A Large DNA-binding Nuclear Protein with RNA Recognition Motif and Serine/Arginine-rich Domain*

(Received for publication, February 6, 1996, and in revised form, March 11, 1996)

Hidetoshi Inagaki†, Yuichi Matsumiša†, Kazuyasu Nakamura‡, Mikiko Ohshima‡, Tatsuhiko Kadowaki†, and Yasuo Kitagawa†

From the †Graduate Program of Biochemical Regulation, School of Agricultural Sciences, Nagoya University and Nagoya University BioScience Center, Nagoya 464-01, Japan

The nucleus is a highly organized structure, permeated by a proteinous “matrix” and including several subcompartments. Lamin are concentrated near the periphery of the nucleoplasm. By coiled-coil self-association (1–3), lamins A, B, and C can form a meshwork throughout the nucleoplasm. Lamin B interacts with chromatin (4) at clusters of adenosine- or thymidine-rich sequences called “matrix-attachment regions” (MARs) (5). The DNA of chromatin is organized into con-...
system model 370A of Applied Biosystems. The 5' terminal sequence was determined by the rapid amplification of cDNA ends (RACE) method (39) using three nested 30-mer primers complementary to the known 5'-end sequence of human NP220 and two anchor primers of 5'-GGAATTCTCGAGTCGACATCG with and without A(T)17-3'.

Definition of the Domain Essential for DNA Binding—A series of partial human NP220 polypeptides was prepared by subcloning the inserted sequence of clone K1 into pKK223-3, digesting it with PstI or HindIII, and ligating pk1, pk1-H, and pk1-P in Fig. 1. A Southwestern blot of the polypeptides with the fragment of mitochondrial promoter region was performed mainly as described (40). Hybridization was carried out for 1 h in TED buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 1 mM dithiothreitol, and the filter was washed twice for 30 min with TED buffer containing 50 mM NaCl.

Sequence Selectivity of Human NP220 for DNA Binding—The preferential sequence of human NP220 for DNA binding was studied by the modified selected and amplified binding sequence (SAAB) method (41). For this, the product of pk-1 (Fig. 1) was separated by SDS-polyacrylamide gel electrophoresis (42), renatured, and transferred to a nitrocellulose filter. The corresponding part of the filter was treated with 5%, 10%, and 20% (v/v) dI/dC/poly(dI/dC) with an 11-mer sequence (5'-TTGCTCACTCGAGA-3') corresponding to the 5' terminal sequence

RESULTS AND DISCUSSION

Cloning of Human NP220 cDNA and Its Sequence—In an attempt to isolate cDNA clone(s) encoding mitochondrial promoter binding protein(s), we screened a HeLa cell cDNA library constructed in λgt11 with a dsDNA fragment encompassing two human mitochondrial promoters (HSP and LSP; see Ref. 48) and obtained clone K1 (Fig. 1). Since Northern hybridization of poly(2)-RNA from HeLa cells, human epidermoid carcinoma A431 cells, and human hepatoma Hep G2 cells showed a band of ~6 kilobases, we continued the screening of HeLa and Namalwa cell cDNA libraries in λZAPII with clone K1 and obtained clones M5 and N9, respectively (Fig. 1). By further screening of a human heart cDNA library, clones HK1 and HK2 were isolated (Fig. 1). This series of clones covered a region of 0.1 mM sodium tetrathionate, 1.0 mM phenylmethylsulfonyl fluoride, and 15 kallikrein-inactivating unit/ml aprotinin. The fractions were separated by SDS electrophoresis (42) using 4% (for NP220) or 4–20% (for lamin B1) (w/v) acrylamide gel. As a control of proteolytic degradation of NP220 during the fractionation, packed HeLa cells were directly lysed in the SDS sample buffer and separated in parallel. Western blotting was performed as described (47) using anti-human NP220 antibody or monoclonal antibody against lamin B1 (Matrech Inc.).

Fig. 1. Domain organization of human NP220 and its cDNA clones. A, R5, a domain rich in arginine and serine; MH1, MH2, and MH3, domains homologous to rat matrix 3; PstI-HindIII, a domain essential for DNA binding; Acridic repeat, a domain where sequences with a consensus of LVTDEIILEEDL repeat nine times. AA, amino acids. B, regions covered by a series of λgt11 or λZAPII cDNA clones originally isolated. RACE, 5'-terminal sequence cloned by the rapid amplification of cDNA ends method. Numbers in parentheses are nucleotide numbers in Fig. 2. C, expression plasmids obtained by subcloning the original cDNA clones in B. pk1, pk1-H, and pk1-P are in pk2K223-3 plasmid and used for the study of DNA binding activity (Fig. 5). pk1-BD is in pGEX-3X plasmid and used for the antibody preparation.
signal of AATAAA at 6473 nt and 73 bp of poly(A) tail starting at 6499 nt (Fig. 2). At the 5\' terminus, there are several ATG codons preceded by sequences similar to the ribosome recognition sequence (49). We assume that the ATG codon at 316 nt is the initiation codon since the open reading frames starting at upstream ATGs are very short. The open reading frame encodes a sequence of 1978 amino acids with a calculated molecular mass of 220,617. Since many lines of evidence showed this is a nuclear protein localized in the interchromatin space of various human cell lines, we refer to it as NP220. Human NP220 has unusually high contents of glutamic acid (10.4%), lysine (9.3%), and serine (11.0%). The contents of acidic (Glu and Asp) and basic (Lys, Arg, and His) amino acids are both \( \frac{16}{10} \)%; they tend to form clusters. NP220 is a hydrophilic protein without any large hydrophobic domain.

**Domain Structure of Human NP220—**

Human NP220 has three types of internally repeated amino acid sequences (Figs. 1 and 2). The first type is the sequence rich in arginine and serine (RS domain) at residues 471–574, where 58 out of 104 amino acids were either arginine or serine (Fig. 3A). RS domains are found in pre-mRNA splicing regulator of Drosophila (21–23), mammalian U1 snRNP 70-kDa protein (50, 51), and many non-snRNP splicing factors (24–27). RS proteins detected with a monoclonal antibody against a common epitope range in size from 20 to 75 kDa (28). Additional RS proteins detected with a new monoclonal antibody in reconstituted spliceosomes in vitro include one that is larger than 200 kDa (29).

The second type of repeat is a 76-amino acid sequence...
repeated three times at residues 677–753, 906–981, and 1010–1084 (Fig. 4C). Since a homologous sequence is found in rat matrin 3 (52), we refer to it as a MH2 domain (see below). Together with the sequence in rat matrin 3, the MH2 repeats constitute RRM similar to the RRM of hnRNPs I and L (53). hnRNP I, also known as polyadenylic tract-binding protein (54, 55), binds to hnRNA through this RRM.

The third type of repeat is at the C terminus of human NP220 (Figs. 1 and 2), where characteristic sequences repeat nine times (Fig. 3B). Since 6 out of 13 amino acids in the consensus sequence are acidic, we refer to it as the acidic repeat. Since the acidic repeats contain many amino acids with an oxygen atom capable of interacting with metals as in EF hand (56), we tested the calcium binding ability of this domain by expressing the inserted sequence of clone M5 (Fig. 1) in Escherichia coli, separating the product by SDS electrophoresis, blotting to nitrocellulose, and probing with $^{45}\text{CaCl}_2$ (57). Although the product gave a radioactive signal (results not shown), the binding could be demonstrated only at calcium concentrations above 0.1 mM.

As summarized in Fig. 4, human NP220 shares three types of domains (MH1, MH2, and MH3 domains) with matrin 3, which had been cloned by Belgrader et al. (52) from a rat cDNA library. In the MH1 domain, more than 70% of the amino acids are identical or similar. In the MH2 domains, ~50% of the amino acids are similar, and both NP220 and matrin 3 retain the core sequences of RRM found in hnRNPs I and L, suggesting that they form a subfamily within the large superfamily of RNA-binding proteins (35). In MH3, 42% of the amino acids are identical or similar.

DNA Binding Activity of Human NP220—The original cDNA clone (K1) of human NP220 was obtained due to the DNA binding activity of the product. In order to define the domain essential for this DNA binding, we prepared a series of pKK223–3 plasmids with clone K1 sequences progressively truncated at the 3' end (pK1, pK1-H, and pK1-P in Fig. 1). Southwestern blotting with a $^{32}$P-labeled fragment of the mitochondrial promoter region as probe showed binding for pK1 and pK1-H but not for pK1-P (Fig. 5). This shows that residues

**Fig. 3.** Internally repeated amino acid sequences in human NP220. A, amino acid sequence of RS domain (a domain rich in arginine and serine). B, amino acid sequences of acidic repeats are compared. Numbers are the amino acid number in Fig. 2.

**Fig. 4.** Domains of human NP220 homologous with rat matrin 3. Dot matrix plot of NP220 (abscissa) against matrin 3 (ordinate) with minimum averaged score of 1.7 over 20 amino acids (63) is shown in A. Homologies of three paired amino acid sequences (MH1, MH2, and MH3) of NP220 and matrin 3 are compared in B–D with identical or similar amino acids in reversed color. Numbers are the amino acid number of human NP220 in Fig. 2 or of rat matrin 3 in Ref. 52.
1353–1477 (Figs. 1 and 2) are essential for DNA binding. Except for the relatively high content of serine (17.6%), this domain has no characteristic motif found in other DNA-binding proteins.

A modified SAAB method employing the pK1 product showed that NP220 preferentially binds to cytidine clusters in either strand of dsDNA. Thus, after six rounds of successive selection and amplification of oligonucleotides having 20 bp of random sequence, the fragments were cloned into pBluescript. Sequences of either strand of the inserted fragments in randomly selected clones are arranged to give maximal matching. Nucleotides in random and cassette regions are written with uppercase and lowercase letters, respectively. The consensus sequence is presented at the bottom of Fig. 6A.

### Fig. 6. dsDNA fragments selected by pK1 product

- **A**. Synthesized dsDNA fragments having 20 bp of random sequences between 18 bp each of two cassette sequences for amplification and cloning (5'-TTGCTCAGTCAACACCCCGC-CAGTGTTGAC-3') were selected by binding to the pK1 product and then amplified by polymerase chain reaction. After six rounds of successive selection and amplification, the fragments were cloned into pBluescript. Sequences of either strand of the inserted fragments in randomly selected clones are arranged to give maximal matching. Nucleotides in random and cassette regions are written with uppercase and lowercase letters, respectively. The consensus sequence is presented at the bottom of Fig. 6A.

- **B**. Mitochondrial Promoter

**HSP**

CCCTGCTCAACCTGCAGTTACACAA

**LSF**

GACTATTTAACAACCCCTACCTTTAACA

**Consensus**

| G | C | C | C | C | C | C | C |
|---|---|---|---|---|---|---|---|
| T | G | C | A | T | A | C | T |

| G | 3 | 1 | 2 | 1 | 2 | 7 |
|---|---|---|---|---|---|---|
| T | 0 | 2 | 1 | 1 | 1 | 1 |
| C | 12 | 12 | 12 | 13 | 11 | 7 |

**Fig. 5.** Definition of the domain in human NP220 essential for DNA binding. A series of fragments of NP220 was expressed in E. coli by subcloning the inserted sequence of K1 into pkK223-3, digesting it with PstI or HindIII and ligating to yield pk1, pk-H, and pk-P in Fig. 1. Extracts of E. coli expressing pk1 (lanes a), pk-H (lanes b), and pk-P (lanes c) were separated by SDS electrophoresis, and the transblotted filters were protein stained or hybridized to the fragment of mitochondrial promoter region. Arrows indicate the migration of the specific recombinant products.

**Fig. 7.** Western blot analysis of subcellular fractions from HeLa cells. A homogenate of HeLa cells was separated into nuclear and cytoplasmic fraction, and nuclei were further fractionated into LIS-soluble (supernatant) and insoluble (nuclear matrix) fractions. The whole cell fraction was prepared by immediate lysis of living cells with SDS sample buffer. After SDS electrophoresis, proteins on the transblotted filters were detected by protein staining or with the antibody against human NP220 in A or with monoclonal antibody against lamin B1 in B. The lanes under HMW (high molecular weight) and LMW (low molecular weight) contained size markers of 200, 116.2, 97.4, 66.2, and 43 kDa, and 97.4, 66.2, 43, 30, 20.1, and 14.1 kDa, respectively.

1353-1477 (Figs. 1 and 2) are essential for DNA binding. Except for the relatively high content of serine (17.6%), this domain has no characteristic motif found in other DNA-binding proteins.

A modified SAAB method employing the pK1 product showed that NP220 preferentially binds to cytidine clusters in either strand of dsDNA. Thus, after six rounds of successive selection and amplification of oligonucleotides having 20 bp of random sequence, fragments having the consensus sequence of CCCCC(G/C) were selected (Fig. 6A). Since both mitochondrial promoters (HSP and LSP) have such cytidine clusters (Fig. 6B), it is reasonable that clone K1 was isolated as a binding protein of the mitochondrial promoter region. It is worthwhile to note that this preferential DNA target of NP220 is distinct from the A- and T-rich sequences in MARs. This shows that the DNA binding specificity of NP220 is different from that of ARBP (8), SATB1 (9), and SAF-A (10).

Intranuclear Localization of Human NP220—We prepared a polyclonal antibody directed against human NP220 by subcloning the XbaI-HindIII fragment (4034–4746 nt) into pGEX-3X (pK1-BD in Fig. 1) and immunizing a rabbit with the product. Fig. 7 shows a Western blot with this antibody of subcellular fractions prepared from a HeLa cell homogenate. A signal was seen exclusively in the nuclear fraction (Fig. 7A), as for lamin B1 as an endogenous marker (Fig. 7B). Most of the signal remained associated with the LIS-insoluble so-called nuclear matrix fraction (45) with a minor signal in the LIS-soluble fraction, as for lamin B1. The staining pattern of subcellular fractions suggested some degradation of NP220. Comparative degradation was observed in whole cell lysates prepared by immediate solubilization of living cells with SDS sample buffer.
NP220 migrated in SDS electrophoresis as a 250-kDa polypeptide. This anomalous migration was probably due to the high content of charged amino acids and their clustering.

Indirect immunofluorescence microscopy of interphase HeLa cells with the anti-NP220 antibody showed a diffuse nucleoplasmic signal excluding nucleoli and concentrated in a punctate or "speckled" pattern (29, 61). This suggests that NP220 may have an unexpected function. The role of acidic repeats next to the DNA-binding domain requires further studies. In addition, the MH1 and MH3 domains include motifs not found in any class of nuclear proteins. Thus, human NP220 is a novel nuclear protein with multiple domains.
55. Patton, L. G., Mayer, S. A., Tempst, P., and Nadal-Ginard, B. (1991) Genes & Dev. 5, 1237–1251
56. Tufty, R. M., and Kretsinger, R. H. (1975) Science 187, 167–169
57. Maruyama, K., Mikawa, T., and Ebashi, S. (1984) J. Biochem. (Tokyo) 95, 511–519
58. Nyman, U., Hallman, H., Hadlaczky, G., Pettersson, I., Sharp, G., and Ringertz, N. R. (1984) J. Cell Biol. 102, 137–144
59. Reuter, R., Appel, B., Bringmann, P., Rinke, J., and Lührmann, R. (1984) Exp. Cell Res. 154, 548–560
60. Matera, A. G., and Ward, D. C. (1993) J. Cell Biol. 121, 715–727
61. Pirlot-Rona, S., Swanson, M. S., Gall, J. G., and Dreyfuss, G. (1989) J. Cell Biol. 109, 2575–2587
62. Ferreira, J. A., Carmo-Fonseca, M., and Lamond, A. I. (1994) J. Cell Biol. 126, 11–23
63. Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961
64. Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., and Pieler, T. (1992) Cell 71, 679–690
65. Dubnau, J., and Struhl, G. (1996) Nature 379, 694–699
66. Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., and Jäckle, H. (1996) Nature 379, 746–749