Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6186–6190, July 1991
Biochemistry

Molecular cloning of matrin F/G: A DNA binding protein of the nuclear matrix that contains putative zinc finger motifs

DAVID J. HAKES* AND RONALD BEREZNEY†

Department of Biological Sciences, State University of New York, Buffalo, NY 14260

Communicated by Sheldon Penman, April 2, 1991

ABSTRACT **We have isolated a 2.7-kilobase rat liver cDNA clone that contains the entire 544-amino acid coding sequence for matrin F/G. This protein has previously been localized to the internal, fibrogranular areas of the nuclear matrix and shown to bind to DNA on nitrocellulose blots. The predicted amino acid sequence from the coding region of this cDNA shows that this protein contains ≈50% hydrophobic amino acids with secondary structure predictions suggesting a large percentage of β-sheet regions. No significant homologies were found with any other known proteins, including the nuclear lamins. The predicted amino acid sequence was also searched for DNA binding motifs. Two putative zinc finger motifs were found. In addition, a 7-mer palindromic sequence (Ser-Ser-Thr-Asn-Thr-Ser-Ser) was discovered within one of these zinc finger DNA binding regions. A possible regulatory role for this element is discussed.**

The nuclear matrix consists of a nuclear lamina, residual nucleoli, and an internal nuclear matrix (1). It is typically prepared by nuclease, salt, and detergent treatments of isolated nuclei (for review, see ref. 2). Many molecular processes have been found associated with the nuclear matrix (2), including DNA replication (3), transcription (4), RNA splicing (5), and attachment of supercoiled DNA loops (6).

The nuclear lamins are the most characterized of the nuclear matrix proteins. Analysis of cloned sequences of lamins A and C (7, 8) and lamin B (9) reveal structural similarity to the intermediate filament proteins. In contrast, our knowledge of the rest of the nuclear matrix proteins is very limited. Appropriate two-dimensional analysis has revealed over 200 proteins in the nuclear matrix including those that are common in a variety of cell lines and those that are both cell type and differentiation state dependent (10–12).

Research in our laboratory is concentrating on the analysis of a limited number of major Coomassie blue-stained proteins that are common to nuclear matrices isolated from a variety of mammalian cells. Polyclonal antibodies generated to many of these proteins all show staining of the fibrogranular matrix by immunofluorescence microscopy (13). We have, therefore, termed these proteins the nuclear matrins to distinguish them from the nuclear lamins, which compose the nuclear lamina structure along the periphery of the nuclear matrix (H. Nakayasu and R.B., unpublished data).

Several of these major nuclear matrix proteins specifically bound DNA on Southwestern blots (40). As a step toward further characterization of the nuclear matrins and their putative role as DNA binding proteins, we have been screening λgt11 cDNA expression libraries with our polyclonal antibodies to these matrin proteins. In this manuscript, we report the isolation and sequencing of a cDNA clone for matrin protein F/G‡. The predicted amino acid sequence shows a strikingly high content of hydrophobic amino acids (≈50%) and predicted β-sheet secondary structure (≈45%) and contains two putative Cys-Cys zinc finger DNA binding motifs (14).

MATERIALS AND METHODS

Materials. pGEM-3Z, T7 and SP6 primers, deoxynucleotide and dideoxynucleotide triphosphates, Klenow fragment, and restriction enzymes were purchased from Promega; [α-^{32}P]dATP was from ICN; nitrocellulose (BA85, 0.45 μm) was from Schleicher & Schuell; Sequenase was from United States Biochemical; and alkaline phosphatase-conjugated secondary antibody was from Kirkegaard and Perry Laboratories (Gaithersburg, MD).

Preparation of Polyclonal Antibodies to Proteins F/G. Nuclear matrix proteins were separated by nonequilibrium two-dimensional polyacrylamide gel electrophoresis (15). The individual protein spot, which we have named matrin F, was cut out of numerous gels. The gel pieces were equilibrated for 1 hr with five changes in 60 mM Tris·HCl (pH 6.8), and the proteins were extracted with 60 mM Tris·HCl, pH 6.8/0.1% SDS. The protein was mixed with Freund's complete adjuvant and injected into chickens. After a booster injection, serum was collected.

Isolation and Preparation of Subcellular Fractions. Rat liver nuclei and nuclear matrices were prepared as described by Berezney and coworkers (16). The proteins from each of the fractions recovered during the isolation were ethanol precipitated and resuspended in electrophoresis sample buffer [60 mM Tris·HCl, pH 6.8/2% SDS/20% (vol/vol) glycerol/5% 2-mercaptoethanol]. One-dimensional minigel electrophoresis of the various fractions was performed based on the method of Laemmli (17), and the proteins were electrophoretically transferred to nitrocellulose paper according to Towbin *et al.* (18).

Western Blot Analysis. Nonspecific binding sites on nitrocellulose blots containing the fractions described above were blocked by overnight incubation of the blot at 4°C in block buffer (10 mM Tris·HCl, pH 7.4/150 mM NaCl/0.5% Tween 20). The blot was incubated with the matrin F polyclonal antibody (1:20 dilution of serum) for 2 hr, washed four times for 15 min each in block buffer, incubated for 1 hr in alkaline phosphatase-conjugated goat anti-chicken IgG secondary antibodies (1:1000 dilution), washed four times for 15 min each in block buffer, and developed according to ref. 19.

Isolation of a cDNA Using Matrin F Polyclonal Antibodies. A bacteriophage λgt11 rat liver cDNA library was prepared (20) and screened with our polyclonal antibody to matrin F essentially as described above. Positive plaques were isolated. Phage DNA was lysogenized, and the production of β-galactosidase fusion protein was induced with isopropyl β-D-thiogalactopyranoside (IPTG). The cells were lysed and the total lysate was run on SDS/7.5% polyacrylamide gels

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance *with 18* U.S.C. §1734 solely to indicate this fact.

Abbreviation: IPTG, isopropyl β-D-thiogalactopyranoside.

*Present address: Department of Biochemistry, Purdue University, West Lafayette, IN 47907.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M64862).

and transferred to nitrocellulose. The cDNA insert was isolated from the phage whose fusion protein showed positive reaction with the matrin F antibody.

Northern Blot Analysis. Total RNA was isolated from rat liver according to the procedure of Glisin *et al.* (21). Ten micrograms of total RNA was denatured with glyoxal and subjected to electrophoresis through a 1% agarose gel in 10 mM NaH_2PO_4 (22). Denatured cDNA insert labeled by nick-translation (5×10^5 cpm/ml; ref. 23) was then hybridized to the blots (24).

Rescreening of cDNA Library. A larger cDNA fragment was isolated by rescreening 600,000 colonies from the same λgt11 library with labeled cDNA obtained from the primary screening (25).

Subcloning and Sequencing of Matrin F/G cDNA Insert. Various restriction fragments of the original cDNAs were ligated into the pGEM-3Z plasmid (Promega) and sequenced according to the dideoxynucleotide procedure of Sanger *et al.* (26). The sequence of each fragment was determined from both strands using the T7 and SP6 primers provided by Promega. The sequence of overlapping fragments was also obtained to verify the sequence alignment.

RESULTS

Subcellular Localization of Matrin F and G. Various subcellular fractions were subjected to electrophoresis on a SDS/10% polyacrylamide gel and transferred to nitrocellulose. Polyclonal antibodies raised against matrin F, one of the major DNA binding proteins of the nuclear matrix (40), were used to probe the nitrocellulose blot containing the various subcellular fractions. The antibody specifically reacts with two protein bands that are highly concentrated in the nuclear matrix fraction (Fig. 1*A*). The two proteins have apparent molecular weights of 75,000 and 65,000 and have been identified as matrins F and G on two-dimensional blots (H. Nakayasu and R.B., unpublished data). No significant reaction was visible with any of the other cytoplasmic or subnuclear fractions except total nuclei, which showed relatively weak staining of the same proteins (Fig. 1*A*).

Screening of λgt11 Library with Matrin F Antibody Probe. The matrin F antibodies, which show specificity for matrin proteins F and G, were used to probe a λgt11 cDNA library. Fusion proteins consisting of the native β-galactosidase and cDNA insert coding regions were induced by using IPTG. Initial screening of 300,000 colonies with the antibody resulted in one positive clone. The phage was isolated and lysogenized, with the lysogens induced to produce fusion protein with IPTG. The total cell lysates were electrophoresed on SDS/7.5% polyacrylamide gels and transferred to nitrocellulose. The antibodies reacted with a protein of M_r ≈140,000 (Fig. 1*B*), which is only found in the induced bacteria. In addition, the antibody did not react with any other proteins in either the induced or uninduced bacteria, indicating that the positive antibody reaction was specific for the fusion protein. As the β-galactosidase protein has a M_r of ≈116,000, the molecular weight portion coded for by the cDNA insert corresponds to ≈24,000.

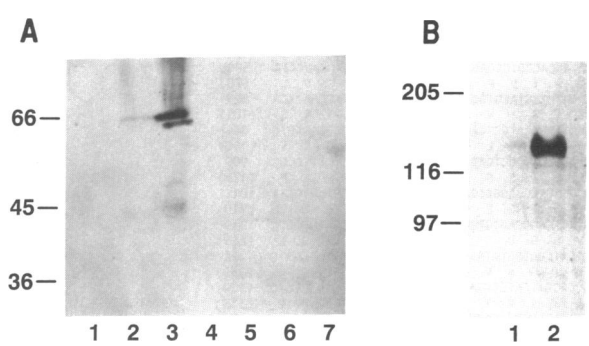


FIG. 1. Specificity of matrin F polyclonal antibodies for matrins F and G and an isolated fusion protein. (*A*) Polyclonal antibodies raised against matrin F were used to probe a nitrocellulose blot containing subcellular fractions of rat liver tissue. Lanes: 1, whole cell homogenate; 2, nuclei; 3, nuclear matrix; 4, cytosol; 5, mitochondria; 6, high-salt nuclear wash; 7, Triton X-100 nuclear wash. (*B*) The same antibodies were used to probe a nitrocellulose blot containing total bacterial proteins. The bacteria were lysogenized with a cDNA identified as a positive clone by immunological screening of a λgt11 library. Lanes: 1, bacterial proteins from uninduced cells; 2, bacterial proteins from cells that were induced with IPTG to produce a fusion protein. This protein consists of β-galactosidase and the coding region of the isolated cDNA. The antibodies specifically recognize a protein of approximate M_r 140,000 present only in the induced bacteria.

Northern Analysis. The cDNA insert isolated from the lysogenized phage had an approximate size of 600 base pairs on agarose gel electrophoresis (data not shown). This is in general agreement with the size necessary to code for a peptide of M_r ≈24,000. Northern blot analysis was then performed to determine the size and specificity of the messenger RNA(s) corresponding to this cDNA insert. A ^{32}P-labeled insert was hybridized to a blot containing 10 μg of total rat liver RNA. As shown in Fig. 2, the insert hybridizes to a band whose approximate size is 3.0 kilobases (kb).

Sequencing of Matrin F/G. A ^{32}P-labeled cDNA insert was used as a probe to rescreen the λgt11 library. A total of four distinct clones were isolated, with the largest insert being ≈2.7 kb (data not shown). It was determined by restriction mapping that the three other clones represented a sequence internal to the 2.7-kb clone. The restriction map of this 2.7-kb insert, along with the sequencing strategy, is presented in Fig. 3*A*. Only those enzymes whose restriction sites are found in the multiple cloning site of the pGEM-3Z sequencing vector are indicated. A large number of overlapping restriction enzyme fragments were subcloned into the pGEM-3Z plasmid and sequenced from both strands.

The sequence of the 2.7-kb insert is shown in Fig. 3*B*. There is one large open reading frame that encodes a 544-amino acid polypeptide with a calculated M_r of 59,724 and a pI of 10.19. A consensus polyadenylylation addition site (27) has not been found in this clone, indicating that we do not have the full-length 3′ untranslated region. A search of various data bases demonstrated no significant similarities to any known sequences (28).

Analysis of Matrin F/G Sequence. The protein sequence predicted from our DNA sequence was run through the sequence analysis program provided by International Biotechnologies (29), which predicts the hydrophilicity/hydrophobicity according to the axiom of Kyte and Doolittle

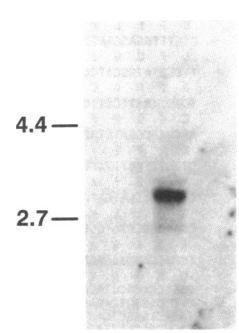


FIG. 2. Northern blot analysis. Ten micrograms of total RNA was treated with glyoxal, subjected to electrophoresis through a 1% agarose gel, and blotted onto nitrocellulose by capillary action. Hybridization was carried out with the 2.7-kb cDNA fragment. The position of RNA standards (in kb) is shown.

(30). As shown in Fig. 4, the protein has several large hydrophobic domains, which are punctuated by short hydrophilic domains. Overall, ≈50% of the total amino acids have hydrophobic side chains. A secondary structure prediction based on the Chou–Fasman algorithm (31) indicates that the protein has the potential to form ≈45% β-sheet, 25% α-helix, and 20% β-turns (Fig. 5).

Identification of a Putative DNA Binding Domain. Since matrins F and G were both identified on nitrocellulose blots as DNA binding proteins, we searched the predicted protein sequence for known DNA binding motifs (14, 32). We discovered two overlapping putative zinc finger domains that have similarity to the Cys-Cys type zinc finger motif (Fig. 5). This domain is located in a region of the protein that does not have a large hydrophobic character, suggesting a potential external location. In addition, there is an intriguing 7-mer palindromic sequence (Ser-Ser-Thr-Asn-Thr-Ser-Ser) located in one of the putative zinc finger domains. It contains a potential glycosylation site, flanked on either side by residues that may serve as phosphorylation sites (Fig. 5).

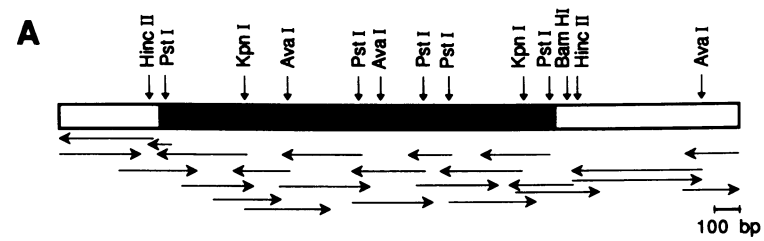
Computer analysis indicates that the second threonine in this palindromic sequence (Thr-380) is a predicted phosphate acceptor site for casein kinase II. The putative zinc finger region also contains another predicted casein kinase II phosphorylation site at Ser-350 that is flanked by two cysteine residues.

DISCUSSION

In this investigation, we report the analysis of the cDNA sequence for matrin F/G. This cDNA sequence includes the entire coding region for a protein with a calculated M_r of 59,724 and a pI of 10.19. These values are in reasonable agreement with the apparent M_rs of matrin F (≈75,000) and matrin G (≈65,000) on SDS/polyacrylamide gels (Fig. 1*A*) as well as their extremely basic character. Further studies are needed to determine whether the cDNA codes for one or both of these two proteins. Similarities in two-dimensional peptide maps and the finding that polyclonal antibodies raised against highly purified matrin F specifically recognize both matrins F and G and vice versa suggests a close relationship, if not identity, between these two proteins (H. Nakayasu and R.B., unpublished data).

The nuclear lamins are currently the most studied of the nuclear matrix proteins. The IFA monoclonal antibody (33), which reacts with all known intermediate filament proteins, has been shown to react with nuclear lamins A, B, and C (34). In addition, cDNA analysis of human lamins A and C (7, 8), as well as *Xenopus* lamin B (9), has demonstrated sequence similarity between the lamin proteins and intermediate filament proteins. Since data base searches using our cDNA sequence and the predicted amino acid sequence revealed no significant homology to any known DNA or protein sequence, this internal matrix protein is structurally distinct from the nuclear lamins and intermediate filament proteins. Further support for this conclusion is provided by the failure of the IFA antibody to stain internal regions of the nucleus (33) and to react with any of the nuclear matrins on Western blots (35).

Studies on the predicted amino acid sequence of the matrin F/G cDNA indicates ≈50% hydrophobic amino acids (Fig. 3). These residues are organized into large hydrophobic domains separated by very short hydrophilic domains (Fig.



B

```
                                                                                CCGGAAGCCCGAAGCACCGGAGTCCCGCAGAACCTGACTCCGGCCTGTCACCACCACCAAAGGCTAGGG    69
GACGTCGCCTCGGTCACTATGGGGCTCCTGCTCAAGCCTGGAGCGCGCCAGGGCAGCGGCACCTCCTCGGTCCCAGACAGACGTTGTCCCCGCTCCGTCTTCAGC   174
AACATTAAGGTATTTGTTCTTTGCCATGGCCTGCTACAGCTCTGCCAGCTGCTCTACAGCGCCTACTTCAAGAGCAGTCTCACCACAATCGAGAAGCGCTTTGGG   279
CTCTCCAGCTCTTCCTCTGGTCTCATCTCCAGTTTGAATGAGATCAGCAACGCTACCCTCATCATCTTCATTAGCTACTTCGGCAGCCGGGTCAACCGCCCACGG   384
ATGATTGGCATAGGGGGTCTCCTCCTGGCTGCAGGGGCCTTTGTCCTCACCCTCCCACACTTCCTGTCAGAGCCCTATCAATACACCTCGACCACGGATGGAAAC   489
 M  I  G  I  G  G  L  L  L  A  A  G  A  F  V  L  T  L  P  H  F  L  S  E  P  Y  Q  Y  T  S  T  T  D  G  N   (35)
AGGAGCAGCTTTCAGACTGACCTCTGTCAGAAGCATTTCGGAGCCCTGCCCCCCAGTAAGTGCCATAGCACCGTGCCAGATACCCACAAGGAGACCAGCAGCCTG   594
 R  S  S  F  Q  T  D  L  C  Q  K  H  F  G  A  L  P  P  S  K  C  H  S  T  V  P  D  T  H  K  E  T  S  S  L   (70)
TGGGGCCTGATGGTGGTTGCTCAACTACTGGCCGGCATTGGGACAGTGCCCATCCAGCCCTTTGGGATCTCCTACGTGGATGACTTTGCCGAGCCTACCAACTCA   699
 W  G  L  M  V  V  A  Q  L  L  A  G  I  G  T  V  P  I  Q  P  F  G  I  S  Y  V  D  D  F  A  E  P  T  N  S   (105)
CCTCTGTATATCTCCATCCTATTCGCCATCGCTGTGTTCGGACCGGCTTTCGGGTACCTGCTGGGCTCAGTCATGCTGAGAATCTTCGTGGACTACGGCAGAGTG   804
 P  L  Y  I  S  I  L  F  A  I  A  V  F  G  P  A  F  G  Y  L  L  G  S  V  M  L  R  I  F  V  D  Y  G  R  V   (140)
GACACTGCTACCGTAAACCTGAGCCCAGGTGACCCTCGGTGGATTGGAGCCTGGTGGCTGGGCCTGCTCATCTCCTCAGGCTTCTTGATTGTCACCTCTTTGCCC   909
 D  T  A  T  V  N  L  S  P  G  D  P  R  W  I  G  A  W  W  L  G  L  L  I  S  S  G  F  L  I  V  T  S  L  P   (175)
TTCTTTTTCTTTCCCCGAGCAATGTCCAGAGGAGCAGAGAGGTCTGTTACCGCAGAGGAAACAATGCAGACGGAGGAGGACAAGTCAAGAGGCTCCCTGATGGAT  1014
 F  F  F  F  P  R  A  M  S  R  G  A  E  R  S  V  T  A  E  E  T  M  Q  T  E  E  D  K  S  R  G  S  L  M  D   (210)
TTCATTAAACGGTTCCCCCGCATCTTCCTGAGGCTGCTGATGAACCCGCTCTTCATGCTGGTGGTCCTGAGCCAGTGTACCTTCTCCTCAGTCATCGCTGGCCTC  1119
 F  I  K  R  F  P  R  I  F  L  R  L  L  M  N  P  L  F  M  L  V  V  L  S  Q  C  T  F  S  S  V  I  A  G  L   (245)
TCCACGTTCCTCAACAAGTTCCTGGAGAAGCAGTATGGAGCCACGGCAGCCTATGCCAACTTCCTCATCGGTGCTGTAAATCTTCCGGCTGCAGCCTTGGGGATG  1224
 S  T  F  L  N  K  F  L  E  K  Q  Y  G  A  T  A  A  Y  A  N  F  L  I  G  A  V  N  L  P  A  A  A  L  G  M   (280)
CTGTTTGGAGGAATCCTCATGAAGCGTTTTGTTTTCCCTCTGCAAACTATCCCCCGAGTGGCTGCCACCATCATCACCATCTCCATGATCCTCTGTGTACCTCTC  1329
 L  F  G  G  I  L  M  K  R  F  V  F  P  L  Q  T  I  P  R  V  A  A  T  I  I  T  I  S  M  I  L  C  V  P  L   (315)
TTCTTTATGGGATGCTCCACATCAGCCGTGGCTGAGGTCTACCCTCCCAGCACATCAAGTTCTATACATCCGCAGCAGCCTCCTGCCTGCCGCAGGGACTGCTCG  1434
 F  F  M  G  C  S  T  S  A  V  A  E  V  Y  P  P  S  T  S  S  I  H  P  Q  Q  P  P  A  C  R  R  D  C  S   (350)
TGCCCAGATTCCTTCTTCCACCCAGTCTGTGGAGACAATGGAGTCGAGTACGTTTCCCCTTGCCACGCCGGCTGCAGCAGCACCAACACAAGCTCAGAAGCTTCT  1539
 C  P  D  S  F  F  H  P  V  C  G  D  N  G  V  E  Y  V  S  P  C  H  A  G  C  S  S  T  N  T  S  S  E  A  S   (385)
AAGGAACCGATCTACTTGAACTGCAGCTGTGTGAGTGGAGGATCGGCGTCACAAGACAGGCTCATGCCCCACGTCCTGCGCGCACTACTGCTCCCGTCCATCTTC  1644
 K  E  P  I  Y  L  N  C  S  C  V  S  G  G  S  A  S  Q  D  R  L  M  P  H  V  L  R  A  L  L  L  P  S  I  F   (420)
CTCATTTCCTTTGCGGCGCTCATTGCCTGCATCTCCCACAACCCGCTCTACATGATGGTCCTTCGCGTGGTGAACCAGGATGAAAAGTCGTTTGCCATTGGGGTA  1749
 L  I  S  F  A  A  L  I  A  C  I  S  H  N  P  L  Y  M  M  V  L  R  V  V  N  Q  D  E  K  S  F  A  I  G  V   (455)
CAGTTCTTGTTGATGCGCTTGCTGGCCTGGCTGCCGGCTCCATCCCTTTATGGCCTCCTCATCGACTCCTCCTGTGTCCGGTGGAACTACCTATGCTCAGGGAGA  1854
 Q  F  L  L  M  R  L  L  A  W  L  P  A  P  S  L  Y  G  L  L  I  D  S  S  C  V  R  W  N  Y  L  C  S  G  R   (490)
CGAGGGGCCTGTGCGTATTATGACAACGATGCTCTCCGAAACAGGTACCTGGGCCTACAGATGGTCTACAAGGCCTTGGGCACACTGCTGCTCTTCTTCATCAGC  1959
 R  G  A  C  A  Y  Y  D  N  D  A  L  R  N  R  Y  L  G  L  Q  M  V  Y  K  A  L  G  T  L  L  L  F  F  I  S   (525)
TGGAGGATGAAGAAGAACAGGGAATACAGCCTGCAGGAGAACACCTCAGGCCTCATCTGACCCTCAGCTGGGACTACTGCCCCACCCCAGAGACTGGATCCTATC  2064
 W  R  M  K  K  N  R  E  Y  S  L  Q  E  N  T  S  G  L  I  U                                                   (544)
CCTTCCACACCTACCTGTATTAACTAATGTCAACATGCCTTCCTCCTCCTCTTCCTCCTCCTCCTTCTTCTTCCTCTCTCTCTCTCTCTCTCTCTCT  2169
CTCTCTCTCTCACACACACACACACACACACACATGAGAGAGAGTTCACTCACCCTTTGAGATCACCTGCCTTTTCTCTTCTGCCTAAAGTCTTAAGGCCTGAAG  2274
TACACTGAGCTGAATGAGCACCGGGCCTGAGAGTTTAGTTTCTCCAAGTCCTTGGAAGGTATCCCCAGCGTAGGCCCTACGTCCTCCAGACAAGATGCCCCATAAT  2379
GAGGCGGCCTCTGTTTTCACCAGTGTCTCAGGAATACTTAATGGAGTGAAAAGAGGGAGTCTTGCCTTCTTGGGCCAGGCAGCCCGGATCTCCTCTGCCTCTGCC  2484
CACACCCAGGAGAGCCAGAGGAGAAGCAGGTAGTTGGTTTCTTATCTGCTCCAGCGGGGCTAAGGGAGCTGGGTGTGTCCACTTTTCATCTGGATTCCGTCTAGC  2589
ATGAAAGCCGTGCCCTCGAGGCTGTTTTGGAAACCACCATTTTGGGAAGTATCCCTCTCTATAAACTATGCCCCGGTATCTGAGGAGGAATGAAGGGAGGAACAA  2694
GGCTGGATCATGGAAAACTGTTCACAGGAACCAGAGGCCTATCCTCCCGTCGGG                                                        2748
```

FIG. 3. (*A*) Restriction map of the 2.7-kb matrin F/G cDNA fragment. The solid area indicates the coding sequence. The sequencing strategy is indicated below the map. (*B*) Sequence of the 2.7-kb matrin F/G cDNA. Sequence determination was performed by using the Sanger (26) dideoxynucleotide chain-termination method. Numbering of the amino acids begins at the predicted initiator methionine. bp, Base pairs.

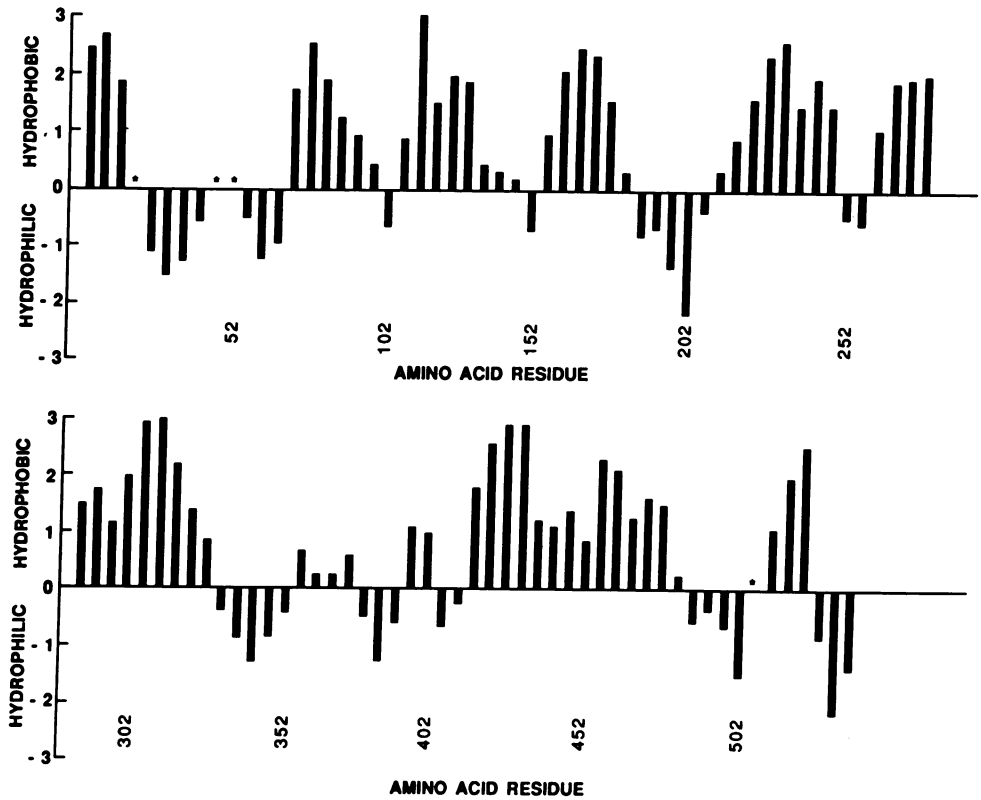


FIG. 4. Hydropathy plot of matrin F/G. The predicted amino acid sequence was run through a computer program [supplied by International Biotechnologies (29)], which predicts hydrophobic and hydrophilic domains based on the algorithm of Kyte and Doolittle (30). A window of 9 amino acid residues was used. Asterisks denote zero values.

4). The large hydrophobic domains may be responsible for the insoluble nature of this protein and may contribute to the overall insoluble nature of the isolated nuclear matrix.

Previous studies have shown that matrins F and G specifically bind to DNA on Southwestern blots (40). Analysis of the predicted amino acid sequence for matrin F/G revealed two putative Cys-Cys zinc finger DNA binding motifs (14, 32). These regions overlap and are located at amino acids 345–396. This is one of the few large regions in the protein that displays a neutral or somewhat hydrophilic character (Fig. 4). It is also flanked by two very large hydrophobic domains. The finger region could extend from the protein with the two hydrophobic flanking regions internally buried and allowing the putative zinc finger(s) to interact with DNA.

Most of the reported Cys-Cys zinc finger domains have two amino acids between each pair of external cysteines (Cys-Xaa_2-Cys-Xaa_n-Cys-Xaa_2-Cys), while the potential zinc finger domains in matrin F/G have three (Cys-Xaa_3-Cys-Xaa_{21}-Cys-Xaa_3-Cys) and one (Cys-Xaa_3-Cys-Xaa_{17}-Cys-Xaa-Cys) amino acid(s). One of the zinc finger domains in the human estrogen receptor, however, contains three and five amino acids between its external cysteines (Cys-Xaa_5-Cys-Xaa_{12}-Cys-Xaa_3-Cys) (36), suggesting that a spacing of two amino acid residues is not absolutely necessary for formation of the finger structure.

A zinc finger domain has been shown to be involved in the specific interaction between the estrogen and glucocorticoid receptors and their target DNA (37). The lack of basic amino

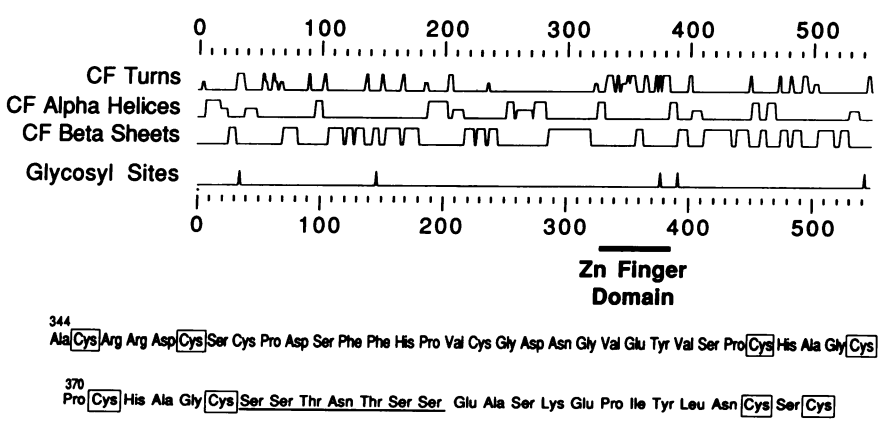


FIG. 5. Secondary structure prediction and identification of a putative zinc finger domain. The predicted amino acid sequence was run through a computer program that predicts secondary structure based on the algorithm of Chou–Fasman (31). The predicted regions of α-helices, β-turns, and β-sheets are indicated. The region containing the putative zinc finger domains is indicated at the bottom of the plot, and the amino acid sequences of these two domains are shown. The cysteine residues that would occupy coordinate positions in the zinc fingers are boxed. A 7-mer palindromic sequence containing possible phosphorylation sites flanking a central potential glycosylation site is underlined.

acids in this region is consistent with the possibility that nonionic interactions are involved in the binding of this protein to DNA. The putative zinc finger motifs in matrin F/G have a similar lack of basic amino acids (Fig. 5). Since nonionic interactions between DNA and proteins are likely to be high-salt resistant, this protein may be involved in determining the high-salt-resistant characteristic of nuclear matrix–DNA associations. In this regard, we have recently demonstrated salt-resistant binding of exogenous DNA to whole nuclear matrices and to individual matrix proteins (including matrins F and G) on Southwestern blots (40).

Within one zinc finger domain, and flanking the other, there is a palindromic sequence of seven amino acids (Ser-Ser-Thr-Asn-Thr-Ser-Ser; see Fig. 5). Computer searches have failed to identify this sequence in any other known protein. This sequence contains a potential glycosylation site surrounded on each side by potential phosphorylation sites. Since nuclear matrix proteins have been shown to be both phosphorylated (38) and glycosylated (39), it is interesting to speculate that this sequence may serve as a regulatory site within the zinc finger domain(s). In this regard, we have confirmed that Thr-380 in the palindromic sequence is a predicted casein kinase II phosphorylation site as is Ser-350 in another region of the zinc finger domain.

We wish to thank Dr. Michael Mueckler for supplying the λgt11 cDNA library, Dr. Hiroshi Nakayasu and Linda A. Buchholtz for preparation of matrin F polyclonal antibodies, Adam Bell for critical reading and help with preparation of the manuscript, and Jim Stamos for the photographs and illustrations. This work was supported by National Institutes of Health Grant GM-23922 (awarded to R.B.).

1. Berezney, R. & Coffey, D. S. (1974) *Biochem. Biophys. Res. Commun.* **60,** 1410–1419.
2. Berezney, R. (1984) *Chromosomal Nonhistone Proteins*, ed. Hnilica, L. S. (CRC, Boca Raton, FL), Vol. 4, pp. 119–180.
3. Berezney, R. & Coffey, D. S. (1975) *Science* **189,** 291–293.
4. Jackson, D. A., McCready, S. J. & Cook, P. R. (1981) *Nature (London)* **292,** 552–555.
5. Zeitlin, S., Parent, A., Silverstein, S. & Efstratiadis, A. (1987) *Mol. Cell. Biol.* **7,** 111–120.
6. Volgelstein, B., Pardoll, D. M. & Coffey, D. S. (1980) *Cell* **22,** 79–85.
7. Fisher, D. Z., Chaudhary, N. & Blobel, G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6450–6454.
8. McKeon, F. D., Kirschner, M. W. & Caput, D. (1986) *Nature (London)* **319,** 463–468.
9. Krohne, G., Wolin, S. L., McKeon, F. D., Franke, W. W. & Kirschner, M. W. (1987) *EMBO J.* **6,** 3801–3808.
10. Fey, E. G. & Penman, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 121–125.
11. Dworetzky, S. I., Fey, E. G., Penman, S., Lian, J. B., Stein, J. L. & Stein, G. S. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4605–4609.
12. Stuurman, N., Meijne, A. M. L., van der Driel, R. & van Renswoude, J. (1990) *J. Biol. Chem.* **265,** 5460–5465.
13. Berezney, R., Nakayasu, H., Hakes, D. J. & Belgrader, P. (1989) *J. Cell. Biochem. Suppl.* **13E,** 256.
14. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52,** 1–3.
15. O'Farrell, P. Z., Goodman, H. M. & O'Farrell, P. H. (1977) *Cell* **12,** 1133–1142.
16. Basler, J. B., Hastie, N. D., Pietris, D., Matsui, S., Sandberg, A. & Berezney, R. (1981) *Biochemistry* **20,** 6921–6929.
17. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
18. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4350–4354.
19. Blake, M. S., Johnston, K. H., Russell-Johns, G. J. & Gotschich, E. C. (1984) *Anal. Biochem.* **136,** 173–179.
20. Mueckler, M. & Pitot, H. C. (1985) *J. Biol. Chem.* **260,** 12993–12997.
21. Glisin, V., Crkvenjakov, R. & Byus, C. (1974) *Biochemistry* **13,** 2633–2637.
22. McMaster, G. K. & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4835–4840.
23. Rigby, P., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed., p. 7.52.
25. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed., pp. 8.46–8.49.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5453–5467.
27. Proudfoot, N. J. & Brownlee, G. G. (1974) *Nature (London)* **252,** 359–362.
28. Lipman, D. J. & Pearson, W. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
29. Pustell, J. & Kafatos, F. C. (1982) *Nucleic Acids Res.* **10,** 51–59.
30. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
31. Modrow, S. M. & Wolf, H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5703–5707.
32. Struhl, K. (1989) *Trends Biochem. Sci.* **14,** 137–140.
33. Pruss, R. M., Mirsky, R. & Raff, M. C. (1981) *Cell* **27,** 419–428.
34. Osborn, M. & Weber, K. (1987) *Exp. Cell Res.* **170,** 195–203.
35. Belgrader, P., Siegel, A. J. & Berezney, R. (1991) *J. Cell Sci.* **98,** 281–291.
36. Green, S. & Chambon, P. (1987) *Nature (London)* **325,** 75–78.
37. Green, S., Kumar, V., Theulaz, I., Wahli, W. & Chambon, P. (1988) *EMBO J.* **7,** 3037–3044.
38. Allen, S. L., Berezney, R. & Coffey, D. S. (1977) *Biochem. Biophys. Res. Commun.* **75,** 111–116.
39. Hart, G. W., Haltiwanger, R. S., Holt, G. D. & Kelly, W. G. (1989) *Annu. Rev. Biochem.* **58,** 841–874.
40. Hakes, D. J. & Berezney, R. (1991) *J. Biol. Chem.* **266,** in press.