Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4851–4855, August 1984
Cell Biology

Isolation and nucleotide sequence of the gene encoding human rhodopsin

(amino acid sequence/membrane protein structure/vision/evolutionary conservation)

JEREMY NATHANS AND DAVID S. HOGNESS*

Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305

Contributed by David S. Hogness, May 2, 1984

ABSTRACT **We have isolated and completely sequenced the gene encoding human rhodopsin. The coding region of the human rhodopsin gene is interrupted by four introns, which are located at positions analogous to those found in the previously characterized bovine rhodopsin gene. The amino acid sequence of human rhodopsin, deduced from the nucleotide sequence of its gene, is 348 residues long and is 93.4% homologous to that of bovine rhodopsin. Interestingly, those portions of the polypeptide chain predicted to form loops on the cytoplasmic face of rhodopsin are perfectly conserved between the human and bovine proteins.**

Visual pigments consist of an apoprotein, opsin, covalently linked to a small conjugated chromophore, 11-*cis* retinal, or in some cases 11-*cis* dehydroretinal. Photon absorption by the visual pigments initiates visual excitation by causing an 11-*cis* to all-*trans* isomerization in the chromophore. The variation in absorption spectra observed among different visual pigments that use the same chromophore is hypothesized to arise from differences in the primary structure of the apoprotein. In particular, the visual pigments in the three types of cones that mediate human color vision are thought to differ in their absorption spectra as a result of attachment of 11-*cis* retinal to three structurally distinct cone opsins. As a corollary to this hypothesis, the well-known inherited variations in human color vision are assumed to arise from alterations in the genes encoding these cone opsins. Unfortunately, direct tests of this hypothesis have been hampered by difficulties in isolating and purifying the cone pigments from the human retina.

We wish to test the model that the visual pigments in the three types of cones resemble the rhodopsin of the rods in consisting of an apoprotein covalently linked to a common 11-*cis* retinal chromophore, that differences in the absorption spectra among the visual pigments result from structural differences among the opsins, that these structural differences result from nucleotide sequence differences among the members of a multigene family encoding the opsins, and that the genetic variations in color vision result from mutations in those members that encode the cone opsins. Our strategy consists of isolating and analyzing the human opsin genes from mutant as well as normal individuals. It depends on the assumption that these genes exhibit sufficient sequence homology that a molecular clone of one can be used as a hybridization probe to isolate the others, and that at least one member of this putative multigene family can be identified and isolated by its hybridization to the bovine rhodopsin gene, which we previously isolated and analyzed as the first step of this strategy (1). We report here the results of the second step—namely, the use of the bovine rhodopsin gene as a hybridization probe to isolate the human rhodopsin gene, and the subsequent sequence analysis of that gene.

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.

MATERIALS AND METHODS

Construction and Screening of Human Genomic DNA Libraries. Human germ line DNA was obtained from J.N., who has normal color vision. It was prepared by gently lysing 1 ml of semen in 30 ml of 0.1 M EDTA/0.1 M Tris·HCl, pH 7.5/1% sarkosyl/0.3 M 2-mercaptoethanol/100 μg of proteinase K per ml. After incubating at 50°C for 2 hr, 28 g of CsCl and 10 mg of ethidium bromide were added and the DNA was purified by equilibrium centrifugation (2). Fragments from partial *Sau*3A digests of this DNA were cloned by insertion into the *Bam*HI site of the EMBL 3 λ phage vector (3). The resulting libraries were screened by the method of Benton and Davis (4).

Other Methods. Low criteria hybridization of ^{32}P-labeled cloned DNA probes to recombinant phage plaques after their transfer to nitrocellulose filter replicas (4), or to electrophoretically fractionated fragments from restriction digests of total genomic DNA after their transfer to diazotized paper (5), were carried out at 42°C in 30% formamide/1.0 M NaCl/50 mM sodium phosphate, pH 7.0/5 mM EDTA. Filters and paper were washed under the same conditions.

Polysomal RNA was isolated from human retinas obtained at autopsy by the same method used to isolate polysomal RNA from bovine retinas (1). All other methods not described in *Results* are defined by Nathans and Hogness (1).

RESULTS

Isolation of Genomic DNA Clones Containing the Presumptive Human Rhodopsin Gene. We first examined restriction digests of total human germ line DNA isolated from J.N. for fragments containing sequences homologous to the coding region of the bovine rhodopsin gene. Thus, we hybridized at low criteria the electrophoretically separated fragments from such digests with a ^{32}P-labeled probe (6) consisting of the 1044 base pairs (bp) in this coding region plus the adjacent 40 bp and 24 bp from the 5′ and 3′ untranslated regions, respectively. [This probe represents a subcloned *Sma* I/*Bal* I fragment obtained from the bd20 cDNA segment cloned from bovine rhodopsin mRNA (1).] Both strongly and weakly hybridizing fragments were observed (data not shown). We are concerned here only with the strongly hybridizing fragments, which we initially presumed, and subsequently demonstrated (see below), to contain sequences from the human rhodopsin gene. The *Eco*RI digests yielded two such fragments of 2 and 8 kilobases (kb), while the *Hin*dIII and *Bam*HI digests yielded single fragments of 10 and 21 kb, respectively.

To isolate these fragments as recombinant molecules, genomic libraries were constructed from the germ line DNA of J.N. and screened with the cDNA probe described above.

Abbreviations: bp, base pair(s); kb, kilobase(s) or kilobase pair(s) in single- or double-stranded nucleic acids, respectively.

*To whom reprint requests should be addressed.

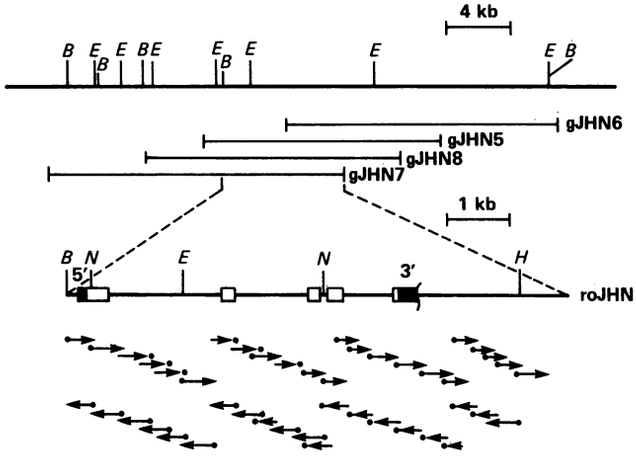

FIG. 1. Sequence homologies in human DNA segments and sequencing strategy. The four genomic segments, gJHN5–8, are aligned beneath a restriction map of the chromosomal DNA from which they derive. The sequences in these segments that are homologous to the coding sequences of the bovine rhodopsin gene are confined to the 7.0-kb *Bam*HI/*Hin*dIII fragment, roJHN, and are indicated by the five open blocks on the map of roJHN. The lines connecting these blocks represent the four introns in this presumptive human rhodopsin gene, and the two closed blocks represent its 5′-untranslated region (see Fig. 3) and a part of its 3′-untranslated region, whose extent has not been determined. *B*, *Bam*HI; *E*, *Eco*RI; *H*, *Hin*dIII; *N*, *Nco* I. The strategy for sequencing both strands of roJHN is shown at the bottom of the figure. Starting with roJHN subclones in both orientations in pBR322 and using a modification of the method of Frischauf *et al.* (7), we constructed two sets of overlapping deletions that fused different internal points in the roJHN segment to a flanking restriction site in the plasmid vector. Bacterial colonies harboring individual members of each set were analyzed by a rapid lysate procedure (2) and those deletions with endpoints at successive intervals of 400 bp were selected for sequence analysis by the method of Maxam and Gilbert (8). DNA was end-labeled with ^{32}P either at a vector restriction site adjacent to the point of fusion, or, more rarely, at a restriction site within the roJHN insert. The arrows in the figure represent the extent of sequencing of strands oriented 3′ to 5′ from base to tip, with dots indicating whether end-labeling was 3′ (dot at base) or 5′ (dot at tip).

Fig. 1 shows four overlapping genomic segments isolated in this manner. They define 30 kb of chromosomal DNA and account for all of the strongly hybridizing fragments. Restriction fragment mapping of the four genomic segments for sequences homologous to those in the cDNA probe demonstrated that such sequences are confined to the central 7.0-kb *Bam*HI/*Hin*dIII fragment, referred to hereafter and in Fig. 1 as roJHN.

roJHN Contains Coding Sequences That Are Highly Homologous with Those in the Bovine Rhodopsin Gene. Nucleotide sequence analysis of genomic and cDNA clones derived from the 6.4-kb bovine rhodopsin gene showed that it contains a 1044-bp coding region that is interrupted by four introns (1). Our first step in determining whether the 7.0-kb roJHN contains the human rhodopsin gene was to quantitate the homology between its sequences and the coding sequences of the bovine rhodopsin gene. To this end, we determined the sequence of the 6952 bp in roJHN by the strategy indicated at the bottom of Fig. 1 and described in its legend.

A computer-generated comparison of all eight-nucleotide sequences in the bovine rhodopsin coding region with all such sequences in roJHN delineated the five highly homologous regions, indicated by the open blocks in the schematic of roJHN shown in Fig. 1. Fig. 2 shows the sequence of 1044 bp in these five regions, along with the 348 amino acids that they encode, after they have been combined by conceptual splicing out of the four introns that separate them in the genomic DNA. The figure also indicates that the sequences at putative intron–exon boundaries are typical of donor and acceptor splice sites (9).

These coding sequences in roJHN are 89.7% homologous with those in the bovine rhodopsin gene, and they generate an amino acid sequence that is 93.4% homologous with the 348 residues in bovine rhodopsin (1). Of the 108 bp substitutions, only 31% are involved in the 23 amino acid replacements shown in Fig. 2 and, of these, only 30% are nonconservative replacements (12). Furthermore, the coding regions of both genes are interrupted at precisely analogous positions by four introns of comparable lengths. (The human and bovine intron lengths are, respectively, 1.783 kb and 1.7 kb for intron 1, 1.205 kb and 1.2 kb for intron 2, 0.116 kb and 0.117 kb for intron 3, and 0.833 kb and 0.85 kb for intron 4.) As expected, conservation of intron sequences is considerably less than for coding sequences, decreasing sharply with distance from the intron–exon boundary. Averaging the sequence homologies for the eight ends of the four introns, we find 85% homology for the 10 bp adjacent to the boundary, 60% for the next 10 bp, and 48% for the next 46 ± 26 bp (range), which represents the limit of the known sequences in the bovine introns (ref. 1; unpublished observations).

The 5′ End and Abundance of the mRNA from roJHN. Additional evidence that roJHN contains the human rhodopsin gene derives from the S1 nuclease protection and cDNA extension experiments (13) shown in Fig. 3, which were designed both to determine the abundance of the roJHN mRNA in human retinas and to map its 5′ end. The DNA probe for the S1 protection experiments consisted of the 5′ proximal *Nco* I/*Bam*HI fragment of roJHN (Fig. 1) that was 5′-end-labeled with ^{32}P at the *Nco* I site located 102 bp downstream from the initiator methionine codon (Fig. 2). After high criteria hybridization of this probe with polysomal RNA from human retinas, as well as with control RNAs from bovine retinal polysomes and *Drosophila* embryos, and subsequent S1 nuclease digestion, the protected ^{32}P-labeled DNA fragments were assayed by autoradiography after gel electrophoresis adjacent to a sequence ladder of the same 5′-end-labeled *Nco* I/*Bam*HI fragment. Fig. 3 shows that only the human retinal RNA efficiently protected the probe from S1 digestion. Given that the bovine rhodopsin mRNA present in the bovine retinal RNA control is 82% homologous with the probe over this region yet did not protect it, we are confident that the protecting polysomal RNA from the human retina is transcribed from roJHN.

A quantitative analysis of the level of S1 protection provided by this mRNA (data not shown) indicates that it accounts for 0.5% of mRNAs in the human retina. This value matches the 0.5% abundance of rhodopsin mRNA in the bovine retina estimated by *in vitro* translation and immunoprecipitation (14). This correspondence indicates that roJHN encodes the human rhodopsin of the abundant rods rather than an opsin from the much rarer cones.

Shorter exposure of the gel used in Fig. 3 reveals two major S1 protected fragments whose lengths differ by two nucleotides and indicate that transcription of the roJHN gene is initiated at each of the two "mRNA start sites" shown in Fig. 2. These two transcription initiation sites were confirmed by the results of the cDNA extension experiments, also shown in Fig. 3. The substrates for reverse transcriptase extension consisted of an 80-nucleotide primer 5′-end-labeled at the *Nco* I site described above and hybridized at high criteria to human retinal polysomal RNA and to the two control RNAs. Only the human retinal RNA provided an adequate template for cDNA extension, which yielded two products whose lengths correspond to the indicated mRNA start sites.

If we take the furthest upstream of these two start sites as the 5′ end of the roJHN gene, then that gene contains a 95-bp 5′-untranslated region (Fig. 2), which is to be compared with

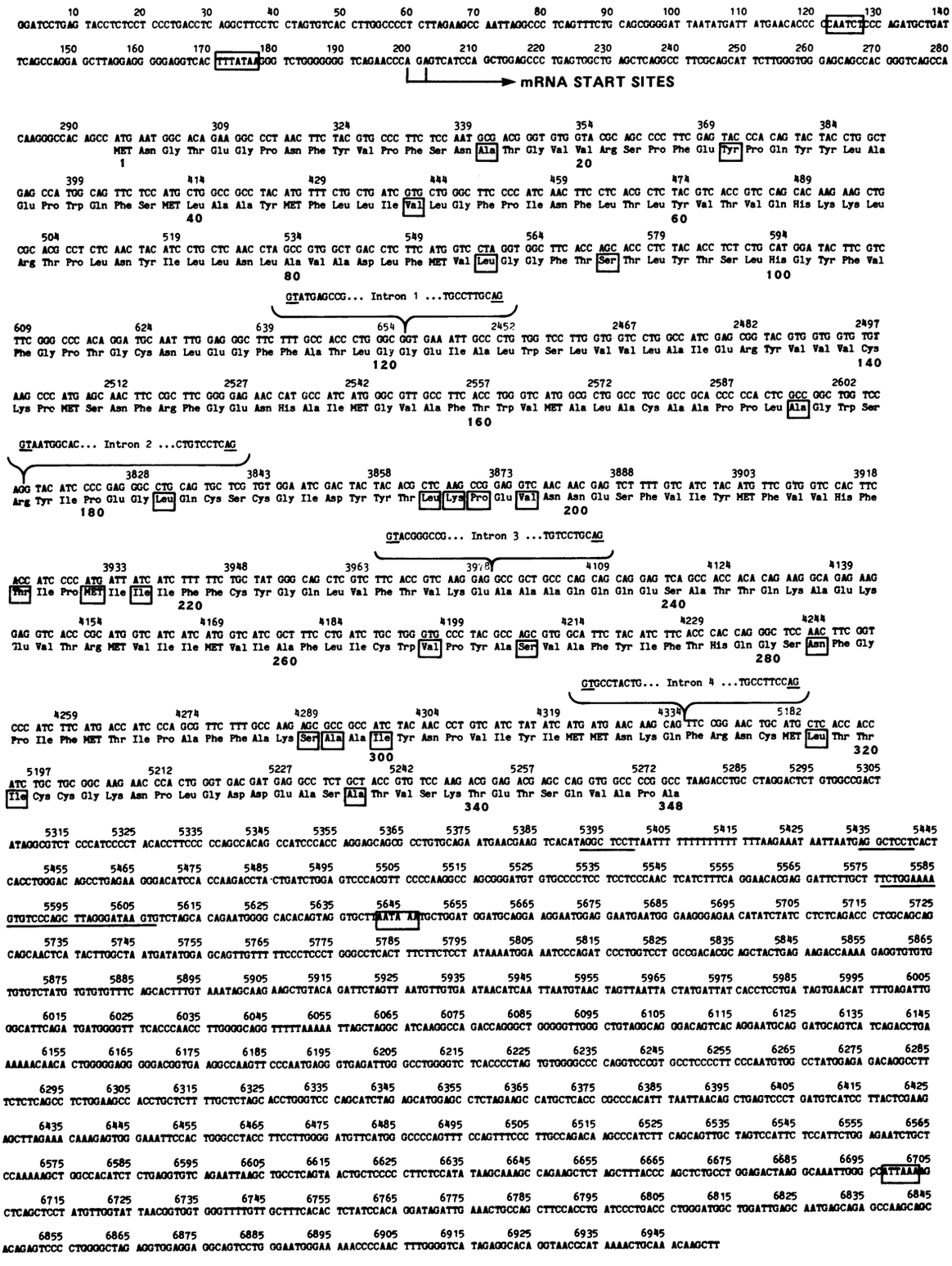

FIG. 2. Sequence of roJHN. Position 1 is the first nucleotide of the roJHN *Bam*HI site, and position 6953 is the last nucleotide of its *Hin*dIII site. To convert this scale to that in which *N*th residue upstream from the 5′ end of the gene (position 200 in the figure) is assigned position $-N$, and the *N*th residue in the gene, counting from its 5′ end, is assigned position $+N$, subtract 200 from position numbers <199 and subtract 199 from position numbers >200. The boxed bases beginning at positions 122 and 171 in the figure are the CAAT box (9) and the Goldberg–Hogness or TATA box (9, 10), respectively. The boxed residues in the deduced amino acid sequence are those that differ from the bovine rhodopsin sequence (1). In bovine rhodopsin, the corresponding amino acids are Lys^{16}, Ala^{26}, Met^{49}, Phe^{88}, Thr^{93}, Val^{173}, Met^{183}, Pro^{194}, His^{195}, Glu^{196}, Thr^{198}, Ile^{213}, Leu^{216}, Val^{218}, Leu^{266}, Gly^{270}, Asp^{282}, Thr^{297}, Ser^{298}, Val^{300}, Val^{318}, Leu^{321}, and Thr^{335}. To save space, only sufficient intron sequences are shown to indicate homology to consensus donor and acceptor splice sequences (9). The complete roJHN sequence may be obtained from the National Institutes of Health–Genbank or will be furnished by us upon request. The underlined sequences starting at positions 5392 and 5434 indicate the direct 8-bp repeats flanking the 34-bp element described in the text and whose polyadenylate tract is in the antisense strand. The other underlined sequence starting at position 5577 represents the perfectly conserved 32-bp element; the boxed sequences starting at positions 5642 and 6698 represent possible 3′ cleavage polyadenylylation signals (11).

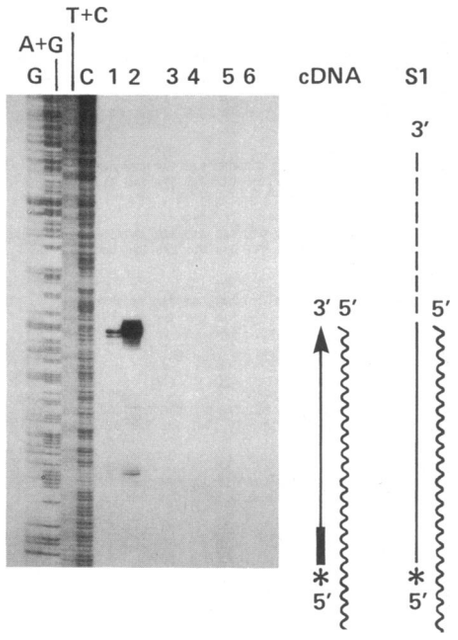

FIG. 3. S1 nuclease protection and cDNA extension analysis. At the left is a Maxam–Gilbert sequence ladder (8) of the 5' proximal *Nco* I/*Bam*HI fragment from roJHN, 5'-end-labeled at the *Nco* I site. Lanes 2, 4, and 6 display the products of S1 nuclease digestion after hybridization of the same 5'-end-labeled *Nco* I/*Bam*HI fragment to 40 μg of human retina polysomal RNA (lane 2), 40 μg of bovine retina polysomal RNA (lane 4), and 40 μg of *Drosophila melanogaster* embryo RNA (lane 6). Lanes 1, 3, and 5 display the reverse transcriptase extension products generated by priming with a 5'-end-labeled 80-nucleotide *Nco* I/*Hae* III fragment, extending upstream from the above mentioned *Nco* I site and annealed to 20 μg of human retina polysomal RNA (lane 1), 20 μg of bovine retina polysomal RNA (lane 3), and 20 μg of *Drosophila melanogaster* embryo RNA (lane 5). Exposure was for 48 hr. Hybridizations were carried out at 50°C for 3 hr in 0.04 M NaCl/0.04 M Pipes, pH 6.4/80% (vol/vol) formamide. See ref. 13 for other aspects of the methods. In the diagram at the right, the RNA and DNA strands are represented by wiggly and straight lines, respectively, with the asterisk indicating the position of the ^{32}P label.

the 96-bp length of that region in the bovine rhodopsin gene (1). These two regions exhibit 77% sequence homology.

DISCUSSION

Comparison of the Human and Bovine Rhodopsin Gene Structures. Our observations on the coding sequence homologies between roJHN and the bovine rhodopsin gene and on the abundance of their mRNAs lead to the conclusion that roJHN encodes human rhodopsin. Under the assumption that the mechanisms regulating the transcription of the human and bovine rhodopsin genes are conserved, we have examined the DNA flanking the 5' ends of the two genes for conserved sequences that may partake in that regulation. To that end, sequence homologies were computed for successive, overlapping 10-bp segments initiated at each of the 161 bp of 5' flanking DNA for which the sequence is known in both genes (Fig. 2; ref. 1 and unpublished data) and plotted against the position in the human DNA of the midpoints of the segments. This plot revealed six significant peaks centered 30 bp (peak 1), 56 bp (2), 77 bp (3), 96 bp (4), 129 bp (5), and 148 bp (6) upstream from the 5' end of the human gene (position 200 in Fig. 2). Each peak contains a perfectly conserved 13 ± 2 bp (range) sequence and is defined by 14 ± 5 (range) contiguous 90%–100% homology values. Peaks 1 and 3 include, respectively, the TATA and CAAT boxes (Fig. 2) found at these positions in many genes (9, 10) and are therefore unlikely to define regulatory sequences specific to rhodopsin gene expression. We suggest that the other four peaks identify such sequences.

The length of the 3'-untranslated region in the human rhodopsin gene has not been defined. Two possible 3' cleavage polyadenylylation signals, A-A-T-A-A-A and A-T-T-A-A-A (11), are located 367 and 1423 bp, respectively, downstream from the 3' end of the coding region (Fig. 2). The longer distance corresponds to the ≈1400-bp length of the 3'-untranslated region in the bovine gene (1). Unfortunately, only the first 500 bp of this region have been sequenced (1), preventing intergenic comparison of its 3'-terminal sequences.

Intergenic sequence comparison of the 500 bp downstream from the coding region reveals a curious 85-bp (bovine) to 88-bp (human) zone of virtually random 27% homology that interrupts the 71% sequence homology observed for the rest of this region. In the human gene, this zone begins 125 bp downstream from the coding region with a 34-bp element that contains a 16-residue polyadenylate tract near one end and is flanked by 8-bp direct repeats (Fig. 2). This structure is akin to that exhibited by the "processed" class of pseudogenes that are thought to arise by reverse transcription of poly(A)$^+$ mRNA and by insertion of the resulting DNA via a staggered-break mechanism that provides a direct repeat of the target sequence at the ends of the insert (15). Although the random zone undoubtedly arose from multiple evolutionary events, it is tempting to think that one of these was such an insertion event.

Another curious feature of this 500-bp region is a perfectly conserved 32-bp sequence that, in the human gene, is located 34 bp upstream from the A-A-T-A-A-A sequence. In the bovine gene, this 32-bp sequence is found 37 bp upstream from an A-A-T-A-A-A sequence located in approximately the same position 351 bp downstream from the coding region.

Inferences Drawn from the Conservation of Rhodopsin Structure. We previously used the method of Steitz *et al.* (16) to assign amino acid residues to the seven transmembrane segments present in the model for the secondary structure of bovine rhodopsin (1). Virtually identical assignments were obtained when the same method was applied to the amino acid sequence deduced for human rhodopsin from the coding sequence of its gene (Fig. 2). A similar topography

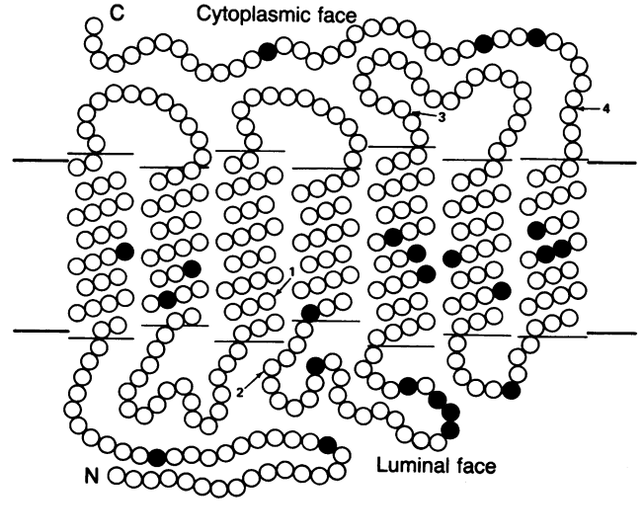

FIG. 4. Schematic representation of rhodopsin in the lipid bilayer. The drawing is a modified version of the schematic of Hargrave *et al.* (17) to depict their model for the secondary structure of bovine rhodopsin. Those positions at which the human and bovine amino acid sequences differ are marked by black solid circles. The numbered arrows indicate the positions of the four introns in both the human and bovine rhodopsin genes.

for the transmembrane segments of bovine rhodopsin was obtained by Hargrave *et al.* (17) and is shown in Fig. 4. The cytoplasmic face of the protein includes (*i*) several sites close to the carboxyl terminus that are subject to light-dependent phosphorylation by rhodopsin kinase (18) and (*ii*) a catalytic site that promotes GTP–GDP exchange by Transducin/G protein (19). Part of the catalytic site may reside in the third cytoplasmic loop, because thermolysin cleavage of bovine rhodopsin at residue 239 inactivates the catalytic activity without altering rhodopsin's spectral properties (20). The black dots in Fig. 4 indicate those residues that differ between the bovine and human proteins. It is apparent that the three cytoplasmic loops and each of the adjacent 10–12 membrane-embedded residues are entirely free of amino acid substitutions. The probability of this occurring by chance, if the variation between the two proteins were distributed at random throughout their length, is 0.0005. This localized conservation suggests an important role for the cytoplasmic loops in visual pigment function.

The human and bovine opsins also display a high degree of conservation at those residues that have been shown to be covalently modified in bovine rhodopsin (21). The human rhodopsin sequence includes Lys^{296}, the site of attachment of 11-*cis* retinal, six out of seven serine and threonine residues near the carboxyl terminus that are targets of phosphorylation, and the two asparagine-X-threonine glycosylation sites, the second of which differs with respect to its central amino acid.

We previously noted that introns 2, 3, and 4 of the bovine rhodopsin gene are located immediately distal to the codons for predicted transmembrane segments (1), as is shown in Fig. 4. Those segments, although differing in primary structure, share a common pattern: each consists of a stretch of ≈20 hydrophobic amino acids followed by a single positively charged residue (either lysine or arginine). The bovine rhodopsin gene sequence shows that intron 2 interrupts the codon for Arg^{177}, the residue that terminates the fourth transmembrane segment, and that introns 3 and 4 begin 3 bp beyond the codons for Lys^{231} and Lys^{311}, respectively, residues that terminate the fifth and seventh transmembrane segments. This intron–exon arrangement is precisely conserved in the human rhodopsin gene (Fig. 2). The same pattern is also found at the 3′ junctions of those exons encoding membrane-anchoring segments of surface immunoglobulins and histocompatibility antigens (22, 23). This arrangement of exons encoding integral and transmembrane proteins supports the hypothesis of gene evolution by juxtaposition of exons encoding structural domains (24). Should this pattern hold generally for genes encoding transmembrane proteins it may aid in secondary structure prediction.

We thank Drs. Lubert Stryer, Jonathan Rothbard, David Austin, and Douglas Brutlag for enthusiastic discussions. This work was supported by a National Institutes of Health grant to D.S.H., and J.N. is supported by a fellowship from the Medical Scientist Training Program of the National Institutes of Health.

1. Nathans, J. & Hogness, D. S. (1983) *Cell* **34,** 807–814.
2. Davis, R. W., Botstein, D. & Roth, J. R. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
3. Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
4. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
5. Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R. & Wahl, G. M. (1979) *Methods Enzymol.* **68,** 220–242.
6. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
7. Frischauf, A.-M., Garoff, H. & Lehrach, H. (1980) *Nucleic Acids Res.* **8,** 5541–5549.
8. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
9. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
10. Goldberg, M. L. (1979) Dissertation (Stanford Univ., Stanford, CA).
11. Nevins, J. R. (1983) *Annu. Rev. Biochem.* **52,** 441–465.
12. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), pp. 14–16; 168–175.
13. Muskavitch, M. A. T. & Hogness, D. S. (1982) *Cell* **29,** 1041–1051.
14. Schechter, I., Burstein, Y., Zemell, R., Ziv, E., Kantor, F. & Papermaster, D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2654–2658.
15. Hollis, G. F., Hieter, P. A., McBride, O. W., Swan, D. & Leder, P. (1982) *Nature (London)* **296,** 321–326.
16. Steitz, T. A., Goldman, A. & Engelman, D. M. (1982) *Biophys. J.* **37,** 124–125.
17. Hargrave, P. A., McDowell, J. H., Curtis, D. R., Wang, J. K., Juszczak, E., Fong, S.-L., Mohana Rao, J. K. & Argos, P. (1983) *Biophys. Struct. Mech.* **9,** 235–244.
18. Kühn, H., Cook, J. H. & Dreyer, W. J. (1973) *Biochemistry* **12,** 2495–2502.
19. Fung, B. K.-K. & Stryer, L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2500–2504.
20. Kühn, H. & Hargrave, P. A. (1981) *Biochemistry* **20,** 2410–2417.
21. Hargrave, P. A. (1982) in *Progress in Retinal Research*, eds. Osborne, N. & Chador, G. (Pergamon, Oxford), pp. 1–51.
22. Yamawaki-Kataoka, Y., Nakai, S., Miyata, T. & Honjo, T. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2623–2627.
23. Kvist, S., Roberts, L. & Dobberstein, B. (1983) *EMBO J.* **2,** 245–254.
24. Gilbert, W. (1978) *Nature (London)* **271,** 501.