Characterization of Developmentally Regulated and Retina-specific Nuclear Protein Binding to a Site in the Upstream Region of the Rat Opsin Gene*

Maria A. Morabito‡, Xi Yu, and Colin J. Barnstable§

From the Department of Ophthalmology and Visual Science, Yale University School of Medicine, New Haven, Connecticut 06510

(Received for publication, October 12, 1990)

DNase I protection and gel retardation assays have identified a sequence 5' to the transcription start site of the rat opsin gene that interacts with nuclear proteins from mammalian retinas but not from a variety of other neural and non-neural tissues. Following sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transfer to nitrocellulose the protein(s) responsible for this binding were identified with an oligonucleotide probe and were found to migrate with an apparent molecular size of 40 kilodaltons. The binding complex eluted from fast protein liquid chromatography gel filtration as a peak centered at 100 kilodaltons, suggesting the presence of more than one subunit. Binding activity could be detected in postnatal day 1 retinal extracts and increased over the next 2 weeks of development, a time course coincident with opsin gene expression and maturation of rod photoreceptors.

Synthetic oligonucleotides with altered sequences showed that the binding was dependent upon residues in a CTAAT motif and was facilitated by surrounding GGCCC sequences. The specificity of the binding interaction was measured by inhibition of complex formation in a gel retardation assay. The unaltered sequence was over 2 orders of magnitude more effective than an unrelated DNA sequence or a consensus sequence corresponding to a known CCAAT box binding protein NF1.

One of the earliest cell type-specific genes expressed by differentiating rod photoreceptors is the visual pigment protein opsin. Opsin can be detected immunocytochemically in the plasma membranes of immature rods shortly after they become postmitotic (1, 2). The interval between final mitosis and expression of immunocytochemically detectable opsin has been estimated as 48 h (3). In rats and mice, rods become postmitotic between embryonic day 20 (E20) and postnatal day 8 (PN8) with the peak of cell generation between PN1 and PN2 (4–6). Although additional forms of regulation cannot be excluded, the expression of opsin is regulated at the transcriptional level. Both dot-blot analysis of retinal RNA and nuclear run off experiments have detected opsin gene expression at PN1 or PN2 rising to adult levels by PN10–12 (7).

The transcription of many eukaryotic genes is regulated by the combination of multiple sequence-specific DNA-protein(s) interactions. In many cases the DNA elements involved in these interactions are located upstream of the genes. Some of these cis-acting sequences are common to many genes, whereas others have a more limited distribution (8, 9). By identifying sequences required for the correct expression of an early neural gene such as opsin, the proteins interacting with them, and the ways in which these proteins are regulated, it will be possible to move closer to an understanding of the events occurring between the period of cell type commitment and initial differentiation.

As one step in analyzing the molecular mechanisms in the temporal and cell type-specific regulation of opsin expression, we have isolated DNA upstream of the rat opsin transcription start site and have used DNase I protection to identify a sequence that interacts with nuclear protein(s) found only in retina. Making a series of deletions and substitutions indicate that residues within a CTAAT motif found in this sequence are important for binding. The binding activity of the protein(s) interacting with this sequence increases over postnatal retinal development, suggesting that this DNA/protein interaction may be one of the important steps in the regulation of opsin expression.

MATERIALS AND METHODS

Isolation and Characterization of the Rat Opsin Gene—The bovine opsin cDNA subclone rOPlps (7) was used to screen an EcoRII rat genomic Charon 4A library (Clontech) under stringent hybridization conditions (10). From the isolated clones, one was further characterized by restriction enzyme mapping and sequence analysis. The isolated genomic DNA was subcloned in pBS M13+ (Stratagene). Subclones containing DNA 5' to the opsin gene were sequenced on both strands using the dyeoxy chain termination method (11).

Preparation of Nuclear Extracts—Fresh tissues from adult rats were homogenized and nuclear extracts were prepared as described by Digman et al. (12). Nuclei from brain, cerebellum, liver, and thymus homogenates were isolated by sucrose gradient centrifugation prior to the high salt extraction as described by Graves et al. (13). Bovine retina extract was prepared in the same way as the rat extracts except that the dissections were performed under dim red light and the photoreceptor outer segments were removed by mild agitation before homogenization. Protein concentrations were determined by the Lowry method with bovine serum albumin as standard (14).

DNase I Protection Analysis—DNase I protection analysis (15) was performed using 3’ end-labeled fragments (5,000–10,000 cpm/0.05–0.1 ng) incubated in the binding reaction buffer in the presence...
of 80–100 μg of nuclear extract in a total volume of 50 μl. After 40 min on ice, MgCl₂ and DNase I were added to a final concentration of 5 mM and 5–15 μg/ml, respectively, and the reactions were then incubated for 10 min on ice. The samples were isolated by phenol/ chloroform extractions followed by ethanol precipitation and were then analyzed on denaturing 6 and 8% polyacrylamide gels. The "G + A" reactions on the same samples were performed by the method of Maxam and Gilbert (16) and used as size markers.

**Gel Retardation Analysis**—Synthetic oligonucleotides were synthesized from the appropriate 3′-cyanoethyl phosphoramidites using an Applied Biosystems 391 DNA synthesizer. Sequences were chosen such that after annealing the molecules had single stranded ends that could be labeled using the Klenow fragment of DNA polymerase I and [³²P]dNTPs. The labeled oligonucleotides (10,000 cpm corresponding to 0.1–0.5 ng in different experiments) were then used in binding assays in the presence of 5–20 μg of nuclear extracts from various tissues (17, 18). The binding reactions were performed in a total volume of 25 μl containing 10 mM Tris, pH 7.5, 5% (v/v) glycerol, 50 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol, supplemented with 5 μg of poly(dI-dC) as a nonspecific DNA competitor. The reaction mixtures were subjected to PAGE analysis and the gels were then dried without fixation and autoradiographed for 8–12 h.

**Southwestern Blot Analysis**—Protein fractions were fractionated by electrophoresis on a 10% sodium dodecyl sulfate-polyacrylamide gel (19) and transferred to nitrocellulose filters at 4 °C (20). Transferred marker proteins were detected by staining with Ponceau S. The nitrocellulose filter was incubated with blocking buffer (25 mM HEPES, pH 7.9, 50 mM KCl, 3 mM MgCl₂, 5 mM dithiothreitol) containing 5% (w/v) non-fat dried milk at 4 °C for 2 h and then in the same buffer with 0.25% non-fat dried milk, 20 μg/ml salmon sperm DNA, and 10 μg/ml ³²P-labeled oligonucleotide probe for 8–12 h at 4 °C. The filter was then washed with three changes of binding buffer for 5 min at 4 °C and exposed to x-ray film at −80 °C overnight with an intensifying screen.

**Fast Protein Liquid Chromatography Gel Filtration**—3–10 μg of bovine retinal nuclear extract was loaded onto a 30-ml Superose™ 6 column in a sample volume of 0.5 ml. Proteins were eluted with 25 mM HEPES, pH 7.9, 100 mM KCl, 12.5 mM MgCl₂, and 2 mM dithiothreitol at a flow rate of 0.5 ml/min and 1-ml fractions collected. 10–15 μl of each fraction was used for gel retardation or Southwestern blot analyzes. The column was calibrated with the marker proteins thyroglobulin, aldolase, bovine serum albumin dimer, bovine serum albumin, and ovalbumin, all eluted with the same buffer.

### RESULTS

The cloning and structural organization of the whole rat opsin gene will be described in a separate paper. The sequence of a Hinfl/Sacl fragment upstream of the coding region is shown in Fig. 1 and is compared with the previously published 5′ sequence of the human rod opsin gene (21). Two localized regions of homology were found, one including the putative TATA box (−36 to −10 in rat) and another including a CTAAT(rat) or CCAAT(human) motif (−140 to −107 in rat). These latter motifs are surrounded by conserved GGCCCC repeats.

End-labeled fragments derived from the Hinfl/Sacl region (see Fig. 1) were used in DNase I protection assays. Binding reactions were carried out in the presence of nuclear extracts from various neural and non-neural tissues and, after treatment with DNase I, were analyzed by denaturing gel electrophoresis. A retina-specific DNase I protection was found between −125 and −110 for the coding strand, and −125 and −107 for the noncoding strand (Fig. 2). The results suggested the existence of a nuclear protein found only in retina that interacted with this region of DNA.

Based on the sequence of this region synthetic oligonucleotides were prepared and used in gel retardation assays using nuclear extracts from a variety of rat tissues. The position and sequence of these oligonucleotides are shown in Fig. 3. Both oligonucleotides Ret 1* (−138 to −102) and the shorter

---

1 The abbreviation used is: HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

![Fig. 1. Comparison of the upstream region of the rat (R) and human (H) opsins genes.](image)

![Fig. 2. DNase I protection of the Ret 1 region in the presence of nuclear extracts from rat retina (R), brain (B), cerebellum (C), liver (L), and thymus (T).](image)

**Nuclear Protein Binding Site Upstream of the Rat Opsin Gene**

**DNase I protection of the Ret 1 region in the presence of nuclear extracts from rat retina (R), brain (B), cerebellum (C), liver (L), and thymus (T).** Bovine serum albumin was used as control (M). The left panel represents the coding strand, the right panel the noncoding strand. The boxes represent the retina-specific binding region from −125 to −110 on the coding strand and −125 to −107 on the noncoding strand. Ret 1 (−136 to −110) contain the CTAAT motif and the proximal and distal GGCCCC repeats. A major complex (labeled c in Fig. 4A) was formed in the presence of nuclear extract from retina but not with extracts of other tissues. With extracts from most tissues two slower complexes (labeled a and b) were formed, although these were often stronger with retina. Bands migrating faster were found with some extracts of all tissues (see lane B in Fig. 4, left panel). These bands were variable and may be due to a variety of nuclear factors that interact with related sequences. Comparison of bovine and rat retinal extracts showed that each formed a complex migrating at the position of complex c in gel retardation assays using oligonucleotide Ret 1 (Fig. 4, right panel). When the assays were performed with an oligonucleotide lacking the proximal GGCCCC sequence (Ret 1a), the retina-specific complex c was formed (Fig. 5A). Complex c formation required larger amounts of retina extract when the oligonucleotide...
Nuclear Protein Binding Site Upstream of the Rat Opsin Gene

Fig. 3. List of the sequences of the oligonucleotides used in the gel retardation assays. Ret 1* contains three AT pairs at each end that were added for ease of labeling. The sequences of Ret 1 through Ret 1f are aligned to demonstrate differences in the ends; individual base changes in Ret 1c through Ret 1f are indicated by bars.

Fig. 4. Left panel, gel retardation assay using oligonucleotide Ret 1* with nuclear extracts from rat retina (R), brain (B), cerebellum (C), liver (L), and thymus (T). The marker lanes (M) contain bovine serum albumin and represent the migration of free probe. a and b indicate the slow migrating complexes, c represents the retina-specific DNA-protein complex. Right panel, gel retardation assay using oligonucleotide Ret 1 with nuclear extracts from rat retina (lane 1) and bovine retina (lane 2). Both species show formation of complex c.

cleotide lacked both GGCCCC sequences (Ret 1b). Changes in the CTACCTAG motif to CTACACTAG (Ret 1c) or CTACCTAG (Ret 1d) also caused a decrease in binding such that formation of the retina-specific complex c was only barely detectable at the highest concentrations of retina extract used (Fig. 5, A and B). Changing the sequence to CTACCTAG (Ret 1e) or GTCATCTAG (Ret 1f) had minor or no effects upon the formation of complex c (Fig. 5B).

To test whether formation of complex c reflected the specific binding of a nuclear protein(s) to the Ret 1 sequence, and because this region contains a pentanucleotide resembling the consensus CCAAT sequence, we estimated the relative affinity of the protein(s) for three DNA sequences, Ret 1, another retina-specific region (Ret 2), and a consensus sequence for a known CCAAT-box factor NF1. A competition assay was set up in which binding of retinal proteins to labeled Ret 1 oligonucleotide was inhibited by unlabeled oligonucleotides. Fig. 6, A and B, shows individual inhibition experiments. The reproducibility of this effect is shown in Fig. 6C in which the averaged results from three separate experiments of this type are plotted as the optical density of complex c against the amount of unlabeled inhibitor. Less than a 5 × excess of unlabeled Ret 1 was sufficient to inhibit 50% of the formation of complex c, whereas a 300 × excess of unlabeled Ret 2, a more upstream retina-specific binding site, gave less than 50% inhibition. An oligonucleotide corresponding to a consensus binding site for NF1 gave partial inhibition, but even a 100 × excess did not reduce binding to less than 50%.

Nuclear extracts of rat retinas at different ages were incubated with oligonucleotide Ret 1 and the products analyzed by gel electrophoresis to determine if the retina specific binding varied in parallel with the transcription of the opsin gene. The results are shown graphically for a single experiment in Fig. 7A. The amounts of complex c formed were measured by densitometric scanning of autoradiograms from a series of such experiments and the results plotted in Fig. 7B. The results are expressed using a constant amount of retinal nuclear extract. Essentially identical results have been obtained when the amount of extract added was normalized by measuring the amount of binding to a presumably housekeeping and invariant sequence corresponding to the binding site for OCT 1 (data not shown). The complex was detected as early as PN1 and the amount increased over 2 weeks of postnatal development.

To estimate the size and complexity of proteins giving rise to complex c, we carried out two types of experiment. In the first, we fractionated rat and bovine retinal nuclear extract by sodium dodecyl sulfate-polyacrylamide electrophoresis and transferred the proteins to nitrocellulose filters. When the
Nuclear Protein Binding Site Upstream of the Rat Opsin Gene

Fig. 6. Competition experiments using labeled oligonucleotide Ret 1 and increasing amounts of unlabeled oligonucleotides in the presence of rat retina extract. The arrow indicates the retina-specific complex c. A, left panel, gel retardation of Ret 1 with the addition of unlabeled Ret 1. Right panel, competition by unlabeled oligonucleotide NF1. C, For a series of experiments, including those shown in A and B, the intensity of complex c formation was measured densitometrically from the autoradiograms. These values were plotted as a percentage of the intensity with no inhibitor added. Points and error bars represent the means and standard deviations of three experiments.

Filters were incubated with radiolabeled oligonucleotide Ret 1, a single band with an apparent molecular size of 40 kilodaltons was detected in both species (Fig. 8A). To test the specificity of this binding, the reaction was carried out in the presence of differing amounts of unlabeled Ret 1 or the unrelated oligonucleotide Ret 2. As shown in Fig. 8B, a 10-fold excess of Ret 1 reduced the intensity of binding and a 200-fold excess completely abolished it, whereas a 200-fold excess of Ret 2 had minimal effect. Other oligonucleotides known to interact with other retinal nuclear proteins (27) also had no effect on Ret 1 binding (data not shown).

To analyze the size of the native proteins giving rise to complex c, we fractionated bovine retina nuclear proteins by fast protein liquid chromatography gel filtration. When individual fractions were assayed by gel retardation, complex c forming activity was found in a broad peak with a molecular size centered at 100 kilodaltons relative to the elution positions of known globular proteins (Fig. 9A). Southwestern blot analysis of the same fractions showed that the 40-kilodalton band coincided with the complex c forming ability (Fig. 9B). Bands corresponding to complexes a and b were formed by proteins with a peak activity in fraction 12 of Fig. 9A, sug-
suggested that these complexes are formed by proteins different to those giving complex c.

**DISCUSSION**

At present little is known about the molecular events occurring between the last mitosis of a neuroblast and the onset of expression of differentiated neural properties. We have identified a developmentally regulated retina-specific DNA:protein interaction in the region immediately upstream of a gene, opsin, that is expressed early in the differentiation pathway of the rod photoreceptor. The protein(s) responsible for this interaction are candidates for tissue-specific factors involved in the regulation of this neural gene.

The retinal specificity of the interaction was defined by both DNase I protection and gel retardation assays. Although the oligonucleotides showed some minor bands with extracts from all tissues examined, only retina gave the major complex (complex c of Fig. 4A). The binding activity responsible for this major gel retardation band migrated as a single band of 40 kilodaltons as detected by Western blots. This binding activity was found in a peak eluted from a gel filtration column with a molecular size of approximately 100 kilodaltons. This suggests that the binding is due to a single species of polypeptide but that the native form of the binding activity is an oligomer of these subunits or the binding subunit in combination with other polypeptide chains.

The DNase I protection extended over the sequence CTAAATTAG and at least one of the flanking GCCCCC direct repeats. Although these repeats were not essential for complex c formation, at least the downstream one may be involved in the interaction, because an oligonucleotide lacking this sequence required much greater quantities of nuclear extract for complex formation (compare the results for Ret 1a and Ret 1b in Fig. 5A). The binding site contains a CTAAATTAG motif that resembles the CCAAT box consensus sequence. Indeed, in the human opsin gene the equivalent site has a sequence CCAATTAG (Fig. 1). The mouse opsin gene also has a CCAATTAG sequence (22). A number of proteins have been described that bind to CCAAT boxes, including the factor NF 1. There are a number of differences between these previously described factors and those binding to this region of the opsin gene. These differences include reduction of binding to a mutated pentanucleotide (CTAAC), the negligible effect of other mutations (to CTCAAT and GATAAT), and the weak inhibition by NF1 oligonucleotide indicate that this is a novel retina-specific binding factor. The only other homology that has been detected for the Ret 1 binding site is with the consensus binding site for certain homeobox proteins, CYYNATTAY. Whether or not the Ret 1 region serves as a functional CCAAT box, or whether the protein binding to Ret 1 is related to any homeobox protein, will require further analysis.

Because over 90% of the cells of the rat retina are rod photoreceptors (7) detection of the binding activity in extracts almost certainly reflects the presence of these factors in rods. We cannot yet determine whether they are also present in other cell types. Both the tissue distribution and the developmental appearance of the proteins have been measured as binding activity. Until we have an independent assay for the presence of this protein(s), we cannot distinguish a regulated appearance of the molecule from selective activation by post-translational modifications such as those known for a number of DNA binding proteins (see for example Refs. 23 and 24).

We have shown previously that opsin RNA transcripts can be detected as early as PN1 (7). That complex c formation is important not just for tissue specificity, but also for the developmental regulation of opsin gene expression, is suggested by the detection of increasing amounts of complex c using extracts from rat retinas at later stages of differentiation. The correlation between transcription rate and amount of complex c formation is not perfect. Nuclear runoff assays have suggested that opsin transcription is at adult levels by PN10-PN14 (7). The adult levels of complex c formation had not been reached by PN20, suggesting that, during development, other proteins and DNA regions become rate-limiting for transcription and the protein(s) involved in complex c formation may also have additional functions.

The retina-specific binding site described here is unlikely to be the only region involved in opsin regulation. We have preliminary evidence that additional retina-specific binding sites can be identified within a 2-kilobase pair region upstream from the opsin gene (27). Although the number of sites regulating the tissue-specific and temporal expression of opsin has yet to be defined precisely, all the necessary elements appear to be within this 2-kilobase pair region. Preliminary results using 2-kilobase pair of human opsin promoter coupled to a lacZ reporter gene inserted into transgenic mice gave rod specific and temporally regulated expression (25).

Other proteins involved in visual transduction are expressed with a very similar time course to that of opsin (7, 26). It will be of great interest to determine whether the opsin gene and these other genes share some transcription factors or have factors whose activity is under common regulation. Similarly, cone opsins are evolutionarily related to rod opsin but are expressed in a different cell type, the cone photoreceptor. Determining the factors that regulate the expression of cone opsins may help clarify the mechanisms involved in the cell type specific regulation of the opsin gene family.

**Acknowledgments**—We thank J. Schiffman and A. Butler for excellent technical assistance and H. Stopka and R. Brown for help with the manuscript.

**REFERENCES**

Nuclear Protein Binding Site Upstream of the Rat Opsin Gene