MOLECULAR AND CELLULAR BIOLOGY, Sept. 1993, p. 5593–5603
0270-7306/93/095593-11$02.00/0

Vol. 13, No. 9

NonO, a Non-POU-Domain-Containing, Octamer-Binding Protein, Is the Mammalian Homolog of *Drosophila nonA*^{diss}

YIH-SHENG YANG,[1] JEFFREY H. HANKE,[1]† LEON CARAYANNOPOULOS,[1] CHERYL M. CRAFT,[2] J. DONALD CAPRA,[1] AND PHILIP W. TUCKER[1]*

Departments of Microbiology[1] and Psychiatry,[2] The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9048

Received 11 January 1993/Returned for modification 17 February 1993/Accepted 14 June 1993

We have cloned the ubiquitous form of an octamer-binding, 60-kDa protein (NonO) that appears to be the mammalian equivalent of the *Drosophila* visual and courtship song behavior protein, no-on-transient A/dissonance (*nonA*^{diss}). A region unprecedently rich in aromatic amino acids containing two ribonuclear protein binding motifs is highly conserved between the two proteins. A ubiquitous form of NonO is present in all adult tissues, whereas lymphocytes and retina express unique forms of NonO mRNA. The ubiquitous form contains a potential helix-turn-helix motif followed by a highly charged region but differs from prototypic octamer-binding factors by lacking the POU DNA-binding domain. In addition to its conventional octamer duplex-binding, NonO binds single-stranded DNA and RNA at a site independent of the duplex site.

All V_H and V_L promoters and the immunoglobulin heavy-chain (IgH) enhancer contain an essential octamer motif (ATGCAAAT). Deletion of the promoter octamer results in a drastic reduction of Ig gene expression, to almost background levels in B cells (10, 32). A heptamer motif (CT CATGA), located 2 to 21 bp upstream of the octamer, contributes to full promoter activity (7, 42), apparently by serving as a low-affinity, cooperative binding site for factors that bind the octamer (23, 29).

Conventional octamer-binding proteins are part of a large gene family that contains a DNA-binding motif, the POU domain. OTF-1 (Oct-1), present in all cell types, and OTF-2 (Oct-2), found primarily in B cells and the brain, share no similarity other than their POU domains (reviewed in references 22, 45, and 51). All POU domain proteins appear to function as positive or negative transcription factors, with the major activating domain(s) of known members located outside the POU domain (22, 45). The findings that OTF-2 is found predominantly in B cells and that the octamer motif is conserved in all Ig promoters led to the hypothesis that this protein is necessary and sufficient for tissue-specific Ig gene expression (28, 48, 53). However, OTF-1 can stimulate an Ig V_H promoter in vitro as efficiently as OTF-2 can (18, 41). Several pre-B-cell lines have no detectable OTF-2 RNA or DNA-binding activity yet transcribe V_H genes as well as do mature B cells, which have moderate levels of OTF-2 (18). This finding raised the possibility that in vivo, OTF-1 and/or another octamer-binding factor is sufficient for Ig expression, particularly during early B-cell differentiation.

We isolated a nuclear protein from a murine B-cell leukemia line (BCL_1) that footprinted the octamer motif indistinguishably from OTF-2 (11). Here we characterize the complementary DNA clone encoding a ubiquitous form of this 60-kDa molecule, which we designate NonO. NonO lacks conventional POU domains, but it possesses distinct properties sufficient to suggest functional roles for it in immunologic and/or neurologic gene regulation. These include a striking homology with no-on-transient A (*nonA*^{diss}), a protein in *Drosophila melanogaster* involved in at least two different abnormalities: reduced visual acuity and aberrant male courtship songs (38). *nonA*^{diss} mutants demonstrate an electrophysiological defect in characteristic on- and off-transients (16, 39). These signals arise from electrical activity in neurons of the lamina, the first optic ganglia postsynaptic to photoreceptor cells of the eye. Extensive behavioral characterization of *nonA*^{diss} flies revealed reduced visual acuity, decreased detection of intensity differences, and impaired optomotor responses (13), phenotypes suggestive of a defect in the processing of visual information. A second mutation, *dissonance* (*diss*), associated with aberrant courtship songs, has many of the visual defects found in *nonA* (26). *diss* and *nonA* alleles exhibit a complex pattern of complementation, leading to the speculation that the same *nonA* gene product (NONA) is used in different locations within the nervous system to affect different regulatory patterns (20, 26). This view was confirmed by the recent cloning of the gene and complementation of mutants (1, 20, 44). The NONA protein has no known function, but it contains consensus RNA-binding sites, prompting the suggestion (1, 44) of a role in RNA metabolism.

MATERIALS AND METHODS

Protein purification and peptide sequencing. The purification of NonO by virtue of binding to an Ig promoter element has been described previously (11). Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE)-separated 60-kDa protein was electroeluted and subjected to trypsin digestion. The resultant peptides were fractionated by reverse-phase high-pressure liquid chromatography (HPLC) using a C_{18} column and a trifluoroacetic acid-acetonitrile gradient. The major peaks were collected and subjected to amino acid sequencing on an Applied Biosystems 477A amino acid sequencer. The following peptide sequences were obtained: FGQAATME, FAQPGSFEYEYAMRW, IVEFSGK?AA, ?K?EMEMEAA, and MGQMAMGGAMG INN.

Identification and sequencing of ubiquitous NonO cDNA. A cDNA library was constructed in an Okayama and Berg

* Corresponding author.
† Present address: Department of Molecular Genetics, Pfizer Inc., Groton, CT 06340.

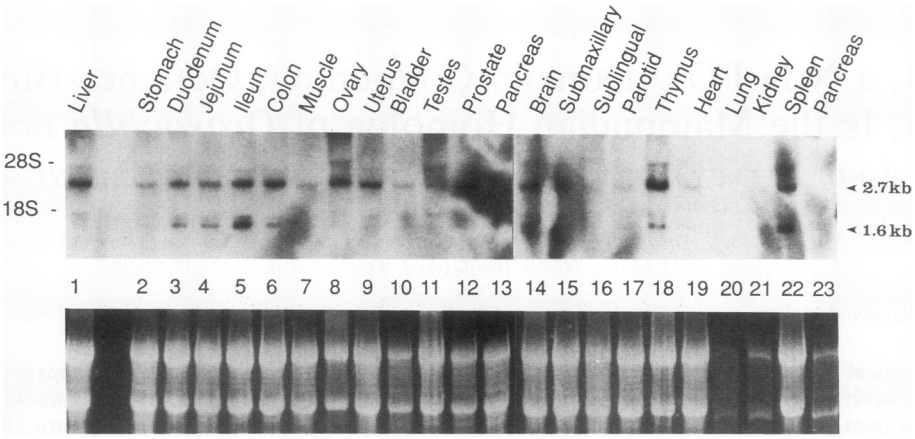

FIG. 1. Two forms of NonO transcripts in various mouse tissues. Total RNA was extracted from the indicated mouse tissues and organs. Five to 10 μg of RNA was fractionated, transferred to a nylon membrane, and probed with a NonO cDNA coding-region fragment. Shown below the autoradiogram is the ethidium bromide-stained gel to assess relative loading among lanes. Indicated are the positions of the ubiquitous and lymphoid cell-specific NonO RNAs and the 28S and 18S rRNAs.

vector (37) from the B-cell leukemia, BCL_1, and screened with a ^{32}P-phosphorylated oligonucleotide, 5′-TTT GG(G/C) CAG GCT GCC AC(A/T/C) ATG GAG-3′, derived from the amino acid sequence of one of the tryptic peptides. A partial cDNA was isolated and used to screen a library of BCL_1 cDNA in λZap (Stratagene). The oligonucleotide screening procedure used tetramethylammonium chloride (46), while screening with the partial cDNA clone was done with a nick-translated fragment. The probe was hybridized with 6× SSC (1× SSC contains 150 mM NaCl and 15 mM sodium citrate) at 65°C. Final washing conditions were 0.1× SSC at the same temperature. A 2.4-kb insert was subcloned into M13, and a series of deletion mutants were generated by exonuclease III digestion (14). The sequence of the complete cDNA clone was achieved by sequencing individual deletion mutants or priming with synthetic oligomers with an automatic DNA sequencer (Genesis 2000; Dupont).

Protein preparation. NonO mRNA was synthesized in vitro with T7 and T3 RNA polymerase (Pharmacia). One microgram of plasmid DNA linearized at the 3′ end of the NonO-5.7 insert was incubated with the enzyme at 37°C in reaction buffer (50 mM Tris-HCl [pH 8.0], 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 30 mM dithiothreitol) for 1 h. Synthesized RNA was purified by phenol extraction, then phenol-chloroform extracted, and finally ethanol precipitated. The mRNAs were used as a template for translation in a rabbit reticulocyte lysate system (Promega). In vitro-translated proteins were internally labeled with [^{35}S]methionine (NEN). Protein products were analyzed on SDS–10% polyacrylamide gels (27).

Bacterial fusion proteins were prepared for wild-type NonO by blunt-end ligation of the *Mse*I fragment of NonO cDNA into the *Sma*I site of pGEX-2T (49), which alone generates a 26-kDa N-terminal glutathione *S*-transferase (GST) fusion. Mutants N-RNP and ΔRNP-C were constructed by polymerase chain reaction-T4 DNA polymerase methods as described previously (54). The fusion gene constructs were transformed into *Escherichia coli*, and the culture was grown to log phase before isopropylthiogalactopyranoside (IPTG) induction (49). Cell pellets were harvested from a 500-ml culture and washed in 10 ml of buffer A (50 mM Tris-HCl [pH 8.0], 25% sucrose, 10 mM EDTA), resuspended in 10 ml of buffer B (10 mM Tris-HCl [pH 7.4], 1 mM EDTA, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 1 μg of leupeptin per ml, 1 μg of antipain per ml, 1 μg of pepstatin A per ml), and sonicated, and the cell debris was centrifuged. The supernatant was mixed with 25 ml of buffer C (20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.6], 100 mM KCl, 20% glycerol, 1 mM EDTA, 1 μg of leupeptin per ml, 1 μg of antipain per ml, 1 μg of pepstatin A per ml), 1/10 volume of 10% Triton X-100, and 4 ml of preswollen glutathione-agarose beads at 4°C for 2 h. The beads were spun and washed four times with 5 ml of buffer C at 4°C. Finally, the fusion protein was eluted from beads with 2.5 ml of 10 mM glutathione (pH 7.5; Sigma) in buffer C.

Mobility shift gel electrophoresis. The electrophoretic mobility shift assay (EMSA) was performed as previously described (11), with some modifications. Five microliters (about 1 μg) of GST or GST-NonO fusion protein purified from affinity beads was used in binding assays. Samples were incubated in the reaction mixture [20 mM HEPES (pH 7.9), 20% glycerol, 25 mM KCl, 0.2 mM EDTA, 10 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, 0.05% Nonidet P-40, 1 μg of poly(dI-dC)] in a final volume of 20 μl at room temperature for 15 min. The probe (final concentration of 3 nM) was added to the mixture described above and incubated at room temperature for 30 min prior to electrophoresis on 4% acrylamide gels in 0.5× TBE (50 mM Tris base, 50 mM boric acid, 0.55 mM EDTA [pH 8.0]) at 4°C. In competition assays, competitors were added 10 min prior to the probe. For the RNA competition assay, 40 U of RNAsin (Promega) was included in the reactions to retard degradation. The dimerized BN oligonucleotide was cleaved at polylinker sites from a plasmid containing subcloned BN oligonucleotide, gel purified, filled in with Klenow enzyme, and labeled as a binding probe.

Western blot (immunoblot) analysis. Proteins were separated by SDS-PAGE and electroblotted onto polyvinylidene difluoride membranes (Millipore). Antisera were collected from rabbits immunized with GST-NonO fusion protein and purified on protein A-Sepharose (12). Pooled IgG eluted from the protein A column was depleted of antibodies reacting with GST by GST affinity chromatography. Remaining antibodies were concentrated and used for immunoblotting according to the ImmunoSelect kit protocol (Bethesda

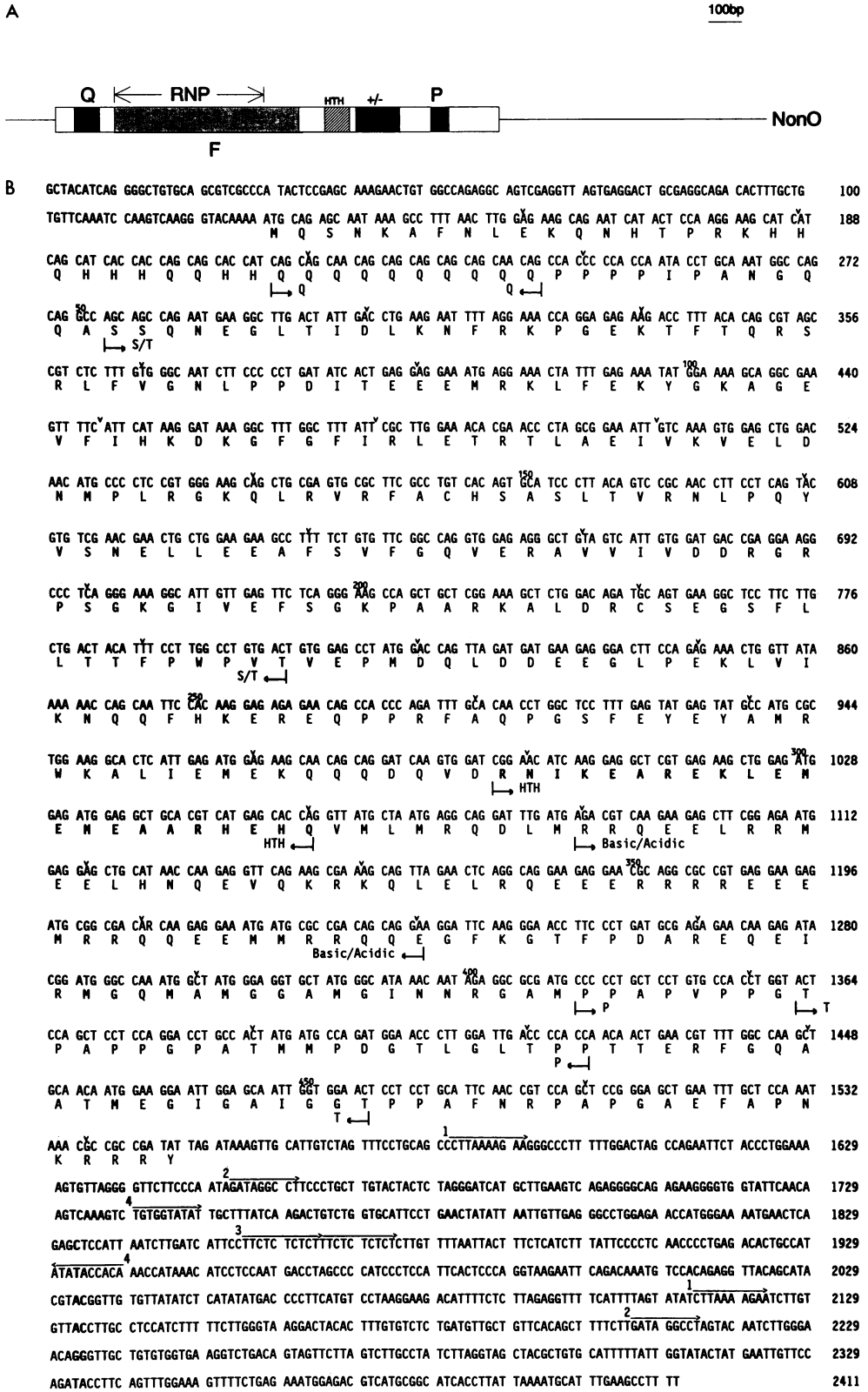

FIG. 2. Structure and sequence of ubiquitous NonO. (A) Schematic representation of NonO cDNA. The open box indicates the open reading frame. The glutamine (Q)-rich region, proline (P)-rich region, HTH domain, and highly charged regions are as labeled. F, phenylanine-rich region. (B) Nucleotide and predicted amino acid sequences of NonO. Both nucleotide and amino acid sequences are numbered from the beginning of the sequences. Amino acids in boldface are the potential HTH domain regions. Putative domains (discussed in the text) are bordered with arrows. The direct and inverted repeats at the 3′ end of the UTR are marked with arrows.

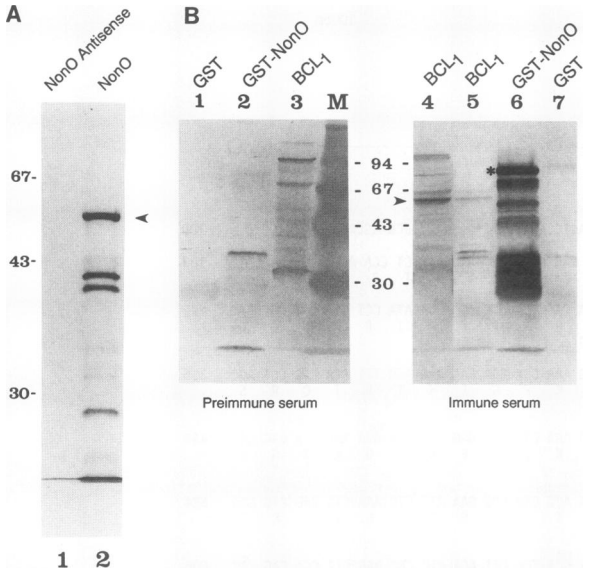

FIG. 3. In vitro and in vivo identification of ubiquitous NonO proteins. (A) In vitro-translated [^{35}S]methionine-labeled NonO products were analyzed on an SDS–10% polyacrylamide gel. Protein markers are as indicated (in kilodaltons). Lanes: 1, in vitro-translated product of full-length antisense RNA; 2, in vitro-translated product of full-length sense RNA. The arrowhead marks the full-length NonO protein. (B) Western blot analysis of NonO from a B-cell leukemia nuclear extract. Nuclear extracts were prepared from BCL_1 cells, fractionated on an SDS–10% polyacrylamide gel, and then transferred for Western blotting. The primary antiserum was prepared from GST-NonO rabbit immunization. Lanes: 1 and 7, GST protein; 2 and 6, GST-NonO fusion protein; 3 to 5, BCL_1 nuclear extract. Lanes 1 to 3 were developed with preimmune rabbit sera and goat anti-rabbit IgG (Sigma); lanes 4 to 7 were developed with immune rabbit sera; lane 4 was developed with the same secondary antibody as in lanes 1 to 3; lanes 5 to 7 were developed with a goat anti-rabbit antibody from Bio-Rad. The arrow denotes the 60-kDa NonO protein in BCL_1 extracts. The asterisk denotes the full-length NonO fusion protein (positive control). The pre- and postimmune rabbit sera were purified as described in Materials and Methods.

Research Laboratories), except that 5% dry milk was substituted for 5% bovine serum albumin as a blocking agent.

DNA and protein sequence analysis. Sequence manipulation, homology analyses, and protein domain analyses were carried out with DNAstar programs (DNAstar Inc.).

Northern (RNA) filter hybridization. Filters containing total RNA from various mouse tissues were kindly provided by Ray MacDonald (University of Texas Southwestern Medical Center, Dallas). Total RNA was extracted from rat tissues as described by Illaria et al. (17) and analyzed as previously described in detail (5). The probe was a nick-translated *Mse*I fragment which includes the complete coding region of NonO. Northern filter hybridization was performed at 42°C overnight in 5× SSC buffer with 50% formamide, 1× Denhardt's solution, 1% SDS, 100 μg of calf thymus single-stranded DNA (ssDNA) 5 μg of poly(A) per ml, 0.1% sodium pyrophosphate, 1 mM EDTA, 10 mM Tris (pH 7.5), and 10^6 cpm of probe per ml. After hybridization, the filters were washed twice in 2× SSC–1% SDS–0.1% sodium pyrophosphate for 15 min at room temperature, four times in the same buffer in 0.1× SSC at 55°C, and finally in 0.1× SSC buffer alone at 55°C for 30 min.

RESULTS

Peptide sequencing and isolation of NonO cDNA clones. DNA-binding proteins were purified from BCL_1 nuclear extracts as reported previously (11). Proteins that had been affinity column purified twice were fractionated by SDS-PAGE, and a 60-kDa protein band was electroeluted and then digested with trypsin. The digested fragments were separated by HPLC and subjected to peptide sequencing. Several peptide sequences were obtained, some of which were used to construct oligomer probes. All of the sequences found in the protein fragmentation studies were subsequently found in the NonO cDNA clone. A cDNA clone of 2.4 kb was further characterized.

Ubiquitous and potentially lymphoid cell-specific forms of NonO mRNA are present in normal mouse tissues. We probed Northern filters containing total RNA isolated from various mouse tissues for NonO transcripts. As shown in Fig. 1, there are two forms of NonO mRNA that accumulate to sufficient abundance to be detected. The larger mRNA is about 2.7 kb, a size consistent with that of our cDNA clone, assuming average polyadenylation. It is ubiquitously expressed in most tissues in various levels, although it is much lower in submaxillary, sublingual, and parotid glands and in the heart. Low signals in lung and kidney are inconclusive because of suboptimal RNA loading (Fig. 1, lower panel). The smaller form of NonO mRNA is about 1.6 kb. Its expression is restricted to certain organs or tissues that are predominantly lymphoid (spleen and thymus) or are sites of high lymphocyte migration or contamination (the small intestinal segments, particularly the ileum). Although preliminary, these results suggest that the 1.6-kb message is lymphoid cell specific.

Characterization of ubiquitous NonO cDNA and deduced protein structure. A schematic of the ubiquitously expressed cDNA and deduced protein is shown in Fig. 2A, and the nucleotide sequence is presented in Fig. 2B. There is an open reading frame encoding a protein of 473 amino acids. The deduced peptide has a molecular mass of 55 kDa. The mRNA contains a long (860-nucleotide) 3′ untranslated region (UTR). Although no consensus polyadenylation signal was found, a poly(A) tract present at the 3′ end of the cDNA clone suggests that the clone was not truncated. The 3′ UTR has three direct repeats and one inverted repeat (Fig. 2B). One of the direct repeats is continuous from nucleotide positions 1855 to 1874, containing only U and C nucleotides. Whether these repeats have any functional significance in mRNA stability or translation efficiency is unknown.

The deduced amino acid sequence (Fig. 2B) revealed some putative domains of interest (summarized in Fig. 2A). The first is a glutamine-rich area near the N terminus. A stretch of 10 glutamines (Q), from amino acid residues 29 to 38, is immediately bordered by 7 histidines (H) and 5 prolines (P).

A second type of motif was located by using an algorithm (6) to detect helix-turn-helix (HTH) DNA-binding motifs in protein sequences. Two segments which have the potential to form the HTH domain structure were identified between amino acids 289 and 310 and between amino acids 320 and 341. The SD scores of these two segments, based on Dodd and Egan's calculation method (6), are 1.8 and 1.7, respectively. The SD score of the 289 to 310 segment with the last three residues removed (EHQ) is 2.4. This score strongly suggests that this segment is an HTH.

The putative HTH domain is followed by a region rich in both basic and acidic amino acids. From residues 320 to 370, 32% (16 of 50) of the residues are basic amino acids (14

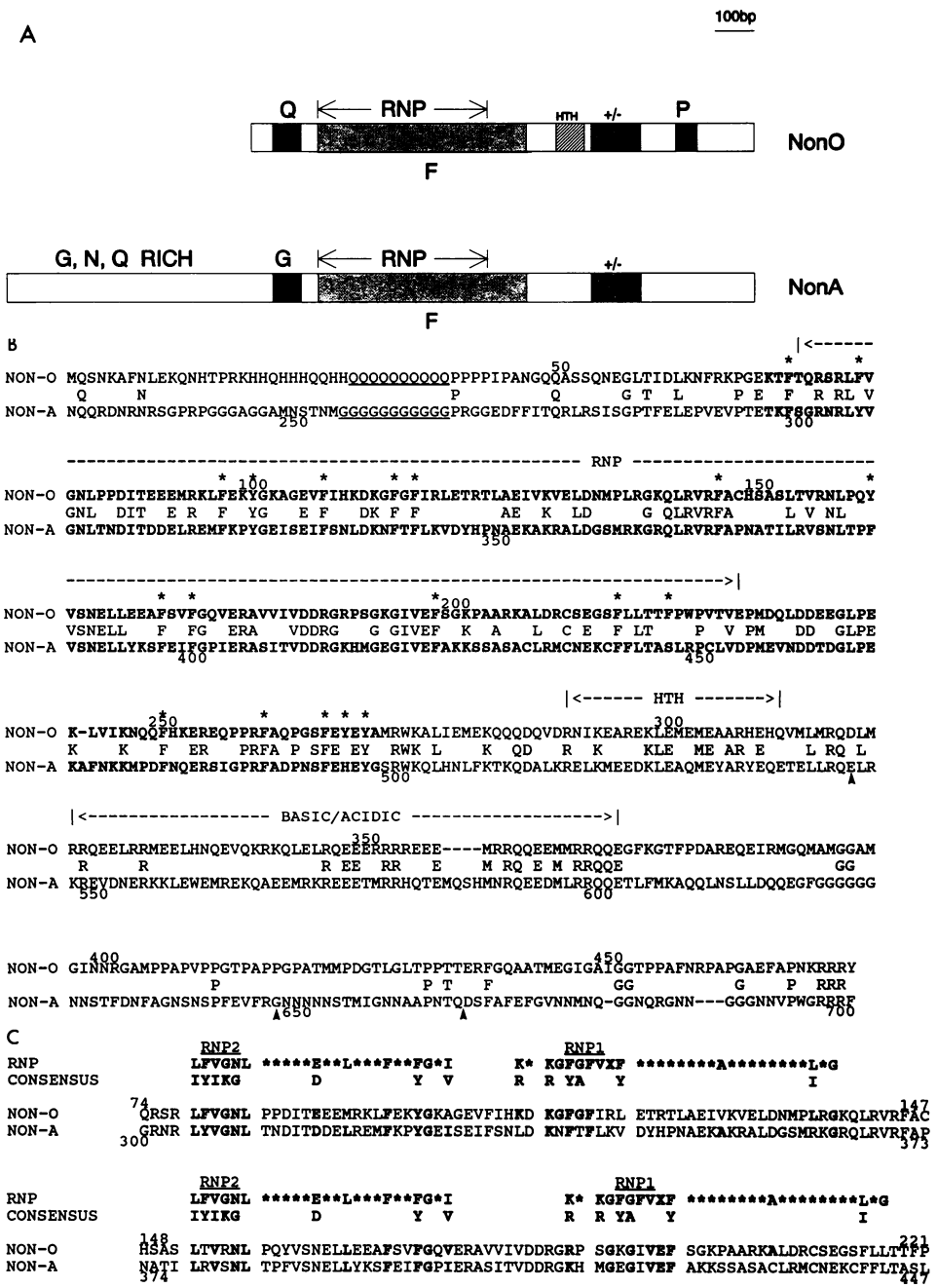

FIG. 4. Sequence homology and RNP consensus sequences of ubiquitous NonO and *nonA*. (A) Schematic comparison of NonO and *nonA* proteins, indicating conserved regions. Amino acids are given in one-letter code. (B) The complete NonO deduced amino acid sequence aligned with the carboxyl two-thirds of the *nonA* sequence. Capital letters between the sequences indicate amino acid identity. Asterisks mark the phenylalanines and tyrosines in NonO within the F domain (boldface). Stretches of glutamine (Q) and glycine (G) are underlined. Arrows denote the locations of introns in the NonA gene. Dashes indicate deletions introduced to maximize alignment. The RNP, HTH, and basic/acidic regions in NonO are indicated. (C) RNP consensus sequence alignment of NonO and *nonA*. RNP1 and RNP2 consensus sequences and other conserved residues are aligned with amino acid positions of RNP domains indicated. Asterisks denote nonconserved amino acid residues. Conservative changes within RNP consensus sequences are indicated in boldface.

arginines and 2 lysines), and 30% (15 of 50) are acidic residues (all glutamic acids). The net charge of this region is positive at physiological pH. Such an arrangement, with an HTH domain followed by a nearby charged region, is characteristic of the DNA-binding domain in OTF-1 and OTF-2 (3, 50).

Another distinctive area was found at the carboxyl-terminal region from amino acids 404 to 432, where 41% (12 of 29) of the residues are proline. Like the N-terminal glutamine-rich region, proline-rich domains may mediate protein-protein interactions essential for transcriptional activation properties (15, 21, 34, 40). Serine/threonine clusters critical for

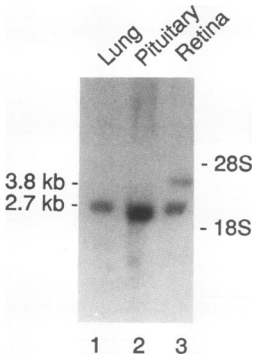

FIG. 5. Ubiquitous and retinal cell-specific NonO transcripts. Total RNA, extracted from rat tissues, was fractionated, blotted to a nylon membrane, and probed with a fragment containing the entire NonO cDNA coding region. The larger NonO band present in retina is about 3.8 kb. The ubiquitous band is 2.7 kb.

transactivation in certain cases (52) are not observed in NonO. But curiously, the vast majority of both of these amino acids are restricted to two regions: residues 51 to 225 and residues 412 to 452 at the C terminus.

In vitro and in vivo characterization of ubiquitous NonO protein. Sense and antisense RNAs from the 2.4-kb ubiquitous cDNA clone were synthesized, and in vitro-translated products of these RNAs were analyzed by SDS-PAGE. As shown in Fig. 3A, the molecular size of in vitro-translated product NonO (55 to 60 kDa) corresponded with our previous estimates derived by purification of the protein (11) and confirmed the predicted open reading frame. The several smaller NonO in vitro products probably result from premature termination or nonspecific initiation of translation.

An *Mse*I fragment, containing the entire coding region sequence minus the N-terminal seven amino acids, was ligated in frame to the carboxyl terminus of the GST gene. The expressed fusion protein, GST-NonO, had the predicted molecular size (data not shown). Attempts to use thrombin to cleave NonO from GST were unsuccessful, as NonO contains several potential thrombin cleavage sites (2). Therefore, the purified fusion protein was used directly. The 60-kDa ubiquitous NonO protein was identified in Western blots of BCL_1 nuclear extracts by heteroantisera raised in rabbits by injection of GST-NonO bacterial fusion protein (see Materials and Methods). The major band detected immunologically (denoted by the arrowhead in Fig. 3B, lanes 4 and 5) is the same size as the protein originally purified by DNA affinity chromatography (11) and is of a size consistent with the predicted size from the cDNA sequence. The minor bands are proteins that react with the secondary antibody (compare lanes 4 and 5, which are immunostained with different preparations of secondary antibody). The only common band is the 60-kDa NonO protein.

Ubiquitous NonO is homologous to a protein, *nonA*, which is essential for visual processing in flies. Using the Lipman and Pearson method (31) to search protein sequence homologies, we found a striking relationship between a portion of ubiquitous NonO and a portion of the *Drosophila* protein *nonA* (Fig. 4A). The amino acid sequence of *nonA* from residue 227 to the C terminus aligns with the entire NonO sequence. Within the first 316 residues, NonO and *nonA* have 40% amino acid sequence identity, achieved with only one insertion in *nonA* at position 468. The sequence similarity between NonO residues 72 to 326 and *nonA* residues 298 to 543 approaches 50% identity. The level of homology focused within this segment implies similar domain structures within the two proteins.

The NonO-*nonA* alignment has several interesting features (Fig. 4B). Near the N terminus, a stretch of glutamines (residues 29 to 38 in NonO) are glycines in *nonA*. However, *nonA* also has a glutamine-rich area further toward its N terminus (20) (Fig. 4A), outside our alignment. This region is also rich in asparagine and glycine.

Within the 200 residues of highest homology, there are 15 phenylalanines and 4 tyrosines in NonO (amino acids 70 to 270) (marked with a asterisks in Fig. 4B). Fifteen of those nineteen aromatic residues align identically in *nonA*. Two exhibit a conservative change to another aromatic residue. Aromatic residues outside of this homology region are not conserved. Like NonO, *nonA* has a corresponding cluster of charged basic and acidic residues (amino acids 547 to 601). However, the other two potentially functional domains predicted for NonO (HTH and proline-rich region) are missing in *nonA*.

It has been suggested (1, 44) that *nonA* may bind RNA by means of tandemly arrayed ribonuclear protein (RNP)-binding motifs (reviewed in reference 24). NonO has retained both RNP-binding motifs within the highly conserved, phenylalanine-rich region (Fig. 4C). Aromatic amino acids within the RNP domain are believed to be involved in ring-stacking interactions between protein and RNA (24). Although neither of the RNP repeats is a perfect match with the consensus, the more N-terminal motifs match very closely and are likely to be responsible for the observed RNA binding of NonO (see below).

A third form of NonO transcript is present in retina tissue. The observed similarity of *nonA* and NonO prompted us to examine NonO expression in retina and pituitary (Fig. 5). Besides the 2.7-kb ubiquitous transcript, retina samples have a third form of 3.8 kb. This transcript was undetectable in other neural tissues examined (Fig. 1 and data not shown).

Ubiquitous NonO fusion protein binds double-stranded DNA (dsDNA) and ssDNA with octamer specificity. The 60-kDa protein was originally purified from a murine B-cell nuclear extract and was monitored by EMSA of an octamer-containing, V_H promoter fragment (11). A similar assay was used with synthetic oligonucleotides (Fig. 6A) to document the binding of the recombinant protein to DNA. The BN oligonucleotide, which contains wild-type octamer and heptamer (a low-affinity octamer site) sequences and was originally used on a DNA affinity column to purify the NonO protein (11), was employed as the probe. As shown in Fig. 6B, four protein-DNA complexes were formed with the double-stranded BN oligonucleotide at increasing protein concentrations. In competition analyses, the most slowly-migrating complex (ds2) was competed by double-stranded oligonucleotides that contain wild-type octamer (BH; Fig. 6C, lane 5) or octamer and heptamer (V1; lane 6) motifs. Little competition was seen with a mutated octamer duplex (BO; lane 4) and an unrelated duplex (Ro; lane 7) showed no competition. The fastest-migrating complex (ds1) did not show octamer specificity in that complete competition was achieved by Ro (lane 7).

The binding of NonO to the BN oligonucleotide appeared complex, and the affinity of the specific binding complex (ds2) was relatively low. Therefore, we dimerized the BN probe in an effort to increase binding affinity. As shown in Fig. 6D (lane 2), a single complex was observed under these circumstances. Competition analyses again confirmed the

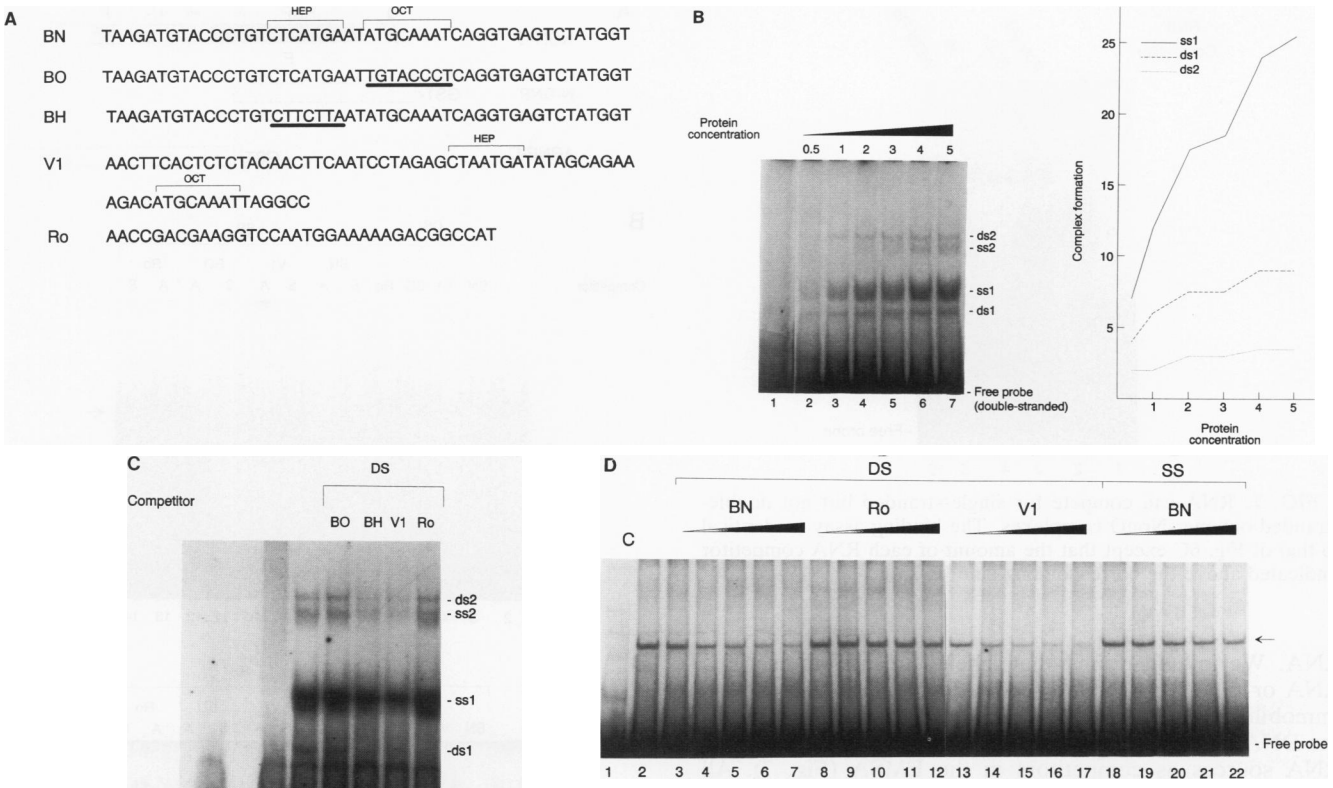

FIG. 6. Binding specificity of ubiquitous NonO for octamer motifs. (A) Synthetic oligonucleotides for DNA binding and competition assays. Sequence motifs within each oligomer (see text) are indicated above wild types (BN and V1), with mutations of octamer (BO) and heptamer (BH) underlined. (B) Concentration-dependent formation of dsDNA-NonO complexes. The probe was double-stranded BN. Protein concentrations are indicated above each lane in microliters (5 μl ≈ 1 μg). NonO-DNA complexes were quantified on a Betascope, and the values are plotted at the right as a function of protein concentration. (C) EMSA competition analysis of double-stranded octamer-containing DNA with double-stranded competitors. The BN duplex oligomer was end labeled as the binding probe. GST affinity-purified NonO fusion protein was incubated with a final concentration of 3 nM probe in the presence of 25-fold molar excesses of various competitors. Lanes: 1, probe only; 2, GST vector control; 3, GST-NonO fusion protein; 4 to 7, with BO, BH, V1, and Ro dsDNAs as competitors. (D) EMSA competition analysis of double-stranded octamer-containing DNA with double- or single-stranded competitors. The BN duplex head-to-tail dimer was end labeled as the binding probe. The binding reactions are the same as described above except that 0.8 nM (final concentration) probe and 0.5 μg of poly(dI-dC) were used in each lane. Specific competitors were used at 50- to 1,000-fold molar excess for double-stranded competitors and 100- to 2,000-fold for single-stranded competitor. The complex indicated by the arrow migrates at the position of ds2. (E) Concentration-dependent formation of ssDNA-NonO complexes. The probe was the sense strand of BN. The binding conditions are as described for panel B.

binding specificity (lanes 3 to 22) for octamer and heptamer motifs. The NonO-BN dimer complex is self-competed for by the BN oligonucleotide (lanes 3 to 7) as well as another oligonucleotide, V1 (lanes 13 to 17), which has only the octamer and heptamer motifs in common with BN (Fig. 6A). Neither an unrelated oligonucleotide (Ro; lanes 8 to 12) nor the single strand of BN (lanes 18 to 22) could compete within the same concentration range as BN or V1.

We noticed that the relative abundances of the four complexes varied among experiments (data not shown). This finding along with the predicted RNP-binding motifs common to NonO and *nonA* prompted us to test whether the complex pattern of Fig. 6B might involve ssDNA binding. Indeed, with use of the BN sense strand as a probe, two (ss1 and ss2) of the four complexes seen with the double-stranded probe were apparent (Fig. 6E). The formation of the ss2 complex was concentration dependent, suggesting that it might be an oligomer of ss1. Binding of neither complex was strictly octamer dependent, as demonstrated with ssDNA competitors (data not shown). Actually, NonO binds to ssDNA with significantly higher affinity (Fig. 6B) than to dsDNA, even though ssDNA binding does not show equivalent sequence specificity.

NonO binds RNA and ssDNA at the same site. The finding of consensus RNP motifs suggested that NonO might bind

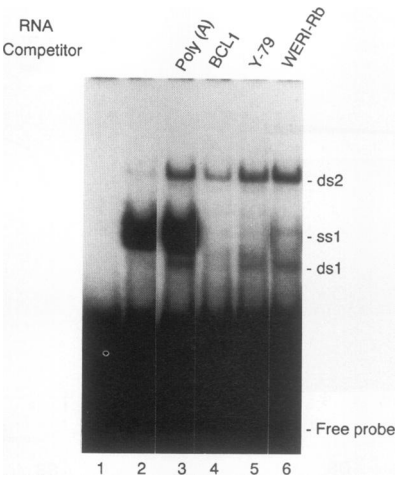

FIG. 7. RNA can compete for single-stranded but not double-stranded octamer-NonO complexes. The binding assay is identical to that of Fig. 6C except that the amount of each RNA competitor (indicated above the corresponding lane) is 2 μg.

RNA. We initially found that labeled in vitro-transcribed RNA or total RNA bound readily to NonO fusion protein immobilized on glutathione beads (data not shown). Given the DNA binding results presented above, we used various RNA sources as competitors in the EMSA (Fig. 7). All RNAs competed efficiently for ssDNA binding but not for dsDNA-protein complexes (lanes 4 to 6). However, simple ribonucleotide polymers such as poly(A) (lane 3) or poly(I-C) (data not shown) did not compete. These data suggest that RNA and ssDNA bind to the same site, independently of dsDNA.

Segregation of dsDNA- and ssDNA/RNA-binding domains. NonO contains tandem RNP domains N-terminal to a putative HTH domain followed by a highly charged region. Given our binding data, we suspected that the RNP domains are responsible for the RNA/ssDNA binding and that the HTH and charged residues are necessary for dsDNA binding. Mutants containing one or the other of these domains were generated for further EMSA analysis (Fig. 8A). As shown in Fig. 8B, N-RNP, which is the N-terminal half of NonO and contains both RNP domains, bound to the BN sense-strand probe. The binding complex was competed for by all of the previously tested ssDNA competitors except for either strand of Ro (Fig. 8B, lanes 7 to 14). The binding complex was not eliminated by any dsDNA competitors (Fig. 8B, lanes 3 to 6). Reciprocally, ΔRNP-C, which contains only the putative HTH and highly charged region, was able to bind the duplex octamer probe (Fig. 8C, lane 2) with sequence specificity (lanes 3 to 6). However, as opposed to the N-RNP binding complex, the ΔRNP-C binding complex was not competed for by any of the ssDNAs (lanes 7 to 14).

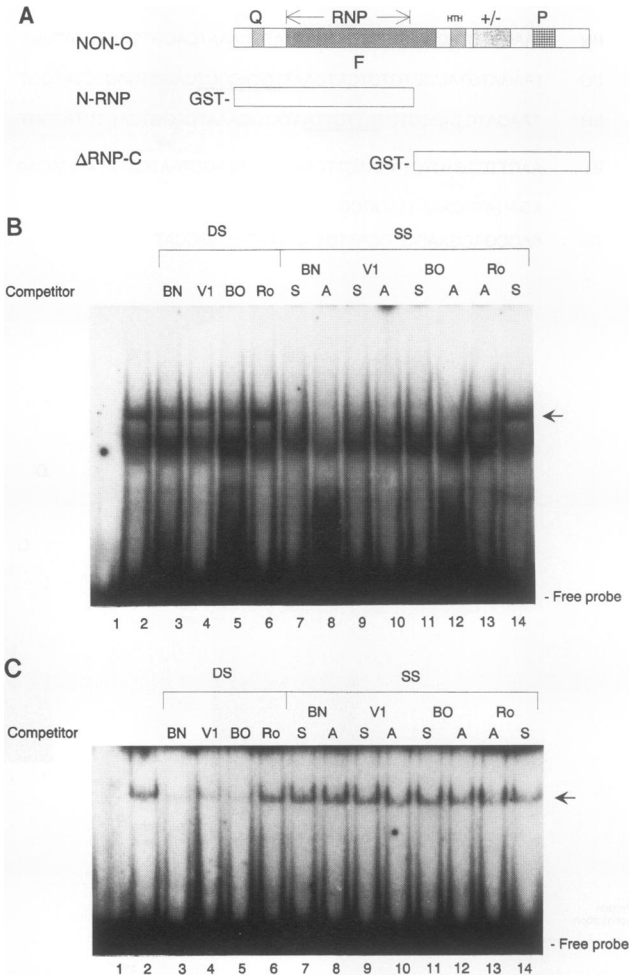

FIG. 8. Different domains in NonO are required for duplex and ssDNA binding. (A) NonO truncation mutants containing separate domains. Each mutant was constructed and expressed as a GST fusion (Materials and Methods). (B) Binding of N-RNP to the single-stranded BN oligonucleotide. The sense strand of BN was labeled as the probe. The binding reactions and competitors are the same as described for Fig. 6. (C) Binding of ΔRNP-C to the double-stranded BN DNA probe. Reaction conditions were as described above. S, sense; A, antisense. The arrow marks the binding complex.

DISCUSSION

We previously isolated a novel 60-kDa protein from a B-cell leukemia (BCL_1) nuclear extract which bound to the murine, IgH promoter-associated octamer (11). In this study, we have characterized the complementary DNA that encodes the ubiquitous form of this protein (NonO) and documented its nucleic acid binding specificity and its striking homology with *Drosophila nonA*. By virtue of conservation of potentially functional domains, aspects of the discussion below are likely to apply to both NonO and *nonA*.

The amino acid sequence of five distinct peptides isolated by enzymatic digestion of the electroeluted NonO protein confirmed that the cDNA corresponds to the purified DNA-binding protein. The molecular weight of recombinant or cellular NonO detected immunologically is indistinguishable from that of our purified protein (11) and is similar to that of the lymphoid cell-specific transcription factor OTF-2. Using a computer algorithm (6), we found a potential HTH motif within NonO. HTH motifs are integral components of all homeoboxes including that of POU (22, 45). However, the NonO potential HTH motif shares no homology with the POU homeobox, nor does the remainder of NonO align with any position of OTF-1, OTF-2, or other POU domain-containing proteins (3, 8, 25, 50). Likewise, NonO shares no similarity with Ku, another non-POU, octamer-binding protein (33, 43).

Northern blots revealed three forms of NonO mRNA: ubiquitous (2.7 kb), lymphoid cell specific (1.6 kb), and retinal cell specific (3.8 kb). On the basis of size, we concluded that the cDNA clone characterized here is the ubiquitous form. Our inability to immunologically detect two species of NonO protein in BCL_1 nuclei (Fig. 3B) could be due to their similar molecular sizes, the absence of shared epitopes, or their presence in different subcellular compartments. The difference in size between the three NonO mRNAs probably results from alternative processing of the same precursor. This view is suggested by our finding (data not shown) of no more than two hybridizing fragments generated in human and mouse genomic DNA, using various restriction enzymes that do not cut within the ubiquitous cDNA. Furthermore, initial sequences of potential lymphoid and retinal cell-specific cDNA candidates matched that of the ubiquitous form perfectly at different segments but completely diverged at both 5′ and 3′ UTRs (data not shown). Cell-type-specific inclusion versus exclusion of exons has been observed for several genes (4, 35, 36). Indeed, at least three alternative splicing patterns giving rise to two polypeptides were observed for the single *nonA* gene (20). Indirect evidence suggests differential tissue accumulation of *nonA* during development (1, 44).

Our previous observations (11) and the presence of a potential HTH domain followed by a region high in charged amino acids predicted that recombinant NonO would have the ability to bind dsDNA. Direct examination by EMSA with octamer-containing oligonucleotides confirmed this hypothesis. However, the binding of NonO is more complex than anticipated, probably in part because of the presence of tandem RNP sites. NonO binding to dsDNA occurs independently of its binding to ssDNA. Given the consensus RNP sites, it was not surprising that NonO bound RNA. However, it also appears to do this independently of dsDNA binding. This conclusion was suggested by the observation that RNA competes with ssDNA but not octamer dsDNA. Further confirmation was provided by our segregation of dsDNA- and ssDNA-binding activities in mutants in which the RNP or the putative HTH/charged region domains were deleted.

While several proteins that bind both forms of DNA have been characterized, TFIIIA (reviewed in reference 55) may be the only precedent for specific DNA and RNA binding. TFIIIA acts as a positive transcription factor by binding to an internal control region in the oocyte 5S RNA gene. It performs a second essential function by binding 5S RNA in the cytoplasm, thereby stabilizing the RNA prior to ribosome assembly. The specificity of RNA binding is controversial (discussed in reference 19); some reports claimed that TFIIIA binding to RNA resembles its DNA binding, while others found no similarity between 5S DNA and 5S RNA binding. However, both nucleic acids interact with the same linear array of nine zinc fingers (30). NonO, on the other hand, has evolved separate binding sites for dsDNA and RNA. A dual-site model is consistent with the fact that neither the RNA (RNP)- nor the DNA (HTH)-binding motifs proposed for NonO have been shown to bind the other nucleic acid.

Given the demonstration of dual and separate binding sites, do both have physiological significance? Clearly duplex octamer binding, although specific, is of relatively low affinity. This may account for our inability to date to footprint bacterially expressed NonO-octamer complexes. Similar failures have been noted for other low-affinity protein-DNA complexes, including the testis-determining factor SRY (9) and the recombination signal sequence-binding protein T160 (47). However, we were able to footprint NonO purified to apparent homogeneity from a B-cell line (11). The discrepancy could indicate the necessity for a posttranslational modification or a protein-protein interaction; neither is provided by bacteria. ssDNA, and predictably RNA, interacts with the NonO RNP domain at significantly higher affinity. The conservation of this region in *Drosophila nonA* suggests that RNA binding is critical to the function of both proteins.

The most critical issue yet to be addressed is the function of NonO and whether this function is shared by the tissue-specific forms. Although we have no functional data for recombinant NonO, previously we showed that the same 60-kDa material, affinity purified from BCL_1 nuclear extract, transactivated the BCL_1 promoter in vitro and that the transactivation was octamer dependent (18). The similar sizes of NonO and OTF-2 raise the possibility that this activity was provided by copurified OTF-2. If so, OTF-2 would have had to be present in amounts below the detection limits of our peptide analysis. This is unlikely, but the interpretation is further complicated by the potential copurification of the lymphoid cell-specific form of NonO. In addition to nucleic acid-binding domains, an essential criterion for a transcription factor is a transactivation domain. Within ubiquitous NonO, there is a stretch of 10 glutamine residues near the N terminus and a proline-rich region near the carboxyl terminus. Both of these domains have been shown to be involved in protein-protein interactions within transcriptional initiation complexes, and they function to enhance transcriptional activity (reviewed in reference 34). An alternative interpretation of our in vitro transcription results (18) is that NonO functioned as a cofactor with conventional octamer-binding proteins to affect transcription.

Although a role for ubiquitous NonO in Ig gene expression remains to be established, its homology with *nonA* and its alternative retinal form raise distinct functional possibilities. Since NonO is the mammalian equivalent of *nonA*, it is conceivable that a retinal cell-specific NonO has a high-affinity binding site for a gene involved in visual processing. Octamer- and degenerate octamer-binding proteins of the POU family act as transcriptional activators in the developing forebrain (reviewed in reference 45). Indeed, knockout experiments (4a) suggest that the most important function for OTF-2 is in the brain, not the immune system.

ACKNOWLEDGMENTS

We thank William Watson for technical assistance. We also thank Richard Baer for discussions and Fred Kruse for preparing the mouse tissue Northern blot. We thank Jeff Hall and Kathy Rendahl for sharing data prior to publication.

This work was supported by NIH grants AI-18499 and GM-31689 (to P.W.T. and J.D.C.) and NS28126, VA research, and the Ella C. McFadden Charitable Trust (to CMC). J.D.C. holds the Edwin L. Cox distinguished chair in immunology and genetics.

REFERENCES

1. **Besser, H. V., P. Schnabel, C. Wieland, E. Fritz, R. Stanewsky, and H. Saumweber.** 1990. The puff-specific Drosophila protein Bj6, encoded by the gene no-on-transient-A, shows homology to RNA binding proteins. Chromosoma **100**:37–47.
2. **Chang, J.-Y.** 1985. Thrombin specificity; requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate. Eur. J. Biochem. **151**:217–224.

3. **Clerc, R. G., L. M. Corcoran, H. J. LeBowitz, D. Baltimore, and P. A. Sharp.** 1988. The B-cell-specific Oct-2 protein contains POU box and homeo box-type domains. Genes Dev. **2**:1570–1581.
4. **Collier, V. L., W. A. Kronert, P. T. Donnell, K. A. Edwards, and S. I. Bernstein.** 1990. Alternative myosin hinge regions are utilized in a tissue-specific fashion that correlates with muscle contraction speed. Genes Dev. **4**:885–895.

4a. **Corcoran, L. M., M. Karrelas, G. J. V. Nossal, Z.-S. Ye, T. Jacks, and D. Baltimore.** 1993. Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. Genes Dev. **7**:570–582.

5. **Craft, C. M., D. H. Whitmore, and L. A. Donoso.** 1990. Differential expression of mRNA and protein encoding retinal and pineal s-antigen during the light/dark cycle. J. Neurochem. **55**:1461–1473.
6. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18**:5019–5026.
7. **Eaton, S., and K. Calame.** 1987. Multiple DNA sequence elements are necessary for the function of an immunoglobulin heavy chain promoter. Proc. Natl. Acad. Sci. USA **84**:7634–7638.
8. **Garcia-Blanco, M. A., R. G. Clerc, and P. A. Sharp.** 1989. The DNA-binding homeo domain of the Oct-2 protein. Genes Dev. **3**:739–745.
9. **Giese, K., J. Cox, and R. Grosschedl.** 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell **69**:185–195.
10. **Grosschedl, R., and D. Baltimore.** 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell **41**:885–897.
11. **Hanke, J. H., N. F. Landolfi, P. W. Tucker, and J. D. Capra.** 1988. Identification of murine nuclear proteins that bind to the conserved octamer sequence of the immunoglobulin promoter region. Proc. Natl. Acad. Sci. USA **85**:3560–3564.
12. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
13. **Heisenberg, M.** 1972. Comparative behavioral studies on two visual mutants of Drosophila. J. Comp. Physiol. **80**:119–136.
14. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
15. **Herr, W., R. A. Sturm, R. G. Clerc, L. M. Corcoran, D. Baltimore, P. A. Sharp, H. A. Ingraham, M. G. Rosenfeld, M. Finney, G. Ruvkun, and H. R. Horvitz.** 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes Dev. **2**:1513–1516.
16. **Hotta, Y., and S. Benzer.** 1970. Genetic dissection of the Drosophila nervous system by means of mosaics. Proc. Natl. Acad. Sci. USA **67**:1156–1163.
17. **Ilaria, R., D. Wines, S. Pardue, S. Jamison, S. R. Ojeda, J. Snider, and M. R. Morrison.** 1985. A rapid microprocedure for isolating RNA from multiple samples of human and rat brain. J. Neurosci. Methods **15**:165–174.
18. **Johnson, D. G., L. Carayannopoulos, J. D. Capra, P. W. Tucker, and J. H. Hanke.** 1990. The ubiquitous octamer-binding protein(s) is sufficient for transcription of immunoglobulin genes. Mol. Cell. Biol. **10**:982–990.
19. **Joho, K. E., M. K. Darby, E. T. Crawford, and D. D. Brown.** 1990. A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA in xenopus. Cell **61**:293–300.
20. **Jones, K. R., and G. M. Rubin.** 1990. Molecular analysis of no-on-transient A, a gene required for normal vision in Drosophila. Neuron **4**:711–723.
21. **Kadonaga, J. T., A. J. Courey, J. Ladika, and R. Tjian.** 1988. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science **242**:1566–1570.
22. **Kemler, I., and W. Schaffner.** 1990. Octamer transcription factors and the cell type-specificity of immunoglobulin gene expression. FASEB J. **4**:1444–1449.
23. **Kemler, I., E. Schreiber, M. M. Muller, P. Matthias, and W. Schaffner.** 1989. Octamer transcription factors bind to two different sequence motifs of the immunoglobulin heavy chain promoter. EMBO J. **8**:2001–2008.
24. **Kenan, D. J., C. C. Query, and J. D. Keene.** 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. **16**:214–220.
25. **Ko, H.-S., P. Fast, M. McBride, and L. M. Staudt.** 1988. A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. Cell **55**:135–144.
26. **Kulkarni, S. J., A. F. Steinlauf, and J. C. Hall.** 1988. The dissonance mutant of Drosophila melanogaster, isolation, behavior and cytogenetics. Genetics **118**:267–285.
27. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nature (London) **227**:680–685.
28. **Landolfi, N. F., J. D. Capra, and P. W. Tucker.** 1986. Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences. Nature (London) **323**:548–551.
29. **Landolfi, N. F., X.-M. Yin, J. D. Capra, and P. W. Tucker.** 1988. A conserved heptamer upstream of the IgH promoter region octamer can be the site of a coordinate protein-DNA interaction. Nucleic Acids Res. **16**:5503–5514.
30. **Liao, X., K. R. Clemens, L. Tennant, P. E. Wright, and J. M. Gottesfeld.** 1992. Specific interaction of the first three zinc fingers of TFIIIA with the internal control region of the xenopus 5S RNA gene. J. Mol. Biol. **223**:857–871.
31. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
32. **Mason, J. O., G. T. Williams, and M. S. Neuberger.** 1985. Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. Cell **41**:479–487.
33. **May, G., C. Sutton, and H. Gould.** 1991. Purification and characterization of Ku-2, an octamer-binding protein related to the autoantigen Ku. J. Biol. Chem. **266**:3052–3059.
34. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
35. **Mosthaf, L., K. Grako, T. J. Dull, L. Coussens, A. Ullrich, and D. A. McClain.** 1990. Functionally distinct insulin receptors generated by tissue-specific alternative splicing. EMBO J. **9**:2409–2413.
36. **Munroe, S. H., and M. Lazar.** 1991. Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J. Biol. Chem. **266**:22083–22086.
37. **Okayama, H., and P. Berg.** 1982. High efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
38. **Pak, W. L.** 1991. Molecular genetic studies of photoreceptor function using Drosophila mutants. Prog. Clin. Biol. Res. **362**:1–32.
39. **Pak, W. L., H. Grossfeld, and S. Arnold.** 1970. Mutants of the visual pathway of Drosophila melanogaster. Nature (London) **227**:518–520.
40. **Peterson, M. G., N. Tanese, B. F. Pugh, and R. Tjian.** 1990. Functional domains and upstream activation properties of cloned human TATA binding protein. Science **248**:1625–1630.
41. **Pierani, A., A. Heguy, H. Fujii, and R. G. Roeder.** 1990. Activation of octamer-containing promoters by either octamer-binding transcription factor 1 (OTF-1) or OTF-2 and requirement of an additional B-cell-specific component for optimal transcription of immunoglobulin promoters. Mol. Cell. Biol. **10**:6204–6215.
42. **Poellinger, L., B. K. Yoza, and R. G. Roeder.** 1989. Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter. Nature (London) **337**:573–576.
43. **Reeves, W. H., and Z. M. Sthoeger.** 1989. Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen. J. Biol. Chem. **264**:5047–5052.
44. **Rendahl, K. G., K. R. Jones, S. J. Kulkarni, S. H. Bagully, and J. C. Hall.** 1992. The dissonance mutation at the no-on-transient-A locus of D. melanogaster. Genetic control of courtship

song and visual behaviors by a protein with putative RNA-binding motifs. J. Neurosci. **12**:390–407.

45. **Rosenfeld, M. G.** 1991. POU-domain transcription factors: powerful developmental regulators. Genes Dev. **5**:897–907.
46. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
47. **Shirakata, M., K. Huppi, S. Usuda, K. Okazaki, K. Yoshida, and H. Sakano.** 1991. HMG1-related DNA-binding protein isolated with V-(D)-J recombination signal probes. Mol. Cell. Biol. **11**:4528–4536.
48. **Singh, H., R. Sen, D. Baltimore, and P. A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
49. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione-S-transferase. Gene **67**:31–40.
50. **Sturm, R. A., G. Das, and W. Herr.** 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev. **2**:1582–1599.
51. **Sturm, R. A., and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. Nature (London) **336**:601–604.
52. **Tanaka, M., and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. Cell **60**:375–386.
53. **Wirth, T., L. Staudt, and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. Nature (London) **329**:174–178.
54. **Yang, Y.-S., W. J. Watson, P. W. Tucker, and J. D. Capra.** 1993. Construction of recombinant DNA by exonuclease recession. Nucleic Acids Res. **21**:1889–1893.
55. **You, W., N. Veldhoen, F. Baudin, and P. J. Romaniuk.** 1991. Mutations in 5S DNA and 5S RNA have different effects on the binding of Xenopus transcription factor IIIA. Biochemistry **30**:2495–2500.