The EMBO Journal vol.6 no.2 pp.443–451, 1987

An opsin gene that is expressed only in the R7 photoreceptor cell of *Drosophila*

Karl J.Fryxell and Elliot M.Meyerowitz

Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125, USA

Communicated by M.Ashburner

We have used two techniques to isolate and characterize eye-specific genes from *Drosophila melanogaster*. First, we identified genes whose expression is limited to eyes, photoreceptor cells, or R7 photoreceptor cells by differential screening with [^{32}P]cDNAs derived from the heads of mutant flies that have reduced amounts of these tissues and cells (Microcephalus, glass^3, and sevenless, respectively). Secondly, we identified opsin genes by hybridization with synthetic [^{32}P]oligonucleotides that encode domains that have been conserved between some opsin genes. We found seven clones that contain genes expressed only in the eye or optic lobes of *Drosophila*; three are expressed only in photoreceptor cells. One is expressed only in R7 photoreceptor cells and hybridizes to some of the previously mentioned oligonucleotides. The complete DNA sequence of the R7-specific opsin gene and its 5′ and 3′ flanking regions was determined. It is quite different from other known *Drosophila* opsin genes, in that it is not interrupted by introns and shares only 37–38% amino acid identity with the proteins encoded by these genes. The predicted protein structure contains many characteristics that are common to all rhodopsins, and the sequence differences help to identify four domains of the rhodopsin molecule that have been conserved in evolution.

Key words: Drosophila/eye/photoreceptor/opsin/rhodopsin

Introduction

The *Drosophila* eye is composed of about 800 facets, or ommatidia, packed in a precise hexagonal array. Each ommatidium contains eight photoreceptor cells, which are arranged in a constant, asymmetric pattern (Ready *et al.*, 1976). By convention, each photoreceptor cell has been assigned an identification number (R1–R8) based on its position within the ommatidium. The outer photoreceptors (R1–R6) have similar action spectra, spectral adaptation properties, and sensitivity to certain mutations, while the inner photoreceptors (R7 and R8) are different from each other and from R1–R6 in these respects (Harris *et al.*, 1976). Each photoreceptor cell bears a specialized organelle called the rhabdomere, an elongated stack of densely packed microvilli within which the transduction of light into electrical signals is initiated by the rhodopsin molecule. The three classes of eye photoreceptor cell (R1–R6, R7, and R8) probably each express a different rhodopsin molecule(s) (Quinn *et al.*, 1974; Harris *et al.*, 1976; Nathans *et al.*, 1986a,b; Cowman *et al.*, 1986).

Genetic mosaic and immunofluorescence experiments suggest that cells in the developing *Drosophila* eye disk undergo their final cell division, then become committed to form a particular cell type, and then differentiate in rapid succession (Ready *et al.*, 1976; Zipursky *et al.*, 1984). The developmental determination of cell type may be closely related to rhodopsin expression; thus, the opsin genes of *Drosophila* are of interest both as developmental markers and in studies of visual transduction. Similarly, genes whose expression is limited to other cell types in the eye may be of interest as developmental markers. We have identified genes in each category.

We report here the identification of seven clones that contain genes expressed only in the eye or optic lobes of *Drosophila*. One gene encodes an opsin and two more are specifically expressed in photoreceptor cells. The remaining genes are apparently expressed in other cell types in the eye. We present a detailed characterization of the opsin gene; we find that it is expressed only in the R7 class of photoreceptor cells.

Results

Identification of eye-specific clones

Levy *et al.* (1982) differentially screened a *Drosophila* genomic library with two preparations of [^{32}P]poly(A)$^+$RNA: the first derived from fly heads, and the second derived from fly bodies. They isolated 25 'head-specific' clones and grouped them into classes by cross-hybridization. We identified which of these classes are 'eye-specific' by differential screening with [^{32}P]cDNAs derived from fly heads and from the heads of mutant flies that are nearly eyeless.

Mutations that decrease the eye size of *Drosophila* generally have incomplete expressivity or poor viability (Lindsley and Grell, 1968; Ransom, 1979). We found that *Mc/Tab trans*-heterozygotes (*Mc*: Microcephalus; *Tab*: Transabdominal; see Materials and methods) have much smaller eyes than *Mc*/+ and much better viability than *Mc/Mc*. *Mc* and *Tab* both have cytologically visible breakpoints in 89E. *Tab/Tab* flies do not have small eyes. *Mc/Tab* flies were maintained as a balanced stock and their eye size was further reduced by selection (see Materials and methods). We chose to use *w* flies (*w*: white; Lindsley and Grell, 1968), which have eyes of normal size, as a positive control. We isolated poly(A)$^+$ RNA from the heads of *Mc/Tab* and *w* flies, and synthesized [^{32}P]cDNAs from these templates. We screened bacteriophage plaques (Benton and Davis, 1977) of seventeen head-specific λ clones in duplicate (λcDm506, 507, 512, ... 559; see Materials and methods). At least eight λ clones gave reproducibly stronger signals with one [^{32}P]cDNA or the other.

Seven clones were more strongly labeled by *w* cDNA (Figure 1A and B) and are putatively eye-specific (*Mc/Tab* flies may fail to express optic lobe-specific genes as well as eye-specific genes, since the normal develoment of the optic lobe depends on the transmission of information from the eyes; Meyerowitz and Kankel, 1978). These seven clones (λcDm507, λcDm512, λcDm547, λcDm549, λcDm551, λcDm557, and λcDm559) map to six chromosomal loci (Levy *et al.*, 1982; see Table I). λcDm507 and λcDm547 contain the same or similar genes (see the footnote to Table I).

The eighth clone (λcDm527) was more strongly labeled by

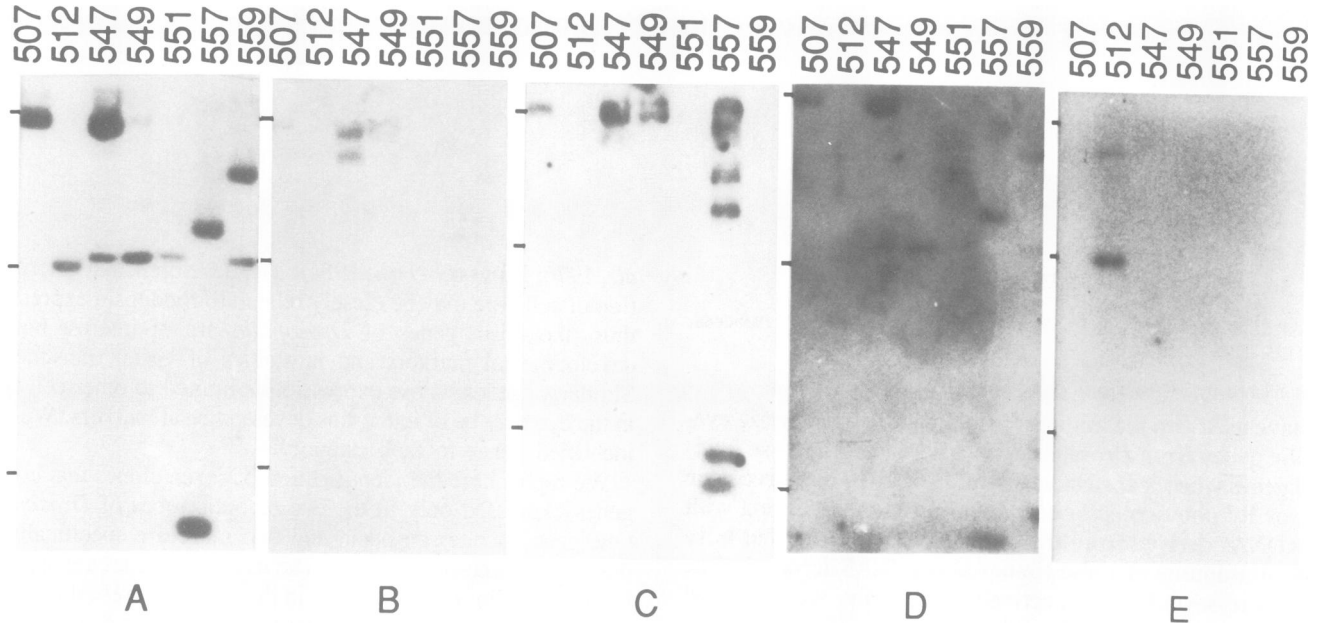

Fig. 1. Characterization of eye-specific genes. Minipreps of recombinant λ clone DNAs were digested with *Bam*HI, subjected to electrophoresis in agarose gels, blotted to nitrocellulose (Southern, 1979), and hybridized with five different [^{32}P]DNA probes: **A**, [^{32}P]cDNA derived from the heads of *w* (Canton-S) *D. melanogaster* adults. **B**, [^{32}P]cDNA from the heads of *Mcp Sab^{2} Mc/p^{p} Mcp Sab^{2} Tab* flies. **C**, [^{32}P]cDNA from the heads of *gl^{3}* flies. Multiple bands are due to incomplete digestion with *Bam*HI. **D**, [^{32}P]cDNA from the heads of *w sev^{LY3}* flies. This nitrocellulose filter was washed 2 × 60 min at 51°C, to eliminate cross-hybridization with the more abundant cDNA derived from the *ninaE* locus. **E**, [^{32}P]D0.3 (an oligodeoxynucleotide: 5′-ACCCTCCGGCACATACCT-3′). Genomic insert numbers are shown above each lane. The mobility of selected molecular length standards (*Eco*RI *Hin*dIII double-digest fragments of λcI857S7) are indicated by marks to the left of each section (from top to bottom: 21.226 kb, 5.148 kb, and 2.027 kb; Daniels *et al.*, 1983).

Mc/Tab cDNA (not shown). λcDm527 may be expressed in a tissue that makes up a greater fraction of the head in *Mc/Tab* flies, such as the mouth parts or the brain.

Identification of photoreceptor cell-specific clones

We identified which classes are 'photoreceptor cell-specific' in a similar way, using the heads of mutant flies that have reduced numbers of photoreceptor cells. We chose the *gl^{3}* mutation (*gl*: glass) for this experiment, since *gl^{3}* eyes do not contain rhabdomeres, send few retinular cell fibers to the brain, do not express several photoreceptor cell-specific antigens, and do not produce an electroretinogram in response to light (Pak *et al.*, 1969; Meyerowitz and Kankel, 1978; Garen and Kankel, 1983; Zipursky *et al.*, 1984). However, the number of photoreceptor cell soma in the *gl^{3}* eye is unknown. This number can be easily determined in the ocelli, where photoreceptor cell soma are reasonably large and pigmented, and are not intermixed with other pigmented cell types (Hertweck, 1931; Toh *et al.*, 1971). These cells were counted in whole mounts under oil immersion; their numbers were reduced ~10-fold and their size was reduced several-fold in *gl^{3}* ocelli (not shown). *gl^{3}* eyes are slightly smaller than normal and contain ommatidia that are irregular but recognizable (Meyerowitz and Kankel, 1978).

When [^{32}P]cDNA derived from *gl^{3}* heads was hybridized with seven clones thought to be eye-specific (λcDm507–559), four of the clones were strongly labeled and three were not detectably labeled (Figure 1C). One unlabeled clone (λcDm512) encodes an opsin gene (this work). Another unlabeled clone (λcDm559) complements a mutation that specifically affects photoreceptor cells (*trp*: transient receptor potential; Montell *et al.*, 1985). In similar experiments, we found another unlabeled clone (λcDm6018, see Table I) that encodes another opsin gene (*ninaE*: neither inactivation nor adaptation E; O'Tousa *et al.*, 1985; Zuker *et al.*, 1985), in agreement with the results of O'Tousa *et al.* (1985). Thus, the *gl^{3}* mutation may be used to provisionally identify genes whose function is limited to photoreceptor cells. The experiment shown in Figure 1C yields five

Table I. Eye-specific clones

Clone	Chromosomal location	Eye-specific	Photoreceptor-specific	R7-specific	Opsin gene
λcDm507	66D	+	–	–	–
λcDm512	92CD	+	+	+	+
λcDm547	66D	+	–	–	–
λcDm549	46E	+	–	–	–
λcDm551	28A	+	+	–	–
λcDm557	72BC	+	–	–	–
λcDm559	99C	+	+	–	–
λcDm6018	92B	+	+	n.t.	+

Tissue- and cell-specificities are based on experiments in which RNA was isolated from the heads and bodies of flies of various genotypes. The identification of opsin genes is based on hybridization to specific oligonucleotides and DNA sequencing. λcDm507–559 were isolated by differential screening (see Results) from a collection of seventeen head-specific clones originally isolated by Levy *et al.* (1982). λcDm507 and λcDm547 contain the same or similar genes, since they are located at 66D on the third chromosome, partially cross-hybridize, encode 1.7 kb mRNA(s) (Levy *et al.*, 1982), have similar patterns of expression (Figure 1), and also contain 1.7 kb and 2.0 kb *Bam*HI fragments (that are not hybridized by the probes shown in Figure 1). λcDm6018 was isolated by screening a genomic library (Maniatis *et al.*, 1978) for clones that hybridize to both D0.1 and D0.3 oligonucleotides (see Materials and methods), and also hybridize to head cDNA but not body cDNA. Clones selected by these criteria (λcDm6018 and λcDm6020) belong to a single class that represents the *ninaE* gene (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985; Table I), based on restriction maps of the clones and the length, specificity and abundance of the transcript that they encode. Chromosomal locations were determined by *in situ* hybridization by Levy *et al.* (1982), O'Tousa *et al.* (1985), and Zuker *et al.* (1985). n.t., not tested.

Fig. 2. Expression of an R7-specific opsin gene. Poly(A)$^+$ RNAs were separated by electrophoresis in a formaldehyde/agarose gel, blotted to nitrocellulose (Seed, 1982), and hybridized to [^{32}P]sDm6034a [that had been labeled by nick-translation (Rigby *et al.*, 1977)]. sDm6034a encodes 105 bp of the coding sequence (at the poorly conserved carboxy terminus), and the entire 3′ untranslated region from Dm6030 (see Materials and methods, and Figure 5). Each lane contains poly(A)$^+$ RNA that was isolated from 40 μg of RNA: **Body**, poly(A)$^+$ RNA isolated from the bodies of decapitated *w* (Canton-S) *D. melanogaster* adults. **Mc**, poly(A)$^+$ RNA from the heads of *Mcp Sab*^{2} *Mc/p*^{p} *Mcp Sab*^{2} *Tab* flies. **gl**, poly(A)$^+$ RNA from the heads of *gl*^{3} flies. **ora**, poly(A)$^+$ RNA from the heads of *ora*^{JK84} flies. **sev**, poly(A)$^+$ RNA from the heads of *w sev*^{LY3} flies.

additional results: λcDm551 is probably photoreceptor-cell specific, and λcDm507, λcDm547, λcDm549, and λcDm557 are likely to be expressed in other cell type(s) in the eye.

Identification of R7-specific clones

Very few R7 photoreceptor cells are present in the eyes of *sev*^{LY3} mutant flies (*sev*: sevenless; Harris *et al.*, 1976; Campos-Ortega *et al.*, 1979; Stark and Carlson, 1985). The R1−R6 and R8 classes of photoreceptor cells appear normal in this mutant. R7 precursor cells apparently develop into cone cells in *sev*^{LY3} flies (Tomlinson and Ready, 1986). In contrast, the *ora*^{JK84} mutation (*ora*: outer rhabdomeres absent) specifically eliminates part of the structure and function of R1−R6 photoreceptor cells: it eliminates R1−R6 opsin mRNA (*ninaE*) and rhabdomeres, but not R1−R6 cell bodies or axons (Harris *et al.*, 1976; Cowman *et al.*, 1986). We hybridized [^{32}P]cDNA, prepared from the heads of *w sev*^{LY3} flies, with seven clones thought to be eye-specific (λcDm507−559). One clone passed this screen: λcDm512 failed to hybridize to *w sev*^{LY3} cDNA (Figure 1D). λcDm512 does hybridize to *w* cDNA (Figure 1A).

Poly(A)$^+$ RNA blots were prepared from fly heads and bodies of various genotypes, and hybridized with [^{32}P]sDm6034a, which encodes the 3′ end of the λcDm512 transcript. *ora*^{JK84} head RNA was strongly labeled, but the following were not: *w sev*^{LY3} head RNA, *gl*^{3} head RNA, *Mc/Tab* head RNA, and *w* body RNA (Figure 2). After longer exposures than shown in Figure 2, *w sev*^{LY3} head RNA showed barely detectable labeling. These results imply that the gene contained within λcDm512 is expressed in R7 photoreceptor cells.

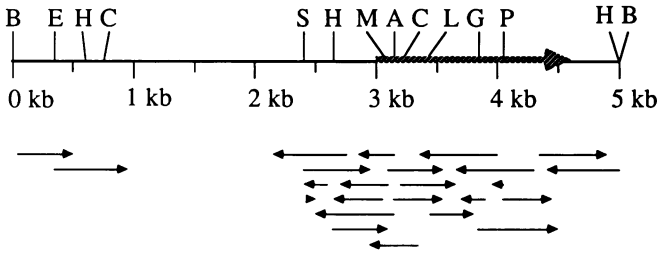

Fig. 3. Structure of an R7-specific opsin gene. The restriction map, transcript and DNA sequence determinations from Dm6030 are shown. DNA sequence determinations are indicated by arrows below the restriction map, and transcript is shown as a large striped arrow within the restriction map. Restriction sites were mapped by single and double restriction digests. The map location and structure of the transcript were determined by probing DNA blots with [^{32}P]cDNA, by electrophoresis of the mRNA, by probing RNA blots with strand-specific probes, by DNA sequencing, and by primer extension (see Results). A, *Ava*I; B, *Bam*HI; C, *Cla*I; E, *Eco*RI; G, *Bgl*I; H,*Hin*dIII; L, *Bgl*II; M, *Mlu*I; P, *Pst*I; S, *Sal*I; no sites for *Kpn*I.

Identification of opsin clones

We found that an antisense bovine opsin RNA probe hybridizes to an eye-specific, poly(A)$^+$ *Drosophila* RNA, while a sense bovine opsin probe does not; however, the hybrid has a very low melting temperature (similar results were independently obtained by Zuker *et al.*, 1985). Because of the low sequence conservation within the opsin gene family, we synthesized oligonucleotide hybridization probes that encode domains of the *Drosophila ninaE* opsin that are homologous to bovine opsin (Nathans and Hogness, 1983; O'Tousa *et al.*, 1985; Zuker *et al.*, 1985). For convenience, we will refer to four of these oligonucleotides as D0.1, D0.3, D0.5, and D0.6 (see Materials and methods).

We hybridized [^{32}P]D0.1 with bacteriophage plaques of 17 head-specific clones (λcDm506−559) and none of the clones were labeled. We hybridized [^{32}P]D0.3, [^{32}P]D0.5, and [^{32}P]D0.6 at a lower temperature (24°C, see Materials and methods) with DNA blots of seven clones thought to be eye-specific (λcDm507−559). A 5.0 kb *Bam*HI fragment in Dm512 hybridized to D0.3 and D0.5, but not D0.6 (Figure 1E); no other restriction fragment hybridized to more than one oligonucleotide. This *Bam*HI fragment (Dm6030, see Materials and methods) encodes a transcript that is specifically expressed in R7 photoreceptor cells (Figures 1−3), so the homology between *ninaE* and Dm6030 may be significant. In fact, it is significant, as shown by DNA sequencing (see below).

Dm6030: transcript structure and DNA sequence

Single and double restriction digests of Dm6030 were subjected to electrophoresis, blotted to nitrocellulose (Southern, 1979), and hybridized to [^{32}P]cDNA derived from the heads of *w* flies. A 2.4 kb *Hin*dIII fragment was labeled, while the adjacent 2.0 kb *Hin*dIII fragment was not (Figure 3). Restriction fragments on both sides of the *Ava*I site were labeled; restriction fragments on both sides of the *Pst*I site were also labeled (Figure 3). RNA blots (Seed, 1982) did hybridize to [^{32}P]cRNA synthesized from sDm6034a but not from sDm6034b, indicating that transcription is rightward (see Materials and methods, and Figure 3).

The DNA sequence of a 3.8 kb portion of Dm6030 was determined (Figure 3), including the entire transcribed region and 5′ and 3′ flanking regions (Figure 4). This sequence was compared to the *Drosophila ninaE* opsin gene (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985), by scoring a dot (within a dot matrix) if at least 12 out of 18 nucleotides were identical. Significant homology

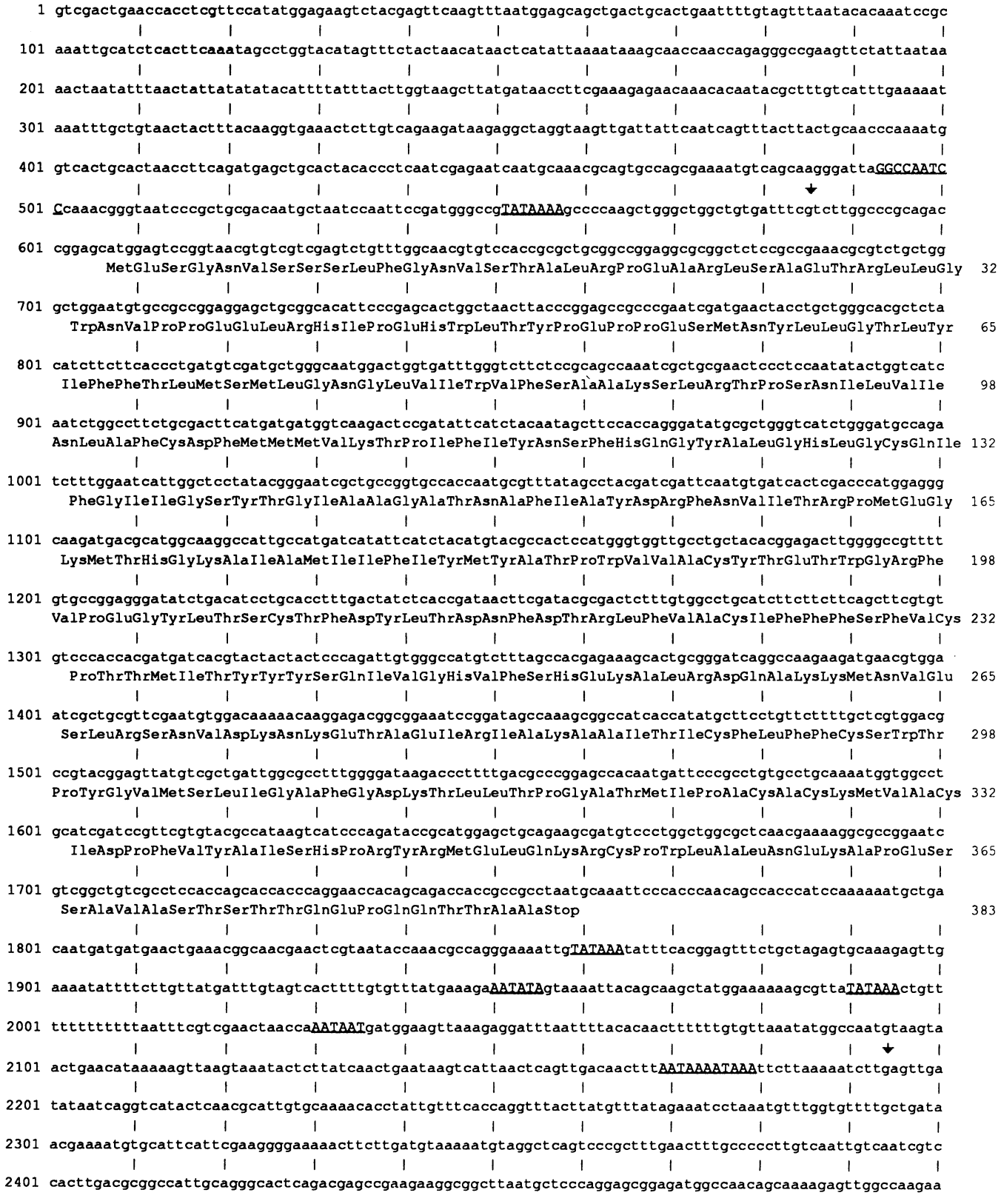

Fig. 4. Complete DNA sequence of an R7-specific opsin gene and its flanking regions. The derived amino acid sequence is shown below the DNA sequence and is numbered separately (on the right). The DNA sequence starts at the *SalI* site in Figure 3 and proceeds rightward for 2500 nucleotides on the mRNA-congruent strand. The entire sequence was determined on both strands. Consensus DNA sequences (see Results) are capitalized and underlined. The approximate locations of the 5′ and 3′ ends of the transcript are indicated by arrows (see Results).

Table II. Comparison of amino acid sequences deduced from *Drosophila* opsin genes

Opsin gene pair[a]	First sequence		Second sequence		Both sequences	
	# of gaps	Total gap length	# of gaps	Total gap length	Identical pairs	Percent identity
R7:R8	4	4	5	6	147	38%
R7:R1−R6	2	2	6	12	140	37%
R8:R1−R6	0	0	3	8	258	68%

Pairs of deduced amino acid sequences were aligned with a computer program, using essentially the algorithm of Gotoh (1982).
[a]R7 denotes the amino acid sequence deduced from Dm6030 (Figure 4), R8 denotes the amino acid sequence deduced from a *Drosophila* opsin gene that is predominantly expressed in R8 photoreceptor cells (Cowman *et al.*, 1986), and R1−R6 denotes the amino acid sequence deduced from a *Drosophila* opsin gene that is predominantly expressed in R1−R6 photoreceptor cells (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985).

to *ninaE* was confined to about 900 bp within Dm6030 and is consistent with rightward transcription (not shown). The longest open reading frame in Dm6030 (383 codons) generally follows *Drosophila* codon usage (Cherbas *et al.*, 1986), and corresponds closely in length and orientation to the *ninaE* homology. Another long open reading frame (294 codons) does not follow typical *Drosophila* codon bias, and is on the opposite DNA strand.

Alignment of the deduced amino acid sequences of *ninaE* and Dm6030 (using essentially the algorithm of Gotoh, 1982) suggested that translation begins at nucleotide 607 (Figure 4). This alignment was extended at the DNA level by the same algorithm, and nucleotide 573 (Figure 4) aligned with the transcription initiation site in *ninaE* (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985). In primer extension experiments, the length of the reverse transcript primed by the oligonucleotide D0.8 (see Materials and methods) corresponded to mRNA initiation at nucleotide 584; minor products were 1−2 nucleotides longer. Since eukaryotic mRNAs generally contain a purine after their 5′ cap (Breathnach and Chambon, 1981), the mRNA encoded by Dm6030 probably begins at nucleotide 585, or possibly at nucleotide 580 (Figure 4).

An exact match to the consensus TATA sequence (TATAAAA: Breathnach and Chambon, 1981) begins at nucleotide −34 with respect to the probable mRNA start site (Figure 4). TATA sequences are present at about −26 to −34 within most eukaryotic genes that are transcribed by RNA polymerase II, and if present specify the mRNA start site (Breathnach and Chambon, 1981; Dierks *et al.*, 1983; Snyder and Davidson, 1983). A close match to the consensus CAAT sequence (GGPyCAATCT: Benoist *et al.*, 1980) begins at −92 (Figure 4). Homology to the CAAT sequence is present at about −70 to −90 within some eukaryotic genes, including a *Drosophila* opsin gene (*ninaE*: O'Tousa *et al.*, 1985); this sequence is required for maximal transcription of mammalian globin genes (Dierks *et al.*, 1983).

No introns are evident within the coding sequence, by three criteria. First, alignment of the deduced amino acid sequence with that deduced from *ninaE* cDNAs (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985) does not reveal gaps larger than 6 codons (Table II). Second, the reading frame is not interrupted by frame shifts or stop codons (Figure 4). Third, no close match to any consensus for intron splice donor sequences (Mount, 1982; Snyder *et al.*, 1982) is present within the coding sequence. Such matches are present outside the coding sequence (i.e. nucleotides 355−363 and 2092−2100 in Figure 4). However, these sequences are not likely to represent introns, based on the length of the 5′ untranslated region of the mRNA, the location of polyadenylation signals, the location of possible intron splice acceptor sequences, and the length of the mRNA (see below).

The length of the mRNA was determined by electrophoresis in formaldehyde/agarose gels, using RNAs of known sequence as standards. We obtained a length of 1.5 kb for both ora^{JK84} (Figure 2) and *w* (Canton-S) flies, by a least squares fit to the mobility versus log(length) of the standards (Figure 2). Many eukaryotic mRNAs terminate in a poly(A) tail, which may be 50 nucleotides in length in insects (Vournakis *et al.*, 1974), and is usually added about 20 nucleotides after the DNA sequence AATAAA (Nevins, 1983; Figure 4, nucleotides 2169−2174 and 2174−2179). This implies that the mRNA encoded by Dm6030 is 1.65 kb in length, in reasonable agreement with the length estimated by electrophoresis (Snyder and Davidson, 1983). Sequences differing from AATAAA by a single A→T transversion apparently function as polyadenylation signals in a few genes (Nevins, 1983), and are present at four sites within the 3′ untranslated region (Figure 4).

Dm6030: deduced amino acid sequence

The amino acid sequence deduced from the Dm6030 coding sequence is shown in Figures 4 and 5. This sequence contains 383 amino acids, with a total mass of 41 150 daltons. Nearly half of the amino acids (48%) are non-polar, and are likely to form seven transmembrane α-helical segments (Figure 5), as shown by normalized hydropathicity plots (Eisenberg *et al.*, 1984; see the legend to Figure 5). The amino acid sequences deduced from two known *Drosophila* opsin genes are also hydrophobic, and are similar in length (373 and 381 amino acids, respectively: O'Tousa *et al.*, 1985; Zuker *et al.*, 1985; Cowman *et al.*, 1986).

When the amino acid sequences deduced from the above *Drosophila* opsin genes were aligned with the amino acid sequence shown in Figure 4, we observed 37−38% amino acid identity (Table II). 31% of the amino acids in the Dm6030 sequence are identical in all three genes (Figure 5). The Dm6030 sequence contains characteristics that are common to all rhodopsins, including seven transmembrane segments; conserved sequences in the first cytoplasmic loop, the second extracellular loop, and the seventh transmembrane segment (particularly a central lysine residue); and a carboxy-terminal segment that is serine- and threonine-rich (Nathans *et al.*, 1986a; see also Figure 5). Further, Dm6030 encodes a transcript that is specifically expressed in one class of *Drosophila* photoreceptor cells. For these reasons, we conclude that Dm6030 contains a *Drosophila* opsin gene. We will refer to this gene as *Rh92CD*, reflecting its chromosomal location (Levy *et al.*, 1982; see Table I).

Discussion

R7 photoreceptor cells constitute a small fraction of the fly head. Why would an R7-specific opsin gene pass a differential screen that requires abundant transcription (Levy *et al.*, 1982)? In part, this is due to the remarkable abundance within photoreceptor cells of rhodopsin and its messenger RNA (Harris *et al.*, 1977; O'Tousa *et al.*, 1985; Zuker *et al.*, 1985; Cowman *et al.*, 1986). In part, it is due to slight cross-hybridization with the more abundant *ninaE* transcript (we found that single-stranded, but not double-stranded, probes cross-hybridize unless higher stringency washes are used; see the legend to Figure 1).

We did not find any introns in *Rh92CD*. Two other *Drosophila* opsin genes have three and four introns, respectively, but share only one intron position (Cowman *et al.*, 1986). Intron positions are not conserved within the *Drosophila* opsin gene family.

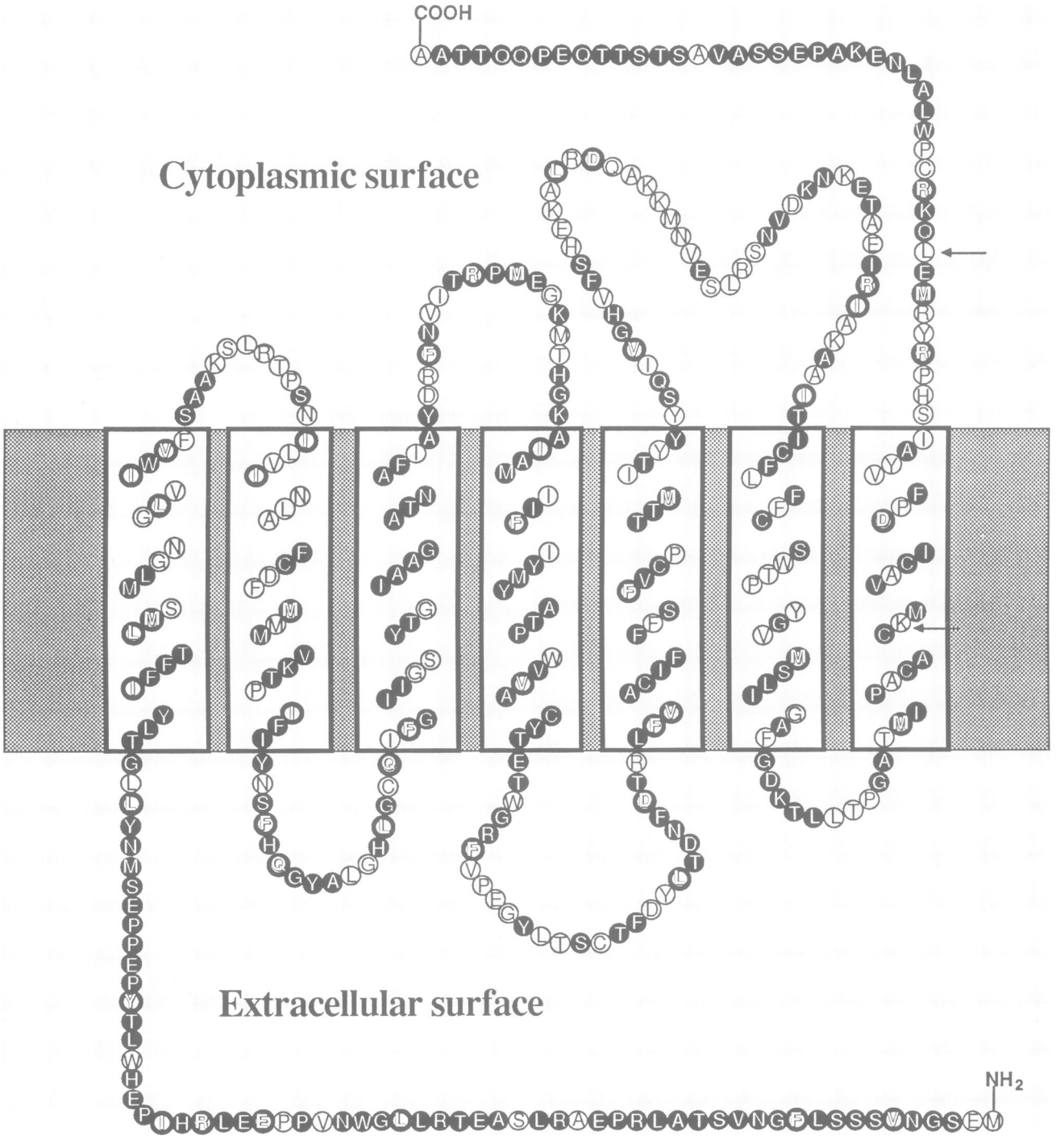

Fig. 5. An R7-specific opsin. The deduced amino acid sequence is shown in the one-letter code. Conserved amino acids are represented by light circles, conservative substitutions by partially filled circles, and non-conserved amino acids by solid circles. Putative transmembrane α-helices are enclosed within large black boxes. The chromophore attachment site (Lys^{328}) and the codon (Leu^{349}) that corresponds to the *Pst*I site in Dm6030 (Figure 3) are indicated by arrows. Each amino acid was compared to the corresponding position in two other *Drosophila* opsin genes (Table II), and scored by the comparison in which it was least conserved. Amino acid substitutions were considered conservative if they are favored by natural selection, as shown by comparison of closely related proteins (Dayhoff *et al.*, 1978). These substitutions are limited to the following groups: (Met,Ile,Leu,Val); (Leu,Phe); (Phe,Tyr); (Lys,Arg); (Arg,Trp); (Arg,His); (His,Gln); (His,Asn); (Asn,Asp); and (Asp,Glu,Gln). The location and number of transmembrane α-helices was predicted with a computer program, using the decision rules and hydrophobicity values of Eisenberg *et al.* (1984). The carboxyl terminus of the protein is assumed to be located on the cytoplasmic side of the membrane, as shown for bovine rhodopsin (Hargrave *et al.*, 1983).

However, the human opsin gene family has four conserved intron positions (Nathans *et al.*, 1986a), one of which is present in a *Drosophila* opsin gene (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985). Similarly, intron positions are not conserved within the *Drosophila* actin gene family (Fyrberg *et al.*, 1981), while three intron positions are conserved within the mammalian actin gene family (Zakut *et al.*, 1982; Hamada *et al.*, 1982), as if introns had been added and removed more frequently in the line of descent leading to *D. melanogaster* than in the line of descent leading to modern mammals.

Functionally important parts of genes change relatively slowly in evolution, and may be identified by sequence comparisons

in favorable cases (Dayhoff *et al.*, 1972). Comparison of several moderately diverged genes is likely to be the most informative (closely related genes may share relatively unimportant sequences, while distantly related genes may differ within important sequences). Given the moderate homology between *Rh92CD* and two other *Drosophila* opsin genes, we should be able to identify which domains of *Drosophila* rhodopsin are most important for function(s) common to all three opsins.

Figure 5 and Table II summarize the results of alignments between *Rh92CD* and two other *Drosophila* opsins. We have highlighted amino acids in Figure 5 (white circles) if they are conserved in comparison to both genes, and have used a rather restricted definition of 'conservative substitutions' (see legend to Figure 5). Nevertheless, small 'conserved' segments may not be significant. We plotted homology scores for all possible 10 amino acid segments of each alignment, and found four consistent peaks of homology that represent conserved domains within the *Drosophila* opsin gene family. The boundaries of these domains are approximately as follows: Lys^{87} to Phe^{105} (the first cytoplasmic loop, extending into the membrane); Val^{199} to Cys^{207} (part of the second extracellular loop); His^{251} to Ser^{269} (part of the third cytoplasmic loop); and Pro^{335} to Leu^{349} (the base of the carboxy terminus, extending into the membrane; Figure 5). The three cytoplasmic conserved domains have the highest homology scores, have a similar length (15−19 amino acids) and apparent net charge (+1 to +3), and in two cases include nonpolar segments that may extend into the lipid bilayer (Figure 5). Perhaps these domains are folded together and have a single function.

Overall, the cytoplasmic loops are the best conserved region of *Drosophila* opsins (Figure 5) and human opsins (Nathans *et al.*, 1986a). The chromophore attachment site (Lys^{328}) is strictly conserved, but the adjacent amino acids are not (Figure 5). In fact, Pappin and Findlay (1984) observed a small variable domain adjacent to lysine in the seventh transmembrane segment of rhodopsin, based on amino acid sequences from four mammalian species. Although the seventh transmembrane segment has been rather well conserved over the long evolutionary period that separates flies and cows (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985), sequence comparisons within species suggest that, apart from the central lysine, only the domain located in the 'cytoplasmic' part of the membrane is strictly conserved (Figure 5; see also Nathans *et al.*, 1986a).

Human rhodopsin genes are not randomly distributed in the genome: there is a cluster of 2−4 opsin genes on the X chromosome (the remaining two opsin genes are on separate autosomes; Nathans *et al.*, 1986b). The high frequency of red-green color blindness among humans is caused, in part, by this chromosomal distribution (and the resulting intergenic recombination, etc.), as shown by an elegant series of molecular experiments (Nathans *et al.*, 1986a,b). Similarly, *Drosophila* opsin genes are not randomly distributed: three opsin genes are present on the right arm of the third chromosome between 91D and 92CD (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985; Cowman *et al.*, 1986; this work). There are probably more than three opsin genes in *Drosophila*. In fact, several physiological subclasses of R7 photoreceptor cells have been discovered within the past ten years (Harris *et al.*, 1976; Hardie, 1986), and a different opsin gene may be expressed by each subclass. *D. melanogaster* can be trained to move toward particular colors of light (Quinn *et al.*, 1974). The genetics of color preference may help to elucidate the function of the various *Drosophila* opsin genes and photoreceptor cells.

Drosophila genetics can be used to delete tissues, or cells, that would be difficult to remove in any other way (Harris *et al.*, 1976). One may use this approach to isolate genes (by differential screening) that would be difficult to isolate in any other way. However, the effects of a developmental mutation on the transcription of other genes may be difficult to predict. We chose three developmental mutations on the basis of their morphological phenotypes, and examined the effect of these mutations on the mRNA levels of seven genes (Table I). The genes that are not expressed by *Mc*, gl^3, and sev^{LY3} mutants fall into completely overlapping subsets, as if *Mc*, gl^3, and sev^{LY3} simply delete the eye, photoreceptor cells, and R7 photoreceptor cells, respectively.

In this work, we have used specific oligonucleotides to isolate genes on the basis of limited homology, and differential screens between specific mutants to isolate genes on the basis of their tissue specificity. Both techniques are likely to be applicable to the isolation of other genes.

Materials and methods

Drosophila strains

w (Canton-S), *w* sev^{LY3}, and ora^{JK84} (Harris *et al.*, 1976) *D. melanogaster* were obtained from S.Benzer. *Mcp* Sab^2 *Mc*/p^p *Mcp* Sab^2 *Tab* flies (Lindsley and Grell, 1968; Celniker and Lewis, 1984) and gl^3 flies (Lindsley and Zimm, 1985) were from the Caltech *Drosophila* Stock Collection. The *Mcp* Sab^2 *Mc*/p^p *Mcp* Sab^2 *Tab* stock is balanced, primarily because *Mc* and *Tab* are (almost always) lethal in homozygotes and have cytologically visible breakpoints within 89E. We selected flies with the smallest number of ommatidia in the *Mcp* Sab^2 *Mc*/p^p *Mcp* Sab^2 *Tab* stock for several generations. This experiment did increase the expressivity of *Mc*, as if one or more additional loci influence its expressivity.

Recombinant DNA clones

This laboratory uses unique symbols for all λ and plasmid vectors, for each cloned DNA insert, and for the species from which the insert was derived. This system of recombinant clone nomenclature was invented by D.S.Hogness (see Meyerowitz and Martin, 1984; Chang and Meyerowitz, 1986). We have added symbols for the following vectors: 'my' for M13mp18, 'n' for pUC18 (Norrander *et al.*, 1983), and 's' for DOA 3.8 [a 3.0 kb plasmid that confers kanamycin resistance, contains an SP6 promoter 154 bp from a pUC8 polylinker, and allows detection of inserts with a Lac^- phenotype (C.Chang, R.E.Pruitt and E.M.M., unpublished results)].

λcDm506, 507, 512, ... 559 were obained from J.Manning, and retain their original insert numbers (Levy *et al.*, 1982). All genomic λ clones in this report are random shear fragments of *D. melanogaster* (Canton-S) DNA, inserted into the *Eco*RI site of Charon 4 (Maniatis *et al.*, 1978). The 5.0 kb *Bam*HI fragment from λcDm512 was subcloned into pUC18 (nDm6030) and M13mp18 (myDm6030). Within nDm6030, a 1.0 kb *Pst*I fragment (at the extreme right in Figure 3) was subcloned from nDm6030 to DOA 3.8 (sDm6034), in both orientations: antisense RNA can be synthesized by *in vitro* transcription of sDm6034a with SP6 RNA polymerase, while sense RNA is synthesized from sDm6034b.

Oligonucleotides

All synthetic oligodeoxynucleotides were 18 nucleotides in length, were synthesized with an Applied Biosystems 380A synthesizer in the Caltech Microchemical Facility (Hunkapiller *et al.*, 1984), and were used without further purification. The oligodeoxynucleotides D0.1, D0.3, D0.5, and D0.6 encode parts of the first cytoplasmic loop and the second extracellular loop of the *Drosophila* opsin, *ninaE* (O'Tousa *et al.*, 1985; Zuker *et al.*, 1985). These domains are conserved between *ninaE* and bovine rhodopsin. The oligodeoxynucleotide sequences are as follows:

D0.1: 5′-TCGTGCGGTATTGACTAC-3′
D0.3: 5′-ACCCTCCGGCACATACCT-3′
D0.5: 5′-GGGCGTGCGCAGTGATTT-3′
D0.6: 5′-GTTAGCGGGCGTGCGCAG-3′

Oligodeoxynucleotides were end-labeled with T4 polynucleotide kinase (Maniatis *et al.*, 1982), and separated from [γ-^{32}P]ATP by centrifugation through a column of Sephadex G-25-300 (Neal and Florini, 1973). End-labeled oligonucleotides were hybridized to nitrocellulose DNA blots (Southern, 1979) at a concentration of 1−2 × 10^6 c.p.m./ml for 18−24 h at 24°C in 1.08 M NaCl, 6 mM EDTA, 60 mM $NaPO_4$ pH 7.4, 0.5 mg/ml SDS, 0.2 mg/ml Ficoll, 0.2 mg/ml polyvinyl pyrrolidone, 0.2 mg/ml bovine serum albumin, washed 4 × 20 min at 24°C in 1.08 M NaCl, 6 mM EDTA, 60 mM $NaPO_4$ pH 7.4, and autoradiographed at −80°C with Cronex 4 film and a Cronex Lightning-Plus intensifying screen (DuPont).

RNA analysis

Fly heads and bodies were separated by sieving frozen fly fragments through nylon screens (Oliver and Phillips, 1970). RNA was isolated by homogenization in 5 M guanidinium thiocyanate, followed by ultracentrifugation through 5.7 M CsCl (Chirgwin *et al.*, 1979). Poly(A)$^+$ RNA was isolated by affinity chromatography on oligo(dT) cellulose columns (Maniatis *et al.*, 1982). [^{32}P]cDNA was synthesized from poly(A)$^+$ RNA with reverse transcriptase (from avian myeloblastosis virus), in the presence of actinomycin D (Verma *et al.*, 1972) and RNasin (de Martynoff *et al.*, 1980), and separated from [α-^{32}P]dATP by centrifugation through a column of Sephadex G-50-80 (Neal and Florini, 1973). [^{32}P]cDNA was hybridized to DNA blots (Southern, 1979) at a concentration of 5×10^5 c.p.m./ml for 18–24 h at 42°C in 50% formamide, 1 M NaCl, 50 mM Tris pH 8.0, 10 mg/ml SDS, 5% dextran sulfate, 0.1 mg/ml denatured salmon sperm DNA, washed 4 × 20 min at 24°C in 0.36 M NaCl, 2 mM EDTA, 20 mM $NaPO_4$ pH 7.4, 2 mg/ml SDS, then washed 2 × 30 min at 50°C in 0.09 M NaCl, 0.5 mM EDTA, 5 mM $NaPO_4$ pH 7.4, 5 mg/ml SDS, then washed 2 × 20 min at 24°C in 0.018 M NaCl, 0.1 mM EDTA, 1 mM $NaPO_4$ pH 7.4, and autoradiographed at −80°C with Cronex 4 film and a Cronex Lightning-Plus intensifying screen (DuPont).

RNA samples were denatured with formamide and formaldehyde, separated by electrophoresis in formaldehyde/agarose gels, and blotted to nitrocellulose (Seed, 1982). Linearized plasmids whose sequence is known were transcribed *in vitro* with SP6 RNA polymerase, to produce [^{35}S]RNA length standards: jAt3011 (Chang and Meyerowitz, 1986) digested with *Hin*dIII gives a 2.566 kb transcript; jAt3011 digested with *Xba*I gives a 1.753 kb transcript; sDm6034a (this work) digested with *Eco*RI gives a 1.165 kb transcript.

Primer extension

The location of the 5′ end of the R7 opsin messenger RNA was identified by primer extension (Ghosh *et al.*, 1980; Snyder *et al.*, 1982). The primer was D0.8, a synthetic oligodeoxynucleotide (5′-CAGGTAGTTCATCGATTC-3′) that hybridizes near the 5′ end of the messenger RNA. The length of the resulting end-labeled cDNA was measured on an 80 cm sequencing gel, using a dideoxy DNA sequence ladder (primed by D0.8) as a length standard, and correcting for the extra phosphate group at the 5′ end of the cDNA (Sollner-Webb and Reeder, 1979).

DNA sequence analysis

DNA sequences were determined by the chemical degradation method of Maxam and Gilbert (1980) and the chain termination method of Sanger *et al.* (1977). Chain termination sequencing was performed with the vectors M13mp18 and M13mp19 (Norrander *et al.*, 1983) and [α-^{35}S]dATP (Biggin *et al.*, 1983). Subclones for sequencing were constructed by directional cloning (Maniatis *et al.*, 1982) and the sequences extended with specific oligodeoxynucleotide primers (Strauss *et al.*, 1986).

Acknowledgements

We thank J.Manning for λcDm506–559; R.Baker, C.Chang and P.Pang for technical suggestions; C.Martin and R.Pruitt for computer programs; S.Scherer for T4 DNA ligase; R.Mundy for technical assistance; S.Horvath for synthetic oligonucleotides; and our colleagues for comments on the manuscript. K.J.F. is a fellow of the Helen Hay Whitney Foundation. This work was supported by Program Project Grant GM20927 from the NIH.

References

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127–142.

Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.

Campos-Ortega,J.A., Jurgens,G. and Hofbauer,A. (1979) *Wm. Roux Arch.*, **186**, 27–50.

Celniker,S.E. and Lewis,E.B. (1984) *Genetics*, **107**, s17.

Chang,C. and Meyerowitz,E.M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1408–1412.

Cherbas,L., Schulz,R.A., Koehler,M.M.D., Savakis,C. and Cherbas,P. (1986) *J. Mol. Biol.*, **189**, 617–631.

Chirgwin,J.M., Przybla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.

Cowman,A.F., Zuker,C.S. and Rubin,G.M. (1986) *Cell*, **44**, 705–710.

Daniels,D.L., Schroeder,J.L., Szybalski,W., Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F., Petersen,G.B. and Blattner,F.R. (1983) In Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds), *Lambda II.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 519–676.

Dayhoff,M.O., Hunt,L.T., McLaughlin,P.J. and Barker,W.C. (1972) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found., Washington DC, vol. 5, pp. D99–D111.

Dayhoff,M.O., Schwartz,R.M. and Orcutt,B.C.(1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found., Washington DC, vol. 5, supplement 3, pp. 345–352.

de Martynoff,G., Pays,E. and Vassart,G. (1980) *Biochem. Biophys. Res. Comm.*, **93**, 645–653.

Dierks,P., van Ooyen,A., Cochran,M.D., Dobkin,C., Reiser,J. and Weissmann,C. (1983) *Cell*, **32**, 695–706.

Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984) *J. Mol. Biol.*, **179**, 125–142.

Fyrberg,E.A., Bond,B.J., Hershey,N.D., Mixter,K.S. and Davidson,N. (1981) *Cell*, **24**, 107–116.

Garen,S.H. and Kankel,D.R. (1983) *Dev. Biol.*, **96**, 445–466.

Ghosh,P.K., Reddy,V.B., Piatak,M., Lebowitz,P. and Weissman,S.M. (1980) *Methods Enzymol.*, **65**, 580–595.

Gotoh,O. (1982) *J. Mol. Biol.*, **162**, 705–708.

Hamada,H., Petrino,M.G. and Kakunaga,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5901–5905.

Hardie,R.C. (1986) *Trends Neurosci.*, **9**, 419–423.

Hargrave,P.A., McDowell,J.H., Curtis,D.R., Wang,J.K., Jusczak,E., Fong,S.L., Mohanna Rao,J.K. and Argos,P. (1983) *Biophys. Struct. Mech.*, **9**, 235–244.

Harris,W.A., Stark,W.S. and Walker,J.A. (1976) *J. Physiol.*, **256**, 415–439.

Harris,W.A., Ready,D.F., Lipson,E.D., Hudspeth,A.J. and Stark,W.S. (1977) *Nature*, **266**, 648–650.

Hertweck,H. (1931) *Z. Wiss. Zool.*, **139**, 559–663.

Hunkapiller,M., Kent,S., Caruthers,M., Dreyer,W., Firca,J., Giffin,C., Horvath,S., Hunkapiller,T., Tempst,P. and Hood,L. (1984) *Nature*, **310**, 105–111.

Levy,L.S., Ganguly,R., Ganguly,N. and Manning,J.E. (1982) *Dev. Biol.*, **94**, 451–464.

Lindsley,D.L. and Grell,E.H. (1968) *Genetic variations of Drosophila melanogaster*. Carnegie Institution of Washington publication no. 627.

Lindsley,D.L. and Zimm,G. (1985) *The genome of Drosophila melanogaster part 1: genes A–K*. Drosophila Information Service vol.62.

Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–559.

Meyerowitz,E.M. and Kankel,D.R. (1978) *Dev. Biol.*, **62**, 112–142.

Meyerowitz,E.M. and Martin,C.H. (1984) *J. Mol. Evol.*, **20**, 251–264.

Montell,C., Jones,K., Hafen,E. and Rubin,G. (1985) *Science*, **230**, 1040–1043.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.

Nathans,J. and Hogness,D.S. (1983) *Cell*, **34**, 807–814.

Nathans,J., Thomas,D. and Hogness,D.S. (1986a) *Science*, **232**, 193–202.

Nathans,J., Piantanida,T.P., Eddy,R.L., Shows,T.B. and Hogness,D.S. (1986b) *Science*, **232**, 203–210.

Neal,M.W. and Florini,J.R. (1973) *Anal. Biochem.*, **55**, 328–330.

Nevins,J.R. (1983) *Annu. Rev. Biochem.*, **52**, 441–466.

Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.

Oliver,D.V. and Phillips,J.P. (1970) *Dros. Inf. Serv.*, **45**, 58.

O'Tousa,J.E., Baehr,W., Martin,R.L., Hirsh,J., Pak,W.L. and Applebury,M.L. (1985) *Cell*, **40**, 839–850.

Pak,W.L., Grossfield,J. and White,N.V. (1969) *Nature*, **222**, 351–354.

Pappin,D.J.C. and Findlay,J.B.C. (1984) *Biochem. J.*, **217**, 605–613.

Quinn,W.G., Harris,W.A. and Benzer,S. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 708–712.

Ransom,R. (1979) *J. Embryol. Exp. Morphol.*, **53**, 225–235.

Ready,D.F., Hanson,T.E. and Benzer,S. (1976) *Dev. Biol.*, **53**, 217–240.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Seed,B. (1982) In Setlow,J.K. and Hollaender,A. (eds), *Genetic Engineering: Principles and Methods*. Plenum Press, New York, NY, **4**, 91–102.

Snyder,M. and Davidson,N. (1983) *J. Mol. Biol.*, **166**, 101–118.

Snyder,M., Hunkapiller,M., Yuen,D., Silvert,D., Fristrom.J. and Davidson,N. (1982) *Cell*, **29**, 1027–1040.

Sollner-Webb,B. and Reeder,R.H. (1979) *Cell*, **18**, 485–499.

Southern,E. (1979) *Methods Enzymol.*, **68**, 152–176.

Stark,W.S. and Carlson,S.D. (1985) *Dros. Inf. Serv.*, **61**, 164–166.

Strauss,E.C., Kobori,J.A., Siu,G. and Hood,L.E. (1986) *Anal. Biochem.*, **154**, 353–360.

Toh,Y., Tominaga,Y. and Kuwabara,M. (1971) *J. Electron Microsc.*, **20**, 56–66.

Tomlinson,A. and Ready,D.F. (1986) *Science*, **231**, 400–402.
Verma,I.M., Temple,G.F., Fan,H. and Baltimore,D. (1972) *Nature New Biol.*, **235**, 163–167.
Vournakis,J.N., Gelinas,R.E. and Kafatos,F.C. (1974) *Cell*, **3**, 265–273.
Zakut,R., Shani,M., Givol.,D., Neuman,S., Yaffe,D. and Nudel,U. (1982) *Nature*, **298**, 857–859.
Zipursky,S.L., Venkatesh,T.R., Teplow,D.B. and Benzer,S. (1984) *Cell*, **36**, 15–26.
Zuker,C.S., Cowman,A.F. and Rubin,G.M. (1985) *Cell*, **40**, 851–858.

Received on 13 November 1986

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00043.