HMGN3a and HMGN3b, Two Protein Isoforms with a Tissue-specific Expression Pattern, Expand the Cellular Repertoire of Nucleosome-binding Proteins*

Katherine L. West‡‡, Yuko Ito‡, Yehudit Birger, Yuri Postnikov, Hitoshi Shirakawa‡, and Michael Bustin
From the Protein Section, Laboratory of Metabolism, Division of Basic Science, NCI, National Institutes of Health, Bethesda, Maryland 20892

Received for publication, February 23, 2001, and in revised form, May 10, 2001
Published, JBC Papers in Press, May 16, 2001, DOI 10.1074/jbc.M101692200

HMGN1 (HMG-14) and HMGN2 (HMG-17) are nuclear proteins that bind specifically to nucleosomes, reduce the compactness of the chromatin fiber, and enhance transcription from chromatin templates. Here we report that many vertebrates contain an additional type of HMGN protein named HMGN3 (Trip 7). The human HMGN3 gene is located on chromosome 6 and spans 32 kilobase pairs, which is nearly 10-fold longer than the closely related HMGN2 gene. However, the intron/exon boundaries of the HMGN3 gene are identical to those of HMGN1 and HMGN2. Unique within the HMGN family, the HMGN3 transcript undergoes alternative splicing and generates two different variants, HMGN3a and HMGN3b. The shorter variant, HMGN3b, arises from an additional splice site that truncates exon V and causes a frameshift. The resulting HMGN3b protein lacks the majority of the C-terminal chromatin-unfolding domain. Both splice variants are found in many vertebrates from frogs to man and are expressed in many tissues. The pattern of tissue-specific expression differs considerably from those of HMGN1 and HMGN2 at both the mRNA and the protein level. Our results expand the multiplicity of the HMGN protein family and raise the possibility that these nucleosome-binding proteins function as co-activators in tissue-specific gene expression.

The DNA in eukaryotic nuclei is organized into chromatin by wrapping around histone octamers to form nucleosomes, and the nucleosomal array is compacted by the subsequent binding of linker histones. This organization inhibits the access of protein complexes that carry out activities such as transcription, replication, repair, and recombination (1). Several cellular mechanisms, including modification of the core histone N-terminal tails and ATP-dependent nucleosome remodeling, are known to unfold chromatin and facilitate access to the DNA (2). In addition, all mammalian and most vertebrate cells contain the multiplicity of the HMGN protein family and raise the possibility that these nucleosome-binding proteins function as co-activators in tissue-specific gene expression.

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.

‡ Both authors contributed equally to this work.
§ To whom correspondence and reprint requests should be addressed: Bldg. 37, Rm. 3D12, NCI, National Institutes of Health, Bethesda, MD 20892. Tel.: 301-496-5235; E-mail: klmarsch@pop.nci.nih.gov.
‡‡ Present address: Laboratory of Metabolism, Division of Life Sciences, Graduate School of Agricultural Science, Tohoku University, 1-1 Tsutsumidori-Amamiya-cho, Aoba-ku, Sendai 981-8555, Japan.

The revised nomenclature for HMG proteins is as follows: HMGN1 was HMG-14. HMGN2 was HMG-17. HMGN3 was Trip7. HMGB2 was HMG-2. HMGA1a was HMG-I. HMGA1b was HMG-Y.

katherine.l.west@nih.gov

This paper is available on line at http://www.jbc.org

25959

‡ The abbreviations used are: NLS, nuclear localization signal; TRβ1, thyroid hormone receptor β1; NBD, nucleosome binding domain; UTR, untranslated region; PCA, polymerase chain reaction; PAB, perchlorylic acid; kb, kilobase pair; nt, nucleotide; RAPD, random amplified polymorphic DNA.

‡‡ The abbreviations used are: NLS, nuclear localization signal; TRβ1, thyroid hormone receptor β1; NBD, nucleosome binding domain; UTR, untranslated region; PCA, polymerase chain reaction; PAB, perchlorylic acid; kb, kilobase pair; nt, nucleotide; RAPD, random amplified polymorphic DNA.
participate in the regulation of transcription and other DNA-dependent activities by unfolding chromatin.

The hallmark of HMGN proteins is the nucleosome-binding domain, which was originally found only in the canonical HMGN1 and HMGN2 proteins. Recently, we discovered a new member of the HMGN family, NBP-45 (also named NSBP1), that also binds specifically to nucleosome core particles (31). NBP-45 is unique in that in addition to the nucleosome-binding domain and the chromatin-unfolding domain, it has a large (321 amino acids), negatively charged C-terminal domain that can activate transcription in a reporter assay.

In addition, a cDNA coding for a closely related protein named Trip 7 (renamed HMGN3), was detected in a HeLa cell library by a yeast two-hybrid assay using the ligand binding/dimerization/transactivation domain of rat thyroid hormone receptor β1 (TRβ1) as bait (32). In the two-hybrid assay, HMGN3 interacted with both the thyroid hormone receptor and with retinoid X receptor in a ligand-dependent manner. In view of the similarity between HMGN3 and the other known members of the HMGN family, and considering that these proteins bind to nucleosomes, it is possible that HMGN3 promotes the interaction of hormone receptors with the chromatin template. However, to date, neither the protein nor the gene coding for HMGN3 has been characterized. In GenBankTM, the annotation of the partial HMGN3 cDNA clone indicates the presence of additional residues N-terminal to the PKRK motif that is the consensus start sequence of all HMGN proteins, raising the possibility that in fact HMGN3 is a novel type of HMGN protein.

Here we report the first characterization of the HMGN3 gene, transcript, and protein. We demonstrate that the protein is encoded by an unusually large gene located on human chromosome 6 and is expressed in a tissue-specific manner. Unlike other members of the HMGN family, HMGN3 exists as two splice variants, named HMGN3a and HMGN3b, and the latter lacks most of the C-terminal chromatin-unfolding domain. The presence of additional members of the HMGN family within eukaryotic cells raises the possibility that these proteins have specific roles in tissue-specific gene regulation.

**EXPERIMENTAL PROCEDURES**

**Data Base Searching and Sequence Manipulation—**HMGN3 ESTs were found by searching GenBankTM with the L40351 sequence using the basic BLAST 2.1 tool on the NCBI website. Sequences were manipulated and aligned using Biowire Jellyfish version 1.21 or Multialin, and BLAST searched of the working draft of the human genome on the NCBI website. The genomic sequence of human HMGN3 was found by a reverse phase high pressure liquid chromatography as described previously (35). Mouse tissues were homogenized in phosphate-buffered saline, and then PCA extracts were prepared as described previously (36). Overexpressed His6-tagged protein was purified using TALON metal affinity resin (CLONTECH) as described in the manufacturer’s instructions.

**Antibody Production—**Antibodies were produced by Quality Biologicals, Inc. Two antibodies were raised in rabbit against keyhole limpet hemocyanin conjugates of the following peptides from hHMGN3: antibody 2751, RKSPENTEGKDGSKVTKQEPT; antibody 2752, KT-SAKKEPGAKISRGA. Antibody 2859 was raised against chicken against the following peptides: APSENGETKAAEAKTSEVDNEQG. The antibodies were affinity purified using the same peptide sequences.

**Protein Purification—**Overexpression of recombinant proteins, expression vectors were transformed into BL21-DE3lysS cells, and expression was induced with 0.4 mM isopropyl-1-thio-β-D-galactopyranoside. Recombinant HMGN proteins and 5% PCA tissue culture extracts were prepared as described previously (35). Mass Spectrometry—HeLa cell 5% PCA extracts were further purified by reverse phase high pressure liquid chromatography as described previously (35). Fractions containing the high mobility band that was believed to correspond to HMGN3b were identified by Western analysis using 15% SDS-polyacrylamide gels (Bio-Rad) and blotted onto polyvinylidene difluoride membrane. After blocking in PBST (phosphate-buffered saline with 0.1% Tween 20) with 5% milk, membranes were washed in PBST and then incubated at 4 °C overnight with primary antibody in PBST and 1% milk. The final antibody concentrations used were as follows: 2751 and 2752, 6 ng/ml; 2859, 50 ng/ml; 2859, 1 ng/ml (37). Membranes were washed three times in PBST and then incubated with anti-rabbit or anti-chicken IgG-peroxidase conjugate (Pierce) in PBST containing 5% milk. Membranes were washed and then the bound antibodies were detected with ECL plus detection reagent (Amersham Pharmacia Biotech).

**Gel Retardation Analysis—**Nucleosome core particles were purified from chicken red blood cell nuclei as described previously (18, 38). Recombinant HMGN protein (20–200 nM) was incubated with 50 nM core particles in 5 μl of 2× TBE with 2% Ficoll for 15 min on ice. Samples were then loaded directly onto a 5% polyacrylamide gel made in 2× TBE and electrophoresed at 4 °C. A parallel lane containing bromphenol blue and xylene cyanol dyes was run to measure the mi-
After electrophoresis, gels were stained with SYBR gold stain (Molecular Probes) and photographed using a yellow photographic filter.

Analysis of the Nucleosome Sequence Preferences of HMGN1 and HMGN3—His 6-tagged HMGN1 and HMGN3a (20 μg) were incubated with 80 μl of a 25% slurry of TALON resin in 2× TBE on a rotating wheel overnight at 4 °C. The slurry was washed 5 times with 2× TBE and resuspended in a total volume of 80 μl of 2× TBE. The efficiency of HMGN binding to the resin was assessed by SDS-polyacrylamide gel electrophoresis. The HMGN resin (400 ng of protein) was incubated with 10 μg of nucleosome core particles in 100 μl of binding buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.1% Tween 20, 5% glycerol, 2 mM dithiothreitol, 0.5 mg/ml bovine serum albumin). The total volume of resin slurry added was equalized to 5 μl in each reaction. Samples were incubated for 2 h at 4 °C on a rotating wheel and then the HMGN resin with bound core particles was pelleted and washed 3 times with binding buffer. To extract the DNA from the pelleted resin, 150 μl of stop buffer (0.2 M NaCl, 1 mM EDTA, 1% lithium dodecyl sulfate, 0.4 M LiOAc) was added, and then the mixture was extracted twice with phenol/chloroform/isoamyl alcohol prior to ethanol precipitation in the presence of glycogen. As a control, DNA was also extracted from the unselected nucleosome pool. The DNA sequences were analyzed by a modified version of the Random Amplified Polymorphic DNA (RAPD) method, which is a PCR-based technique (39). Each RAPD reaction contained one set of three 10-base oligonucleotide primers (obtained from Operon Technologies, Inc.), which generated an average of 60 PCR products. In a previous study of whether HMGN1 and HMGN2 display sequence preference when binding to nucleosomes, 30 sets of 3 primers each were used to screen the nucleosomal DNA (39). Here, three of the primer sets were used. Set 1 contained primers W11, AE18, and BE11 (Operon Technologies, Inc.). Set 2 contained K06, AM03, and AD13. Set 3 contained primers AD08, AR11, and F08. PCRs were performed with amplitaq DNA polymerase under the manufacturer's recommended conditions (PerkinElmer Life Sciences) and contained one set of end-labeled primers and 50 ng of extracted nucleosomal DNA. A control reaction without template DNA was also performed. The PCR conditions were 30 cycles of 30 s at 95 °C, 30 s at 45 °C, and 30 s at 72 °C. Samples were electrophoresed on a sequencing gel, which was then dried and visualized by PhosphorImager analysis (Molecular Dynamics).

Northern Hybridization—Human and mouse RNA master dot blots and Northern blots (CLONTECH) were probed with 32P-labeled open reading frame of human HMGN3 according to the manufacturer's protocol. The human RNA master blot was also probed with nucleotides 632–1015 from the 3'9-UTR of the HMGN2 cDNA. The autoradiograms were scanned using a Molecular Dynamics densitometer and analyzed with ImageQuant software (Molecular Dynamics). The samples on the dot blots have been normalized to eight different housekeeping genes to adjust for RNA loading differences.
allow the accurate determination of relative mRNA expression levels, whereas the samples on the Northern blots are normalized to β-actin (CLONTECH).

The RNA samples on the human RNA master dot blot in Fig. 7 are as follows: A1, whole brain; A2, amygdala; A3, caudate nucleus; A4, cerebellum; A5, cerebral cortex; A6, frontal lobe; A7, hippocampus; A8, medulla oblongata; B1, occipital lobe; B2, putamen; B3, substantia nigra; B4, temporal lobe; B5, thalamus; B6, subthalamic nucleus; B7, spinal cord; C1, heart; C2, aorta; C3, skeletal muscle; C4, colon; C5, bladder; C6, uterus; C7, prostate; C8, stomach; D1, testis; D2, ovary; D3, pancreas; D4, pituitary gland; D5, adrenal gland; D6, thyroid gland; D7, salivary gland; D8, mammary gland; E1, kidney; E2, liver; E3, small intestine; E4, spleen; E5, thymus; E6, peripheral leukocyte; E7, lymph node; E8, bone marrow; F1, appendix; F2, lung; F3, trachea; F4, placenta; G1, fetal brain; G2, fetal heart; G3, fetal kidney; G4, fetal liver; G5, fetal spleen; G6, fetal thymus; G7, fetal lung; H1, yeast total RNA; H2, yeast tRNA; H3, Escherichia coli tRNA; H4, E. coli DNA; H5, poly(rA); H6, human C1 1 DNA; H7, human DNA 100 ng; and H8, human DNA 500 ng.

The RNA samples on the mouse RNA master dot blot in Fig. 7 are as follows: A1, brain; A2, eye; A3, liver; A4, lung; A5, kidney; B1, heart; B2, skeletal muscle; B3, smooth muscle; C1, pancreas; C2, thyroid; C3, thymus; C4, submaxillary gland; C5, spleen; D1, testis; D2, ovary; D3, prostate; D4, epididymis; D5, uterus; E1, 7-day embryo; E2, 11-day embryo; E3, 15-day embryo; E4, 17-day embryo; F1, yeast total RNA; F2, yeast tRNA; F3, E. coli tRNA; and F4, E. coli DNA.

**RESULTS**

The mRNA for HMGN3 Exists as Two Splice Variants—HMGN3 was detected in a yeast two-hybrid assay through its ligand-dependent interaction with the thyroid hormone receptor, TR b1 (32). The annotated sequence of the partial HMGN3 cDNA clone in the GenBank™ data base (accession number L40357) indicates the presence of several amino acids upstream to the PKRK motif that is the consensus N terminus of all HMGN proteins. Thus, it is possible that translation of the HMGN3 mRNA initiates upstream of this partial cDNA, generating a novel type of HMGN protein with additional N-terminal residues. A BLAST search with the L40357 sequence yielded over 60 EST clones for human HMGN3, indicating that this gene is widely expressed. Whereas there is some sequence variability in the 5′ regions of these clones, probably as a result of sequence or cloning errors, the EST with the longest sequence upstream of the putative open reading frame is clone BE791451. This clone does not contain any additional translation initiation codons, suggesting that the start of the protein is at the PKRK motif. To confirm that this is the full transcript, the 5′-UTR was also cloned by PCR from a CapSite cDNA™ library (Fig. 1A). The generation of the CapSite cDNA™ li-
The open reading frame of 440–900 nt (Fig. 2) is 917–150 nt, an open reading frame of 237–276 nt, and a longer UTR of HMGN1 (40) and HMGN2 (41), as each has a short 5'-splice forms. HMGN3b, respectively, to indicate that they are different full-length and truncated proteins are termed HMGN3a and HMGN3b. The complete sequence of the human HMGN3 mRNA, which has been deposited in GenBankTM, was deduced by comparing the sequences of several EST clones and the CapSite cDNA™ clone (Fig. 1A). The mRNA is 854 nt long, encodes a protein of 273 amino acids, and lacks introns and is 93% identical to the mRNA sequence of HMGN3a, AW823915; and HMGN3b, BE506630. The regulation of the HMGN3 gene is intron I, which is 18 kb, and each subsequent intron is shorter than the one before it.

Two retropseudogenes for HMGN3 were identified, both located on chromosome 1 in band 1q21.3. Both retropseudogenes lack introns and are 93% identical to the mRNA sequence of HMGN3a from bases 1 to 735. The first retropseudogene, starting at position 117,353 of contig NT_004441.3, has no typical regulatory elements in its 5'-region, whereas the 5'-UTR is less conserved, sharing 43–48% identity with those of HMGN1 and N2, and the 3'-UTRs are only 35–39% identical.

The Structure of the HMGN3 Gene—The HMGN3 gene was identified by a BLAST search of the human genome data base using the consensus HMGN3 mRNA sequence. The gene is located on chromosome 6 in band 6q27 at 83.8 megabases, and the genomic sequence confirms the mRNA sequence shown in Fig. 1A. The HMGN3 gene has six exons, as do the HMGN1 and HMGN2 genes (Fig. 2B) (42, 43), and the intron/exon boundaries are conserved between all three genes (Fig. 3A). Exon I of the HMGN3 gene encodes the 5'-UTR and the first four amino acid residues. Exon VI encodes the final 12 residues and all of the 3'-UTR. The genomic sequence reveals that the shorter splice variant HMGN3b arises due to a truncation of exon V, and all the other exons are the same between the two variants. In contrast to the relatively short genes for HMGN1 (6.8 kb) and HMGN2 (3.5 kb), the HMGN3 gene spans over 32 kb (Fig. 2B). The longest intron in the HMGN3 gene is intron I, which is 18 kb, and each subsequent intron is shorter than the one before it.

Two retropseudogenes for HMGN3 were identified, both located on chromosome 1 in band 1q21.3. Both retropseudogenes lack introns and are 93% identical to the mRNA sequence of HMGN3a from bases 1 to 735. The first retropseudogene, starting at position 117,353 of contig NT_004441.3, has no typical regulatory elements in its 5'-region, whereas the 5'-sequence for the second retropseudogene, starting at position 287,580 of contig NT_004441.3, is not available. The lack of introns or regulatory elements points to a retroviral mechanism of origin, although no direct repeats could be identified for either retropseudogene. Neither of these retropseudogenes could produce functional proteins if they were transcribed, as the first has crucial mutations in the nuclear localization signal, and the second has a frameshift shortly after the start of the open reading frame.
FIG. 4. HMGN3a and HMGN3b proteins are expressed in vitro. A, antibodies raised against peptides from HMGN3 are specific for this protein. Western blot analysis of recombinant HMGN1, HMGN2, and HMGN3a with antibodies 2751 (lanes 5–7), 2752 (lanes 8–10), and 2859 (lanes 11–13) against HMGN3, or with a polyclonal antibody against HMGN2 (lanes 14–16). Lanes 1–4 show a Coomassie Blue-stained gel of the same proteins and molecular mass size markers. For the Western blots, 100 ng of each recombinant protein was loaded, whereas 500 ng of each protein was loaded on the Coomassie Blue-stained gel. B, two forms of HMGN3 are expressed in HeLa cells. Western blots of recombinant HMGN3a (lanes 1 and 4), 5% PCA extract from HeLa cells (lanes 2 and 5), and recombinant HMGN3b (lane 3). The blots were probed with antibody 2752 (lanes 1–3) or 2859 (lanes 4 and 5). C, HMGN3 is expressed in a cell line-specific manner. Recombinant HMGN3a (lane 1) and 5% PCA extracts from HeLa (lane 2), MCF-7 (lane 3), Hepa 1 (lane 4), and CV-1 cells (lane 5) were loaded. The top panel shows a Western blot probed with antibody 2752; the middle panel shows a blot probed with anti-HMGN2; and the bottom panel shows part of the equivalent Coomassie Blue-stained gel in the region of H1 as a loading control.

Like most housekeeping genes (44), the genes for HMGN1 and HMGN2 have CpG islands that span the entire promoter region and terminate at the start of exon II (42, 43). The CpG island program from the European Bioinformatics Institute defined the HMGN2 CpG island as spanning nucleotides −279 to +795, with a G + C content of 73% and a CpG dinucleotide content of 11.4%. The CpG island for HMGN1 stretches from −556 to +857, has a C + G content of 70%, and a CpG content of 9.5%. The HMGN3 gene also has a CpG island from −160 to +922, but it is less prominent, with a G + C content of 58.5%, and a CpG content of 6%. A search for potential transcription factor-binding sites in the promoters for HMGN1, HMGN2, and HMGN3 revealed very few sites that are present in all three genes. The major element common to all three genes is a CCAAT box between −59 and −101, which conforms well to the 9-base pair consensus sequence for the transcription factor NF-Y (45). The CCAAT box is one of the most ubiquitous promoter elements, and NF-Y is the major factor that recognizes it (46). Sp1 consensus sites, which can bind members of a large family of Sp1-related transcription factors, are found in a very high number of promoters (47, 48), and both HMGN1 and HMGN2 have several Sp1 sites. HMGN3 does not have either a TATA box or Sp1 consensus sites. The only other potential binding site that is common to all three genes is that for SRY. The lack of conserved transcription factor-binding sites between the promoters of the HMGN genes and the less prominent CpG island for HMGN3 suggests that the expression of each of these genes is regulated differently.

The Amino Acid Sequence of HMGN3 Is Highly Related to Those of HMGN1 and HMGN2—The amino acid sequence of HMGN3a is 41 and 54% identical to those of HMGN1 and HMGN2, respectively (Fig. 3A). Significantly, all the functional domains characteristic of the HMGN protein family, the bipartite nuclear localization signal (NLS1 and -2), the NBD, and the chromatin unfolding domain are present in HMGN3a (Fig. 3A). The NBD, which is rich in prolines and positively charged residues and is encoded mainly by exons III and IV, is very highly conserved between the three proteins. In particular, 9 of the 10 residues encoded by exon III are invariant in all HMGN proteins. The second residue in this domain, Thr-23 in HMGN3, varies between the different HMGN classes, being Lys in all HMGN1 proteins and Gln for nearly all HMGN2 proteins. The bipartite nuclear localization signal in exons I (NLS1) and V (NLS2) is also highly conserved between the three proteins. The C-terminal domain of HMGN3 differs to a greater extent, however, being only 25 and 43% identical to those of HMGN1 and HMGN2, respectively. It does contain the peptide PSENGETKAE, however, which is well conserved in all other HMGN proteins (49). The C-terminal domain of HMGN3a has a net negative charge of −8, which is greater than the charges of −7 and −2 for HMGN1 and HMGN2, respectively. The C-terminal domain is responsible for the chromatin unfolding and transcriptional activation activities of HMGN proteins, and the overall negative charge is characteristic of the transcriptional activation domain in many other proteins (50), including all members of the HMGB (HMG box) family (49). Most of the C-terminal domain is absent in the splice variant HMGN3b, as the final 24 residues, including the PSENGETKAE peptide, are replaced by the residues Glu and Asn.

HMGN3 Is Present in Many Vertebrate Species—ESTs for HMGN3a were found in cDNA libraries from a variety of vertebrates, including mouse, rat, cow, and frog (Fig. 3B). A 915-nt...
HMGN1 and HMGN2, with the intention of generating antibodies specific for HMGN3. Western blots of recombinant proteins revealed that all three antibodies are indeed specific for HMGN3, as they recognize HMGN3a but do not recognize HMGN1 or HMGN2 (Fig. 4A, lanes 5–13). Furthermore, a polyclonal antibody raised against full-length HMGN2 only weakly recognizes HMGN3 (Fig. 4A, lane 16), despite the 54% identity between the two proteins.

Western blots of a 5% PCA extract from HeLa cells probed with antibodies 2752 (Fig. 4B, lane 2) or 2751 (not shown) reveal two bands. The upper band has the same mobility as recombinant HMGN3a (Fig. 4B, lane 1) and corresponds to the full-length HMGN3a. These data support the above conclusion that translation begins at the PKRK motif, rather than at an unidentified upstream ATG codon, which would generate a larger protein. The lower band in the HeLa extract has the same mobility as recombinant HMGN3b, the splice variant (lane 3). Antibody 2859, which was raised against a C-terminal peptide, recognizes the full-length protein but not the smaller band (Fig. 4B, lane 5). Furthermore, matrix-assisted laser desorption ionization/time of flight spectrometry of the lower band confirmed that this protein is indeed a C-terminally truncated variant of HMGN3. These data are consistent with the lower band corresponding to the splice variant that is truncated at the C-terminal end and indicate that both splice forms are expressed in HeLa cells. Extracts from the human breast carcinoma line, MCF-7, contain much less HMGN3 than HeLa cells (Fig. 4C, lane 3), and no HMGN3 was detected in extracts from the monkey kidney cell line, CV-1 (Fig. 4C, lane 5), or the mouse liver cell line, Hepa-1 (Fig. 4C, lane 4). In contrast, HMGN2 had high expression in HeLa, MCF-7, and CV-1 cells but lower expression in the Hepa-1 cell line. It can be concluded that both HMGN3a and HMGN3b are translated in vivo and that their levels of expression vary considerably between cell lines.

**HMGN3 Binds to Nucleosomes with Little or No Sequence Selectivity**—The hallmark of HMGN proteins is the nucleosome-binding domain that enables them to bind specifically to chromatin, and sequence comparisons indicate that HMGN3 also possesses a nucleosome-binding domain. To determine whether this domain is functional, the ability of recombinant HMGN3a to bind to nucleosomes was analyzed by gel retardation assays (Fig. 5). HMGN3a was able to bind nucleosome core particles in a similar manner to HMGN1, and with comparable affinity (Fig. 5, compare lanes 5 and 10). This is consistent with the observation that the nucleosome-binding domain is highly conserved between HMGN1, HMGN2, and HMGN3 (Fig. 3A).

To test whether HMGN3 recognizes specific chromatin subunits, we used a recently developed approach for examining the DNA sequence of nucleosomes bound by HMGN proteins, which is based on the random amplified polymorphic DNA method (RAPD) (39). This sensitive technique is suitable for detecting differences in the sequence composition of different DNA pools. Histidine-tagged HMGN3a or HMGN1 was incubated with a large excess of nucleosome core particles, and the fraction of core particles that bound to the HMGN protein was isolated using the histidine tag. DNA from the bound core particles was extracted and amplified in three different RAPD PCRs, each of which contained one set of three 10-base oligonucleotide primers. The RAPD reaction generates a mixture of differently sized products depending on where the oligonucleotides anneal within the nucleosomal DNA population. The three primers sets used here are a subset of the 30 primer sets used in a previous study of whether HMGN1 and HMGN2 display sequence preferences when binding to nucleosomes (39). After electrophoresis on sequencing gels, the pattern of consensus sequence for mouse HMGN3 mRNA was deduced by comparison of several ESTs (Fig. 1A). It is 80% identical to the human mRNA overall, and the open reading frame is 91% identical to that of human HMGN3. In addition, splice variants homologous to human HMGN3b were identified in murine and frog EST libraries (Fig. 3B). The nucleotide sequences surrounding the splice sites at either end of the 41 nt deletion, including the AG and GT dinucleotides noted earlier, are absolutely conserved between the human, mouse, and frog sequences. The same sequences are found in the rat and cow ESTs, suggesting that the HMGN3b splice variant also exists in these species. At the amino acid level, the cow, mouse, and frog HMGN3a share 91, 86, and 71% identity with human HMGN3a, respectively (Fig. 3B). An alignment of the human, mouse, rat, cow, and frog HMGN3a protein sequences reveals that 61 of the 98 residues are absolutely conserved between these species (Fig. 3B, residues in bold). Many of these residues are also conserved with HMGN1 and HMGN2, but 18 of them do not occur in either human HMGN1 or HMGN2 and thus represent a unique amino acid signature for HMGN3 (Fig. 3B, underlined residues). For example, the variable residue encoded by exon III is Thr in all species of HMGN3. These observations imply that HMGN3a and HMGN3b are distinct HMGN proteins that are well conserved in all vertebrates and thus may play important roles within the cell.

**HMGN3a and HMGN3b Are Translated in Vivo**—To determine whether the mRNA for HMGN3 is indeed translated in vivo, Western blot analysis was performed on cell extracts using the following three antibodies raised against peptides from HMGN3: o2751, o2752, and o2859 (Fig. 3A). The peptides were chosen for their lack of conservation with HMGN1 and
bands from the HMGN3a-bound core particles (Fig. 6, lanes 4 and 8) was compared with those from unbound (lanes 2 and 6) and HMGN1-bound core particles (lanes 3 and 7). No significant differences were found between the three DNA samples for any of the primer sites tested (Fig. 6 and data not shown). The data indicate that HMGN3a does not have a detectable sequence preference when binding to nucleosomes. This concurs with previous data showing that neither HMGN1 nor HMGN2 have strong DNA sequence preferences (39).

**HMGN3 Has a Tissue-specific Expression Pattern**—The expression of HMGN3 in a wide range of human and mouse tissues was analyzed by RNA dot blots (Fig. 7). In these blots, the amounts of RNA in each spot are normalized to the transcription levels of 8 housekeeping genes; therefore, the intensity of the spot is indicative of the relative mRNA abundance in a tissue. Quantitative analysis of the spots indicates tissue-specific variations in the mRNA levels of each of the transcripts. HMGN3 is expressed in nearly all the tissues tested, but there is considerable variation in the level of expression between different tissues. The highest expression of human
HMGN3 occurs in the pancreas, the pituitary gland, and the heart, whereas expression was low in the brain. In contrast, the expression of mouse HMGN3 is very low in the pancreas but high in the brain and prostate. The highest expression in mouse is in the eye, however, where it is expressed at a 3-fold higher level than in the brain. Intermediate expression of HMGN3 is found in many tissues in both mouse and human, including the thyroid, kidney, and testis, whereas expression in the liver is intermediate in humans but very low in mouse. The data indicate that whereas the expression of HMGN3 is comparable between mouse and human in many tissues, there are some major differences between the HMGN3 expression patterns in the two species.

The pattern of expression of HMGN3 is significantly different from that of HMGN2, which also displayed tissue specificity (Fig. 7E). For example, in the pancreas HMGN3 is expressed at a high level, whereas HMGN2 is only expressed at an intermediate level. Conversely, in the thymus there is high expression of HMGN2 but intermediate expression of HMGN3.

To confirm the tissue-specific expression pattern revealed by the RNA dot blots, human and mouse multiple tissue Northern blots were probed with hHMGN3 (Fig. 8). Three transcripts of 0.9, 1.5, and 2.7 kb were observed, with the smallest transcript being most abundant in all the tissues tested. These transcripts have different sizes than those of HMGN1 and HMGN2, which are 1.2 kb in human and 1.1 kb in mouse. The size of the smallest transcript, 0.9 kb, corresponds to the mRNA sizes of 854 and 915 bases deduced from human and mouse EST sequences, respectively (Fig. 1A). It is likely that the longer mRNAs have extended 3'UTRs due to transcription readthrough. The Northern blot for mouse HMGN3 reveals a similar pattern of tissue-specific expression as the mouse RNA dot blot, with high expression in the brain, intermediate expression in the kidney and testis, and low expression in the liver and skeletal muscle. The lower panel shows the mouse Northern blot probed for β-actin as a loading control.

**Fig. 8.** Northern blot of HMGN3 mRNA in mouse and human tissues. The multiple tissue Northern blot was probed for HMGN3 (upper panel) or β-actin (lower panel).

HMGN3 and HMGN2 are expressed in a tissue-specific manner. Western blot analysis was performed on extracts from six different mouse tissues (Fig. 9). The splice variant, HMGN3b, was present at significantly higher levels than the full-length form (HMGN3a) in all the tissues, but the ratio between HMGN3a and HMGN3b appeared to be constant in all tissues. Consistent with the mRNA expression data, the highest protein expression was in the brain, with intermediate expression in the kidney and low expression in the liver. However, HMGN3 protein levels were significantly lower in the testis than in the kidney, whereas the mRNA expression is similar in both tissues (Fig. 7C and 8), suggesting that in some tissues the levels of HMGN3 proteins may be regulated post-transcriptionally. In contrast to he wide variation in HMGN3 protein levels, HMGN2 was expressed at a fairly constant level in the tissues analyzed (Fig. 9).

**DISCUSSION**

Our results show that many cell types contain two new nucleosome-binding proteins named HMGN3a and HMGN3b, which are splice variants of a transcript originating from a gene located on chromosome 6. The structures of the HMGN3a and HMGN3b proteins are very similar to those of the canonical members of the HMGN family, HMGN1 and HMGN2. Both splice variants contain the bipartite nuclear localization signal and the nucleosome-binding domain. HMGN3a also contains a typical chromatin-unfolding domain. However, HMGN3b has a C-terminal truncation in which a critical portion of the chromatin-unfolding domain is deleted. Both HMGN3 splice variants are expressed in a tissue-specific manner that differs from that of HMGN2. ESTs for HMGN3a and HMGN3b are found in a wide range of vertebrates, including frog, and the protein sequence is over 70% conserved between species. Our findings expand the cellular repertoire of nucleosome-binding proteins and raise the possibility that different members of the HMGN protein family participate in the regulation of specific sets of genes.

The **HMGN3 Gene**—The HMGN3 gene spans 32 kb and is much longer than the 6.8-kb HMGN1 gene or the 3.5-kb HMGN2 gene. However, the intron/exon boundaries of the HMGN3 gene are identical to those of the canonical HMGN1 and HMGN2, and all three genes have CpG islands in their 5' region, although the CpG island for HMGN3 is less prominent than those of HMGN1 and HMGN2. The HMGN3 promoter area, spanning ~500 nucleotides 5' to the start of transcription, shares little sequence homology with those of HMGN1 and HMGN2, and the only major element common to all three genes is a CCAAT box that fits the consensus sequence for the transcription factor NF-Y (45, 46). The lack of conserved elements in the HMGN3 promoter compared with those for HMGN1 and HMGN2, and its less prominent CpG island,
indicate that the regulation of expression of HMGN3 is likely to be quite different from those of HMGN1 and HMGN2 and may account for the distinct tissue-specific expression patterns of these proteins.

The HMGN3a and HMGN3b Splice Variants—The major HMGN3 transcript is 854 nt long and, unlike other HMGN transcripts, undergoes alternative splicing in which 41 nt at the 3’ end of exon V are removed, resulting in the deletion of 22 amino acids from the C terminus of the protein. At the amino acid level, the full-length variant, HMGN3a, shares 41 and 54% identity with HMGN1 and HMGN2, respectively. The most conserved regions are the nucleosome-binding domain and the two nuclear localization signals. The functionality of the nucleosome-binding domain was confirmed by gel shift analysis using recombinant HMGN3a and nucleosome core particles. The nuclear localization of HMGN3 was confirmed by microscopy of cells transfected with a plasmid expressing GFP-tagged HMGN3.3

Although the chromatin-unfolding domain of HMGN3a is less conserved than other domains of the protein, it still has a net negative charge and possesses the conserved peptide PSENGETKAE that is found in all HMGN family members. It is interesting that the C-terminal 22 residues, including the PSENGETKAE peptide, are absent from the splice variant, HMGN3b. C-terminal deletion mutants of HMGN1 or HMGN2 truncated at the same position are unable to unfold chromatin or to activate transcription but can still bind to nucleosomes (4, 6). Replacement of the deleted region with the acidic activation domains of HMGB2 (HMG-2)4 or GAL4 restored the chromatin unfolding and transcription activation abilities of HMGN1, indicating the importance of the negative charge in the C-terminal domain (4). These data indicate that the full-length protein, HMGN3a, is likely to unfold chromatin and activate transcription, whereas the shorter splice variant may be inactive in this respect. Furthermore, it is conceivable that the splice variant could act as a dominant negative inhibitor by binding to chromatin and preventing the access of full-length HMGN proteins (6). This could be an important function, as HMGN3b is present at levels similar to or greater than those of HMGN3a. The ratio between the two proteins did not vary significantly between different mouse tissues, suggesting that the alternative splicing is not subject to tissue-specific regulation. There are many reports of splice variants with functional differences, including a recent study on the splice variants HMGA1a (HMGI) and HMGA1b (HMGY),1 which differ by 11 amino acid residues between the first and second AT-hook DNA binding domains (51, 52). The two variants modulate the expression of distinctive sets of genes, and only HMGA1b can promote tumor formation and metastasis in nude mice (53).

Tissue and Species Specificity of HMGN3 Splice Variants—HMGN3 is expressed in most tissues, although at apparently lower levels than its closest homologue HMGN2. A rough indication of the overall level of expression is given by the observation that in the human EST data base there are ~65 ESTs for HMGN3, compared with over 470 ESTs for HMGN2. It is relevant that all previous protein purification and cDNA cloning detected HMGN1 and HMGN2 but not HMGN3 proteins. Taken together, the data suggest that HMGN3 is less abundant than HMGN1 or HMGN2.

The level of HMGN3 expression varies considerably between tissues, and the tissue-specific expression pattern is distinct from those of HMGN2 or HMGN1. For example, human HMGN3 is expressed most highly in the pancreas and pituitary gland, whereas human HMGN2 is expressed most highly in the thymus and thyroid gland. In comparison, the highest expression of HMGN3 is in the kidney and thymus.4 Mouse HMGN3 has a different expression pattern than human HMGN3 and is particularly highly expressed in the eye and brain. Preliminary immunolocalization studies with brain slices indicate that HMGN3 but not HMGN2 localizes to regions that are enriched in antigens specific for glial cells.3

Like HMGN1 and HMGN2, both variants of HMGN3 are present in a wide range of vertebrates from amphibians to human, and the degree of sequence conservation with human HMGN3a varies from 71% in frog to 91% in cow. The level of sequence conservation is comparable to those of HMGN1 and HMGN2, which share 65–76% identity between their human and frog counterparts. It is notable that 18 of the HMGN3a residues are totally conserved between all the species studied yet do not occur in either HMGN1 or HMGN2. Thus, the HMGN3 splice variants are clearly a novel type of HMGN protein and may perform functions that are distinct from those of HMGN1 or HMGN2.

Cellular Function of HMGN3 Proteins—The discovery of additional HMGN proteins raises the possibility that the different family members have distinct roles within the cell, an hypothesis supported by the strong evolutionary conservation and distinct tissue-specific expression patterns of each HMGN protein. Recent studies have shown that HMGN proteins are highly mobile within the nucleus (23) and that they are components of multiprotein complexes (39). Indeed, it is likely that the HMGN proteins are targeted to their sites of action in chromatin by association with specific protein partners, rather than by recognizing particular DNA sequences (39). We speculate that the different HMGN isoforms are components of different protein complexes and are thus specifically localized to the distinct chromatin regions where they are required.

The original finding that HMGN3a binds to TRβ1 and retinoid X receptor in a ligand-dependent fashion (32) points to one possible function for this protein. Chromatin modification is an important aspect of the mechanism of action of the thyroid hormone receptor (54, 55). It is possible that TRβ1 recruits HMGN3 to assist in chromatin unfolding and promote transcriptional activation. We have been unable to confirm the interaction of HMGN3 with TRβ1 by using recombinant proteins in vitro, however, and we are currently investigating the possibility that additional proteins and/or chromatin are needed to facilitate the interaction in vivo.

Acknowledgments—We thank Ulla Hansen and Han-Fei Ding for the expression vector for His6-HMGN1.

REFERENCES
1. Wolfe, A. P., and Kurumizaka, H. (1998) Prog. Nucleic Acids Res. Mol. Biol. 61, 379–422
2. Wolfe, A. P., and Hayes, J. J. (1999) Nucleic Acids Res. 27, 711–720
3. Bustin, M., Trieschmann, L., and Postnikov, Y. V. (1995) Semin. Cell Biol. 6, 237–255
4. Ding, H. F., Bustin, M., and Hansen, U. (1997) Mol. Cell. Biol. 17, 5843–5855
5. Trieschmann, L., Postnikov, Y. V., Rickers, A., and Bustin, M. (1995) Mol. Cell. Biol. 15, 6663–6669
6. Trieschmann, L., Alfonso, P. J., Crippa, M. P., Wolfe, A. P., and Bustin, M. (1995) EMBO J. 14, 1478–1489
7. Weigmann, N., Trieschmann, L., and Bustin, M. (1997) DNA Cell Biol. 16, 1207–1216
8. Vestner, B., Bustin, M., and Gruss, C. (1998) J. Biol. Chem. 273, 9409–9414
9. Paranjape, S. M., Krumm, A., and Kadonaga, J. T. (1995) Genes Dev. 9, 1978–1991
10. Bustin, M. (1999) Mol. Cell. Biol. 19, 5237–5246
11. Hock, R., Wilde, F., Scheer, U., and Bustin, M. (1998) EMBO J. 17, 6992–7001
12. Mardian, J. K., Paton, A. E., Bunick, G. J., and Olins, D. E. (1988) Science 209, 1534–1536
13. Sandeen, G., Wood, W. I., and Felsenfeld, G. (1988) Nucleic Acids Res. 8, 3757–3778
14. Postnikov, Y. V., Trieschmann, L., Rickers, A., and Bustin, M. (1995) J. Mol. Biol. 252, 423–432

3 Y. Ito, unpublished observations.
4 M. Bustin, unpublished observations.
New Members of the HMGN Nucleosome-binding Protein Family

25969

15. Postnikov, Y. V., Herrera, J. E., Hock, R., Scheer, U., and Bustin, M. (1997) J. Mol. Biol. 274, 454–465
16. Alfonso, P. J., Crippa, M. P., Hayes, J. J., and Bustin, M. (1994) J. Mol. Biol. 236, 189–198
17. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
18. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
19. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
20. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
21. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
22. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
23. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
24. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
25. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
26. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
27. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
28. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
29. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
30. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
31. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
32. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
33. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
34. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
35. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
36. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
37. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
38. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
39. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
40. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
41. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
42. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
43. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
44. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
45. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
46. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
47. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
48. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
49. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
50. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
51. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
52. Trieschmann, L., Martin, B., and Bustin, M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5468–5473
HMGN3a and HMGN3b, Two Protein Isoforms with a Tissue-specific Expression Pattern, Expand the Cellular Repertoire of Nucleosome-binding Proteins
Katherine L. West, Yuko Ito, Yehudit Birger, Yuri Postnikov, Hitoshi Shirakawa and Michael Bustin

J. Biol. Chem. 2001, 276:25959-25969.
doi: 10.1074/jbc.M101692200 originally published online May 16, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M101692200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 55 references, 24 of which can be accessed free at
http://www.jbc.org/content/276/28/25959.full.html#ref-list-1