Genomics **71,** 21–33 (2001)
doi:10.1006/geno.2000.6397, available online at http://www.idealibrary.com on IDEAL®

Identification and Characterization of a SET/NAP Protein Encoded by a Brain-Specific Gene, MB20

Hsin-Hsin Shen,* A-Mei Huang,* Jörg Hoheisel,† and Shih-Feng Tsai*,[1]

*Institute of Genetics, National Yang Ming University, Taipei 112, Taiwan, Republic of China; and †Molecular-Genetic Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120, Germany

Received May 31, 2000; accepted September 19, 2000

A new member of the NAP/SET gene family, named MB20, was isolated from a mouse brain cDNA library by virtue of its CAG trinucleotide repetitive sequence and a brain-specific gene expression pattern. The complementary DNA sequence predicted an open reading frame of 545 amino acids, with four copies of an 11-amino-acid direct repeat. The consensus sequence for these repeats, PKE-P--K-EE, is present in the largest subunit of murine neurofilament (NF-H). The MB20 protein sequence is homologous to nucleosome assembly proteins of several species, and its C-terminus is homologous to SET proteins. Immunoblot analysis revealed that MB20 protein is expressed in the brain. Transient transfection and immunofluorescence microscopy demonstrated that MB20 is distributed in the cytoplasm as well as in the nucleus. Deletion of the N-terminal end imparts the complete localization of MB20 protein to the nucleus. The ability of MB20 to bind histone proteins was analyzed by sucrose gradient sedimentation and by retention of histone proteins by immobilized MB20 protein. On the basis of its expression pattern, predicted sequence, and protein properties, we propose that MB20 plays a unique role in modulating nucleosome structure and gene expression during brain development.

INTRODUCTION

In eukaryotes, nucleosomal structures undergo dynamic changes to adapt to the cellular activities of DNA replication, mRNA transcription, and cell cycle progression. The modulation of nucleosome assembly and disassembly, therefore, is critical to many important biological processes. A eukaryotic genome is organized in the basic unit of the nucleosome. A nucleosome core consists of two sets of the four histone proteins, H2A, H2B, H3, and H4, and approximately 146 bp of DNA (Kornberg, 1977). Other proteins, such as H1 histone and nonhistone proteins, also contribute to the high-order structure of a functional chromosome. Although the structure of the nucleosome core has been extensively studied (Kornberg, 1977; Luger *et al.*, 1997), the dynamic aspect of the nucleosome has been elucidated only recently (Widom, 1998; Wolffe and Kurumizaka, 1998).

Several lines of evidence suggest that nucleosome assembly is coupled to DNA replication (Worcel *et al.*, 1978; Smith *et al.*, 1984; Kim *et al.*, 1988). On the other hand, the regulation of gene expression is dependent on transient alteration of the nucleosome in eukaryotic cells. Many expressed genes have nucleosome-free regions in their promoters, as identified by hypersensitivity to endonuclease digestion (Elgin, 1988). Biochemical analysis of the nucleosome assembly process has led to the identification of several histone-binding proteins, such as nucleoplasmin (Laskey *et al.*, 1978), N1/N2 (Kleinschmidt and Franke, 1982; Kleinschmidt *et al.*, 1985), and NAP-1 (Fujii-Nakata *et al.*, 1992). Collectively, these proteins appear to function as histone transfer vehicles that facilitate deposition of histones to DNA in an ATP-independent manner (Gruss and Sogo, 1992; Walffe, 1995). In addition to the histone-binding activity, NAP-1 proteins of *Xenopus* and yeast have been shown to interact specifically with B-type cyclins (Kellogg *et al.*, 1995). Genetic analysis also demonstrated that Clb2 (B-type cyclin of yeast) could not carry out its full range of functions in a NAP-1 null background (Kellogg and Murray, 1995). These observations extend the roles of NAP-1 from regulating gene expression, through histone transfer, to propelling cell cycle progression by interacting with components of the cell cycle regulators.

While it is well accepted that nucleosome modulators are universal, specific mechanisms exist among different tissues in higher animals. The brain, the major part of the central nervous system, is unique in that its major cellular constituents, neurons, cease to divide

The HGMW-approved symbol for the gene described in this paper is NAP1L3.

Sequence data reported in this article have been deposited with the GSDB, DDBJ, EMBL, and NCBI Data Libraries under Accession No. AB010711.

[1] To whom correspondence should be addressed at Institute of Genetics, National Yang Ming University, Shih-Pai, Taipei 112, Taiwan. Telephone: 886-2-8146-1041. Fax: 886-2-2826-4930. E-mail: ympetsai@ym.edu.tw.

0888-7543/01 $35.00

during much of adulthood. Despite its low proliferation potential, the neuron is metabolically and transcriptionally active. How are these genes regulated in neurons? Brain-specific factors and regulatory mechanisms involved in gene expression control should be of considerable interest.

Considering the large number of reports (Amato *et al.*, 1993; THDCRC, 1993; Orr *et al.*, 1993; Takiyama *et al.*, 1993; Koide *et al.*, 1994) relating "CAG" triplet repeats to neural functions, we set out to isolate "CAG" trinucleotide repetitive sequences containing genes that are specifically expressed in the brain. It is reasonable to assume that these brain-specific CAG triplet sequence-containing genes will be related to specific mechanisms required for maintaining normal neuronal functions. A 30-mer oligonucleotide with 10 copies of CA(G/A) was used as a probe to isolate 48 clones from a mouse fetal brain cDNA library and 326 clones from a *Drosophila* early embryo cDNA library. Using individual mouse CAG-containing clones as a probe, we screened the *Drosophila* CAG-containing clone pool to isolate murine clones that cross-hybridized with specific *Drosophila* clones. One of the mouse clones thus identified, MB20, was found to be specifically expressed in the brain. In this article, we report the identification and characterization of MB20. The predicted amino acid sequence of the MB20 protein is highly homologous to that of NAP proteins. Furthermore, we identified unique properties of the MB20 protein involved in histone binding and trafficking, reflecting its role as a chaperon in governing the availability of histone proteins. On the basis of its structural features and the previous observation of enhanced expression during neuronal differentiation (Fan *et al.*, 1998), we speculate that MB20 plays a role in sustaining neuronal function during brain development.

MATERIALS AND METHODS

Isolation of mouse brain cDNA clones containing a [CA(G/A)] trinucleotide repeat. $[CA(G/A)]_{10}$ oligonucleotides were end-labeled by T4 polynucleotide kinase (New England Biolabs) with [γ-^{32}P]ATP and used as probes to screen a mouse neonatal brain λgt11 cDNA library (Stratagene). Hybridization was carried out at 55°C in a solution of 5× SSPE, 5× Denhardt's, 100 μg/ml salmon sperm DNA, and 0.5% SDS. Sixty positive clones were isolated from a total of 120,000 clones after two rounds of plaque purification. These [CA(G/A)] trinucleotide repeats containing phage clones were then subjected to *in vivo* excision of the pBluescript SK(−) phagemid. The plasmid DNA of these [CA(G/A)] trinucleotide repeat-containing clones was prepared by a standard alkaline lysis method and the clones were named MB1 to MB60 because they were obtained from mouse brain. Southern blot analysis using $[CA(G/A)]_{10}$ oligonucleotides as probes was then applied to confirm the presence of [CA(G/A)] trinucleotide repeats in these clones. Among the 60 primary isolates, 48 clones were shown to be positive and were further characterized.

Cross-species hybridization. Three hundred twenty-six *Drosophila* embryonic cDNA clones that contain a CA(G/A) repeat were used as targets for cross-species hybridization. An individual mouse brain CA(G/A) repeat-containing cDNA clone was used as a probe and labeled with [α-^{32}P]dCTP (Amersham) with a random primer DNA labeling kit (Boehringer Mannhein). To block the triple repeats, $CA(G/A)]_{10}$ oligonucleotides were added to the cDNA probe, and the mixture was denatured by being boiled for 5 min. After the denaturing step, the probe and oligonucleotide mixture was cooled to 37°C, further incubated for 15 min, and then added to a hybridization solution (0.5 M sodium phosphate, pH 7.2, 7% SDS, 1 mM EDTA, and 0.1 mg/ml yeast tRNA). Hybridization was carried out at 45°C, followed by washing three times (15 min each) at the same temperature.

Northern blot analysis. The 3′-region of MB20 cDNA without a [CA(G/A)] trinucleotide repeat was used as a probe in Northern blot analysis. The DNA fragment was labeled with [α-^{32}P]dCTP (Amersham) with a random primer DNA labeling kit (Boehringer Mannhein). The poly(A)$^+$ RNA filters with multiple newborn mouse and fetal human tissues (Clontech) were probed to reveal the mRNA expression pattern of MB20. Hybridization was carried out according to a standard procedure recommended by the manufacturer. Control for RNA loading was carried out by probing with either β-actin or G3PDH probe.

Sequence determination and analysis. Restriction fragments of cDNA clones were subcloned into the pBluescript SK vector (Stratagene). Overlapping cDNA sequences were determined on both strands by standard dideoxy sequencing. Initially, T7 and T3 primers were used, and then oligonucleotides were synthesized and used as sequencing primers on the basis of the finished cDNA sequence. Sequences were initially analyzed with the computer program MicroGenie (Beckman), and later searched against current nucleotide and protein databases with the Blast program (Altschul *et al.*, 1990).

Recombinant protein production and purification. DNA sequences encoding the complete form and four different deleted forms of MB20 protein were subcloned into pET-23a vector (Novagen). These constructs could generate T7-tag-MB20-His-tag fusion proteins and they were designated pET-23a-MB20.a (full length), pET-23a-MB20.b, pET-23a-MB20.c, pET-23a-MB20.d, and pET-23a-MB20.e.

Recombinant fusion proteins were expressed in *Escherichia coli* strain BL21(DE3) carrying pLysS (Studier and Moffatt, 1986). Cultures of 500 ml were grown at 37°C to A_{600} = 0.5–1.0. IPTG was then added to a final concentration of 0.4 mM and further incubated for 3 h. The induced cells were harvested and resuspended in His-Bind Resin binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris–HCl, pH 7.9). After sonication for 20 min, insoluble debris was pelleted at 12,000 rpm in an SS34 rotor for 15 min. The full-length MB20 fusion protein (pET-23a-MB20.a) and three fusion proteins (pET-23a-MB20.b, pET-23a-MB20.c, and pET-23a-MB20.d) were insoluble and formed inclusion bodies. To purify fusion proteins from inclusion bodies, 6 M urea was added to the His-Bind Resin purification solutions. Only pET-23a-MB20.e-expressed protein was soluble. A standard protocol provided by Novagen was used to purify soluble fusion protein (pET-23a-MB20.e). Purified proteins were routinely analyzed by SDS–PAGE to determine the integrity and purity of recombinant proteins.

Anti-MB20 antibodies. Purified MB20 fusion protein, pET-23a-MB20.e, and a synthetic peptide, MB20.1, were used as antigens to immunize rabbits. Peptide MB20.1 corresponds to residues 221–234 (PEEVPEAKVEEEEA) of the MB20 polypeptide. Keyhole limpet hemocyanin was coupled to MB20.1 using Imject Activated Immunogen Conjugation Kits (Pierce). To enhance specificity, rabbit antisera were purified using an affinity column packed with purified pET-23a-MB20.e fusion protein-coupled His-Bind Resin (Novagen). MB20-specific antibodies were eluted by 0.2 M glycine, pH 2.5, and immediately neutralized by 1 M Tris–HCl, pH 8.0. The specificity of anti-MB20 antibodies was demonstrated by peptide blocking experiments.

Tissue extract preparation and Western blot analysis. Adult mouse brain and liver were homogenized in 1× PBS buffer containing 0.5% Triton X-100 and a cocktail of protease inhibitors (250 μM *p*-amidinophenylmethanesulfonyl fluoride, 5 μg/ml leupeptin, 5

μg/ml aprotinin, 5 μg/ml pepstatin A, 1 mM sodium metabisulfite, and 1 mM benzamidine). Crude extracts were mixed with an equal volume of loading dye and boiled for 20 min. Boiled samples were then centrifuged at 14,000 rpm for 15 min before electrophoresis in 10% polyacrylamide gel containing 0.1% sodium dodecyl sulfate. Gel-resolved proteins were electrophoretically transferred to a nitrocellulose filter, and the filter was blocked in TBST buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 0.1% Tween 20) containing 5% nonfat milk powder at room temperature for 1 h. Thereafter, filters were incubated with different dilutions of anti-MB20 antibodies at 4°C overnight. The membranes were washed with TBST buffer (three times, 15 min each time) and then incubated in a solution containing AP-conjugated secondary antibodies (Sigma). After being washed with TBST, proteins that reacted with the antibodies were detected with the NBT/BCIP kit (Promega).

MB20 eukaryotic expression plasmids. DNA sequences encoding the complete form and three deleted forms of MB20 polypeptides were subcloned into pFLAG-CMV vectors (Kodak) for expression in cell lines. These recombinant plasmids were named MB20.a, MB20.ak, MB20.b, and MB20.d (see Fig. 8 for details). Of these, only MB20.a is a full-length construct. These expression plasmids could generate FLAG-MB20 fusion proteins, and commercially available M5 or M2 monoclonal antibodies (Knappik and Pluckthun, 1994) are able to recognize them through the FLAG epitope.

Transfection and immunofluorescence microscopy. Human fibroblast 3T3 cells were maintained at 37°C in DMEM (Dulbecco's modified Eagle's medium) supplemented with 10% calf serum. MB20 expression constructs were transfected into 3T3 cells by a standard calcium phosphate precipitation procedure (Stenberg *et al.*, 1982). Forty-eight hours after transfection, cells were washed with 1× PBS buffer and then fixed and permeabilized with methanol:acetone (1:1) solution at −20°C for 5 min. The samples were blocked in 1% BSA in 1× PBS buffer at 37°C for 60 min. Rabbit sera against MB20 protein were diluted in 0.5% BSA in 1× PBS buffer (anti-MB20.1 peptide was a 1:100 dilution; anti-pET-23a-MB20.e fusion protein was a 1:200 dilution), and samples were incubated with antibody solutions at 37°C for 60 min. Thereafter, samples were washed with 1× PBS buffer and followed by a 60-min incubation with FITC-conjugated secondary antibodies. Finally, samples were mounted with glycerol and visualized with a fluorescence microscope.

Purification of histone proteins. Histone proteins were isolated from HeLa cells based on the method of Simon and Felsenfeld (1979). HeLa cells were maintained in DMEM containing 10% fetal bovine serum. After trypsinization, cells were collected by centrifugation at 3000 rpm for 5 min. The pellet was washed twice with 1× PBS and once with hypotonic buffer (10 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM PMSF, and 0.5 mM DTT). The cleaned cells were then resuspended in hypotonic buffer and placed on ice for 60 min. These cells were homogenized in a douncer and the nuclei were then precipitated by centrifugation at 5000 rpm for 15 min and subsequently resuspended in 1 mM Tris–HCl, pH 8.0, and 0.1 mM Na_2 EDTA. DNA was sheared by passing through a homogenizer, and the final concentration of DNA in the solution was adjusted to 1 mg/ml.

Sheared nuclear extract was later dialyzed in buffer A, which contained 0.63 N NaCl and 0.1 M potassium phosphate, pH 6.7. A hydroxyapatite column (BioGel, bed volume is 135 ml) was used to bind sheared chromatin (24 mg). After being washed with buffer A, H2A and H2B histone proteins were eluted by a buffer of 0.93 N NaCl and 0.1 M potassium phosphate, pH 6.7. The column was then washed with a 0.1 M potassium phosphate buffer (pH 6.7) containing a 0.93 to 1.2 N NaCl gradient. Finally, H3 and H4 histone proteins were eluted with 2 N NaCl in 0.1 M potassium phosphate, pH 6.7. The quality of prepared histones was assessed by SDS–PAGE.

Sucrose sedimentation analysis. Purified histone proteins (8 μg) were incubated with an equal amount of MB20 fusion proteins in 100 μl buffer containing 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl, and 0.1% Triton X-100. The histone binding reaction was carried out at 4°C overnight, and the mixture was then loaded onto a linear gradient of sucrose containing the same buffer. Sedimentation was performed in a tabletop centrifuge at 45,000 rpm for 18 h (TLA 100.3, Beckman). Aliquots of each fraction were analyzed by immunoblot analysis. An anti-histone antibody (Boehringer Mannheim) and an-T7 tag antibody were used to monitor histone proteins and MB20 fusion protein, respectively. The anti-histone antibody preferentially stained H2A/H2B.

Affinity chromatography. The binding reaction of MB20 and histones was carried out at 4°C overnight in a solution (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl, and 0.1% Triton X-100) containing 8 μg each of MB20 fusion protein and purified histone proteins. His-resin (50 μl) was then added to the reaction. After being washed 12 times with 1 ml of 1× PBS buffer, proteins captured by the beads were analyzed by immunoblots using anti-T7-tag and anti-histone antibodies.

RESULTS

Isolation and Characterization of cDNA Clones Containing CAG Trinucleotide Repeats

CAG repeat-containing genes were isolated from *Drosophila melanogaster* and mouse. Using a 30-mer oligonucleotide containing 10 copies of a CA(G/A) repeat as a probe, we isolated 326 genes from a *Drosophila* early embryonic cDNA library (Brown and Kafatos, 1988). Through DNA sequencing and hybridization analysis, this pool of CA(G/A)-containing genes was estimated to include more than 150 unique genes (Tsai, unpublished results).

To isolate CAG triplet repeat sequences from mouse, we screened several cDNA libraries of mouse tissues and concluded that the CAG triplet repeat-containing genes were most abundant in the brain, consistent with a previous report (Riggins *et al.*, 1992). Sixty cDNA clones containing a CAG triplet repeat were then isolated from a total of 120,000 plaques from a newborn mouse brain cDNA library. These clones were subjected to *in vivo* excision to release the pBluescript SK(−) phagemid. Forty-eight of the 60 clones were confirmed positive with the CAG triplet repeat probe by Southern blot analysis.

Using the two pools of CAG triplet repeat-containing clones from mouse brain and *Drosophila* embryos, we further classified these genes based on evolutionary conservation. An individual mouse clone in which CAG triplet repeats were depleted was used as a probe to screen the *Drosophila* CAG gene pool for homologous sequences (Fig. 1). Two mouse clones were found to cross-hybridize with several *Drosophila* clones. One of them was subsequently shown to be specifically expressed in the brain and was named MB20.

Expression Patterns of MB20 and Its Homologous Genes

Northern analysis of poly(A)$^+$ RNA from multiple mouse tissues (Clontech) was carried out to determine the expression pattern of the MB20 gene. As shown in Fig. 2A, the MB20 transcript was detected in the brain but not in any other tissues, including heart, placenta, lung, liver, skeletal muscle, kidney, and pancreas. The size of the MB20 transcript approximates 3.3 kb. Ad-

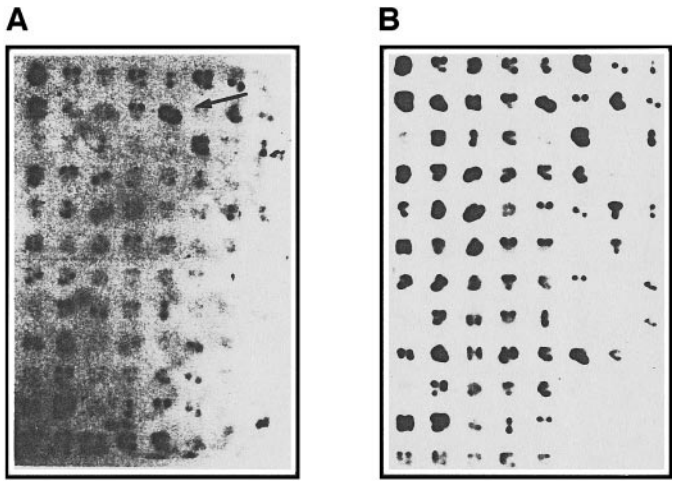

FIG. 1. Cross-hybridization of MB20 with *Drosophila* cDNA clones. Mouse clone MB20 was used as a probe to hybridize with filters containing *Drosophila* cDNA clones with CAG/CAA triplet repeats. (**A**) MB20 probe preannealed with $[CA(G/A)]_{10}$ and $[C/T(TG)]_{10}$ oligonucleotides to block the repetitive sequence. (**B**) MB20 probe. Arrow indicates enhanced signal with MB20 probe depleted of CAG/CAA sequences.

ditionally, using MB20 as a probe, we detected a transcript of the same size in human brain (data not shown). This finding indicates that the MB20 gene has a homologous human gene, which is also expressed specifically in the brain (Fan *et al.*, 1998).

RT-PCR was used to examine further the expression pattern of the MB20 gene. Total RNA of adult mouse cerebrum and cerebellum was isolated and used as templates for RT-PCR. The MB20 transcripts were detected in both tissues (data not shown). Moreover, RNA from fetal (E18) and adult rat tissues was applied in Northern blot analysis. As shown in Fig. 2B, transcripts of the MB20 homologous gene were detected at the E18 stage of development and in adult cerebrum and cerebellum of rat.

MB20 Encodes a Protein Related to the SET/NAP Family Members

The insert of the MB20 cDNA clone was completely sequenced on both strands. The sequence of the original clone predicted an open reading frame without a translation start site. We then carried out additional library screening and isolated six clones that overlap with the original MB20 clone. The overall length of the nucleotide sequence of these overlapping clones is 2837 bp, and it predicts a protein of 544 amino acids (Fig. 3). The CAG triplet repeat sequence is at nucleotide position 514 to 621, and it encodes a polyserine stretch in the N-terminal portion of the predicted polypeptide. When compared to known sequences in the protein database, the MB20 protein exhibits a significant homology to nucleosome assembly proteins (NAPs) (Ishimi and Kikuchi, 1991) and a proto-oncogene, SET (Adachi *et al.*, 1994). Multiple sequence alignment of MB20, NAPs, and SET proteins is shown in Fig. 4. The N-terminal portion of the MB20 polypeptide sequence, from amino acids 106 to 180, is highly homologous to that of NAP proteins. In this region, they share 53% identity and 78% similarity with the NAP consensus sequence. In addition, the C-terminal portions of MB20, NAP, and SET proteins are similar to one another. From amino acid position 337 to 513 of MB20, the identity and similarity between MB20 and NAP are 49 and 64%, respectively. Overall, the identity and similarity to NAP and SET proteins are 20 and 75% over the entire MB20 coding region. On the basis of sequence conservation, MB20 is considered to be a new member of the SET/NAP gene family. To be consistent with the nomenclature system of the human NAP gene family, we suggest that MB20 should be named mouse nucleosome assembly protein 1-like 3 (Nap1l3).

PEEK Repeat

Matrix analysis of the predicted MB20 amino acid sequence revealed several direct but incomplete tandem repeats in the middle portion of the amino acid sequence, which has little homology to NAP or SET proteins. This 11-amino-acid repeat unit (designated PEEK, hereafter) is reiterated four times in the MB20 peptide sequence. We derived a consensus sequence,

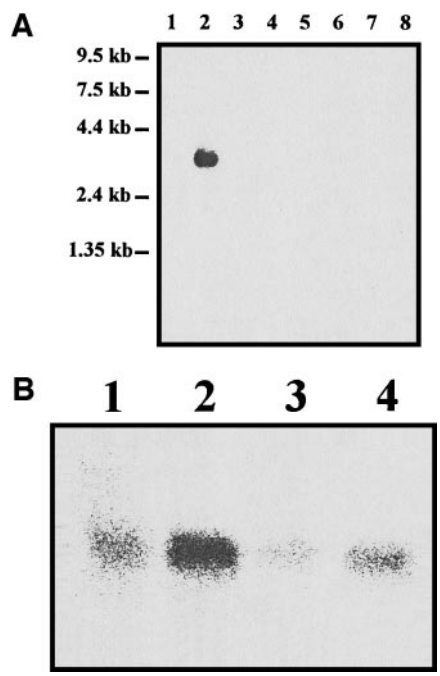

FIG. 2. Northern analysis of MB20. (**A**) Poly(A)$^+$ RNA from multiple mouse tissues was probed by MB20. Lane **1,** heart; **2,** brain; **3,** spleen; **4,** lung; **5,** liver; **6,** skeletal muscle; **7,** kidney; **8,** testis. Size markers are indicated on the left. (**B**) Expression of a rat homologue of MB20 in adult and embryonic brains. RNA of the cortex of 18-day embryos, adult cerebrum, and adult cerebellum was isolated from rat and applied for Northern analysis. Lane **1,** cerebrum; **2,** cerebellum; **3** and **4,** cortex of 18-day embryos.

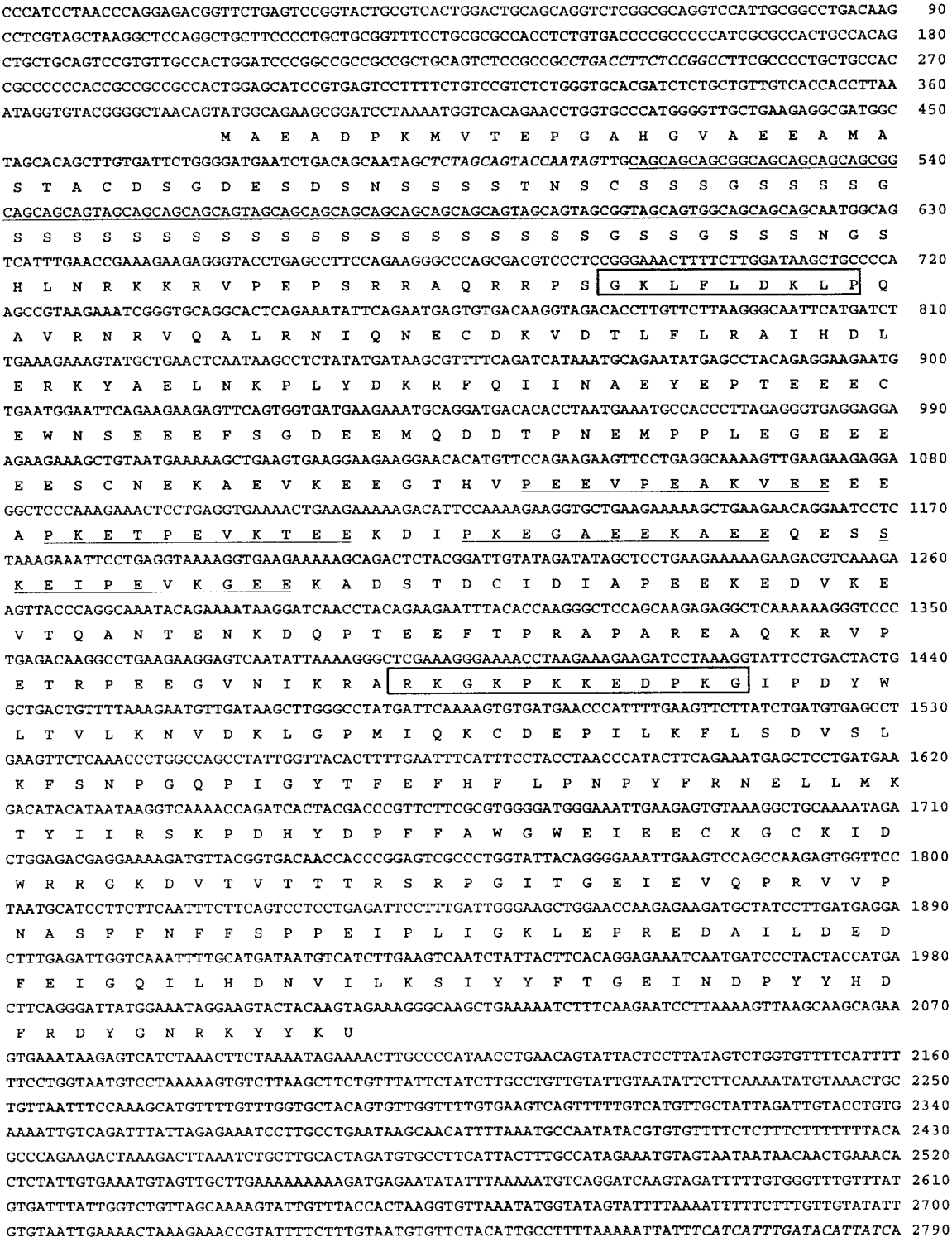

FIG. 3. The nucleotide and predicted amino acid sequences of MB20. The nucleotide sequence of the MB20 cDNA is shown, and the predicted amino acid sequence is given in single-letter code. A nuclear export signal (codons 104 to 112) and a nuclear localization signal (codons 337 to 348) are indicated by boxes. The CAG triplet repeat encoding polyserine and the 11-amino-acid PEEK repeats are underlined.

PKEXPEXKXEE (X is an aliphatic amino acid), and used it as a query to search the protein database. A peptide sequence similar to PEEK was found in the large subunit of murine neurofilament (NF-H) (Fig. 5).

MB20 Protein Is Expressed in the Brain

An anti-NAP monoclonal antibody, 4A8, originally generated against the NAP-1 protein from HeLa cells,

```
            1                                                                                                  80
RATSETA     .......... .......... .......... .......... .......... .......... .......... .....MAPKR
RATSETB     .......... .......... .......... .......... .......... .......... .......... ..........
HUNSET      .......... .......... .......... .......... .......... .......... .......... ..........
DSET        .......... .......... .......... .......... .......... .......... .......... ..........
MB20P       MAEADPKMVT EPGAHGVAEE AMASTACDSG DESDSNSSSS TNSCSSSGSS SSGSSSSSSS SSSSSSSSSS SSSGSSGSSS
HB20P       .......... .......... .MASSTSDSG EESD...... .....SSSSS SSTSDSSSSS STSGSSSGSG SSSSSSVSTS
MUS         .......... .......MAD IDNKEQSELD Q......... .......DLE DVEEVEEEET GEETKIKAR. .....QLTVQ
HUMN        .......... .......MAD IDNKEQSELD Q......... .......DLD DVEEVEEEET GEETKLKAR. .....QLTVQ
DNAPA       .......... .......... .......... .......... ..........  .......... .......... ..........
YESNAP      .......... .......... .MSDPIRTKP KSSMQIDNAP TPHNTPASVL NPSYLKNGNP VRAQAQEQDD KIGTINEEDI
ARAB        .......... .......... .......... .......... .......... ..KSLSSPNP QKMSNDKDSF NVSDLTAALK
URCHINNAP   .......... ........MD MDNPSENIVD QSAAGDTVVD MDPGPQGDGQ DLEGVEVIDT STAQKINSPD AIALSELKSQ

            81                                                                                                 160
RATSETA     QSAILPQPKK PRPVAAPKLE DKSASPGLPK GEKEQQEAIE HIDEVQNEID RLNEQASEEI LKVEQKYNKL RQPFFQKRSE
RATSETB     .......MSA PTAKASKKEL NSNHDGADET SEKEQQEAIE HIDEVQNEID RLNEQASEEI LKVEQKYNKL RQPFFQKRSE
HUNSET      .......MSA PTAKASKKEL NSNHDGADET SEKEQQEAIE HIDEVQNEID RLNEQASEEI LKVEQKYNKL RQPFFQKRSE
DSET        .......MSS VPKRAKLDGA PDGNTSAAAG NNEEESEALE QIDACQNEID ALNEKASEEI LKVEQKYNKL RKPCYEKRSE
MB20P       NGSHLNRKKR VPEPSRRAQR RPSGKLFLDK LPQAVRNRVQ ALRNIQNECD KVDTLFLRAI HDLERKYAEL NKPLYDKRFQ
HB20P       SRSRLYRKKR VPEPSRRARR APLGTNFVDR LPQAVRNRVQ ALRNIQDECD KVDTLFLKAI HDLERKYAEL NKPLYDRRFQ
MUS         MMQNPQILAA LQERLDGLVD TP..TGYIES LPKVVKRRVN ALKNLQVKCA QIEAKFYEEV HDLERKYAVL YQPLFDKRFE
HUMN        MMQNPQILAA LQERLDGLVE TP..TGYIES LPRVVKRRVN ALKNLQVKCA QIEAKFYEEV HDLERKYAVL YQPLFDKRFE
DNAPA       .......... MPAYMNSVMR RQYLQQMVKM LPAPVQNRIV FLKNLQLQHL NIEAQFFEDV YKLEQKYQVQ YQPLFDKRRE
YESNAP      LANQPLLLQS IQDRLGSLVG QD..SGYVGG LPKNVKEKLL SLKTLQSELF EVEKEFQVEM FELENKFLQK YKPIWEQRSR
ARAB        DEDRAGLVNA LKNKLQNLAG QR..SDVLEN LTPNVRKRVD ALRDIQSQHD ELEAKFREER AILEAKYQTL YQPLYVKRYE
URCHINNAP   VMQNPKALAA LQGHLDSLVG QR..SGYIES LPKVVKRRIK ALKNYQVKHT QLEAKFYEEV HSLECKYAKL YDGLYDKRRE
                                              *** *****  ******  *   **  *   *  ******* *   ** ****

            161                                                                                                240
RATSETA     LIAK...... .......... .......... .......... .......... .......... .......... ..........
RATSETB     LIAK...... .......... .......... .......... .......... .......... .......... ..........
HUNSET      LIAK...... .......... .......... .......... .......... .......... .......... ..........
DSET        LVKR...... .......... .......... .......... .......... .......... .......... ..........
MB20P       IINAEYEPTE EECEWNSE.E EFSGDEEMQD DTPNEMPPLE GEEEEESCNE KAEVKEEGTH VPEEVPEAKV EEEEAPKETP
HB20P       IINAEYEPTE EECEWNSEDE EFSSDEEVQD NTPSEMPPLE GEEEEN.... .......... .......... .....PKENP
MUS         IINAIYEPTE EECEWKPDEE D......... .......... .......... .......... .......... .......E..
HUMN        IINAIYEPTE EECEWKPDEE D......... .......... .......... .......... .......... .......E..
DNAPA       IIEGKVDPAE EKPQWKEPES S......... .......... .......... .......... .......... .......TDN
YESNAP      IISGQEQPKP EQIA...... .......... .......... .......... .......... .......... ..........
ARAB        IVNGTTEV.. .......... .......... .......... .......... .......... .......... ..........
URCHINNAP   VVNGVVEPTD SDCEWESSDD E......... .......... .......... .......... .......... .......EEE
            **** ***** *****  *** *                                                                  *

            241                                                                                                320
RATSETA     .......... .......... .......... .......... .......... .......... .......... ..........
RATSETB     .......... .......... .......... .......... .......... .......... .......... ..........
HUNSET      .......... .......... .......... .......... .......... .......... .......... ..........
DSET        .......... .......... .......... .......... .......... .......... .......... ..........
MB20P       EVKTEEKDIP KEGAEEKAEE QESSKEIPEV KGEEKADSTD CIDIAPEEKE DVKEVTQANT ENKDQP.TEE FTPRAPAREA
HB20P       EVKAEEKEVP KEIPEVKDEE KEVAKEIPEV KAEEKADSKD CMEATPEVKE DPKEVPQVKA DDKEQPKATE AKARAAVRET
YESNAP      .......... .KGQEIVESL NETELLVDEE EKAQNDSEE. .......... .......... .......... ..........
MUS         .......... .....VSEEL KEKAKIEDEK KDEEK..... .......... .......... .......... ..........
HUMN        .......... .....ISEEL KEKAKIEDEK KDEEK..... .......... .......... .......... ..........
DNAPA       EADAE..... ....HFREAL SSLKSIP... .......... .......... .......... .......... ..........
ARAB        .......... .........EL APEDDTKVDQ GEEKT..... .......... .......... .......... ..........
URCHINNAP   EEDEDMEEGK KEVNALAGEL KGKAKVTD.. .GEEE..... .......... .......... .......... ..........
            *   *   *   *   *   *   *       *
```

FIG. 4. Alignment of amino acid sequences of the MB20, NAPs, and SETs from different species. Asterisks beneath the alignment indicate identical or conserved amino acid residues between MB20 and NAP proteins. Asterisks above the alignment indicate identical or conserved amino acid residues of MB20 and the SET/NAP family (RATSETA, rat SET A; RATSETB, rat SET B; HUNSET, human SET; DSET, *Drosophila* SET; MUS, mouse NAP; HUMN, human NAP; DNAPA, *Drosophila* NAP A; YESNAP, yeast NAP; ARAB, *Arabidopsis* NAP; URCHINNAP, sea urchin NAP).

was used to test the relation of MB20 to other NAP proteins by the conserved signature epitope. Western blot analysis (Fig. 6A) revealed that the 4A8 monoclonal antibody reacted with the T7-tag-MB20-His-tag fusion protein produced in *E. coli*. Noticeably, recombinant proteins could form dimers at high concentrations (lanes 2 and 3 of Fig. 6A).

To facilitate a biochemical study of MB20, we raised antibodies to recombinant proteins expressed in *E. coli* and to a synthetic peptide derived from the PEEK repeat region. Immunoblot analysis indicated that both polyclonal antibodies detected MB20 fusion proteins at the nanogram sensitivity level (not shown).

A notable feature of the MB20 gene is that it lacks any intron (Shen and Tsai, unpublished results). It raises the concern that MB20 might represent a processed pseudogene that is transcribed but not translated. While MB20 transcripts are abundantly ex-

```
           321                                                                                                  400
                                                  *** *** *  *   *   *    **  *  *   * *
RATSETA    .......... .......... .......... IPNFWVTTFV NHPQVSALLG EEDEEALHYL TRVEVTEFED IKS..GYRID
RATSETB    .......... .......... .......... IPNFWVTTFV NHPQVSALLG EEDEEALHYL TRVEVTEFED IKS..GYRID
HUNSET     .......... .......... .......... IPNFWVTTFV NHPQVSALLG EEDEEALHYL TRVEVTEFED IKS..GYRID
DSET       .......... .......... .......... IPNFWVTSFI NHPQVSGILD EEEEECLHAL NKLEVEEFED IKS..GYRIN
MB20P      QKRVPETRPE EGVNIKRARK GKPKKEDPKG IPDYWLTVLK NVDKLGPMIQ KCDEPILKFL SDVSLKF.SN PGQPIGYTFE
HB20P      HKRVPEERLQ DSVDLKRARK GKPKREDPKG IPDYWLIVLK NVDKLGPMIQ KYDEPILKFL SDVSLKF.SK PGQPVSYTFE
MUS        .......... .......... .....EDPKG IPEFWLTVFK NVDLLSDMVQ EHDEPILKHL KDIKVKF.SD AGQPMSFVLE
HUMN       .......... .......... .....EDPKG IPEFWLTVFK NVDLLSDMVQ EHDEPILKHL KDIKVKF.SD AGQPMSFVLE
DNAPA      .......... .......... .....KDAKG IPGFWLTVFR NTAIMSEMVQ PHDEPAIRKL IDISIKY.DN GH...SYTLE
YESNAP     .......... .......... .....EQVKG IPSFWLTALE NLPIVCDTIT DRDAEVLEYL QDIGLEYLTD GRPGFKLLFR
ARAB       .......... .......... .....AEEKG VPSFWLTALK NNDVISEEVT ERDEGALKYL KDI..KW.CK IEEPKGFKLE
URCHINNAP  .......... .......... .....ESPSG IPGFWMTIFK NVDILGEMVQ DHDEPILSHL NDIRVKF.HE GPQ.MGFTLE

           401                                                                                                  480
           ***  * **  * ** *                 *  *   *     * **   ** *          *
RATSETA    FYFDENPYFE NKVLSKEFHL N......... ..ESGD.PSS KSTEIKWK.S GKDLTKRSSQ TQNKASRKR. .....QHEEP
RATSETB    FYFDENPYFE NKVLSKEFHL N......... ..ESGD.PSS KSTEIKWK.S GKDLTKRSSQ TQNKASRKR. .....QHEEP
HUNSET     FYFDENPYFE NKVLSKEFHL N......... ..ESGD.PSS KSTEIKWK.S GKDLTKRSSQ TQNKASRKR. .....QHEEP
MB20P      FHFLPNPYFR NELLMKTYII RSKPDHYDPF FAWGWEIEEC KGCKIDWR.R GKDVTVTTTR SRPGITGEIE VQP..RVVPN
DSET       FHFDENPYFE NKVLTKEFHL NSAAA..... .SENGDWPAS TSTPIKWK.E GKNLLK.LLL TKPYGNKKK. .....RNSEY
HB20P      FHFLPNPYFR NEVLVKTYII KAKPDHNDPF FSWGWEIEDC KGCKIDWR.R GKDVTVTTTQ SRTTATGEIE IQP..RVVPN
MUS        FHFEPNDYFT NEVLTKTYRM RSEPDDSDPF SFDGPEIMGC TGCQIDWK.K GKNVTLKTIK KKQKHKGRGT VRTVTKTVSN
HUMN       FHFEPNEYFT NEVLTKTYRM RSEPDDSDPF SFDGPEIMGC TGCQIDWK.K GKNVTLKTIK KKQKHKGRGT VRTVTKTVSN
DNAPA      FHFDKNEYFS NSVLTKQYVL KSTVDPNDPF AFEGPEIYKC TGCTINWE.K KMNLTVKTIR KKQKHKERGA VRTIVKQVPT
YESNAP     FDSSANPFFT NDILCKTYFY QKELGYSGDF IYDHAE.... .GCEISWKDN AHNVTVDLEM RKQRNKTTKQ VRTIEKITPI
ARAB       FFFDTNPYFK NTVLTKSYHM IDE....... ..DEPLLEKA MGTEIDWY.P GKCLTQKILK KKPK.KGSKN TKPITKLEDC
URCHINNAP  FFFTANDYFT NTVLNKSYQM KAEPDESDPF SFEGPEIIGS QGCEIDWK.K NKNVTVKVIK KKQKHKGRGT TRLVTKTVQT

           481                                                                                                  560
            ***                                  *** * * * *    *  **
RATSETA    ESFFTWFTDH SDAGAD.... .......... .....ELGEV IKDDIWPNPL QYYLVPDMDD EEGEAEDDDD DDEEEEGLED
RATSETB    ESFFTWFTDH SDAGAD.... .......... .....ELGEV IKDDIWPNPL QYYLVPDMDD EEGEAEDDDD DDEEEEGLED
HUNSET     ESFFTWFTDH SDAGAD.... .......... .....ELGEV IKDDIWPNPL QYYLVPDMDD EEGEAEDDDD DDEEEEGLED
DSET       KTFFDWFSDN TDPVND.... .......... .....EIAEL IKDDLWPNPL QYYLVPDIEV EPEDEEDNED ND..EEAFDD
MB20P      ASFFNFFSPP EIPLIG.... .KLEPREDAI LDEDFEIGQI LHDNVILKSI YYFTGEINDP YYHDFRDYGN RKYYK.....
HB20P      ASFFNFFSPP EIPMIG.... .KLEPREDAI LDEDFEIGQI LHDNVILKSI YYYTGEVNGT YYQFGKHYGN KKYRK.....
MUS        DSFFNFFAPP EVPENG.... .DLDDDAEAI LAADFEIGHF LRERIIPRSV LYFTGEAIED DDDDYDEEGE EADEEGEEEG
HUMN       DSFFNFFAPP EVPESG.... .DLDDDAEAI LAADFEIGHF LRERIIPRSV LYFTGEAIED DDDDYDEEGE EADEEGEEEG
DNAPA      DSFFNFFSPP EVPSDQ.... EEVDDDSQQI LATDFEIGHF LRARIIPKAV LYYTGDIVTM RTMRTKKSMT RTRRMSTTTT
YESNAP     ESFFNFFDPP KIQNEDQD.. EELEEDLEER LALDYSIGEQ LKDKLIPRAV DWFTGAALEF EFEEDEEEAD EDEDEEEDDD
ARAB       ESFFNFFSPP EVPDEDEDID EERAEDLQNL MEQDYDIGST IREKIIPRAV SWFTGEAMEA EDFEIDDDEE DDIDEDEDEE
URCHINNAP  DSFFNFFNPP KAPEEEVDGE IEEDEETEAL LSADFEIGHL IRERIIPRAV LYFTGEAIE. .DDEYEEEAE EDDQEGEEGD
               ***                           *  *                    **

           561                                                                                                  640
RATSET     IDEEGDEDEG EEDDDEDEGE EGEEDEGEDD .......... .......... .......... .......... ..........
RATSETB    IDEEGDEDEG EEDDDEDEGE EGEEDEGEDD .......... .......... .......... .......... ..........
HUNSET     IDEEGDEDEG EEDDDEDEGE EGEEDEGEDD .......... .......... .......... .......... ..........
DSET       EDGEDGEGEE EEEDEDDK.. .......... .......... .......... .......... .......... ..........
MB20       .......... .......... .......... .......... .......... .......... .......... ..........
HB20       .......... .......... .......... .......... .......... .......... .......... ..........
MUS        DE.ENDPDYD PKKD.QNPAE CKQQ...... .......... .......... .......... .......... ..........
HUMN       DE.ENDPDYD PKKD.QNPAE CKQQ...... .......... .......... .......... .......... ..........
DNAPA      MHRHQRAPNQ PVSRSSRPTT ARISRLMWLD YTCVMYIHNS NLI....... .......... .......... ..........
YESNAP     HGLEDDDGES AEEQDDFAGR PEQAPECKQS .......... .......... .......... .......... ..........
ARAB       DEEDEEDDDD EDEEESKTKK KPSIGNKKGG RSQIVGEGKQ DERPPECKQQ *SFTTLYQT* KDCVKI*FRS FSVHQE*GLR
URCHINNAP  EEDDNDPDFE PGKDGQKPPE CNQQ...... .......... .......... .......... .......... ..........
```

FIG. 4—*Continued*

pressed in the brain, the corresponding protein might not be translated at all. To address this issue, anti-MB20 polyclonal antisera were initially purified through an affinity column made by the pET-23a-MB20 fusion protein and His-resin. Purified antiserum was then used for immunoblot analysis of mouse brain and liver tissue extracts. In Fig. 6B, anti-hsp70 detected specific protein bands in both tissues, while anti-NF-H demonstrated a brain-specific expression pattern. Using purified anti-MB20 antibodies, two positive bands of approximately 80 kDa were detected in the brain but not in the liver. The MB20 native protein bands are significantly larger than the predicted molecular mass (64.6 kDa) of MB20, but at a similar position in SDS–PAGE to the pET-23a-MB20 recombinant protein (80 kDa). The presence of these two bands could be due to phosphorylation modification of MB20 protein, since it has a stretch of polyserine sequence. The nature of the additional band at around 60 kDa is unknown. To determine whether the 80-kDa doublet could be attributed to phosphorylation modification, we carried out phosphotase treatment

```
MB20   221 P E E V P E A K V E E 231 (E E A)
       235 P K E T P E V K T E E 245 (K D I)
       249 P K E G A E E K A E E 259 (N E S)
       263 S K E I P E V K G E E 273 (K A D)

Consensus:P K E X P E X K X E E
          | | |   |     |   | |
NF-H      P K E A P K P K V E E
```

FIG. 5. PEEK repeat of MB20 and NF-H. Four copies of a direct amino acid repeat of MB20, designated the PEEK repeat, are aligned with one another. A similar amino acid sequence was found in the largest subunit of neurofilament (NF-H). X denotes aliphatic amino acids.

experiments. Under conditions where control NF-H proteins displayed a mobility shift in immunoblots, the proportion and mobility of the two MB20 bands did not change (data not shown).

In Vitro Binding of MB20 Protein to Histones

By token of the sequence homology between NAP proteins and MB20, we speculated that MB20 is a brain-specific NAP protein, functioning as a histone chaperon. To test this contention, we carried out *in vitro* binding assays of MB20 and histones.

Sucrose gradient sedimentation and affinity interaction were used to analyze the MB20/histone complex formed in *in vitro* binding reactions. As shown in Fig. 7A, histone alone forms a sediment in the top fractions. Addition of MB20 proteins shifted the histone proteins

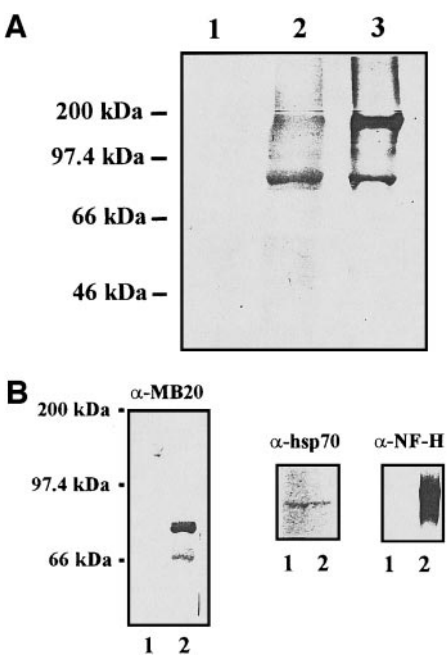

FIG. 6. Immunoblot analysis of MB20 proteins. (**A**) Detection of MB20 recombinant proteins by the anti-NAP-1 monoclonal antibody, 4A8. Lane 1, uninduced crude *E. coli* lysate; **2,** induced crude *E. coli* lysate; **3,** purified MB20 fusion protein. (**B**) Expression of MB20 proteins in mouse brain and liver tissues. Immobilized proteins (100 μg) on nitrocellulose filters were detected by the following antibodies: purified anti-MB20 rabbit serum (α-MB20), anti-hsp70 rabbit serum (α-hsp70), anti-NF-H mono-clonal antibodies (α-NF-H). Lane **1,** liver extract, **2,** brain extract.

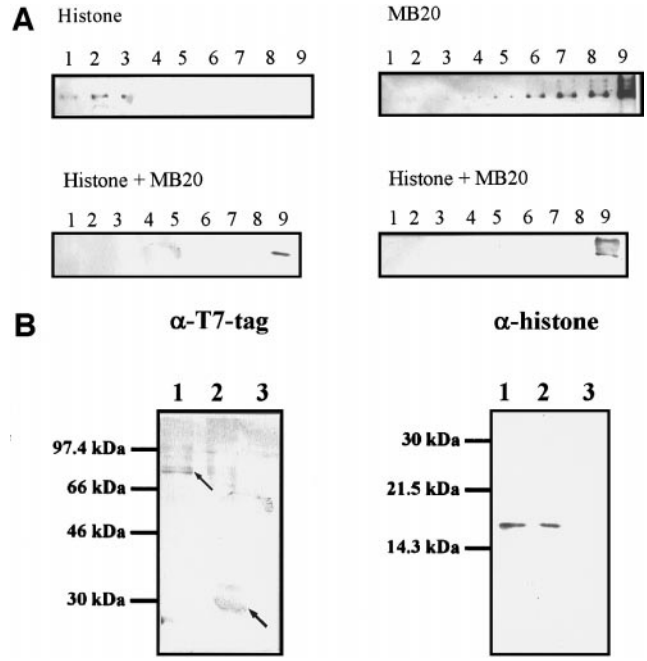

FIG. 7. Histone-binding activity of MB20. (**A**) Sucrose gradient sedimentation analysis of MB20/histone complex. Immunoblots were used to detect the presence of recombinant MB20 and histone proteins in the sucrose gradient fractions by anti-histone antibody (**top left, bottom left**) and by anti-T7 tag antibody (**top right, bottom right**). (**B**) Affinity interaction of MB20/histone complex. Immunoblots using anti-T7-tag antibody (**left**) and anti-histone antibody (**right**) were applied to detect the protein complex. Lane **1,** full-length MB20 proteins plus histones; lane **2,** N-terminally truncated MB20 proteins plus histones; lane **3,** histone only. Arrows indicate full-length and truncated MB20 proteins. Only H2A/H2B was stained by the anti-histone antibodies.

to a fraction that is at the bottom in sedimentation. Anti-MB20 antibody detected the presence of MB20 protein, in mononer and dimer forms, in the same bottom fraction. A truncated form of MB20 protein also formed a complex with histones in the bottom fraction (not shown).

Additionally, we examined the histone-binding activity by affinity interaction. Histone proteins were incubated with MB20 fusion proteins and then added to the affinity resin His-resin beads. In this assay system, we found that histone proteins were retained by the affinity beads (Fig. 7B). Furthermore, deletion of the N-terminal portion from amino acid 1 to 187 and the C-terminus from amino acid 364 to 545 did not remove the histone-binding domain of the MB20 protein (Fig. 7B, lane 2). This finding is consistent with the fact that the same truncated protein also forms complexes with histone proteins in the sucrose sedimentation analysis. These results, together, suggest that the MB20 protein has histone-binding activity and that the histone-binding domain is located in the region from amino acid 188 to 363 of the protein. Although we could not formally reject the possibility that the histone-binding activity of MB20 is nonspecific in nature, a previous study has reported histone-binding activity of truncated yeast NAP proteins (Fujii-Nakata *et al.*, 1992).

Cellular Distribution of the MB20 Protein

Previous studies of the NAP-1 proteins have reported their distribution in the nuclei of HeLa (human) cells (Ishimi *et al.*, 1983, 1984) as well as primarily in the cytoplasm of yeast (Kellogg and Murray, 1995). Ito *et al.* (1996) investigated the cellular localization of dNAP-1 in early *Drosophila* embryos and found that it is present in the nucleus during S phase and is predominantly cytoplasmic during the G_2 phase. As suggested by its sequence homology to NAP proteins, the cellular distribution of MB20 protein could reflect its function. We, therefore, examined its localization in the cells by transfection and immunofluorescence microscopy.

We constructed four expression plasmids driven by a CMV promoter, directing the expression of a full-length MB20 polypeptide sequence and three different deletion derivatives. These constructs were named MB20.a (full length), MB20.ak, MB20.b, and MB20.d (Fig. 8A). After transient tranfection into 3T3 cells, the intracellular distributions of full-length and truncated MB20 proteins were detected by immunofluorescence microscopy. Control experiments with preimmune sera or preabsorbed antibodies were included to demonstrate that the staining patterns are specific to MB20 proteins. Interestingly, we found that MB20.a and MB20.ak polypeptides are mostly cytoplasmic, and only a small proportion showed nuclear staining (typically 10%) (Fig. 8B, a; Fig. 8C, a). In contrast, the distribution of MB20.b and MB20.d polypeptides was exclusively nuclear (Fig. 8C, c and d).

DISCUSSION

MB20, a Brain-Specific Nucleosome Modulator Protein?

In this paper, we report the cloning and characterization of MB20, a brain-specific member of the SET/NAP gene family. The SET/NAP gene family has been identified only recently, and the function of each member awaits further clarification through biochemical and genetic studies. Nevertheless, it is now accepted that NAP proteins have a high affinity to histones, and some of them have been shown to have *in vitro* nucleosome assembly activity. Therefore, NAP proteins have been postulated to be involved in the DNA synthesis process. On the other hand, in studying mechanisms by which transcription factors approach nucleosomal DNA, Walter *et al.* (1995) indicated that the yeast NAP-1 protein stimulated transcription factor binding to nucleosomal DNA and induced core histone displacement. This report extended the possible role of NAP proteins to regulating gene expression. Activities other than NAP proteins have been shown to modulate nucleosome structure and transcription factor binding. These include histone-binding protein nucleoplasmin (Walter *et al.*, 1995), purified SWI/SNF complexes from yeast and human (Cote *et al.*, 1994; Kwon *et al.*, 1994), and ATP-dependent activities in *Drosophila* embryo extracts (Tsukiyama *et al.*, 1995; Tsukiyama and Wu, 1995). Together, these findings indicated that there are multiple factors that are involved in modulating structural states of chromatin and that thus play a role in transcriptional regulation. Conceivably, tissue-specific factor that enhance transcription factor binding to nucleosomal DNA or facilitate the disruption of nucleosome structure might be called upon to achieve a definitive regulation of tissue-specific genes. MB20, a brain-specific member of the NAP protein family, could fulfill such a role in the regulation of brain-specific gene activities.

CAG Repeat in MB20

MB20 was initially identified through its CAG triplet repeat. Like other CAG repeat-containing genes, the CAG triplet repeat is within the translated region of the MB20 gene. However, it encodes 32 serine residues interposed with 4 glycine residues. This is in sharp contrast to glutamine tracts, commonly encoded by the CAG triplet-containing genes associated with neurodegenerative diseases. In a search for MB20 homologous proteins, we found that polyserine tracts are present in other polypeptide sequences that have diverse functions, and they have been isolated from different species, for example, human transcription factor E2F4 (Ginsberg *et al.*, 1994); human AF-9 protein, which is involved in chromosomal translocation in acute leukemia (Nakamura *et al.*, 1993); yeast suppressor protein SRP40 (Bou *et al.*, 1993); and the *Xenopus* vitellogenin A2 precursor in which the serine-rich portion is assumed to be phosphorylated to a level of about 80% (Walker *et al.*, 1984). Although the significance of the polyserine sequence in MB20 function remains unclear, it is possible that the serine stretch is a domain that could be subjected to phosphorylation modification.

PEEK Repeats of MB20

Several incomplete direct repeats at position 221 to 276 of MB20 were identified. They were named PEEK repeats, reflecting the inclusion of proline, glutamic acid, and lysine residues in the repeat unit. The consensus sequence of PEEK repeats was also found in the largest subunit of neurofilament (NF-H) (Shneidman *et al.*, 1988). It is possible that this domain could be related to the unique expression of MB20 protein in the nervous system and that PEEK repeats might be critical to the functions of MB20 and NF-H in neuron cells. This finding is consistent with the notion that the function of MB20 is brain-specific and multifarious.

Control of MB20 Cellular Distribution

In analyzing the distribution of MB20 proteins in the cell, we found that MB20 is either cytoplasmic or nu-

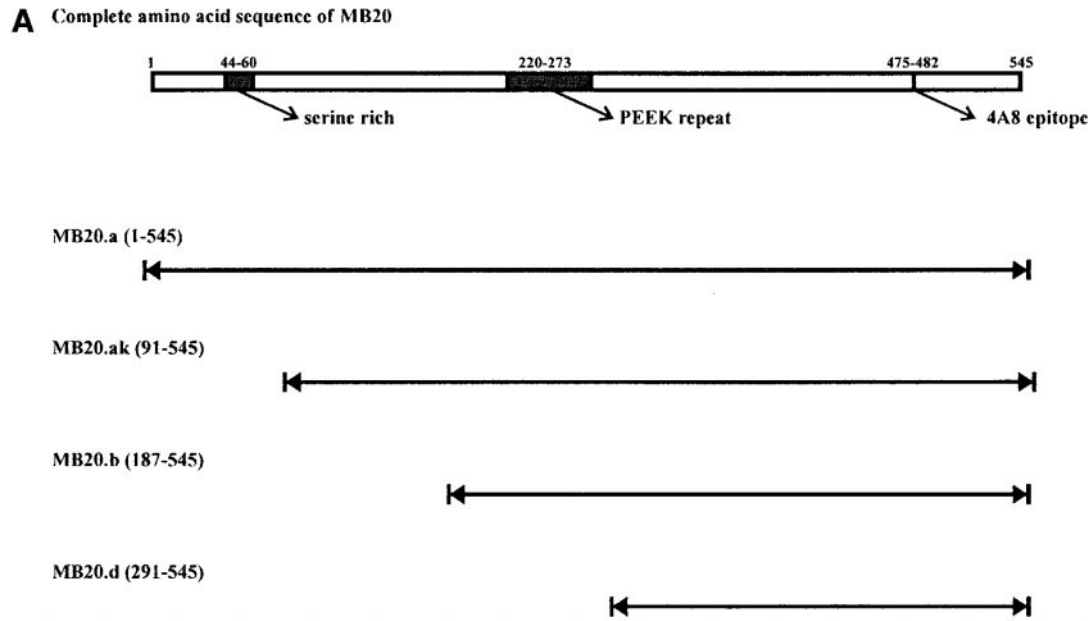

FIG. 8. Immunofluorescence staining of MB20 proteins. (**A**) Eukaryotic expression constructs of MB20. The positions of the serine-rich domain, the PEEK repeat, and the 4A8 epitope are indicated. (**B**) Immunofluoresence microscopy. The distribution of full-length MB20 protein in transfectant cells was revealed by the anti-MB20 antibody. (**a**) 100×; (**b**) 400×; (**c**) stained with preimmune sera; (**d**) stained with MB20-absorbed antisera. (**C**) Immunofluorescence microscopy. N-terminally truncated MB20 proteins detected by anti-MB20. (**a**) MB20.ak, 100×; (**b**) MB20.ak; 400×; (**c**) MB20.b; (**d**) MB20.d.

clear. The wildtype MB20 protein, when overexpressed in 3T3 cells, is distributed in the cytoplasm and/or the nucleus. Deletion of its N-terminal region restricted the MB20 protein to the nucleus. Our data indicated that some key domains are present in the N-terminus as well as the C-terminus to regulate the cellular distribution of MB20 proteins. This led us to examine further the amino acid sequence of position 91 to 186 that is commonly shared by expression constructs MB20.a and MB20.ak. These two constructs expressed proteins that are located in the cytoplasm and the nucleus. A segment of the MB20 protein at position 104–112, which has three leucine residues (see Fig. 3), was found to be highly homologous to the nuclear export signal (Gerace, 1995). On the other hand, the sequence of amino acid residues at position 334–343 of MB20 (see Fig. 3) resembles the nucleus localization signal (Dingwall and Laskey, 1991).

A recent study of the cellular localization of *Drosophila* NAP-1 reported that it occurs in the nucleus during S-phase and in the cytosol during G_2-phase (Ito *et al.*, 1996). Somewhat unexpectedly, it was also found that the *Saccharomyces cerevisiae* and *Xenopus* NAP-1 bind either directly or indirectly to B-type cyclins (Kellogg *et al.*, 1995). Furthermore, on the basis of a genetic analysis of yeast strains carrying various combinations of mutations in genes encoding NAP-1 and mitotic cyclins, it has been proposed that NAP-1 affects the function of cyclin B/p34^{cdc2} kinase complexes (Kellogg and Murray, 1995). NAP-1, therefore, might have multiple functions and participate in different processes at different phases of the cell cycle. Considering the regulatory role of the segment at amino acid position 91–186 in determining MB20 localization, it is possible that a specific trafficking mechanism might be operating to control the cellular distribution and function of MB20 proteins.

MB20 in Brain Development

In this study we confirmed the existence of a brain-specific NAP member in several mammalian species, including humans. We have previously reported the isolation of a human homologue, named HB20, from a human fetal brain cDNA library (Fan *et al.*, 1998). The expression of HB20 in the fetal brain is evidenced by an unusually high frequency of HB20-positive clones in the library. HB20 expression is abundant in the adult brain but is reduced in brain tumor tissues (Fan and Tsai, unpublished results). Using an *in vitro* culture system, we also demonstrated up-regulation of HB20 gene expression upon neuronal differentiation induced by retinoic acid and cytosine arabinoside treatment. It is noticeable that NT2 was originally isolated from a teratocarcinoma, and it represents a committed neuronal precursor that is still capable of cell proliferation (Pleasure *et al.*, 1992; Pleasure and Lee, 1993). Low-level HB20 expression was observed in the uninduced NT2 cells but the HB20 transcripts increased at least 3.5-fold when the cells acquired neuronal phenotypes and ceased dividing (Fan *et al.*, 1998). On the basis of *in vitro* and *in vivo* expression patterns, we speculated that HB20, and MB20 also, represents a modular marker in the transition from proliferating neural precursors to fully matured neural cells. The fact that NAP can complex with cyclin B in yeast and *Xenopus* agrees with the notion that MB20 could act in a similar

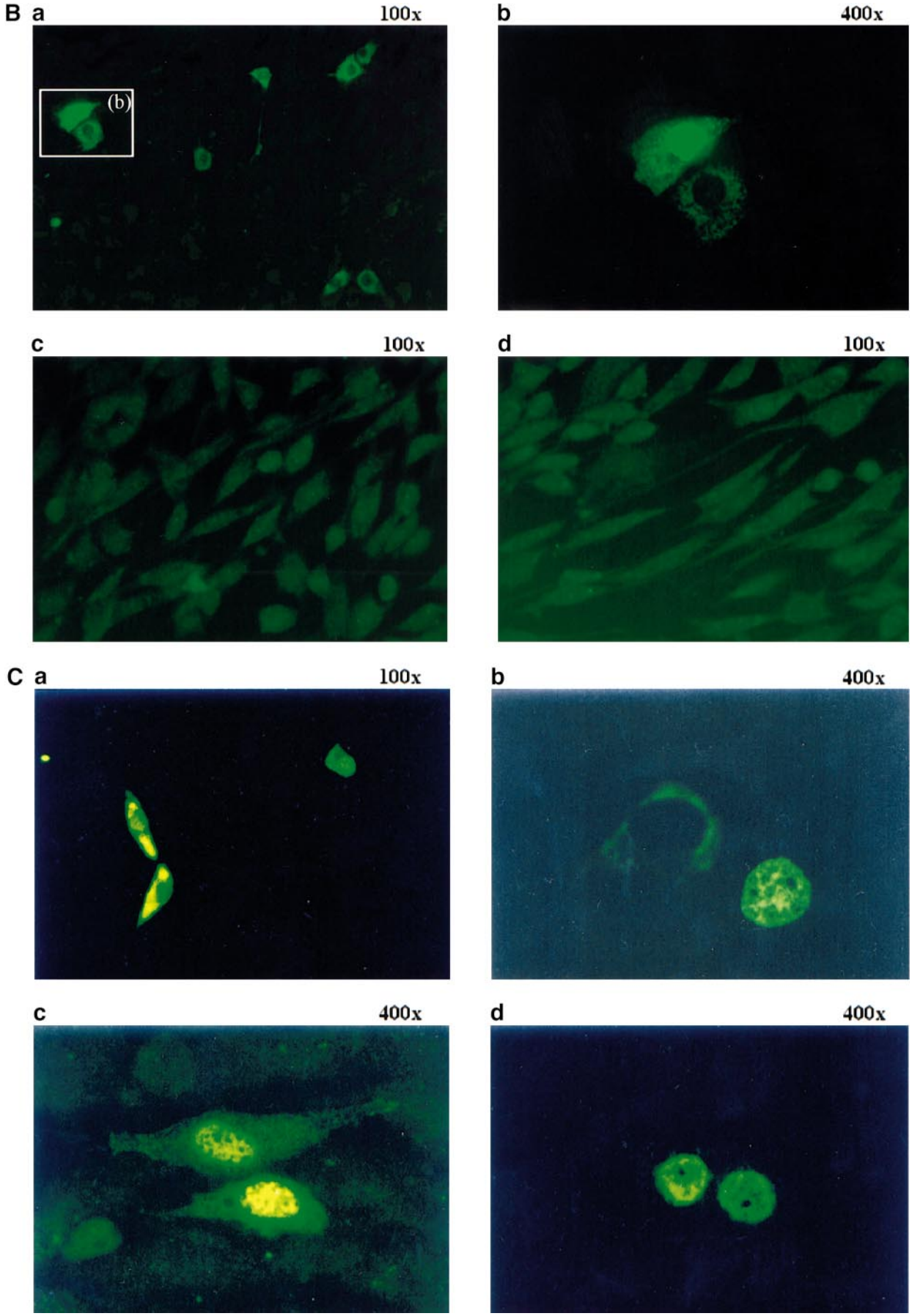

FIG. 8—*Continued*

way in regulating cell cycle progression during maturation of the nervous system. Moreover, by shuffling histone proteins and by modulating chromatin structure, MB20 could function during DNA replication or transcription activation. All these molecular events could be linked to the cellular regulation processes that lock the neuronal component into a maturing nervous system during fetal brain development.

ACKNOWLEDGMENTS

The authors thank Yun-Chia Chou for providing animal tissues, and Hsian-Guey Hsieh and Cheng-Po Hu for helping with antibody preparation. We also appreciate the assistance of Yan-Hwa Wu Lee, Ting-Fen Tsai, and Frank S. Fan in critically reviewing the manuscript. This work was supported by Grants NSC87-2314-B-010-015 and NSC88-2314-B-010-005 (to S-F. Tsai).

REFERENCES

Adachi, Y., Pavlakis, G. N., and Copeland, T. D. (1994). Identification and characterization of SET, a nuclear phosphoprotein encoded by the translocation break point in acute undifferentiated leukemia. *J. Biol. Chem.* **269:** 2258–2262.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Amato, A. A., Prior, T. W., Barohn, R. J., Snyder, P., Papp, A., and Mendell, J. R. (1993). Kennedy's disease: A clinicopathologic correlation with mutations in the androgen receptor gene [See comments]. *Neurology* **43:** 791–794.

Bou, G., Esteban, P. F., Baladron, V., Gonzalez, G. A., Cantalejo, J. G., Remacha, M., Jimenez, A., Del Rey, F., Ballesta, J. P., and Revuelta, J. L. (1993). The complete sequence of a 15,820 bp segment of *Saccharomyces cerevisiae* chromosome XI contains the UBI2 and MPL1 genes and three new open reading frames. *Yeast* **9:** 1349–1354.

Brown, N. H., and Kafatos, F. C. (1988). Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* **203:** 425–437.

Cote, J., Quinn, J., Workman, J. L., and Peterson, C. L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265:** 53–60.

Dingwall, C., and Laskey, R. A. (1991). Nuclear targeting sequences—A consensus? [See comments]. *Trends Biochem. Sci.* **16:** 478–481.

Elgin, S. C. (1988). The formation and function of DNase I hypersensitive sites in the process of gene activation. *J. Biol. Chem.* **263:** 19259–19262.

Fan, F. S., Shen, H. H., Tseng, W. P., Chen, P. M., and Tsai, S. F. (1998). Molecular cloning and characterization of a human brain-specific gene implicated in neuronal differentiation. *Brain Res. Mol. Brain Res.* **54:** 113–123.

Fujii-Nakata, T., Ishimi, Y., Okuda, A., and Kikuchi, A. (1992). Functional analysis of nucleosome assembly protein, NAP-1. The negatively charged COOH-terminal region is not necessary for the intrinsic assembly activity. *J. Biol. Chem.* **267:** 20980–20986.

Gerace, L. (1995). Nuclear export signals and the fast track to the cytoplasm. *Cell* **82:** 341–344.

Ginsberg, D., Vairo, G., Chittenden, T., Xiao, Z. X., Xu, G., Wydner, K. L., DeCaprio, J. A., Lawrence, J. B., and Livingston, D. M. (1994). E2F-4, a new member of the E2F transcription factor family, interacts with p107. *Genes Dev.* **8:** 2665–2679.

Gruss, C., and Sogo, J. M. (1992). Chromatin replication. *BioEssays* **14:** 1–8.

The Huntington's Disease Collaborative Research Group (THDCRG) (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. [See comments]. *Cell* **72:** 971–983.

Ishimi, Y., Hirosumi, J., Sato, W., Sugasawa, K., Yokota, S., Hanaoka, F., and Yamada, M. (1984). Purification and initial characterization of a protein which facilitates assembly of nucleosome-like structure from mammalian cells. *Eur. J. Biochem.* **142:** 431–439.

Ishimi, Y., and Kikuchi, A. (1991). Identification and molecular cloning of yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. *J. Biol. Chem.* **266:** 7025–7029.

Ishimi, Y., Yasuda, H., Hirosumi, J., Hanaoka, F., and Yamada, M. (1983). A protein which facilitates assembly of nucleosome-like structures in vitro in mammalian cells. *J. Biochem.* **94:** 735–744.

Ito, T., Bulger, M., Kobayashi, R., and Kadonaga, J. T. (1996). *Drosophila* NAP-1 is a core histone chaperone that functions in ATP-facilitated assembly of regularly spaced nucleosomal arrays. *Mol. Cell. Biol.* **16:** 3112–3124.

Kellogg, D. R., Kikuchi, A., Fujii-Nakata, T., Turck, C. W., and Murray, A. W. (1995). Members of the NAP/SET family of proteins interact specifically with B-type cyclins. *J. Cell. Biol.* **130:** 661–673.

Kellogg, D. R., and Murray, A. W. (1995). NAP1 acts with Clb1 to perform mitotic functions and to suppress polar bud growth in budding yeast. *J. Cell Biol.* **130:** 675–685.

Kim, U. J., Han, M., Kayne, P., and Grunstein, M. (1988). Effects of histone H4 depletion on the cell cycle and transcription of *Saccharomyces cerevisiae*. *EMBO J.* **7:** 2211–2219.

Kleinschmidt, J. A., Fortkamp, E., Krohne, G., Zentgraf, H., and Franke, W. W. (1985). Co-existence of two different types of soluble histone complexes in nuclei of *Xenopus laevis* oocytes. *J. Biol. Chem.* **260:** 1166–1176.

Kleinschmidt, J. A., and Franke, W. W. (1982). Soluble acidic complexes containing histones H3 and H4 in nuclei of *Xenopus laevis* oocytes. *Cell* **29:** 799–809.

Knappik, A., and Pluckthun, A. (1994). An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. *BioTechniques* **17:** 754–761.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A., Hayashi, T., *et al.* (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat. Genet.* **6:** 9–13.

Kornberg, R. D. (1977). Structure of chromatin. *Annu. Rev. Biochem.* **46:** 931–954.

Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E., and Green, M. R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex [See comments]. *Nature* **370:** 477–481.

Laskey, R. A., Honda, B. M., Mills, A. D., and Finch, J. T. (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275:** 416–420.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution [See comments]. *Nature* **389:** 251–260.

Nakamura, T., Alder, H., Gu, Y., Prasad, R., Canaani, O., Kamada, N., Gale, R. P., Lange, B., Crist, W. M., Nowell, P. C., *et al.* (1993). Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. *Proc. Natl. Acad. Sci. USA* **90:** 4631–4635.

Orr, H. T., Chung, M. Y., Banfi, S., Kwiatkowski, T. J., Jr., Servadio, A., Beaudet, A. L., McCall, A. E., Duvick, L. A., Ranum, L. P., and Zoghbi, H. Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **4:** 221–226.

Pleasure, S. J., and Lee, V. M. (1993). NTera 2 cells: A human cell line which displays characteristics expected of a human committed neuronal progenitor cell. *J. Neurosci. Res.* **35:** 585–602.

Pleasure, S. J., Page, C., and Lee, V. M. (1992). Pure, postmitotic, polarized human neurons derived from NTera 2 cells provide a system for expressing exogenous proteins in terminally differentiated neurons. *J. Neurosci.* **12:** 1802–1815.

Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. A., Sherman, S. L., Wilkinson, K. D., and Warren, S. T. (1992). Human genes containing polymorphic trinucleotide repeats [Published erratum appears in *Nat. Genet.* 1993, Mar;**3**(3):273]. *Nat. Genet.* **2:** 186–191.

Shneidman, P. S., Carden, M. J., Lees, J. F., and Lazzarini, R. A. (1988). The structure of the largest murine neurofilament protein (NF-H) as revealed by cDNA and genomic sequences. *Brain Res.* **464:** 217–231.

Simon, R. H., and Felsenfeld, G. (1979). A new procedure for purifying histone pairs H2A + H2B and H3 + H4 from chromatin using hydroxylapatite. *Nucleic Acids Res.* **6:** 689–696.

Smith, P. A., Jackson, V., and Chalkley, R. (1984). Two-stage maturation process for newly replicated chromatin. *Biochemistry* **23:** 1576–1581.

Stenberg, K., Oberg, B., and Chattopadhyaya, J. B. (1982). Precipitation of nucleotides by calcium phosphate. *Biochim. Biophys. Acta* **697:** 170–173.

Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189:** 113–130.

Takiyama, Y., Nishizawa, M., Tanaka, H., Kawashima, S., Sakamoto, H., Karube, Y., Shimazaki, H., Soutome, M., Endo, K., Ohta, S., *et al.* (1993). The gene for Machado–Joseph disease maps to human chromosome 14q. *Nat. Genet.* **4:** 300–304.

Tsukiyama, T., Daniel, C., Tamkun, J., and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* **83:** 1021–1026.

Tsukiyama, T., and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83:** 1011–1020.

Walffe, A. P. (1995). "Chromatin: Structure and Function," Academic Press, San Diego.

Walker, P., Germond, J. E., Brown-Luedi, M., Givel, F., and Wahli, W. (1984). Sequence homologies in the region preceding the transcription initiation site of the liver estrogen-responsive vitellogenin and apo-VLDLII genes. *Nucleic Acids Res.* **12:** 8611–8626.

Walter, P. P., Owen-Hughes, T. A., Cote, J., and Workman, J. L. (1995). Stimulation of transcription factor binding and histone displacement by nucleosome assembly protein 1 and nucleoplasmin requires disruption of the histone octamer. *Mol. Cell. Biol.* **15:** 6178–6187.

Widom, J. (1998). Structure, dynamics, and function of chromatin in vitro. *Annu. Rev. Biophys. Biomol. Struct.* **27:** 285–327.

Wolffe, A. P., and Kurumizaka, H. (1998). The nucleosome: A powerful regulator of transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **61:** 379–422.

Worcel, A., Han, S., and Wong, M. L. (1978). Assembly of newly replicated chromatin. *Cell* **15:** 969–977.