reasonable, however, to suspect that the substrate binding properties of the A-T-hook motif are significantly altered when fused to a large, structured ectopic protein partner.

Unfortunately, it is difficult to assess whether or not the A-T-hook DNA-binding motifs found in naturally occurring chimeric tumor proteins are functional since little is known about either the biological function or structure of these fusion hybrids. To directly test the functionality of the A-T-hook motif in the context of a structured fusion partner, we performed in vitro domain swap experiments (see Fig. 1) in which the second, and strongest DNA-binding, A-T-hook motif of the HMG-I(Y) protein (9) was inserted in place of a corresponding length peptide at the N terminus of the HMG-1 B-box DNA binding domain (1). The HMG-1 B-box peptide was chosen as a "model" fusion partner for our domain swap experiments with HMG-I(Y) because NMR studies have demonstrated that the native wild type B-box peptide is highly structured in solution (Fig. 1) (19, 20), and because an extensive literature also exists on the possible biological function(s) of proteins that contain one or more HMG-1 box DNA-binding motifs.

Here we report on the biophysical and substrate binding properties of recombinant chimeric HMG-I(Y) A-T-hook/HMG-1 B-box proteins. Proper folding of both the hybrid B-box (hereafter referred to as the "correctly folded hybrid") and the wild type (WT) B-box domains were verified using circular dichroism and intrinsic fluorescence measurements. As a control, a non-folded hybrid B-box protein, also containing a single A-T-hook, was likewise generated and analyzed in comparison to both the correctly folded hybrid and the WT B-box proteins using electrophoretic mobility shift assays. These experiments demonstrate that both the correctly folded and the non-folded hybrid B-box proteins have acquired DNA-binding specificities more closely resembling those of the HMG-I(Y) proteins than those of either the full-length HMG-1 protein or those of the isolated WT HMG-1 B-box domain. The ability to alter the substrate binding preferences of the WT HMG-1 B-box protein by motif exchange with an A-T-hook DNA-binding peptide suggests that the chimeric fusion proteins observed in vivo in human tumors also retain HMG-I(Y)-like substrate binding specificities. In addition, the results presented provide new insight into the nature of interactions of both the HMG-1 B-box and HMG-I(Y) A-T-hook with DNA.

MATERIALS AND METHODS

5' → 3' Sequence of Polymerase Chain Reaction (PCR) Primers used for Expression Vector Construction—Sequences are as follows: primer 1, ATGATCAGACAGCTGAATG; primer 2, TGAATTCCTTGGCACGATGATGCT; primer 3, CGTTGGGCTCTCGCTGCTTCT; primer 4, TGGATCCCCAACACCTAAG; primer 5, GAAGAATGCCGAAGGTCGAC- TGGGATCCCCAACACCTAAG; primer 5, GAAGAATGCCGAAGGTCGAC-TGGGATCCCCAACACCTAAG; primer 6, CGACCCAAGAAGTTTTTCCTATTCTGC; primer 7, TTCCCGGTGGATCTGCTAG; primer 8, AAAATACCTATA-TGGTTCAACACT; primer 9, GAATAGGAAAAACCTTGGTCGACC; primer 10, TGGATATCCATGTTCAAGATGCTC. Construction of Glutathione S-Transferase (GST) Fusion Protein Expression Vectors for Producing Wild Type and Correctly Folded Hybrid HMG-1 B-box Recombinant Proteins—The isolated WT HMG-1 B-box domain, corresponding to residues 89–164 of the 214-amino acid full-length HMG-1 protein, was constructed by PCR amplification of a human HMG-1 cDNA template with primers 8 and 9. An N-terminal extended segment (residues 89–98) with primers 3 and 2, hereafter referred to as the "B-box w/o N-term" fragment; 2) PCR primers 4 and 5 were used to amplify the region of the human HMG-I(Y) cDNA (5) coding for the "second" A-T-hook DNA-binding motif (i.e. TFPKPRGPRK) corresponding to aa residues 62–61 of full-length HMG-I; 3) the correctly folded hybrid B-box construct was then created by annealing the isolated PCR DNA fragments coding for the A-T-hook motif and the B-box w/o N-term and extending the annealed products by PCR using primers 4 and 2. The 5' region of the B-box w/o N-term corresponding to primer 3 contained sequence complimentary to the 3' portion of A-T-hook motif, and the 3' portion of primer 2 contained sequence complimentary to the 5' portion of the B-box w/o N-term construct. Specific nucleotide base mutations in the third base wobble position of the original HMG-I(7C) and HMG-1 cDNA clones that do not alter amino acid coding sequence were engineered into the PCR primers to counteract primer-dimer formation during the amplification reaction. The amplified PCR products were purified and ligated into a pGEX-2T plasmid (Pharmacia Biotech, Uppsala, Sweden) at the BamHI and EcoRI restriction sites. The proper sequence of the resulting constructs was confirmed by dideoxynucleotide sequencing prior to transformation into Esherichia coli BL21(DE3)/plysS for isopropyl-1-thio-β-D-galactopyranoside-induced production of recombinant GST fusion proteins.

Construction of Hexahistidine-tagged (HIS-tagged) WT and Non-folded Hybrid HMG-1 B-box Expression Vector—The non-folded hybrid B-box was constructed in the same manner as described above for the correctly folded hybrid B-box but with slight modifications. The DNA coding for an HMG-1 B-box lacking a 13-amino acid N-terminal extended segment was constructed in a series of steps. 1) The non-folded B-box w/o N-term (aa residues 88–100 of HMG-1) was constructed by PCR amplification using primer 9. 2) The vector was transformed with primers 6 and 7; 2) DNA coding for 13 amino acid residues of the second A-T-hook motif of the HMG-I(Y) protein (i.e. VPTPKPRGPRK-K, corresponding to aa residues 51–62) was constructed by PCR amplification of human HMG-I(Y) cDNA template with primers 8 and 9. It should be noted that during this construction the last lysine residue of this 13-amino peptide was changed from the original glycine residue present in the native HMG-1 protein so that it now corresponds to the "consensus" A-T-hook motif of the HMG-I(Y) family of proteins (8); 3) the non-folded hybrid B-box was then constructed by annealing and extending the 13-amino acid consensus A-T-hook motif with the non-folded B-box w/o N-term and primers 8 and 7. As before, the 5' region of primer 6 is complimentary to the 3' region of the consensus A-T-hook motif, and the 3' region of the primer 9 is complimentary to the 5' region of the non-folded B-box w/o N-term construct. Specific nucleotide base mutations in the third base wobble position of the original HMG-I(7C) cDNA clone that do not alter amino acid coding sequence were engineered into the PCR primers to counteract primer-dimer formation. The HIS-tagged WT HMG-1 B-box (corresponding to residues 88–164 of the 214-amino acid full-length WT HMG-1 protein; corresponding to residues 89–164 of the 214-amino acid full-length WT HMG-1 protein) was constructed by PCR amplification of human HMG-I(Y) cDNA template with primers 10 and 7. In each case, dideoxynucleotide sequencing confirmed that DNA of the correct nucleotide sequence had been produced. The amplified PCR products were purified and ligated into the bacterial plasmid expression vector pET-24b (Novagen, Madison, WI) between the Ndel and XhoI restriction sites and transformed into E. coli BL21(DE3)/plysS for production of recombinant proteins as subsequently described below.

Expression and Purification of Recombinant Proteins—GST fusion WT and correctly folded hybrid B-box proteins were expressed and purified as described by Read et al. (22), with slight modifications. Prior to induction with 0.4 mM isopropyl-1-thio-β-D-galactopyranoside, the LB culture was heat-shocked at 42 °C for 5 min to induce chaperone proteins. After heat shock, the cultures were grown for an additional 3 h before processing. The cells were sonicated, and Nonidet P-40 was added to a final concentration of 1%. After centrifugation to pellet cellular debris, the GST fusion protein containing supernatant was made 2 mM with respect to CaCl2 and MgCl2, and the GST fusion proteins attached to the beads were washed extensively with 300 mM NaCl, 50 mM Tris/HCl, pH 8.0, 2 mM DTT solution (22). The washed solution was made 2 mM with respect to CaCl2 and MgCl2, and the GST fusion proteins attached to the beads were cleaved with thrombin overnight at 4 °C. The beads were centrifuged, the supernatant containing the free recombinant protein cleaved of its GST moiety was removed, and fresh wash solution, CaCl2, MgCl2, and thrombin were incubated again with the beads overnight at 4 °C. This procedure was repeated twice with SDS-polyacrylamide gel electrophoresis, or, no, GST fusion proteins remaining attached to the beads. The released recombinant proteins were further purified away from the thrombin by reverse-phase high pressure liquid chromatography. The purified recombinant proteins were lyophilized and reconstituted in 10 mM Tris/HCl, pH 7.8, 1 mM DTT.

HIS-tagged WT and non-folded hybrid B-box recombinant proteins

Role of the A-T-hook Motif in Hybrid Proteins

16537
Role of the A'T-hook Motif in Hybrid Proteins

Fig. 1. Peptide domain-swap strategy for exchanging the extended N-terminal region of the HMG-1 B-box with the second A'T-hook motif from HMG-I(Y). The sequences of the two 10-aa peptides involved in the exchange are shown. The horizontal cross-hatched bar between the sequences indicates the area of the palindromic "core" residues of the A'T-hook peptide with a glycine (G) at its center. The italicized letters denote the identical spacing of the trans configuration proline residues in the two peptides. Only 5 of the 10 aa from the second (II) DNA-binding motif of HMG-I(Y), that correspond to the "core" region (PBGRP), are modeled with and without DNA to illustrate the planar, crescent-shaped structure formed upon substrate interaction (modeling done using Rasmol; Ref. 40). The three-dimensional L-shaped structure of the hydrophobic hinge region of the correctly folded molecule. In addition, key hydrophobic residues in the C-terminal third helix are modeled to illustrate the existence of a hydrophobic interface formed with the extended N terminus. The model of the B-box was generated using MOLSCRIPT (41).

The gels were dried and analyzed using PhosphorImager analysis tools (Molecular Dynamics, Sunnyvale, CA).

DNase I Footprinting on BLT DNA—DNase I footprinting followed standard protocols (28, 29). 5 μM amounts of either the correctly folded hybrid or the WT B-box proteins were bound to BLT DNA (0.1 ng) in the footprinting reactions. Both the concentration and time of digestion of DNase I were empirically determined to obtain optimal results. Products of Maxam-Gilbert chemical cleavage reactions (30) of BLT DNA served as reference standards in the sequencing gels used in the footprinting experiments.

Circular Dichroism and Intrinsic Fluorescence Determinations of Protein Structure—In all circular dichroism analyses, the absorption spectrum of recombinant protein preparations was measured at least three times, with the average being used to calculate the molar ellipticity. The background spectrum, consisting of buffer only, was subtracted from each absorption spectrum. The correctly folded hybrid B-box was measured under reducing conditions in 20 mM KH2PO4, pH 6.0, and 1 mM Tris-(2-carboxyethyl)-phosphine hydrochloride (Molecular Probes; Eugene, OR) while the non-folded hybrid and both WT B-box proteins were measured under non-reducing conditions in 20 mM KH2PO4, pH 6.0. Data were collected with nitrogen purging on an AVIV 62DS CD instrument (AVIV Inc., Lakewood, NJ). Intrinsic fluorescence measurements of the B-box proteins were done on a Shimadzu 5300 spectrofluorometer (Shimadzu Corp., Kyoto, Japan). Equal concentrations of proteins, solubilized in 10 mM Tris/HCl, pH 7.8, 1 mM DTT, were used for the analysis. The excitation wavelength was 280 nm, and the fluorescence emission was collected from 290 to 400 nm.

RESULTS

Rationale for Hybrid B-box Protein Construction and Analysis—For the reasons outlined above, the HMG-1 B-box peptide was selected as the "highly structured" fusion partner for incorporation into a "model" hybrid protein in which the N-terminal segment of the B-box has been replaced with the A'T-hook DNA-binding peptide motif (Fig. 1). As illustrated in Fig. 1, the approximately 75 amino acid residues of the WT HMG-1 B-box exist in solution as a twisted L-shaped structure consisting of three stabilizing α-helices and an "extended" N...
Role of the AT-hook Motif in Hybrid Proteins

Expression and Analysis of Recombinant Wild Type and Hybrid B-box Proteins—Hybrid and WT B-box proteins were expressed both as GST fusion and HIS-tagged recombinant proteins. The correctly folded hybrid and WT B-box proteins were expressed as soluble recombinant GST fusion proteins, which were subsequently processed by cleavage with thrombin to remove the GST moiety. To promote the formation of inclusion bodies, prevent bacterial proteolytic degradation, and enhance the ease of rapid purification, the non-folded hybrid and WT B-box proteins were expressed with a HIS tag on their C termini.

As shown in Fig. 2, the GST-expressed WT and correctly folded hybrid B-box proteins have slight differences in amino acid composition at their N- and C-terminal ends relative to the HIS-tagged WT and non-folded hybrid B-box proteins. After thrombin cleavage and subsequent purification, both the GST-derived WT and correctly folded hybrid B-box proteins are 84 amino acids in length (Fig. 2A, rows b and c). In both of these isolated recombinant proteins, eight amino acids (i.e., GS on the N terminus and EFIVTD on the C terminus) are derived from the pGEX-2T plasmid expression vector. The HIS-tagged proteins are similar in length at 83 amino acids but have six histidine residues attached on their C termini (Fig. 2A, rows a and d). Regardless of these minor differences, the various recombinant WT B-box proteins had similar biophysical properties and also behaved identically in all experimental assays (data not shown), indicating that their slight differences in amino acid composition did not have any observable influence on their substrate-binding properties under our experimental conditions. It can therefore be argued that any biophysical differences that are observed between the HIS-tagged, non-folded B-box hybrid and the correctly folded recombinant B-box hybrid are due to variations in the state of folding of their respective peptide domains and are unrelated to their mode of preparation.

As an experimental control for the correctly folded B-box proteins, site-specific mutations were introduced into the hybrid B-box that interfered with the protein’s ability to correctly fold into its normal tertiary configuration. The engineered differences in this non-folded hybrid B-box, as compared with the correctly folded hybrid, are two lysine residues that have been introduced at the C-terminal end of the AT-hook motif (aa 12 and 13) that replace both a serine and a highly conserved alanine residue (32) (Fig. 2B; compare rows c and d in Fig. 2A). The positioning of these two positively charged lysine residues, in close proximity to the hydrophobic hinge region of the B-box (Fig. 1), disrupts the hydrophobic packing interactions and thereby prevents proper tertiary folding of the mutant protein. This predicted structural disruption was confirmed by subsequent structural analyses (see below).

Circular Dichroism and Intrinsic Fluorescence Structural Analysis of Recombinant Hybrid and WT B-box proteins—Circular dichroism (CD) gives information concerning the relative amount of secondary structure in a molecule and was used to verify that both the correctly folded hybrid and the WT B-box proteins contained their normal a-helical secondary structural elements. In addition, CD analysis verified the predicted structural disruption of the non-folded hybrid B-box. As shown in Fig. 3 (A and B), the spectra at 25 °C for both the WT B-box and correctly folded hybrid B-box proteins derived from GST fusion products resemble those previously reported for correctly folded HMG box peptides (22, 33). The CD data were deconvoluted (34) to yield approximate percent values of α-helix, β-turn, and random coil for the WT and correctly folded hybrid proteins that, within experimental error, are quite similar to each other (see Fig. 3, A and B). Fig. 3 (A and B) also demonstrates that upon heating to 80 °C, the CD spectra for both proteins shift toward a shape characteristic of an irregular or random coil. The HIS-tagged WT B-box protein was also examined using CD at both 25 °C and 80 °C and is nearly identical to the hybrid and WT proteins derived from GST fusion products (data not shown). Together, these data sets indicate that all three of these recombinant proteins have refolded properly at 25 °C and contain α-helical folds characteristic of a WT B-box domain (19, 20, 22). The slight differences in the shape of the two CD spectra for the correctly folded hybrid and the WT B-box proteins at 25 and 80 °C (Fig. 3, A and B) are attributable to slight differences in buffer conditions (see Materials and Methods). The intrinsic fluorescence (IF) measurements shown in Fig. 3 (C and D) likewise confirm that the tertiary configurations of the correctly folded hybrid and the
two WT B-box proteins are very similar, if not identical. These data indicate that the tryptophan residues in the hydrophobic hinge regions of these three B-box peptides are equally protected from solvent accessibility. Importantly, the IF spectra for all three of the proteins are also almost superimposable, confirming the structural similarities between the correctly folded hybrid and the WT B-box proteins. Panel D, the red shift and decreased quantum yield of the IF spectrum of the non-folded hybrid (HIS-tagged), resulting from solvent accessibility to the hydrophobic pocket of the domain, relative to the WT B-box protein (also HIS-tagged), confirms a lack of proper folding in this chimeric protein.

In contrast to the situation for the properly folded recombinant proteins noted above, the CD spectrum at 25 °C for the non-folding hybrid B-box protein (Fig. 3B) is very similar to the random coil profile of proteins denatured at 80 °C (Fig. 3A and B). Consistent with this interpretation, the IF spectrum for the non-folded hybrid B-box (Fig. 3D) is quite different from that of properly folded B-box proteins. The non-folded hybrid B-box exhibits a red shift of nearly 20 nm for the wavelength showing maximum fluorescence (350 nm versus 330 nm) and a significant decrease in the overall quantum yield of fluorescence relative to the three other B-box proteins (Fig. 3, C and D). In combination with the CD data, the IF measurements indicate that the non-folded hybrid B-box is not forming α-helical secondary structural elements to the same degree as the correctly folded hybrid or WT B-box proteins. These data also demonstrate that the non-folded hybrid B-box exhibits physical characteristics more closely resembling those of denatured, rather than native, proteins.

Hybrid B-box Proteins Preferentially Bind to A-T-rich BLT DNA—Fig. 4 shows a comparison of various recombinant proteins binding to a well characterized A-T-rich duplex DNA substrate, BLT, the 3′-untranslated tail region of the bovine interleukin-2 cDNA, which contains multiple A-T-rich binding sites for the HMG-I(Y) protein (7, 11). Panel A, full-length wild-type HMG-I(Y) protein. Panel B, full-length wild-type HMG-1 protein. Panel C, correctly folded hybrid B-box protein. Panel D, wild-type B-box protein. Panel E, non-folded hybrid B-box protein. Note that the specific BLT-binding characteristics of the HMG-I(Y) protein (A) are mimicked in the both the correctly folded (C) and non-folded (E) hybrid B-box proteins. In contrast, both the full-length HMG-1 protein (B) and WT B-box protein (D) bind only nonspecifically as indicated by smearing.

![Image](https://example.com/image1.png)  
**Fig. 3.** CD and IF measurements of B-box proteins. Panel A, the structural elements of the WT B-box protein (GST-expressed), as monitored by CD, were measured at 25 °C and 80 °C. Note the loss of α-helical structure when the protein is denatured at 80 °C. Panel B, the correctly folded hybrid B-box protein exhibits a similar CD pattern to that of the WT B-box protein at both 25 and 80 °C. However, at 25 °C, the non-folded hybrid B-box protein has a CD spectrum similar to that seen for the denatured correctly folded hybrid at 80 °C. Panel C, IF measurements (excitation = 280 nm; emission = 290–400 nm) confirmed the structural similarities between the correctly folded hybrid and the WT B-box proteins. Panel D, the red shift and decreased quantum yield of the IF spectrum of the non-folded hybrid (HIS-tagged), resulting from solvent accessibility to the hydrophobic pocket of the domain, relative to the WT B-box protein (also HIS-tagged), confirms a lack of proper folding in this chimeric protein.

![Image](https://example.com/image2.png)  
**Fig. 4.** Both HMG-I(Y) and hybrid B-box proteins specifically bind to A-T-rich B-form DNA. EMSAs were performed using various recombinant proteins and the A-T-rich BLT DNA as a substrate. BLT DNA is the 300-bp 3′-untranslated tail region of the bovine interleukin-2 cDNA, which contains multiple A-T-rich binding sites for the HMG-I(Y) protein (7, 11). Panel A, full-length wild-type HMG-I(Y) protein. Panel B, full-length wild-type HMG-1 protein. Panel C, correctly folded hybrid B-box protein. Panel D, wild-type B-box protein. Panel E, non-folded hybrid B-box protein. Note that the specific BLT-binding characteristics of the HMG-I(Y) protein (A) are mimicked in the both the correctly folded (C) and non-folded (E) hybrid B-box proteins. In contrast, both the full-length HMG-1 protein (B) and WT B-box protein (D) bind only nonspecifically as indicated by smearing.
from the nonspecific complexes formed by either the full-length HMG-1 or the WT B-box proteins. It should be noted, however, that there are several orders of magnitude difference in the binding affinity of HMG-I(Y) and the correctly folded hybrid B-box protein for BLT DNA. This difference in binding affinity is at least partially explained by the differing number of DNA-binding motifs present in the two proteins. While the full-length HMG-I(Y) protein, which has three independent A-T-hook motifs that interact cooperatively with A-T-rich substrates (8, 9), binds BLT DNA in the nanomolar range (Fig. 4A), the correctly folded hybrid B-box protein, which has only a single A-T-hook motif and consequently has a lower affinity for this substrate (8), binds in the micromolar range. In addition, constraint of the normally flexible A-T-hook motif in the confines of the rigid B-box configuration likely contributes to the reduced binding affinity of the correctly folded hybrid (9). Consistent with this interpretation, the non-folded hybrid B-box, as shown in Fig. 4E, also forms discrete complexes with BLT DNA at a higher affinity (in the submicromolar range) than the correctly folded hybrid B-box. This is reasonable considering that the A-T-hook motif in the non-folded hybrid B-box is not confined by a rigid structural scaffolding and is therefore quite flexible, similar to the situation in the tightly binding HMG-I(Y) protein itself (9). Nevertheless, both hybrid proteins are able to form specific protein-DNA complexes, suggesting that the ability to bind specifically to A-T-rich sequences has been conferred to the fusion proteins by the A-T-hook DNA binding motif. Importantly, these data demonstrate that the functional A-T-hook peptide structure is not seriously compromised by the presence of the rigid structural elements of the HMG-1 B-box in the correctly folded hybrid protein.

A series of DNA competition experiments were performed to verify, by an independent technique, that the correctly folded hybrid B-box protein is recognizing the narrow minor groove structure associated with stretches of A-T-rich DNA. Fig. 5 shows the results of experiments in which the correctly folded hybrid B-box protein was incubated with radiolabeled BLT DNA in the presence of increasing concentrations of one of two different non-labeled DNA competitors: dG-dC, a nonspecific competitor DNA, or dI-dC, a specific competitor for A-T-rich DNA sequences. As seen in Fig. 5 (A and B), when the correctly folded hybrid B-box is bound to labeled BLT DNA without any added non-labeled competitor there is smearing, characteristic of nonspecific protein-DNA interactions. However, as shown in Fig. 5A, in the presence of low concentrations of dG-dC (e.g. 0.1 µg reaction), these nonspecific protein-DNA interactions are abated, and it is possible to resolve both a specific protein-DNA complex and the unbound free DNA on the gel. Furthermore, as expected with any nonspecific competitor, at considerably higher concentrations of dG-dC (e.g. 0.5 µg reaction) the specific hybrid B-box-DNA complex is also eliminated (Fig. 5A). On the other hand, in titration reactions containing increasing concentrations of unlabeled dI-dC that compete with A-T-DNA for binding (Fig. 5B), no specific protein-DNA complexes are observed under any conditions. Important for the interpretation of these competitions is the fact that dI-dC has been shown to effectively compete with the HMG-I(Y) protein for binding to the A-T-rich sequences present in BLT DNA and other substrates (8, 10). These results suggest that the correctly folded hybrid B-box protein recognizes the structural characteristics of a narrow minor groove such as that associated with A-T-rich DNA.

Correctly Folded Hybrid B-box Proteins Footprint to A-T-rich Sequences—Fig. 6 shows the results of DNase I footprinting experiments in which a direct comparison was made between the binding of equimolar concentrations (indicated by the + lanes) of the correctly folded hybrid B-box protein (panel A) and the GST-expressed WT B-box protein (panel B) on A-T-rich BLT DNA. As indicated by the vertical lines adjacent to the + lane in panel A, the correctly folded hybrid B-box protein clearly shows protection of several A-T-rich regions in BLT DNA, relative to the nuclease cleavage of naked DNA shown in the − lane. In contrast, the WT B-box protein does not show protection of the substrate relative to the − lane (Fig. 6B). In fact, the pattern of DNase I protection for the correctly folded hybrid B-box protein to these A-T-rich regions in the BLT DNA is similar, if not identical, to the previously published binding sites for the HMG-I(Y) protein on this same substrate (7, 11). These footprinting results, combined with the previously described results from the EMSA and substrate competition experiments, conclusively demonstrate that the specificity of the correctly folded hybrid B-box protein has been altered to closely resemble that of the HMG-I(Y) protein for A-T-rich regions of B-form DNA. Both HMG-I(Y) and Hybrid B-box Proteins Bind to Nucleosome Core Particles—It is well established that, in chromatin, the HMG-1 protein binds to the DNA linker region separating adjacent nucleosomes but does not bind in vitro to isolated nucleosome core particles that lack linker DNA (1). In marked contrast, as illustrated in Fig. 7A, HMG-I(Y) can form up to four or more specific complexes with the random sequence nucleosome core particles isolated from chicken erythrocytes. These data are consistent with previous observations (27) and are thought to be a consequence of the A-T-hook motifs of the HMG-I(Y) protein recognizing certain distorted structural features of the DNA as it wraps around the histone octamer (27, 29). In a similar fashion, as indicated by the arrows in Fig. 7B, the correctly folded hybrid B-box also forms a number of distinct complexes with nucleosome core particles, although with a distinctly reduced binding affinity relative to the HMG-I(Y) protein (i.e. micromolar versus nanomolar). As mentioned previously, this reduced core particle binding affinity is most likely a consequence of there being only a single A-T-hook peptide
The present results also raise intriguing questions about the structural features of both the B-box and the A'T-hook that permit the fusion product of these two very different peptides to form a correctly folded chimeric protein, which, on the one hand, has the rigid tertiary structure of the B-box of the HMG-I(Y) protein family and yet, on the other, retains the substrate binding specificity of the HMG-I(Y) protein family. Considerable insight into this problem comes from analysis of computer modeling studies based on the known structures of both the HMG-box and A'T-hook peptides as they exist either free in solution (12, 19, 20, 38) or as co-complexes with DNA interactions, increased tumor metastatic potential, and overt neoplastic malignancy. What remains unclear, however, is the precise molecular role(s) played by the HMG-I(Y) proteins, or their derivatives, in these tumorigenic processes.

Considerable uncertainty surrounds the possible physiological roles, if any, played by the translocation-derived, A'T-hook-containing, chimeric fusion proteins associated with a wide variety of benign tumors and overtly malignant cancers (Table I). Here, employing domain-swap experiments in a well characterized in vitro model protein system, we have demonstrated that the A'T-hook motif of the HMG-I(Y) protein family maintains its A'T-DNA-binding, and its nucleosome core particle binding, specificity in both an unstructured non-folded hybrid protein and in a structurally rigid correctly folded hybrid protein. These findings are important because they strongly suggest that A'T-hook-containing chimeric tumor proteins likewise retain many of the substrate-binding properties of the HMG-I(Y) protein regardless of whether or not these hybrids are folded in vivo into a distinct tertiary configuration. Although the structural characteristics of the folded ectopic peptide partners fused to A'T-hooks are unlikely to radically alter the substrate binding specificity of the A'T-hook motifs in vivo, they may, nevertheless, considerably reduce the affinity of the hybrid proteins for their substrate binding sites. The conferral of HMG-I(Y)-like substrate binding characteristics to chimeric tumor proteins in vivo thus provides a potential focus for the development of a new category of cancer therapeutic drugs aimed at selectively disrupting such aberrant protein-DNA interactions.
of the hybrid transcripts, alterations in the stability and/or intracellular localization of the hybrid proteins, as well as others. Thus, overexpression of A·T-hook-containing, but otherwise defective, hybrid fusion proteins could be expected to lead to abnormal gene transcriptional expression, as well as major alterations in chromatin structure, in tumor cells. Speculation as to the in vivo role played by these chimeric proteins often tends to focus on the type of protein partner fused to the A·T-hook motif when, in fact, the most important aspect of these hybrid tumor proteins may be the abnormal functioning of A·T-hook motifs themselves. Thus, processes related to the A·T-hook motif such as constitutive overexpression, and/or regulation of expression of A·T-hook-containing proteins are likely to be at least as important as the nature of the ectopic fusion partner per se. Future characterization of the structure and DNA-binding properties of individual chimeric proteins in specific tumors should provide more insight into the role and functionality of the A·T-hook motif in tumorigenesis.

Acknowledgments—We thank Dr. David Baker and Dr. Michael A. Kennedy for use of their circular dichroism instruments, Mark Nissen for his advice, and Gerhard Munske for the synthesis of primers used in PCR.

REFERENCES

1. Bustin, M., and Reeves, R. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 10530–10535
2. Nicolas, F. J., Cayuela, J. L., Martinez-Argudo, I., Ruiz-Vazquez, R. M., and Gronenborn, A. M. (1997) Nat. Struct. Biol. 4, 657–665
3. Lohmueller, K., Colman, R. J., and Zeleznik-Le, N. J. (2001) Trends Cell Biol. 11, 128–135
4. Enders, M. W., and Zeleznik-Le, N. J. (2003) J. Biol. Chem. 278, 38453–38458

Role of the A·T-hook Motif in Hybrid Proteins

| Example of fusion proteins that contain the A·T-hook DNA binding motif are shown. The names and/or regions of the translocated genes are separated by a colon (:). ALL, acute lymphoblastic leukemia; MLL, mixed lineage leukemia; HRX, human trithorax protein. | HMGIC:acidic serine/threonine-rich domain (45 amino acids) (15) | HMGIC:acidic serine/threonine-rich domain (45 amino acids) (15) | HMGIC:mitochondrial ADH2 (10 amino acids) (42) | HMGIC:FHIT (31 amino acids) (43) | HMGIC:NKQDSQEQ (7 amino acids) (44) | HMGIC:NFIB (5 amino acids) (45) | HMGIC:LM (3 amino acids) (46) | HMGIC:LaM (127 amino acids) (47) |
|---|---|---|---|---|---|---|---|---|
| HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) | HMGIC:LAM (3 amino acids) (46) |
| ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) | ALL:1:ENL (approximately 2,523 amino acids) (16,50) |

Solution NMR studies of a co-complex of the DNA-binding domains of HMG-I(Y) with a synthetic substrate indicate that the A·T-hook undergoes a disordered-to-ordered structural transition upon binding that is necessary for selective association with the minor groove of A·T-rich sequences. As evidenced by the results presented here, the positioning of the A·T-hook in the extended peptide region of the correctly folded hybrid B-box is not likely to significantly disrupt this critical substrate-associated structural transition. Indeed, molecular modeling studies predict that in the correctly folded hybrid both the peptide backbone and the side chains of the arginine residues in the palindromic “core” sequence of the A·T-hook motif (i.e. PRGPRP; Fig. 1) can assume the physical conformation necessary for interactions with the minor groove of A·T-DNA, thus conferring the appropriate substrate specificity on the chimeric B-box protein (data not shown). On the other hand, molecular modeling also indicates that the peptide backbone and the side chains of the lysine and arginine residues in the N-terminal extended domain of the WT B-box (19) are unable to assume such a physical configuration thus restricting the ability of the WT B-box to specifically bind to A·T-DNA sequences. Together these results provide a structural rationale for explaining the substrate binding differences between these otherwise quite similar proteins. There are numerous biological implications from the present work with regard to the possible etiology and maintenance of human tumors associated with chromosomal translocations involving the A·T-hook motif. The first, and most obvious, of these is that many of the chimeric A·T-hook-containing fusion proteins probably have acquired HMG-I(Y)-like DNA and nucleosome binding properties in vivo, an inference supported by observations made with the translocated A·T-hook motifs of the mixed lineage leukemia gene (17). But perhaps just as importantly, the present results are consistent with the idea that these novel tumor-associated hybrid fusion proteins may have also lost other properties that are necessary for normal in vivo functioning of the HMG-I(Y) proteins. Included among these likely aberrant properties of the chimeric proteins are: weakening of the normally high affinity binding of the A·T-hook motifs to A·T-rich DNA substrates, loss of cell-cycle-regulated transcriptional expression of the hybrid gene, loss of a linkage of expression of the hybrid gene to the differentiated state of the cell, loss of the ability of the hybrid protein to make specific associations with other proteins, loss of specific sites for secondary biochemical modifications in the hybrids, loss of normal mechanisms regulating stability and/or translational efficiency of the hybrid proteins.
Role of the A·T-hook Motif in Hybrid Proteins

37. Zhou, X., Benson, K. F., Ashar, H. R., and Chada, K. (1995) Nature 376, 771–774
38. Giererstanger, B. H., Volkman, B. F., Kremer, W., and Wemmer, D. E. (1994) Biochemistry 33, 5347–5355
39. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) Cell 81, 705–714
40. Sayle, R., and Milner-White, E. J. (1995) Trends Biochem. Sci. 20, 374
41. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
42. Kazmierczak, B., Hennig, Y., Wanschura, S., Rogalla, P., Bartnitzke, S., van de Ven, W. J. M., and Bullerdiek, J. (1995) Cancer Res. 55, 6038–6039
43. Geurts, J. M. W., Schoenmakers, E. F. P. M., Ruijer, E., Stenman, G., and van de Ven, W. J. M. (1998) Oncogene 16, 872
44. Kools, P. F. J., and van de Ven, W. J. M. (1996) Cancer Genet. Cytogenet. 91, 1–7
45. Xiao, S., Lux, M. L., Reeves, R., and Hudson, T. J. F. J. A. (1997) Am. J. Pathol. 150, 801–910
46. Domer, P. H., Fakharzadeh, S. S., Chen, C.-S., Jockel, J., Johansen, L., Silverman, G. A., Kersey, J. H., and Korsmeyer, S. J. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 7884–7888
47. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M., and Canaani, E. (1992) Cell 71, 701–708