1

Opsin genes, cone photopigments, color vision, and color blindness

Lindsay T. Sharpe, Andrew Stockman, Herbert Jägle, and Jeremy Nathans

In this chapter, we introduce the molecular structure of the genes encoding the human cone photopigments and their expression in photoreceptor cells. We also consider the consequences that alterations in those genes have on the spectral sensitivity of the photopigments, the cone photoreceptor mosaic, and the perceptual worlds of the color normal and color blind individuals who possess them. Throughout, we highlight areas in which our knowledge is still incomplete.

Trichromacy. Human color vision is trichromatic; this has three consequences. First, as was recognized in the eighteenth century (e.g., Le Blon, 1722; see Birren, 1963, 1980), but only formally postulated (Grassman, 1853) and verified (Maxwell, 1855, 1860) in the nineteenth century, the number of independent variables in color vision is three. That is, all colors can be matched by just three parameters: either by the three primaries of additive light mixture – typically, violet, green, and red – or by the three primaries of subtractive pigment mixture – typically, cyan, yellow, and magenta.

Second, as intimated by Palmer (1777, 1786; see also Voigt, 1781; Walls, 1956; Mollon, 1997), definitively stated by Young (1802, 1807), and revived by Helmholtz (1852), trichromacy is not a physical property of light but a physiological limitation of the eye: All color perceptions are determined by just three physiological response systems.

Third, as pointed out by Maxwell (1855) and applied by König and Dieterici (1886), a linear transform must exist between the tristimulus color matching properties of the eye, as established by the three primaries of additive light mixture, and the spectral sensitivities of the three physiological systems mediating the matches (see Chapter 2).

The three physiological response systems are universally acknowledged to be the three types of retinal photoreceptor cell, each containing a different photopigment: the short (S)-, middle (M)-, and long (L)-wave sensitive cones.[1] These have distinct, spectral sensitivities (Fig. 1.1A) or absorption spectra (Fig. 1.1B), which define the probability of photon capture as a function of wavelength. The absorbance spectra of the S-, M-, and L-cone photopigments overlap considerably, but have their wavelengths of maximum absorbance (λ_{max}) in different parts of the visible spectrum: ca. 420, 530, and 558 nm, respectively. When estimated in vivo, the λ_{max}'s are shifted to longer wavelengths (ca. 440, 545, and 565 nm, respectively) by the transmission properties of the intervening ocular media: the yellowish crystalline lens and the macular pigment of the eye (see Chapter 2).

The individual cone photopigments are blind to the wavelength of capture; they signal only the rate at

[1]The fourth type of photoreceptor cell, the rods, contain rhodopsin as their photopigment. They are by far the most prevalent in the human retina, constituting more than 95% of all photoreceptor cells. However, they do not contribute to color vision, except under limited, twilight conditions (see section on rod monochromacy). Under most daylight conditions, where we enjoy color vision, the rod photoreceptor response is saturated by excessive light stimulation.

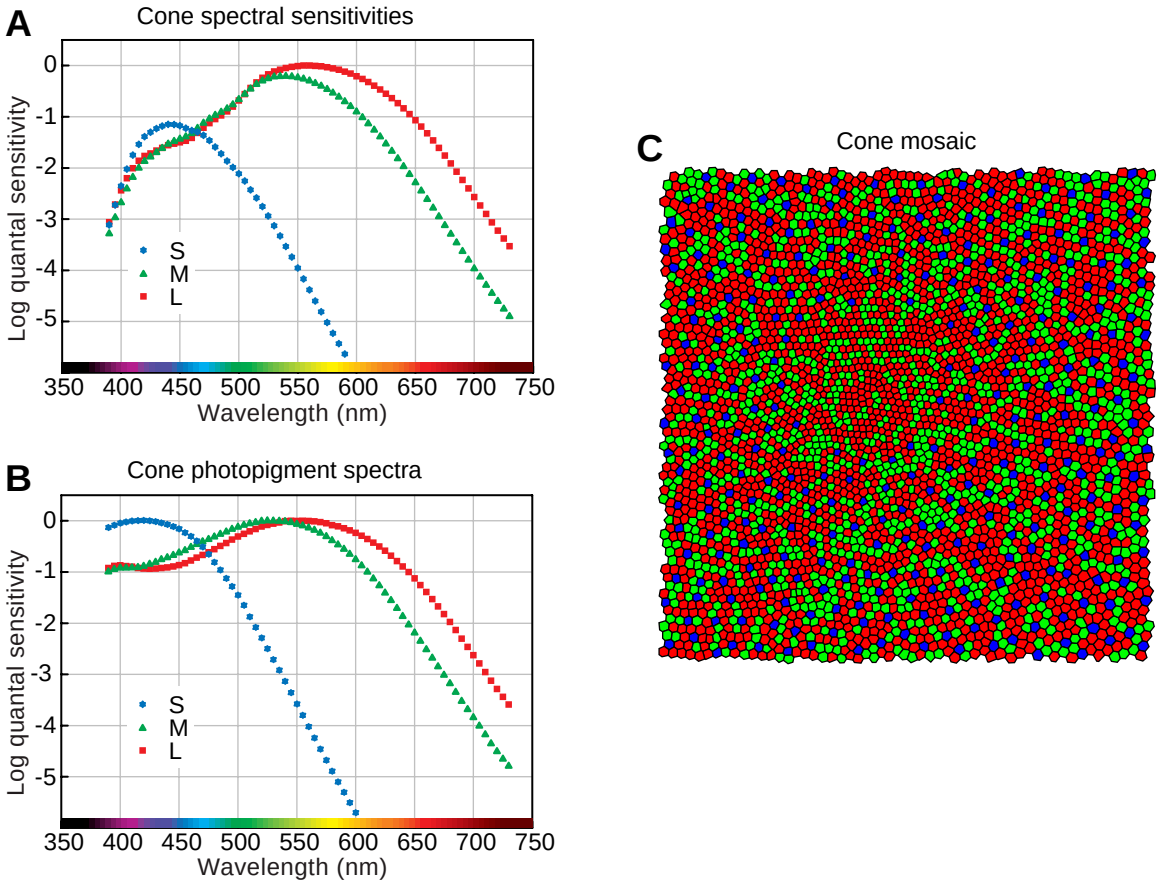

Figure 1.1: Cone spectral sensitivities and their representations in the photoreceptor mosaic. (A) Estimates of the light absorbing properties of the L- M-, and S-cones, measured at the cornea, as a function of wavelength (see Chapter 2, Table 2.1, for values). The heights of the curves have been adjusted according to the assumption that the relative cone sensitivities depend on the relative numbers of the different cone types; namely, that 7% of all cones contain the S-cone pigment and that, of the remaining 93%, those containing the L-cone pigment are 1.5 times more frequent than those containing the M-cone pigment (see Chapter 2). (B) The cone pigment absorption spectra. These were determined from the cone spectral sensitivity functions, by correcting the latter for the filtering of the ocular media and the macular pigment and for the self-screening of the pigment in the outer segment (see Chapter 2, Table 2.1, for values). (C) The cone mosaic of the rod-free inner fovea of an adult human retina at the level of the inner segment (tangential section). Superior is at the top and nasal to the left. The region is ca. 1 deg of visual angle in diameter (ca. 300 μm). The center coordinates of the cone cross sections shown were obtained from the retina of a 35-year-old male (Curcio & Sloan, 1992). The outer dimensions of the cone cross sections have been defined mathematically by Voronoi regions and computer-colored according to the following assumptions: (1) only three cone opsin genes, those encoding the S-, M-, and L-cone pigments are expressed; (2) the inner roughly circular area (ca. 100 μm or 0.34 deg in diameter), displaced slightly to the upper left quadrant of the mosaic, is free of S-cones (Curcio et al., 1991); (3) S-cone numbers in the rest of the retina do not exceed 7% and are semiregularly distributed (Curcio et al., 1991); and (4) there are approximately 1.5 times as many L- as M-cones in this region of the retina and they are randomly distributed (see Chapter 5). The diameters of the cross sections in the center are slightly smaller than those at the outer edge to allow for close packing.

which photons are caught (cf. Rushton, 1972). Lights of different spectral distributions, therefore, will appear identical, if they produce the same absorptions in the three cone photopigments, and different, if they do not (see Chapter 2). Thus color vision – the ability to discriminate on the basis of wavelength – requires comparisons of photon absorptions in different photopigments. And, accordingly, trichromatic color vision requires three such independent comparisons. Merely summing the absorptions in the three cone photopigments at some later neural stage will permit brightness or contrast discrimination, but not color vision.

The loss of one of the cone photopigments, as occurs in certain congenital disorders, reduces (photopic) color vision to two dimensions or dichromacy. The loss of two further reduces it to one dimension or monochromacy. And, the loss of all three completely extinguishes it. Vision, then, is purely scotopic and limited to the rods.

Cone pigments and visual pathways. In man and the higher primates, the primary visual or retino-geniculostriate pathway has evolved into three postreceptoral neuronal systems for transmitting the cone signals that arise from the photopigment absorptions (see Chapter 11). These have been characterized as: (i) a luminance subsystem, which mainly carries information about luminance contrast by summing the relative rates of quantum catch in the M- and L-cones (and is sensitive to high spatial and temporal frequencies); (ii) a yellow-blue color subsystem, which mainly carries information about color contrast by comparing the relative rate of quantum catch in the S-cones with those in the M- and L-cones; and (iii) a red-green color subsystem, which carries information about color contrast by comparing the relative rates of quantum catch in the M- and L-cones. Roughly, it can be said that the three subsystems allow three kinds of discriminations: light from dark, yellow from blue, and red from green.

The dimensionality of the color information transmitted by these postreceptoral subsystems is, in the first instance, limited by the number of available cone photopigments. If one or more of the three normal cone photopigments is absent, then the dimensionality is correspondingly reduced from trichromacy to dichromacy or monochromacy. On the other hand, if an extra, fourth cone photopigment is present, as occurs in certain heterozygotic carriers of color blindness, full four-dimensional or tetrachromatic color vision does not seem readily possible (see page 38). The limitation may be the inability of the postreceptoral subsystems to convey more than three independent color signals (see Chapter 6).

Molecular genetics of the opsin genes

The spectral sensitivity of the cone photopigments is intimately related to the structure of the cone pigment molecules. These are concentrated in the photoreceptor outer segment, a specialized cilium containing the phototransduction machinery (see Fig. 1.2 and Chapter 3). Each pigment molecule consists of a transmembrane opsin (or apoprotein) covalently linked to the same, small conjugated chromophore (11-*cis*-retinal), which is an aldehyde derivative of vitamin A.

All opsins are heptahelical proteins, composed of seven transmembrane helices that are linked together by intra- and extracellular loops. Structural work on the opsin of the rod pigment rhodopsin (Unger & Schertler, 1995), about which we have the most information, indicates that the membrane-embedded helices form a barrel around a central retinal binding pocket (see Fig. 1.2). The binding site of the chromophore in both the cone and rod opsins is located in helix 7, a region that has been relatively conserved during the process of divergent evolutionary change (see the following section).

Photon absorption by the pigment molecules initiates visual excitation by causing an 11-*cis* to all-*trans* isomerization of the chromophore, which activates a transducin G-protein (see Chapter 3). Opsins absorb maximally in the ultraviolet region of the spectrum below 300 nm, whereas retinal absorbs maximally at about 380 nm (Knowles & Dartnall, 1977). It is only by binding together that a broad absorbance band (known as the α-band) in the visible spectrum is created. The λ_{max} of the α-band depends on the geneti-

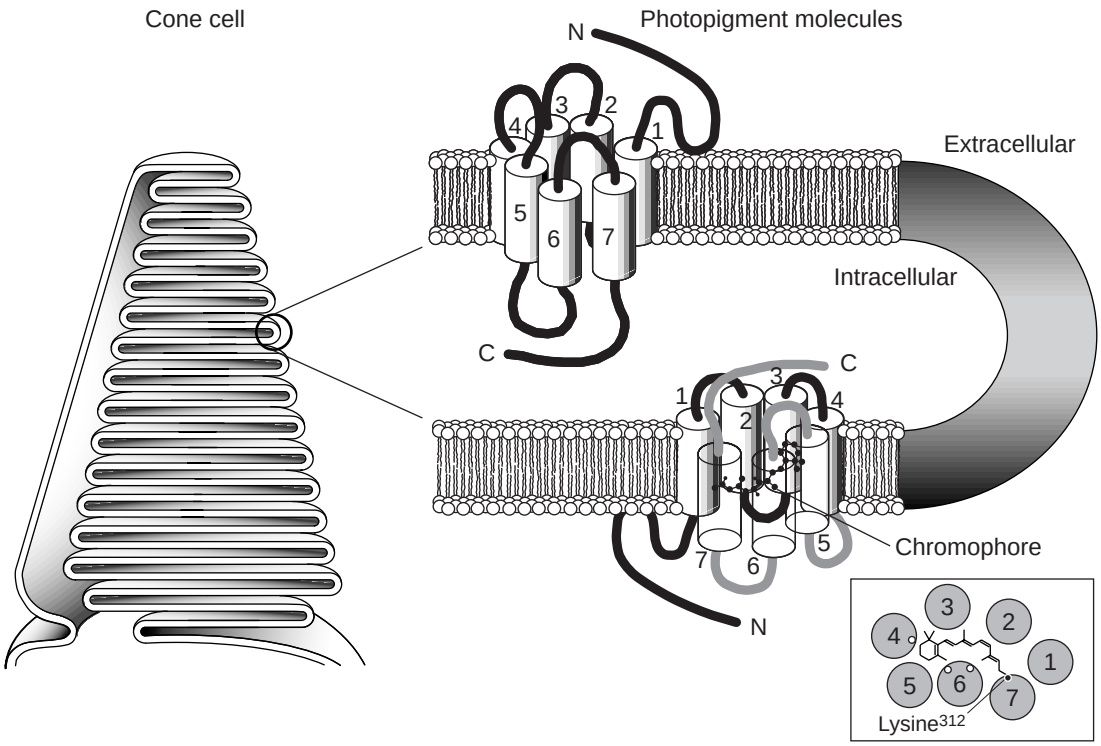

Figure 1.2: Cutaway view of the photopigment molecules (right) packing within the enfolded membrane discs in the outer segments of the cone photoreceptor cells (left). Each molecule consists of a transmembrane opsin bound to a chromophore, 11-*cis*-retinal. The opsin or protein portion of the molecule is a chain of amino acids, running from an amino-terminal end (N), exposed on the external aqueous surface of the membrane discs, to a carboxyl terminal region (C), exposed on the internal aqueous surface of the discs. The chain has seven coils, termed α-helices, spanning the membrane (Hargrave et al., 1984). Linked together by loops in the rest of the chain, the α-helices encircle the chromophore (right, lower cutaway view). The loops are distinguished by whether they occur in the luminal (extracellular) or cytoplasmic (intracellular) face of the cell. The view is from the intracellular surface. It indicates the approximate position of the α-helices and of the three amino acids (open circles) that are believed to have the major influence on the λ_{max} of the pigment (see Fig. 1.5B). The tail of the chromophore is attached by a protonated Schiff base to a charged lysine amino acid residue lying at nucleotide position 312 in the chain of the L- and M-cone opsins (filled circle, see Fig. 1.5B) and at position 293 in the chain of the S-cone opsin (corresponding to position 296 in rhodopsin). Features critical to the function of the opsin are well conserved in all known mammalian species, with the interhelical loops being, on average, more conserved than the transmembrane helical regions.

cally determined amino acid sequence of the opsin and the relationship that the opsin establishes with the chromophore. A second, but lower, absorbance band, known as the β-band, may also be present, which is due to the *cis*-band of the chromophore. The upturn of the L- and M-cone photopigment spectra at very short wavelengths has been interpreted as indicating the presence of such a β-band (see Fig. 1.1B).

The different opsins of the S-, M-, and L-cone photopigments and of the rod photopigment are encoded by four separate genes. These have been formally identified by the HUGO/GDB (genome data base) Nomenclature Committee as the BCP (blue cone pigment), GCP (green cone pigment), RCP (red cone pig-

ment), and RHO (rhodopsin) genes. Visual psychophysicists, however, often refer to them as the S-, M-, L-cone and rod pigment genes.

The genes encoding the S-cone and rod pigments reside alone as single copies. The former is found on the long or q-arm of chromosome 7 (Nathans, Thomas, & Hogness, 1986) within a cytogenetic location between 7q31.3 and 7q32 (Nathans, Thomas, & Hogness, 1986; Fitzgibbon et al., 1994) and the latter on the q-arm of chromosome 3, between 3q21.3 and 3q24 (Nathans, Thomas, & Hogness, 1986). In contrast, the genes encoding the L- and M-cone pigments reside on the q-arm of the X-chromosome at location Xq28 (Nathans, Thomas, & Hogness, 1986) within a head-to-tail tandem array (Vollrath, Nathans, & Davis, 1988; Feil et al., 1990), which may contain as many as six gene copies. In general, the L-cone pigment gene is present in only a single copy and precedes the multiple M-cone pigment genes in the array (see the following section). In addition, the array contains the five nested exons of a complete gene (termed TEX28), the first exon of which is located ca. 700 base pairs downstream of the end of the visual pigment gene cluster (see the section on visual pigment gene structure). Extra, truncated (lacking exon 1) nonfunctional copies may be interdigitated between the opsin genes, filling up most of the intervening area. The TEX28 gene is expressed in testes but not in the cone photoreceptor cells (Hanna, Platts, & Kirkness, 1997). It is transcribed in the orientation opposite to the cone opsin genes (see mRNA transcription).

Pigment gene evolution. The reason for the separate chromosomal locations of the opsin genes is unknown, and their evolutionary development is subject to speculation based on comparisons of their sequence homologies (see the section on opsin gene sequence homologies). One plausible alternative, although by no means the only one, is that the three cone opsin genes and human trichromacy evolved in the following steps (see also Chapters 6 and 7; Goldsmith, 1991):

1. The emergence of a primordial opsin gene on the X-chromosome that encodes a pigment with its λ_{max} in the region conferring the greatest sensitivity to the quantum-intensity–based spectral distribution of sunlight (Dartnall, 1962, but see Lynch & Soffer, 1999) and reflectance of green plants (Lythgoe, 1972). This system formed the basis of the contrast or luminance subsystem of vision, which has a λ_{max} near 555 nm.

2. The emergence of a second opsin gene, about 500 million years ago (Nathans, 1987; Chiu et al., 1994; Hisatomi et al., 1994), located on chromosome 7. Through the accumulation of DNA sequence changes (see the section on opsin gene sequence homologies), it encoded a pigment with its λ_{max} placed at short wavelengths (the S-cone pigment) and expressed it in a different subset of (anatomically distinct) cones from that in which the primordial opsin gene was expressed. The subsequent development of second-order neurons (the yellow-blue opponent color subsystem), which is sensitive to differences in the excitations of the two sets of cones, enabled the discrimination of many forms of natural green vegetation, differing mainly in their reflectances of short-wave light (Hendley & Hecht, 1949; Mollon, 1996).

3. The emergence of a third opsin gene, about 30 to 40 million years ago (Nathans, 1987; Yokoyama & Yokoyama, 1989; Yokoyama, Starmer, & Yokoyama, 1993), as a result of the duplication of the ancestral opsin gene on the X-chromosome. The event copied the transcription unit, but not the locus control region (see the following section). According to one view, the accumulation of DNA sequence changes in the duplicated genes resulted in them encoding distinct M- and L-cone pigments and being expressed in different subsets of (anatomically similar) cones. An alternative view is that the duplication event resulted from unequal crossing over (involving Alu[2] repeats) between two alleles of the ancestral gene that had different spectral sensitivities, so that trichromatic color

[2]An Alu element is a dispersed repetitive DNA sequence that is about 300 bp in length. The name derives from the restriction endonuclease Alu I that cleaves it. The sequence occurs in about 300,000 copies in the human genome and is believed to have no coding value. An Alu repeat element at the site of insertion of the duplicated opsin gene may have been important in promoting crossing over within the array.

vision did not have to await the accumulation of mutations. Initially, changes in the relative excitations of the two pigments caused by changes in wavelength were undifferentiated, but the subsequent recruitment of existing second-order neurons or the development of new ones (the red-green opponent color subsystem) enabled discriminations in the yellow-green to orange-red spectral region. This duplication event may have occurred in our arboreal ancestors, after the divergence of the Old- and New-World monkeys (see Chapters 6 and 7), as an adaptation to frugivory, assisting the detection of fruit amid foilage (Mollon, 1989, 1991; Osorio & Vorobyev, 1996).

This story is necessarily complicated by the coevolution of the rhodopsin gene, which is similar in structure and sequence to the cone opsin genes. It appears to have derived from the S-cone opsin gene, after the divergence of the latter from the common ancestral gene (Okano et al., 1992). The tight clustering of the λ_{max}'s of almost all vertebrate rhodopsins near 500 nm – the human rod spectral sensitivity measured in vivo peaks at 507 nm and the absorbance spectrum at 493 nm – has so far eluded easy explanation (Goldsmith, 1991). It does not directly correspond to the λ_{max} of starlight, moonlight, or twilight (Lythgoe, 1972).

Visual pigment gene structure. Structurally, each visual pigment gene is a large deoxyribonucleic acid (DNA) molecule, consisting of thousands of subunits – nucleotides – linked together. These are the nucleotide base sequences. Each comprises a nitrogenous base (adenine, guanine, thymine, or cytosine), a phosphate molecule, and a sugar molecule (deoxyribose). Owing to the double-helical structure of DNA, the nucleotide in one DNA strand has a complementary nucleotide in the opposite strand. The two are held together, in nucleotide or base pairs (bp), by weak hydrogen bonds. Adenine (A) conjoins with thymine (T) and guanine (G) with cytosine (C); no other combinations are possible. The base sequences can be divided into promoter, noncoding (intron), and coding (exon) sequences (see Fig. 1.3A).

(i) Transcription unit: The term "transcription unit" is often used to refer to the exons and the intervening introns to indicate the region that is actually synthesized into messenger ribonucleic acid (mRNA) before being translated into the opsin. The transcription region begins at the start or cap site at the 5' (upstream or head) end of the gene. It is followed by a short leader sequence – 6 bp long in the S-cone pigment gene and 40 bp long in the M- and L-cone pigment genes – that is not translated into the opsin. Downstream of the leader sequence is the start codon, a trinucleotide sequence, ATG, which specifies the initiation of opsin translation. It is paired with a stop codon, TGA, at the 3' (downstream or tail) end, which specifies the termination of opsin translation. In the transcription region, the stop codon is followed by untranslated tail sequences. These include a signal – the polyadenylation or poly (A) site – for the addition of a string of adenosine residues. The exact function of the residues is unknown, but there is evidence that mRNA degradation occurs from the 3' end and that the poly (A) tail together with the poly (A)–specific RNA-binding proteins increases the half-life of the mRNA during translation (see the section on opsin translation).

(ii) Promoters: Promoters are specific regulatory sequences or boxes upstream of the transcription start site. They bind the enzyme (RNA polymerase) that catalyzes the synthesis of the RNA chain, a reaction that is referred to as transcription (see Fig. 1.3). The first promoter sequence, the TATA regulatory box, is ca. 25 bp upstream of the transcription start site. It is involved in binding RNA polymerase via a TATA binding protein. Another promoter sequence, the CCAAT box, is ca. 70–90 bp upstream of the transcription start site. The promoter sequences also interact with transcriptionally active sequences contained in the upstream locus control region (LCR) to regulate the rate of DNA transcription into RNA and hence the amount of opsin gene expression.

(iii) Introns: The intron sequences are silent or noncoding sequences usually believed to have no apparent function (but see the section on intergenic recombination). The possibility that they contain regulatory sequences involved in gene expression, however, cannot be ruled out. They are delimited by recognition

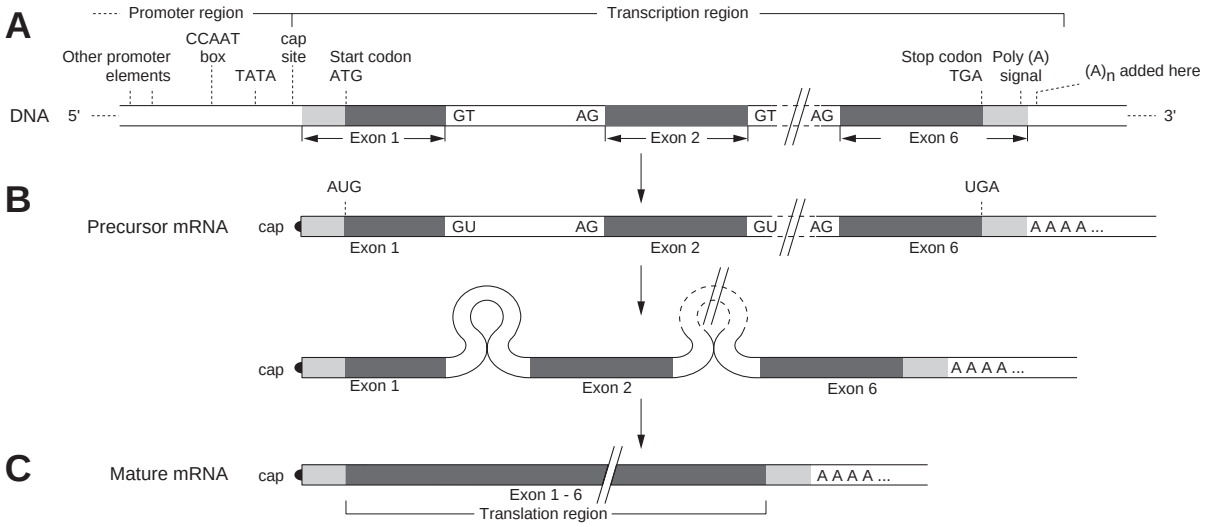

Figure 1.3: Schematic representation of an opsin gene and its transcription into mRNA. (A) The structure of an L-cone opsin gene (DNA), indicating the promoter region, and the untranslated regions at the 5' and 3' ends (gray) the coding regions (black) and the intervening introns (white). (B) and (C) The transcription of the DNA into precursor mRNA and the splicing out of the introns to form mature mRNA. The structure of the mRNA is similar to that of the DNA, except that the sugar molecule is ribose instead of deoxyribose, uracil (U) replaces thymine (T), and the molecule is single-stranded. mRNA is capped or blocked with 7-methylguanosine at its 5' end. There follows a short untranslated region and then the start codon, ATG (AUG), which specifies the initiation of translation. A stop codon, TGA (UGA), indicates the termination of the translated region. Capping and tailing of precursor mRNA precede the splicing out of introns.

sites, which are necessary for identifying and splicing them out from the mRNA precursor (see Fig. 1.3). Introns typically begin with the dinucleotide GT (the splice donor; GU in the precursor mRNA) and end with the dinucleotide AG (the splice acceptor).

(iv) Exons: The opsin-coding sequences are divided into exons, which are separated by the introns and numbered according to their proximity to the 5' end (see Fig. 1.3A). Within the exons, the nucleotide sequences are grouped into triplets – the 3-base sequences or codons – each of which specifies a constituent amino acid (monomer) of the polypeptide chain of the visual pigment opsin (see Table 1.1). There are 64 possible codons (the possible combinations of the four nucleotides), but only 20 unmodified amino acids in the opsin. Thus each amino acid may have more than one codon. Many of the different codons for single amino acids differ only in the third nucleotide of the 3-base sequence.

(v) Gene length: The S-cone pigment gene comprises 5 exons (1,044 bp of which are protein coding) and 4 introns (total length: 2,200 bp). The length of the gene from its mRNA start or cap site (nucleotide base sequence 403) to its poly (A) site (nucleotide base sequence 1,510) is 3,308 bp. (The extra base pairs occur because the exons include 5' and 3' untranslated, nonprotein-coding, sequences that also end up in the mature mRNA; see Fig. 1.3). The M- and L-cone pigment genes each comprise six exons (1,092 bp of which are protein coding) and five introns (total length: 12,036 bp and 14,000 bp, respectively, for the M- and L-cone pigment genes). The length of the M-cone pigment gene is 13,300 bp and that of the L-cone pigment gene is 15,200 bp.

A small, extra exon, encoding only 38 amino acid residues (114 bp), is found at the beginning of the L- and M-cone pigment genes. It may have been added at some point during evolution to the primordial visual

pigment gene to facilitate transcription of more than one gene copy in the tandem array.

mRNA transcription and opsin translation. The base sequences in the DNA are transcribed into RNA, which is subsequently translated to produce the opsin. The primary product in transcription – the mRNA precursor, often called pre-mRNA – contains all of the base sequences, those defining the introns as well as the exons (see Fig. 1.3B). It is blocked or capped with 7-methylguanosine at its 5' end and tailed by a string of adenosine residues at its 3' end. The capping and tailing are believed to permit the export of mRNA from the cell nucleus.

During processing in the cell nucleus, the introns are spliced out, so that the final product in transcription – the mature mRNA (Fig. 1.3C) – only contains the exon sequences. The mature mRNA is exported to the cytoplasm of the photoreceptor cells, where it serves as a template for the synthesis of the opsin from its constituent amino acid residues (see Fig. 1.4). The translation process is complex, involving several stages and a family of transfer RNAs (tRNA), the role of which is to bond with amino acids and transfer them to the ribosome (the site of protein synthesis). The amino acids are assembled sequentially in the growing polypeptide chain of the opsin, from the amino-terminal end to the carboxyl-terminal end, according to the order of codons carried by the mRNA. In the chain, the amino acids are linked by the carboxyl group (COOH) of one amino acid and the amino group (NH_2) of another. Hence, opsins can be identified by their NH_2 (or N) and COOH (or C) ends.

Opsin gene sequence homologies. The nature of the pigment defined by an opsin gene depends on the nucleotide sequences of its exons, which are grouped into triplets (codons) encoding amino acid residues and numbered sequentially beginning with the first codon.

The S-cone pigment gene comprises 348 codons divided over its 5 exons, while the M- and L-cone pigment genes comprise 364 codons divided over their 6 exons. The S-cone pigment gene shows only 43 ± 1%

DNA codons	Amino acid	Abbreviations		Class
GCA, GCC, GCG, GCT	Alanine	ala	A	(1)
AGA, AGG, CGA, CGC, CGG, CGT	Arginine +	arg	R	(4)
GAC, GAT	Aspartic acid -	asp	D	(3)
AAC, AAT	Asparagine	asn	N	(2)
TGC, TGT	Cysteine	cys	C	(1, 2)
GAA, GAG	Glutamic acid -	glu	E	(3)
CAA, CAG	Glutamine	gln	Q	(2)
GCA, GGC, GGG, GGT	Glycine	gly	G	(1)
CAC, CAT	Histidine (+)	his	H	(4, 6)
ATA, ATC, ATT	Isoleucine	ile	I	(1)
TTA, TTG, CTA, CTC, CTG, CTT	Leucine	leu	L	(1)
AAA, AAG	Lysine +	lys	K	(4)
ATG	Methionine (Start)	met	M	(1)
TTC, TTT	Phenylalanine	phe	F	(1, 5)
CCA, CCC, CCG, CCT	Proline	pro	P	(1)
AGC, AGT, TCA, TCC, TCG, TCT	Serine	ser	S	(2)
ACA, ACC, ACG, ACT	Threonine	thr	T	(2)
TGC	Tryptophan	trp	W	(1, 5, 6)
TAC, TAT	Tyrosine	tyr	Y	(2, 5)
GTA, GTC, GTG, GTT	Valine	val	V	(1)
TAA, TAG, TGA	(Stop)			

Table 1.1: The genetic code and the chemical properites of amino acids. The four types of nucleotides forming the deoxyribonucleic acid (DNA) codons are adenylic (A), guanylic (G), cytidylic (C), and thymidylic (T) acid. In ribonucleic acid (RNA) codons, uridylic (U) replaces thymidylic (T) acid. Amino acids belonging to the same class (1–6) are considered homologous and their substitution conservative. (+ or - indicates those amino acids most likely to be positively or negatively charged. The charge of histidine (+) depends on the local environment, and is, therefore, indicated in parentheses.)

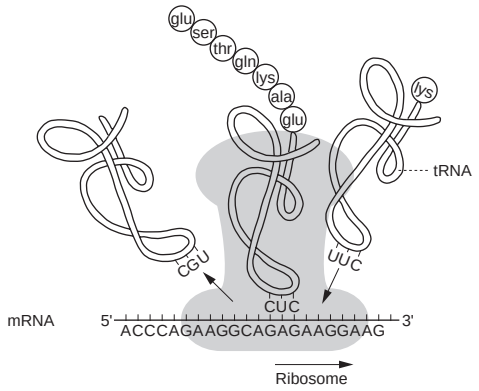

Figure 1.4: The mechanism of translation of the mRNA (from exon 5) into the L-cone opsin. Translation involves tRNA, each of which has a recognition site or triplet nucleotide sequence (anticodon) that is complementary to the triplet nucleotide coding sequence (codon) of mRNA. For example, the tRNA at the right carries lysine and its (RNA) anticodon CUU recognizes the lysine (RNA) codon AAG. Amino acids attach serially to the growing polypeptide chain and the respective tRNAs are jettisoned (as shown at left). The process is mediated by ribosomes (one is shown in outline) moving along the mRNA in a 5' to 3' direction.

amino acid identity with either the M- or L-cone pigment gene (see Fig. 1.5), which is about the same amount of homology with the rod pigment gene ($41 \pm 1\%$). In contrast, the M- and L-cone pigment genes show 96% mutual identity for the 6 exons (they are 98% identical at the DNA sequence level if the introns and 3'-flanking sequences are included). From the sequence homologies, it is possible to estimate the evolutionary divergence of the genes: The greater the identity, the more recent the divergence (see the section on pigment gene evolution). A curiosity is that the noncoding intron sequences of M- and L-cone pigment genes are more homologous than the coding exon sequences, even though the former should be freer to diverge than the latter during the course of evolution (see Mollon, 1997).

(i) L- and M-cone exon sequences: There are only 15 codon differences between the L- and M-cone pigment genes (Fig. 1.5B). They are confined to exons 2–5, which encode the seven membrane-embedded α-helices that together form the chromophore binding pocket (see Fig. 1.2). Three sites are in exon 2 (at codons 65, 111, and 116; for numbering system, see Nathans, Thomas, & Hogness, 1986); two in exon 3 (at codons 153 and 180), three in exon 4 (at codons 230, 233, and 236), and seven in exon 5 (at codons 274, 275, 277, 279, 285, 298, and 309). Six of these differences involve conservative substitutions of hydrophobic residues, which do not influence the interaction of the opsin with the chromophore. Of the remaining nine, one of the sites (codon 116) lies in the first extracellular loop of the molecule; it is therefore unlikely to be involved in a direct interaction with the chromophore. Seven sites, however, lie in the transmembrane helices and may contact the chromophore. These are codons 65, 180, 230, 233, 277, 285, and 309. They involve the substitution of an amino acid residue that lacks a hydroxyl group (a nonpolar or uncharged amino acid) by one that carries a hydroxyl group (a polar or charged amino acid). On theoretical grounds, amino acid substitutions that change the number or locations of polar side chains (e.g., those carrying an hydroxyl group) in the retinal-binding pocket could alter the spectral tuning of the 11-*cis*-retinal chromophore by readjusting its three-dimensional packing or by changing the electrical properties of its immediate environment (Kropf & Hubbard, 1958; Mathies & Stryer, 1976; Hays et al., 1980).

The effect of replacing polar amino acids by nonpolar ones is supported by the results of site-directed mutagenesis experiments in bovine rhodopsin (Chan, Lee, & Sakmar, 1992) and in human hybrid cone pigments (Merbs & Nathans, 1992b; Asenjo, Rim, & Oprian, 1994), as well as by comparisons between electroretinographic (ERG) measurements of primate cone pigment spectral sensitivities and corresponding amino acid sequences (Neitz, Neitz, & Jacobs, 1991). The largest shifts in λ_{max} are produced by substituting alanine for threonine at codon 285 (ca. -14 nm), phenylalanine for tyrosine at position 277 (ca. -7 nm), and alanine for serine at position 180 (ca. -4 nm) (Merbs & Nathans, 1992a). By contrast, substitutions at positions 65, 230, 233, and 309 produce shifts of approximately 1 nm or less at λ_{max} (Merbs & Nathans, 1993).

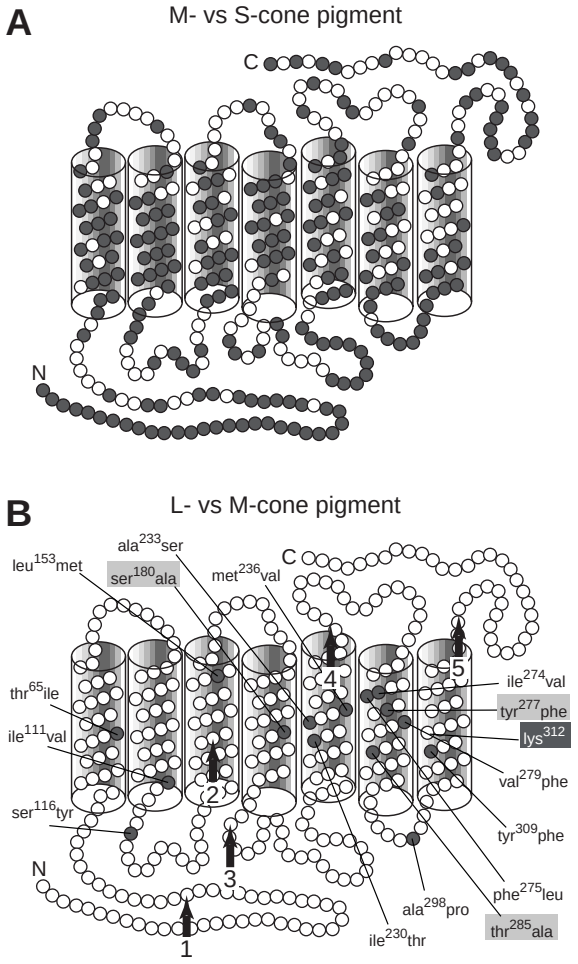

Figure 1.5: Pairwise comparisons of human visual pigment molecules showing amino-acid identities (open circles) and differences (filled circles) (after Nathans, Thomas, & Hogness, 1986). In each representation, the seven α-helices are arranged in a line. When intramembrane regions are optimally aligned, the amino-proximal tails (extracellular face) of the M- (or L-) cone pigments are 16 amino acids longer than for the S-cone pigment. The alignment can be improved by inserting into the M- (or L-) cone pigment sequences gaps of two amino acids and of one amino acid, respectively, at positions 4 residues and 29 residues from the carboxyl terminus. (A) Identity between the M- and S-cone pigments. (B) Identity between the L- and M-cone pigments. The location of lysine^{312}, the site of covalent attachment of 11-*cis* retinal, and the 15 amino acid substitutions are indicated. The start of each of the 5 intron positions are indicated by numbered vertical arrows. The subsitutions at codons 180, 277, and 285 (highlighted) are believed to contribute the majority of the spectral difference between the M- and L-cone pigments.

Such single amino acid substitution data, however, do not explain the nonadditive shifts in λ_{max} that are observed when more than one hydroxyl group is simultaneously substituted (Merbs & Nathans, 1992b; Asenjo, Rim, & Oprian, 1994), nor do they take into account the influence of multiple aliphatic amino acid differences on side-chain packing (Merbs & Nathans, 1993).

(ii) L- and M-cone 5' and intron sequences: L- and M-cone pigment genes differ not only in their coding sequences, but also in their 5' and intron sequences (Nathans, Thomas, & Hogness, 1986; Vollrath, Nathans, & Davis, 1988). In the vast majority of color normal arrays sequenced in Caucasian males, there is only one L-cone pigment gene, which is longer than all of the other gene copies (see the section on the arrangement of the gene array). It is located at the 5' (upstream) end of the array and abuts single copy DNA sequences, which are not found in front of the other, downstream genes. The length difference arises because its intron 1 *typically* contains 1,612 bp extra sequences (comprising 1,284 bp of three Alu elements and 328 bp of intervening unique-sequence DNA), which are also not found in the downstream genes (cf. the L- and M-cone pigment genes in Fig. 1.7A).

Although the extra intron sequences are found in >99% of Caucasian males, in ca. 45, 35, and 2.5 of African, Afro-American, and Japanese males, respectively, the most proximal gene in the array lacks the extra sequences in intron 1 and is the same size as the downstream M-cone pigment genes (Jørgensen, Deeb, & Motulsky, 1990; Meagher, Jørgensen, & Deeb, 1996). A reason for this may be that it contains inserted (exon 2) M-cone pigment-specific sequences (see the next section).

Normal and hybrid pigment genes. The S-cone opsin gene sequence seems to be nearly invariant in the human population. In contrast, the M- and L-cone opsin genes are diversiform, owing to hybrid variants and shared polymorphisms (see Fig. 1.6).

Hybrids are fusion genes containing the coding sequences of both L- and M-cone pigment genes. They are produced by intragenic crossing over: the breaking

during meiosis of one maternal and one paternal chromosome at the opsin gene locus; the exchange of the corresponding sections of nucleotide sequences; and the rejoining of the chromosomes (see the section on intragenic recombination and Fig. 1.15C).

Intragenic crossing over between the M- and L-cone pigment genes is much more likely to occur within intron sequences than within exon sequences, owing to the approximately tenfold greater size of the introns compared with the exons and the paradoxically greater DNA sequence similarity of the M- and L-cone pigment gene introns compared with the exons (Shyue et al., 1995). Thus, in general, hybrid genes contain some number of contiguous exons from one end of an L-cone pigment gene joined to the remaining exons from the other end of an M-cone pigment gene. Those beginning with L-cone exon sequences are known as 5'L-3'M (or 5'red-3'green) hybrid or fusion genes, and those beginning with M-cone exon sequences as 5'M-3'L (or 5'green-3'red) hybrid genes. The 5'L-3'M hybrid genes encode M or M-like anomalous pigments; whereas the 5'M-3'L hybrid genes encode L or L-like pigments. Therefore, a convenient shorthand terminology for referring to normal and hybrid genes is to identify their exon sequences as being M-cone or L-cone pigment-specific (see Fig. 1.6). Two factors, however, complicate this simple picture.

First, exon 3 is more variable than exons 2, 4, or 5 in its amino acid residues (Winderickx, Battisti, et al., 1993; Sharpe et al., 1998), owing to the existence of several shared genetic polymorphisms between the M- and L-cone pigment genes. Genetic polymorphisms (or dimorphisms, if confined to two forms) are allelic variants of a gene occurring with a frequency greater than 1%. Most of the polymorphisms in exon 3 are confined to a dimorphic subsitution of a single nucleotide sequence. These alter the encoded amino acid without apparently affecting the properties of the photopigment. However, one – the substitution of a serine for alanine residue at codon 180 (the only one involving the substitution of a hydroxyl group) – produces a phenotypic variation. It causes a slight red shift (see below). Current estimates in normal observers suggest that the polymorphism is not equally distributed (see

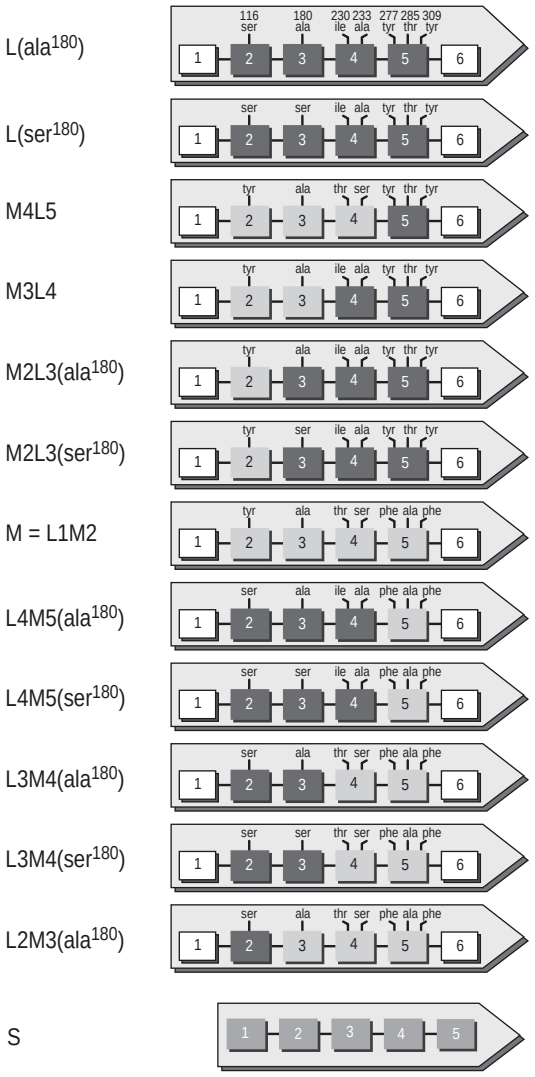

Figure 1.6: Exon arrangement of the S-, M-, L-, 5'L-3'M-hybrid, and 5'M-3'L-hybrid pigment genes. The S-cone pigment gene has one fewer exon, missing from its 5' end, than the X-linked pigment genes. There are no sequence differences between X-linked genes in exons 1 and 6. The 7 amino acid residues indicated above exons 2 to 5 are those responsible for the spectra shift between the normal and anomalous pigments. Dark gray indicates an L-cone pigment gene-specific sequence; light gray, an M-cone pigment gene-specific sequence.

Table 1.2): Among human L-cone pigment genes approximately 56.3% have serine and 43.7% have alanine at position 180, whereas, among M-cone pigment genes approximately 6% have serine and 94% have alanine (Winderickx et al., 1992b; Winderickx, Battisti, et al., 1993; Neitz & Neitz, 1998; Sharpe et al., 1998; Schmidt et al., 1999). However, large variability may occur between groups of different ethnic origin. In one report, 80% of African (N = 56), 84% of Japanese (N = 49), and 62% of Caucasian (N = 49) males had serine at codon 180 (Deeb & Motulsky, 1998). Therefore, it is useful to designate an M- or L-cone pigment gene, by an abbreviation that reflects the identity of the polymorphic residue at position 180 in exon 3.

A further complication is that frequently pigment gene sequences reveal an M-cone pigment gene exon 2 embedded within an L-cone pigment gene or a 5'L-3'M hybrid gene (Sharpe et al., 1998), indicating a complicated history of recombination events, and, therefore, making it pertinent to designate whether exon 2 is derived from an M- or L-cone pigment gene.

Thus, each normal or hybrid gene is more properly referred to by an abbreviation that reflects not only the origin of its various exons, but also the identity of the polymorphic residue at position 180. For example, L-cone pigment genes can be designated L(ala^{180}) or L(ser^{180}) to indicate the presence of alanine or serine, respectively, at position 180. L4M5(ala^{180}) is a hybrid pigment encoded by a gene in which exons 1–4 are derived from an L-cone pigment gene, exons 5 and 6 are derived from an M-cone pigment gene (i.e., the point of crossing over is in intron 4), and position 180 is occupied by alanine. L(M2;ser^{180}) is an L-cone pigment gene in which exon 2 is derived from an M-cone pigment gene and position 180 is occupied by serine. The fact that sequence differences between L- and M-cone pigment genes are confined to exons 2–5 implies that an L1M2 hybrid gene encodes a de facto M-cone pigment and, likewise, that an M1L2 hybrid gene encodes a de facto L-cone pigment.

Protein sequence variation and spectral sensitivity. Several in vitro and in vivo techniques have been applied to studying the variation in normal and, less frequently, hybrid pigment spectral sensitivities (for a review, see Stockman et al., 1999). A partial list, emphasizing studies that have investigated the hybrid pigments, is given in Table 1.3.

Study	Population	Polymorphic residue	Allele frequency
L-cone opsin genes			
Winderickx, Battisti, et al. (1993)	74 normals 35 deutans	ala^{180} ser^{180}	0.440 0.560
Sjoberg et al. (1997), see Neitz & Neitz (1998)	130 normals	ala^{180} ser^{180}	0.485 0.515
Sharpe et al. (1998)	27 deuteranopes	ala^{180} ser^{180}	0.259 0.741
Schmidt et al. (1999)	36 normals 2 deuteranopes	ala^{180} ser^{180}	0.395 0.605
Mean	304	ala^{180} ser^{180}	0.437 0.563
M-cone opsin genes			
Winderickx, Battisti, et al. (1993)	52 normals† 12 deutan 8 protans	ala^{180} ser^{180}	0.959 0.041
Sjoberg et al. (1997), see Neitz & Neitz (1998)	130 normals	ala^{180} ser^{180}	0.930 0.070
Mean		ala^{180} ser^{180}	0.940 0.060

Table 1.2: Polymorphisms in the L- and M-cone opsin genes at codon 180 in human males. For descriptions of deutans, protans, and deuteranopes, see section on color blindness. (†Selected from a larger population of 72; only those were included who had one type of M-cone opsin gene or two types differing by only a single polymorphism.)

The in vitro measurements include ERG, single photoreceptor suction electrode action spectra (see Chapter 4), microspectrophotometry (MSP), and photobleaching difference absorption spectra measurements of recombinant cone pigments produced in tissue culture cells. The in vivo measurements include reflection densitometry, linear transforms of psycho-

Genotype	In vivo	In vitro	
	Sharpe et al. (1998)	Merbs & Nathans (1992a, 1992b)	Asenjo, Rim, & Oprian (1994)
S	418.9 ± 1.5†	426.3 ± 1.0	424.0*
M(ala^{180}) = L1M2(ala^{180})	527.8 ± 1.1	529.7 ± 2.0	532 ± 1.0
L2M3 (ala^{180})	528.5 ± 0.7	529.5 ± 2.6	532 ± 1.0
L3M4 (ser^{180})	531.5 ± 0.8	533.3 ± 1.0	534 ± 1.0
L4M5 (ala^{180})	535.4	531.6 ± 1.8	—
L4M5 (ser^{180})	534.2	536.0 ± 1.4	538 ± 1.0
M2L3 (ala^{180})	—	549.6 ± 0.9	—
M2L3 (ser^{180})	—	553.0 ± 1.4	559 ± 1.0
M3L4	—	548.8 ± 1.3	555 ± 1.0
M4L5	—	544.8 ± 1.8	551 ± 1.0
L (ala^{180})	557.9 ± 0.4	552.4 ± 1.1	556 ± 1.0
L (M2, ala^{180})	556.9	—	—
L (M2, ser^{180})	558.5	—	—
L (ser^{180})	560.3 ± 0.3	556.7 ± 2.1	563 ± 1.0

Table 1.3: Absorbance spectrum peaks (λ_{max} ± SD) of the human normal and hybrid cone pigments. (†Value from Stockman et al., 1999; *value from Oprian et al., 1991.)

physical color matching functions (CMFs), and spectral sensitivity measurements in normal and color-deficient observers of known genotype under conditions chosen to isolate preferentially a single cone pigment (see Chapter 2 for a review of measurements of the normal cone absorption spectra).

(i) S-cone pigment: The λ_{max} (± the standard deviation) of the human S-cone pigment, measured at the retina (see legend to Fig. 1.1), has been placed at: (i) 419.0 ± 3.6 nm by in vitro MSP of human cones (Dartnall, Bowmaker, & Mollon, 1983); (ii) 424 (Oprian et al., 1991) or 426 nm (Merbs & Nathans, 1992a) by in vitro absorption spectroscopy of recombinant cone pigments; (iii) 419.0 and 419.7 nm by in vivo central and peripheral spectral sensitivity measurements, respectively, in normal and blue-cone monochromat (see page 41) observers (Stockman, Sharpe, & Fach, 1999); and (iv) 420.8 nm by transforms of the Stiles and Burch 10-deg CMFs (Stockman, Sharpe, & Fach, 1999). Determination of the λ_{max} of the S-cone pigment in solution is complicated by short-wavelength–absorbing bleaching products that partially overlap the pigment absorbance and must be subtracted from it. On the other hand, the in vivo determinations are complicated by several factors, including individual differences in the absorption of the lens and macular pigment (see Chapter 2). Some variability in the λ_{max} of the S-cone pigment has been suggested (Stockman, Sharpe, & Fach, 1999; see p. 44).

(ii) L- and M-cone pigments: The λ_{max}'s of the normal M-, L(ala^{180})-, and L(ser^{180})-cone pigments have been placed, respectively, at: (i) 530.8 ± 3.5, 554.2 ± 2.3, and 563.2 ± 3.1 nm (558.4 ± 5.2 nm for the mixed L-cone pigments) by MSP[3] of human cones (Dartnall, Bowmaker, & Mollon, 1983); (ii) 529.7 ± 2.0, 552.4 ± 1.1, and 556.7 ± 2.1 nm (Merbs & Nathans, 1992a) or 532 ± 1.0, 556 ± 1.0, and 563 ± 1.0 nm (Asenjo, Rim, & Oprian, 1994) by in vitro spectroscopy of recombinant cone pigments; (iii) 530 and 560 nm (mixed L-cone pigments; Schnapf, Kraft, & Baylor, 1987) or 531 (Kraft, private communication), 559.2, and 563.4 nm (Kraft, Neitz, & Neitz, 1998) by suction electrode data; (iv) 528.6 ± 0.5, 557.5 ± 0.4, and 560.2 ± 0.3 nm by foveal spectral sensitivity measurements in dichromat observers (Sharpe et al., 1998, 1999; Stockman & Sharpe, 2000a); and (v) 530.6 and 559.1 nm (mixed L-cone pigments) by transforms of the Stiles and Burch 10-deg CMFs (Stockman & Sharpe, 2000a).

The in vivo spectral sensitivity measurements obtained from dichromats (Sharpe et al., 1998) show a mean separation of ca. 2.7 nm between the L(ala^{180})- and L(ser^{180})-cone pigments. This value is somewhat less than that which has been obtained from site-directed mutagenesis experiments and other tech-

[3]These values are based on ad hoc subgrouping according to whether the individual λ_{max} lies above or below the group mean and not according to genotype.

niques. However, all of the previous reported values, whether based on inferences from Rayleigh matches (Winderickx et al., 1992b; Sanocki et al., 1993; Sanocki, Shevell, & Winderickx, 1994; He & Shevell, 1994), spectral sensitivities (Eisner & MacLeod, 1981), cloned pigment (Merbs & Nathans, 1992a, 1992b; Asenjo, Rim, & Oprian, 1994), or on the ERG (Neitz, Neitz, & Jacobs, 1995), have shown that serine-containing pigments are red-shifted with respect to alanine-containing pigments. The in vivo estimates accord with other psychophysical measures of the variability of the L-cone λ_{max} in the normal population, based on the analysis of color-matching data by Stiles and Burch (1959), which preclude shifts greater than 3.0 nm (Neitz & Jacobs, 1989, 1990; Webster & MacLeod, 1988; Webster, 1992).

As of yet, no reliable in vivo spectral sensitivity data exist for the two polymorphic variants of the M-cone pigment. However, in vivo comparisons between subjects with a L2M3(ala^{180}) pigment and those with a L3M4(ser^{180}) pigment, for whom the only important amino acid difference is at position 180, show a mean shift of 3.0 nm (see Table 1.3). Further, the spectral shift between the λ_{max}'s of the M(ala^{180})- and M(ser^{180})-cone pigments has been estimated provisionally (awaiting in vivo confirmation) at: (i) 5.9 nm by ad hoc analysis of MSP data (Dartnall, Bowmaker, & Mollon, 1983); and (ii) 4.3 to 4.4 nm (Merbs & Nathans, 1992b) and 2 nm (Asenjo, Rim, & Oprian, 1994) by in vitro spectroscopy of recombinant normal and 5'M-3'L hybrid cone pigments.

(iii) Hybrid cone pigments: Estimates of the λ_{max}'s of the hybrid pigments encoded by 5'L-3'M and 5'M-3'L hybrid genes, which presumably underlie anomalous trichromacy (see color blindness), are summarized in Table 1.3. Differences between the in vivo and in vitro estimates probably reflect the limitations of the measuring techniques (see Chapter 2). For instance, the absorption measurements of visual pigment in vitro are accurate within only about 0.5 to 1.0 log unit of the λ_{max} and thus encompass only a limited range of wavelengths. Additionally, they will differ from the in vivo measurements because they do not account for waveguiding in the photoreceptor.

Nonetheless, the in vivo and in vitro data support one another in indicating that both 5'L-3'M and 5'M-3'L hybrid genes encode a range of pigments with spectral sensitivities, which, in every case so far examined, lie between those of the normal L- and M-cone pigments. The data further indicate that the spectral sensitivity of the hybrid pigment depends on the position of the crossing-over and on the identity of the polymorphic amino acids at position 180. For each exon, the set of amino acids normally associated with the L- or M-cone pigments produce, respectively, spectral shifts to longer or shorter wavelengths, thus producing a monotonic relationship between the λ_{max} and the fraction of the hybrid pigment derived from the L and M parental pigments (see Sharpe et al., 1998).

The primary determinants of the spectral shift are located in exon 5, as seen by the clustering of the λ_{max} of all of the pigments encoded by 5'L-3'M genes within 8 nm of the maxima of the normal M pigments. Further, a comparison of the in vivo measured λ_{max}'s in Table 1.3 indicates that the L/M sequence differences in exon 5 – principally the residues at 277 and 285 – result in spectral shifts of 15–25 nm, the exact value depending on sequences in exons 2–4. The in vivo measured data further suggest that substitutions at the sites confined to exons 2–4 produce much smaller spectral shifts: Exon 2 contributes at most 0–2.0 nm; exon 3, 1.0–4.0 nm; and exon 4, 2.5–4.0 nm. These results are in approximate agreement with the in vitro results (Merbs & Nathans, 1992b; Asenjo, Rim, & Oprian, 1994) and with inferences based on a comparison of primate visual pigment gene sequences and cone spectral sensitivity curves (Neitz, Neitz, & Jacobs, 1991; Ibbotson et al., 1992; Williams et al., 1992).

(iv) Sequence variation and opsin viability: Amino acid substitutions may have other consequences than shifting the spectral sensitivity of the X-chromosome–linked opsins. They also could alter the quantum efficiency or the optical density of the pigment. As Williams et al. (1992) point out, in vitro expression studies have noted that some hybrid pigments may be unstable or of reduced optical density (Merbs & Nathans, 1992b; Asenjo, Rim, & Oprian, 1994). Further, both

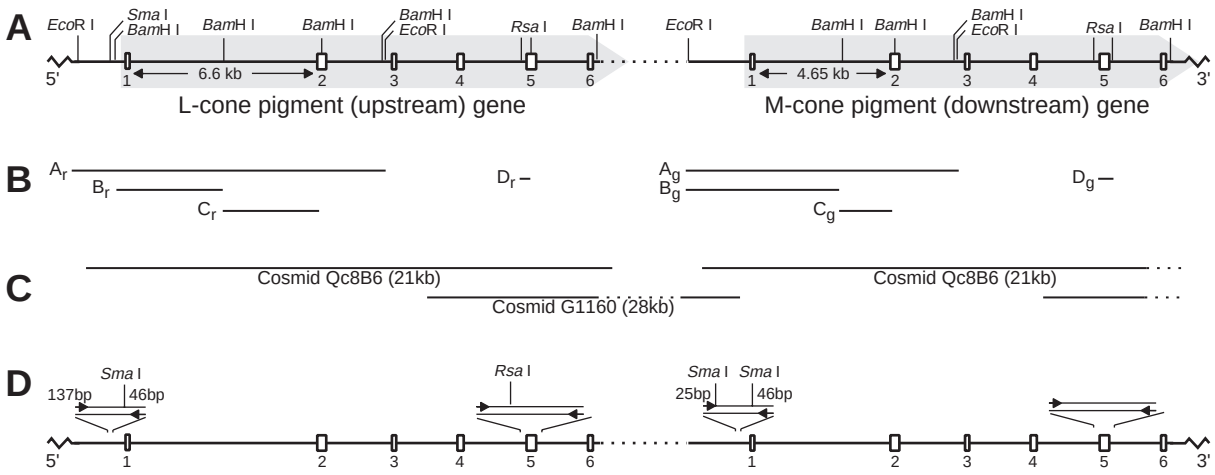

Figure 1.7: A schematic representation of the restriction enzymes and probes used to quantitate or directly visualize opsin genes in the X-chromosome linked visual pigment gene array. (A) The recognition or cleavage sites of the restriction enzymes, *Bam*H I, *Eco*R I, *Hin*d III, *Rsa* I, and *Sma* I, used to define fragment sequences. The exons and intervening introns of a single L-cone pigment gene (upstream) and a single M-cone pigment (downstream) gene are indicated. (B) The L-cone pigment gene restriction fragments, A_r, B_r, C_r, and D_r, and the M-cone pigment gene restriction fragments, A_g, B_g, C_g, and D_g used in the conventional gel electrophoresis/Southern blot hybridization method to estimate the ratios and numbers of L- and M-cone pigment genes in the array (Nathans, Thomas, & Hogness, 1986). (C) Positions of the hybridization pattern of the cosmids, Qc8B6 (Gene Bank Accession number Z68193), and G1160 (Accession number Z46936), used to directly visualize the opsin genes by in situ hybridization. (D) Primer sequences from a 183-bp fragment containing promoter sequences (ca. 50 bp upstream exon 1) and from a ca. 300-bp fragment containing the sequences of exon 5, used by the endlabeled PCR-product method to count the number of genes and the ratio between L-cone and M-cone pigment genes, respectively, in the opsin gene array (Neitz & Neitz, 1995).

psychophysical (Miller, 1972; Smith & Pokorny, 1973; Knau & Sharpe, 1998) and retinal densitometric (Berendschot, van de Kraats, & van Norren, 1996) data suggest that the L- and M-cone pigments, as estimated in deuteranopes and protanopes, respectively (see section on color blindness), differ in optical density.

It is unlikely that other differences between the opsin genes, in particular the extra sequence in intron 1 of the L-cone opsin gene, could contribute to levels of expression and optical density. Although some genes have enhancer sequences in their introns or even 3' of the transcription unit, the available transgenic data for M- and L-cone opsin genes indicate that cone-specific expression is only regulated by the promoter and LCR (however, transgenic expression in the absence of introns does not preclude a role in level of expression). Moreover, most of the extra sequence in intron 1 of the L-cone opsin gene is made up of Alu repeat elements.

The size of the opsin gene array. The opsin gene array on the X-chromosome varies in size, typically containing more than two opsin genes (Nathans, Thomas, & Hogness, 1986). Its variability in the normal population is a subject of controversy; the differences between investigators have been used to challenge the Young–Helmholtz trichromatic theory of color vision as well as current models about the evolution of the human photopigments (Neitz & Neitz, 1995). At the heart of the controversy are the different quantitation and direct techniques (see Figs 1.7 and 1.8) used to assess the total copy number and ratio of L- and M-cone pigment genes within the array (see Table

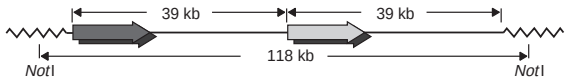

Figure 1.8: The *Not* I fragment that carries the entire opsin gene array, used to determine the total number of genes (Vollrath, Nathans, & Davis, 1988; Macke & Nathans, 1997). It comprises: (i) 40 kb of nonrepeated single-copy flanking DNA; and (ii) repeating units of ca. 39-kb length, each unit consisting of one complete opsin gene (13.2–15.2 kb) and the 24-kb highly conserved flanking region. Thus, the size of the *Not* I fragment will vary in steps of ca. 39 kb; which can be resolved by pulsed field gel electrophoresis. The total gene number will be equal to: (the size of the *Not* I fragment - 40 kb)/39 kb. In the example shown, the 118-kb-long *Not* I fragment contains two genes.

1.4).

(i) Techniques: The techniques rely on bacterial restriction enzymes (endonucleases) that cleave the opsin gene at specific base sequences to produce fragments that differ between the M- and L-cone photopigment genes (Nathans, Thomas, & Hogness, 1986). When isolated on a suitable filter gel, the restriction fragments can be electrically separated according to molecular weight by conventional gel electrophoresis and visualized by Southern blotting by hybridizing them with radioactively labeled probe DNA that recognizes specific nucleotide sequences.

Targeted regions of interest are several restriction fragment length polymorphisms (RFLP). These are pairs of small fragments cleaved by the *Eco*R I, *Bam*H I, and *Rsa* I enzymes (Fig. 1.7A). They include A_r/A_g, B_r/B_g, C_r/C_g, and D_r/D_g (Fig. 1.7B), one of which is specific to the L- (subscript r for red) and the other to the M- (subscript g for green) cone pigment gene. Their labeled intensities can be quantitated by autoradiography or phosphor-imaging to provide information about the relative number of L- and M-cone pigment genes in the array and, by inference, about the total number of gene copies (Nathans, Thomas, & Hogness, 1986).

A targeted large fragment, which is cleaved by the *Not* I enzyme, carries the entire opsin gene array (see Fig. 1.8). It can be resolved by pulsed field gel electrophoresis and visualized by Southern blotting to provide direct information about the total number of genes in the array (Vollrath, Nathans, & Davis, 1988; Macke & Nathans, 1997). *Not* I fragments differing in length will separate in an electrophoresis gel according to the number of genes that they contain. The lengths can then be measured in kilobases relative to size standards (e.g., concatamers of bacteriophage λ DNA).

Alternatively, fragments cleaved by the *Sma* I and *Rsa* I enzymes (Fig. 1.7D) can be amplified by using polymerase chain reaction (PCR). Opposite ends of the targeted region of the fragments that differ between the L- and M-cone opsin genes are annealed with primer pairs that recognize specific nucleotide sequences (end-product labeling). The primers are then extended in opposite directions by using a DNA polymerase (an enzyme that catalyzes the synthesis of DNA) to add nucleotide bases to cover the entire targeted area. Repetition of the cycle generates copies of the target DNA between the primers in an exponential manner. Amplified 183-bp fragments from the M- and L-cone gene promoter sequences and ca. 300-bp fragments from exon 5 of the M- and L-cone pigment genes (Fig. 1.7D) can be resolved by gel electrophoresis and quantitated to provide information about the number of genes and the ratio of genes, respectively, in the array (Neitz & Neitz, 1995).

(ii) Results: The techniques involving either quantitation of RFLPs detected by Southern blot hybridization after conventional gel electrophoresis (Nathans, Thomas, & Hogness, 1986; Drummond-Borg, Deeb, & Motulsky, 1989; Schmidt et al., 1999) or quantitation of sequence differences by denaturing electrophoresis that resolves DNA fragments based on nucleotide sequence as well as size (single-strand conformation polymorphism electrophoresis; SSCP) after PCR of M- and L-cone pigment gene promoter sequences (Yamaguchi, Motulsky, & Deeb, 1997) yield smaller copy number estimates than those relying on quantitation of end-labeled restriction products after PCR amplification (Neitz & Neitz, 1995; Neitz, Neitz, & Grishok, 1995; Schmidt et al., 1999). The former studies report an average of three pigment genes, with only a single L-cone pigment copy present

Study	Method	Technique	No.	Mean (± SD)	Range
Nathans, Thomas, & Hogness (1986)	RFLP quantitation	gel electrophoresis/ Southern blotting	18	3.1 ± 0.6	2–4
Drummond-Borg et al. (1989)	RFLP quantitation	gel electrophoresis/ Southern blotting	134	3.2 ± 1.0	2–6
Neitz & Neitz (1995)	RFLP quantitation	endlabeled PCR products	27	4.3 ± 1.9	2–9
Neitz, Neitz, & Grishok (1995)	RFLP quantitation	endlabeled PCR products	26	4.5 ± 1.9	2–9
Yamaguchi et al. (1997)	SSCP quantitation	PCR and SSCP	51	2.9 ± 0.8	2–5
Macke & Nathans (1997)	sizing of the *Not* I fragment	pulsed field gel electrophoresis	67	2.9 ± 0.9	1–5
Wolf et al. (1999)	*in situ* visualization	fiber FISH	8	3.5 ± 1.7†	1–6
Schmidt et al. (1999)	sizing of the *Not* I fragment	pulsed field gel electrophoresis	35*	3.3 ± 1.0	2–6
	RFLP quantitation	gel electrophoresis/ Southern blotting	35*	3.9 ± 1.0	2–6
	RFLP quantitation	endlabeled PCR products	35*	4.3 ± 1.4	2–9

Table 1.4: Typical number of genes in the X-chromosome–linked visual pigment gene arrays of white Caucasian males of unselected or color normal (*) phenotype. Some of the methods overestimate the number of genes (see text). († The population was skewed to include extremes.)

and not more than five M-cone pigment copies, whereas the latter studies (Neitz & Neitz, 1995; Neitz, Neitz, & Grishok, 1995) suggest that nearly 50% of all subjects carry two or more (up to four) L-cone pigment genes, or 5'M-3'L hybrid genes, with some having as many as nine gene copies in total (Table 1.4).

Light on the controversy has been shed, however, by recent developments: (i) the application of direct visualization techniques, including pulsed field gel electrophoretic sizing of *Not* I fragments (Macke & Nathans, 1997; Schmidt et al., 1999; see Fig. 1.8) and the fiber FISH (fluorescent in situ hybridization; Parra & Windle, 1993) protocol (Wolf et al., 1999; see Fig. 1.7C); and (ii) comparisons between the various methods in the same population of individuals (Wolf et al., 1999; Schmidt et al., 1999). The two direct procedures, which agree exactly in their results on the same individuals (Wolf et al., 1999), demonstrate that on average a typical array contains three pigment genes. Further, they suggest that reports of frequent occurrences of larger arrays, including those with two or more (up to four) L-cone-pigment genes (Neitz & Neitz, 1995; Neitz, Neitz, & Grishok, 1995), may reflect technical artefacts that are inherent in PCR methods (Macke & Nathans, 1997; Yamaguchi et al., 1997; Schmidt et al., 1999) and ambiguities arising from an inability to distinguish 5'M-3'L hybrid genes from L-cone pigment genes.

Examples of visual pigment gene arrays differing widely in gene number are shown in Fig. 1.9. It presents digitized images of single DNA fibers that have been subjected to dual-color FISH, using the cosmids[4] Qc8B6 and G1160 as probes (see Fig. 1.7C; Wolf et al., 1999). In the images, identification by gene or exon type is not possible because each gene is pseudocolored red and each intergenic region – the ~25-kb region at the downstream or 3' end of each gene – is pseudocolored green. The first (upper) fiber is from a deuteranope (see color blindness). His hybridized fibers exhibit one red and one green signal, indicating the occurrence of a single gene copy. The second, third, and fourth fibers are from trichromats, whose

[4]Cosmids are artificially constructed cloning vectors containing the cos site of bacteriophage λ. They permit cloning of larger DNA segments than can be introduced into bacterial hosts in conventional plasmid vectors.

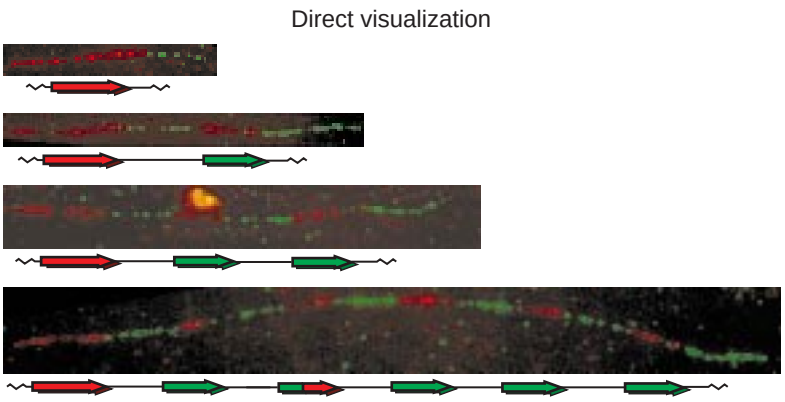

Figure 1.9: Opsin gene array. Single hybridized DNA fibers of four individuals with differently sized opsin gene arrays, containing one, two, three, and six genes (top to bottom). The genes appear red; the intergenic regions, green. The magnification factor is reduced for the six-gene array. For clarity, the location of the genes (arrows) and intergenic regions (straight lines) in each array are schematically depicted below the hybridized images. The color coding has been altered so that red now refers to the L-cone pigment gene, green to the M-cone pigment gene, and red-green to a 5'M-3'L hybrid gene.

hybridized fibers display two, three, and six gene copies, respectively. In each array, the upstream (5') end can be clearly identified because of the characteristic greater length of the most proximal gene.

The arrangement of the gene array. Precise sizing of the number of genes in the array is the important first step in determining the composition of the array. The next steps are to determine the L- to M-cone pigment gene ratio and order of the array. Owing to the high (96%) sequence homology of the L- and M-pigment genes, it is not yet possible to do either of these by direct procedures.

(i) Techniques: Information about the L- to M-cone pigment gene ratio can be obtained by RFLP quantitation of the relative band intensities of fragment pairs. However, the interpretation of these procedures is complicated by the difficulty in distinguishing normal from hybrid genes on the basis of limited restriction fragment pairs (it often requires looking at the D_r/D_g fragments, which encompass exon 5; see Fig. 1.7B) and by potential artefacts such as the background level of radioactivity in the gels and the separation of peaks for the fragment pairs. No procedure is currently available for strictly determining the order of genes in the array, regardless of the array size. A prerequisite for nucleotide sequencing would be the stable propagation in *E. coli* or yeast of large cloned segments with multiple pigment genes. This has not yet been demonstrated. However, very recently Hayashi et al. (1999), using long-range PCR amplification of a 27.4 kb opsin gene fragment, have been able for the first time to completely define the order of gene types in a three gene array (see section on deuteranomaly). This feat is achieved by employing standard techniques to define the most 5' (upstream) gene and also to define the types of genes present without regard to their order in the array; and by employing the long-range PCR to define the most 3' (downstream) gene in the array.

(ii) Results: Generally, the RFLP quantitation methods support the interpretation that there is only a single L-cone pigment gene in the array, occupying the most proximal position, followed by one or more M- or 5'M-3'L hybrid pigment genes (see Yamaguchi et al., 1997; Schmidt et al., 1999). In the Caucasian male population, the range appears to be one to five M-cone pigment gene copies, with a mean of two (Macke & Nathans, 1997; Schmidt et al., 1999). In non-Caucasian populations, both the range and mean are smaller: About one-half of Japanese (48.5%) and Afro-American (42%) males have a single downstream M-cone pigment gene as opposed to about one-fifth (22%) of Caucasian (Jørgensen et al., 1990; Deeb et al., 1992).

The reason for multiple M-cone pigment genes in the array is unclear. It has been speculated that variations in the L- to M-cone ratio in the photoreceptor mosaic (see Fig. 1.1C and Chapter 6) may be related to the number of M-cone pigment genes in the array: The

higher the number of M-cone pigment genes, the greater the number of M-cones. However, Nathans et al. (1992) failed to find any correlation between the relative sensitivities to red and green spectral lights and the number of M-cone pigment genes.

Examples of the variation in pigment gene arrangment are shown in Fig. 1.9 below the hybridized images. The inferred arrangments are based on the FISH results combined with direct sequencing or indirect RFLP quantitation. In the first array, sequencing of exons 2 to 5 establishes that the single gene has only the L-cone pigment-specific sequences (Sharpe et al., 1998). The presence of a single L-cone pigment gene accords with the phenotype, deuteranopia (see page 27). In the second (two gene copies) and third (three gene copies) arrays, RFLP quantitation (based on the A, B, and C fragment pairs) establishes a 1:1 and 1:2 L- to M-cone pigment gene ratio, respectively (Schmidt et al., 1999). Further, the phenotype in both is normal. Therefore, it is reasonable to infer that in both a single L-cone pigment gene is followed by normal M-cone pigment genes. In the fourth (six gene copies) array, RFLP quantitation (based on A and C fragment pairs) establishes a 1:3 L- to M-cone pigment gene ratio; and examination of the B and D fragment pairs reveals a 5'M-3'L hybrid gene (Schmidt et al., 1999). Thus, an L-cone, a 5'M-3'L hybrid, and four M-cone pigment genes are probably present. Because the phenotype is normal, the hybrid gene has been placed at an arbitrary position downstream from a normal M-cone pigment gene (its actual position cannot yet be determined). The placement is consistent with models of gene expression in the array that are described below.

Gene expression in the array. Which opsin genes in the array are actually expressed in the cone photoreceptor cells, quite apart from the number of genes available in the array, can be determined by assessing the ratio of gene mRNA transcripts in extracts of retina. This involves reverse-transcribing cellular RNA into cDNA and then amplifying and quantitating different sequence variants.

mRNA analysis, however, is limited not only by problems inherent to quantitation, but also by two other factors. First, the method can only differentiate whether genes of different types are expressed. That is, if there are five M-cone pigment genes in the array, all with the same sequences, it is not possible to say whether only one or all five are expressed. Second, the method cannot take into account variations in the levels of expression of the gene among individual cone photoreceptor cells. For instance, a difference in the ratio of mRNA transcripts may reflect differences in the number of cones containing L-pigment as compared with the number containing M-pigment, or it may reflect differences in the amount of pigment contained in L-cones as compared with the amount contained in M-cones (e.g., a difference in photopigment optical density; see Chapter 2).

Two issues have been investigated by mRNA transcript analysis: (i) selective expression of the visual pigment genes (Are all genes in complex arrays expressed?) and (ii) differential transcription of the expressed genes (Are L-cone pigment genes more frequently expressed than M-cone pigment genes in the cone photoreceptors?).

(i) Selective expression: Winderickx et al. (1992a) detected in male donor eyes only two retinal mRNA transcripts: one coding for an L-cone pigment and the other for an M-cone pigment. In those donors who had two or more M-cone (or 5'M-3'L hybrid) pigment genes, only one allele was represented in the retinal mRNA. Yamaguchi, Motulsky, and Deeb (1997) and Hayashi et al. (1999) subsequently confirmed and extended these findings. In donors of unknown phenotype with a 5'M-3'L hybrid gene in addition to normal L- and M-cone pigment genes in their array, either (i) the normal L- and M-cone pigment genes were expressed, but not the hybrid gene, or (ii) the normal L-cone and the hybrid pigment genes were expressed, but not the normal M-cone pigment gene. What is decisive is the position of the M and 5'M-3'L hybrid pigment genes in the array. None of these studies found evidence of the presence or expression of more than one L-cone pigment gene (see deuteranomaly). On the other hand, Sjoberg, Neitz, Balding, and Neitz (1998) reported that about 10% of men express more

than one L-cone pigment gene. It is unclear, however, whether the authors, when they refer to extra L-cone pigment genes, are describing L-cone pigment genes or 5'M-3'L hybrid pigment genes that express an L-cone–like pigment. Regardless, they reject the hypothesis that only two opsin genes from one X-chromosome array can be expressed.

(ii) Differential transcription: Yamaguchi, Motulsky, and Deeb (1997) found, in extracts of whole retina, that the ratio of expressed L- to M-cone opsin retinal mRNA varies widely (from unity to ten times greater L-cone opsin expression, with a mode of four) and is not correlated with the ratio of L- to M-cone opsin genes (Yamaguchi et al., 1997). Hagstrom, Neitz, and Neitz (1997), looking at 6-mm-diameter patches of retina (corresponding to about 20 deg of visual angle), reported that the average ratio of L- to M-cone opsin mRNA in patches centered on the fovea was roughly 1.5:1.0, whereas in patches centered at 12 mm (ca. 41 deg) eccentricity it increased to 3.0:1.0 (see also Hagstrom et al., 1998). There were, however, large individual differences among eyes examined: The L- to M-cone opsin mRNA ratios in the fovea patches differed by a factor of greater than 3.

A favoring of L- over M-cone pigment expression, in both the fovea and retinal periphery, is supported by other indirect evidence, including psychophysical spectral sensitivity measurements (e.g., DeVries, 1948a; Brindley, 1954a; Vos & Walraven, 1971; Kelly, 1974; Walraven, 1974; Smith & Pokorny, 1975; Cicerone & Nerger, 1989; Vimal et al., 1989, 1991; Pokorny, Smith, & Wesner, 1991; Wesner et al., 1991; Cicerone et al., 1994), retinal densitometry (Rushton & Baker, 1964), flicker electroretinography (Shapley & Brodie, 1993; Usui et al., 1998), and MSP of human cones (Bowmaker & Dartnall, 1980; Dartnall, Bowmaker, & Mollon, 1983). Taken together, the mean ratios yielded by these methods suggest that there are roughly twice as many L- as M-cones in the central fovea. However, it should be pointed out that individual ratios that are estimated by such methods are highly variable between observers (ranging from 0.33:1 to 10:1) and further that each method, other than MSP, has serious problems of interpretation (see Chapters 2 and 4).

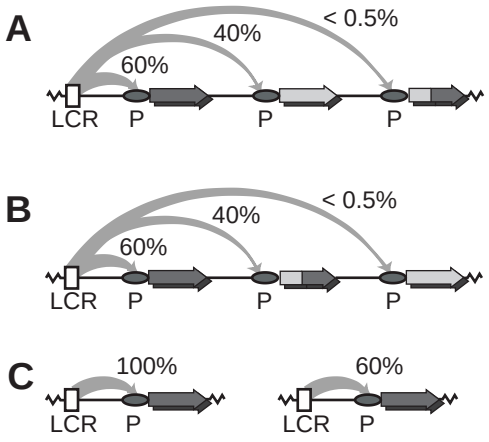

Figure 1.10: Model of opsin gene expression (after Nathans et al., 1989; Winderickx et al., 1992a). (A) Gene expression in an opsin gene array in which a 5'M-3'L hybrid gene (encoding an anomalous L-cone–like pigment) occupies a distal position relative to the normal L- and M-cone pigment genes. An individual with such an array would test as color normal. The expression ratios shown are based on the presumed 1.5 L- to M-cone ratio in the foveal photoreceptor mosaic (see Fig. 1.1C and Chapter 5). (B) Gene expression in an opsin gene array in which the 5'M-3'L hybrid gene occupies a proximal position relative to the normal M-cone pigment gene. An individual with such an array would test as deuteranomalous or deuteranopic, depending on the fusion point of the hybrid gene. This assumes that if the third gene in the array is expressed (< 0.5%), it is not expressed in sufficient amounts to predominate over the upstream genes to enable normal color vision. (C) Two possibilities for gene expression in the reduced, single-gene array of a deuteranope (see section on color blindness). The alternatives are related to the missing cone and replacement cone models in dichromats. (LCR = locus control region; P = promoter region.)

(iii) A model: A possible model of gene expression that incorporates selective expression and differential transcription is shown in Fig. 1.10 (based on Nathans et al., 1989; Winderickx et al., 1992a). The presence of a locus control region (LCR, see blue-cone monochromacy) located between 3.1 and 3.7 kb upstream (5') of the transcription inititation (cap) site of the most proximal gene in the array is known to be required for cone photoreceptor-specific expression (Nathans et al.,

1989; Wang et al., 1992; Nathans et al., 1993). LCRs and other upstream regulatory elements are hypothesized to exert their effects on promoters by a looping mechanism that brings the upstream elements into the vicinity of the transcription initiation site through protein–protein interactions (Knight et al., 1991; Mastrangelo et al., 1991).

Only if the LCR forms a stable transcriptionally active complex with an opsin gene promoter is the gene product expressed in a cone photoreceptor cell. Expression is assumed to be regulated according to a falling gradient: Transcriptionally active complexes between the LCR and the promoters of the most proximal pigment genes are considered more likely than those between the LCR and the promoters of more distal pigment genes. In general, only the two most proximal pigment genes of an array will be expressed in a significant number of cone photoreceptors (for a confirmation of this hypothesis in three-arrays, see page 37), and the first gene in the array (usually an L-cone opsin gene) will be expressed in a higher percentage than the second (usually an M-cone opsin gene).

In accord with the idea that more distal genes are disadvantaged, Shaaban and Deeb (1998) reported that the M-cone pigment gene promoter is two to four times more active than that of the L-cone pigment. Thus, a stronger intrinsic promoter activity may have evolved to offset the distance handicap in competing with the L-cone pigment promoter for coupling to the LCR. However, if LCR complex formation is presumed to be an all-or-none mechanism, promoter strength would not be relevant.

(iv) L- versus M-cone pigment choice: The model allows for the mutually exclusive expression of L- and M-cone pigment genes in single-cone photoreceptor cells, but does not address the question of how, during development, a photoreceptor cell chooses which visual pigment to produce. It simply assumes that L- and M-cones are merely distinguished by the pigment that they contain, and not by other membrane and/or transcriptional factors as well. A mutually exclusive stochastic process that gives preference to a single gene, combined with male hemizygosity and female X-chromosome inactivation (see heterozygotic carriers of protan and deutan defects), allows only one L- and M-cone pigment gene locus to be expressed per cell. The choice is also preserved in all descendant cells, obviating any requirement for coordinating or suppressing promoter–LCR interactions at a second locus (Wang et al., 1992). In short, this "stochastic" model predicts that all the machinery needed for L- versus M-cone segregation is inherent in the gene array, not only in humans but in all mammals who share identical transcriptional regulatory proteins.

An alternative model is that L- versus M-cone specificity in cell to cell contact and information transfer is predetermined during development. This "standard" model assumes that a set of regulatory molecules, including transcriptional factors, differs between L- and M-cone photoreceptor cells and orchestrates the choice of L- and M-cone pigment gene interaction with the LCR as well as the production of any additional proteins that differ between the two cell types. Shared transcriptional factors could also determine the matching up of L- and M-cones with their postreceptoral neurons. Some indirect evidence is consistent with regulatory molecules playing a crucial role in the differentiation of cone photoreceptor cells (Furukawa et al., 1997) and bipolar neurons (Chiu & Nathans, 1994). This model predicts that only primates with trichromatic color vision, among mammals, have evolved the requisite transcriptional regulators to recognize L- and M-cone opsin genes.

To distinguish between the competing models, Wang et al. (1999) investigated whether a mouse, which normally possesses only a single X-linked opsin gene, could support mutually exclusive expression of the human L- and M-cone opsin genes when these are integrated into its genome. They generated transgenic mice carrying a single copy of a minimal human X-chromosome opsin gene array in which the L- and M-cone opsin gene transcription units and 3' intergenic sequences were replaced by alkaline phosphatase (AP) and β-galactosidase (β-gal) reporters, respectively, and then determined the pattern of expression in the cone photoreceptor cells: 63% of expressing cones had AP activity, 10% had β-gal activity, and 27% had activity for both reporters. Thus a large fraction of

cone photoreceptors in the mouse retina can efficiently and selectively express either of the reporter transgenes. The sixfold higher average frequency of the cones expressing AP (the L-cone opsin gene reporter) over β-gal (the M-cone opsin gene reporter) may have arisen from the former's greater proximity to the LCR or from the inherently greater efficiency of activation of the L-cone promoter. The doubly labeled cones could have arisen from simultaneous expression of both reporters or from an occasional switch between expression of one reporter and the other (the phenomenon deserves closer inspection).

The mixed expression aside, these findings tend to support the stochastic model and thereby a simple evolutionary path for the development of trichromacy after visual pigment gene duplication (see pigment gene evolution). They imply that the differentiation of L- and M-cone signals occurs opportunistically at more distal stages of visual processing.

Color blindness

Normal color vision can vary significantly among males and females. A large part of the diversity is due to individual differences in the optical densities of the photopigments and the lens and macular screening pigments (see Chapter 2). A substantial part is also due to polymorphisms in the normal L- and M-cone pigment genes (see page 13), which constitute a hypervariable locus in the human genome (see Neitz & Neitz, 1998).

Trichromacy, however, is not enjoyed by all. Many people are partially color blind or color deficient, confusing colors that full trichromats – regardless of their individual differences in color vision – distinguish easily. A very few are completely color blind, unable to discriminate color at all.

Even among color normals, trichromacy fails under several conditions. A complete collapse is associated with the transition from cone-mediated day vision to rod-mediated night vision: As night falls, colors pale and merge into darker and lighter shades of gray. And a partial reduction to a form of dichromacy – tritanopia or the loss of S-cone function (see page 44) – is associated with small, brief targets: (i) viewed centrally (foveal tritanopia), where S-cones are missing;[5] (ii) viewed in the periphery (small field tritanopia), where S-cones are sparse (see Figs 1.1C and 1.17B); and (iii) viewed immediately after the extinction of strong, yellow, adapting fields (transient tritanopia), which is assumed to polarize the yellow-blue color subsystem (Stiles, 1949; Mollon & Polden, 1977a). Trichromacy also fails, to a limited extent, under several extreme conditions of chromatic adaptation and photopigment bleaching.

Inherited color vision defects. The most common forms of color blindness (a.k.a. color vision deficiency or daltonism) are inherited. They arise from alterations in the genes encoding the opsin molecules, an explanation remarkably guessed at by George Palmer (Voigt, 1781; Mollon, 1997). Either genes are lost (due to intergenic nonhomologous recombination), rendered nonfunctional (due to missense or nonsense mutations or coding sequence deletions), or altered (due to intragenic recombination between genes of different types or possibly point mutations).

Phenotypically, the results of the gene alterations are either: (i) anomalous trichromacy (when one of the three cone pigments is altered in its spectral sensitivity but trichromacy is not fully impaired); (ii) dichromacy (when one of the cone pigments is missing and color

[5]S-cones are completely missing in a central retinal disc, which is about 100 μm (ca. 0.34 deg) in diameter (Curcio et al., 1991). Their absence, confirmed by labeling with anti-blue opsin in whole-mounted retinae (Curcio et al., 1991), accords with much prior psychophysical evidence that the center of the human fovea is tritanopic for very small objects (König, 1894; Wilmer, 1944; Wilmer & Wright, 1945; Wald, 1967; Williams, MacLeod, & Hayhoe, 1981a, 1981b; Castano & Sperling, 1982). An estimate of the blue-blind region obtained from psychophysics, which takes into account the blur introduced by ocular optics and eye movements, is 0.42 deg (ca. 118 μm) in diameter (Williams, MacLeod, & Hayhoe, 1981a,b). The exclusion of S-cones from the central foveola is usually attributed to the need to counteract the deleterious effects of light scattering and axial chromatic aberration on spatial resolution, which cause blurring and/or defocus particularly at short wavelengths (see also Chapter 5).

vision is reduced to two dimensions); or (iii) monochromacy (when two or all three of the cone pigments are missing and color and lightness vision is reduced to one dimension). Under special conditions, color vision may be in some respects tetrachromatic, when an extra cone pigment is present, owing to X-chromosome inactivation (see page 38).

Those inherited alterations affecting a single cone pigment are referred to by the generic names protan (from the Greek *protos*: first + *an*: not), deutan (*deuteros*: second), and tritan (*tritos*: third), to distinguish disorders in the L-, M-, and S-cone pigments, respectively. The ordering is related to the history of discovery and to early assumptions about the nature of the disorders (von Kries & Nagel, 1896; Boring, 1942; Judd, 1943). The suffix "-anomaly", when appended to the generic names, indicates a deviation or abnormality in the function of the L- (protanomaly), M- (deuteranomaly), or S- (tritanomaly) cone pigments. Likewise, the suffix "-anopia" indicates the absence of function of the L- (protanopia), M- (deuteranopia), or S- (tritanopia) cone pigments.

Those inherited alterations resulting in the loss of two cone pigments are referred to as "cone monochromacies" (Pitt, 1944; Weale, 1953). They include blue- or S-cone monochromacy (affecting both the L- and M-cone opsin genes), green or M-cone monochromacy (affecting the S- and L-cone opsin genes), and red or L-cone monochromacy (affecting the S- and M-cone opsin genes).

Other inherited forms of color blindness arise from mutations in genes not encoding the cone opsins, but rather components of cone structure and function. Those resulting in the loss of function of all three cone types are referred to as "complete achromatopsia" or "rod monochromacy."

Acquired color vision defects. The less common forms of color blindness arise from factors other than inherited alterations in the opsin genes. For instance, cerebral achromatopsia or dyschromatopsia (see Chapter 14), a form of total color blindness, can arise adventitiously after brain fever (Boyle, 1688; Mollon et al., 1980), cortical trauma, or cerebral infarction (Critchley, 1965; Meadows, 1974; Damasio et al., 1980; Zeki, 1990; Grüsser & Landes, 1991; Kennard et al., 1995; Rüttiger et al., 1999).

Yet other forms of color blindness may be associated with: (i) disorders of the prereceptoral ocular media; (ii) fundus detachment; (iii) progressive cone dystrophies or degenerations affecting all cone classes with or without involvement of the rods; (iv) macular dystrophies and degenerations; (v) vascular and hematologic diseases; (vi) glaucoma; (vii) hereditary dominant optic atrophy and other optic nerve diseases; (viii) diseases of the central nervous system (e.g., multiple sclerosis) or other organs (e.g., diabetes mellitus); and (ix) toxic agents (e.g., lead, tobacco, alcohol) that affect the retina or the optic tracts (see Birch et al., 1979).

Protan and deutan defects

The most common, hereditary color blindnesses are the loss (protanopia and deuteranopia) and alteration (protanomaly and deuteranomaly) forms of protan (L-cone) and deutan (M-cone; sometimes written as deuteran) defects. Also known as red-green color vision deficiencies, they are associated with disturbances in the X-linked opsin gene array. Their characteristic X-linked recessive pedigree pattern (see Fig. 1.16A) was early remarked upon by Earle (1845) and Horner (1876) and clearly recognized as such by Wilson (1911). Wilson's assignment of the gene(s) responsible for red-green color blindness to the X-chromosome is the event that is often taken to mark the beginning of the mapping of the human genome. That two gene loci, one for protan and one for deutan defects, were involved was subsequently deduced from the characteristics of doubly heterozygous females[6] (Waaler, 1927; Brunner, 1932; Kondo, 1941; Drummond-Borg et al., 1989). Such females were known to carry genes for both types of color blindness but exhibited normal color vision, implying the existence of distinct defective and normal visual pigment genes on each of their X-chromosomes (Vanderdonck & Verriest, 1960; Siniscalco et al., 1964; Drummond-

Borg et al., 1989).

Phenotypes. A variety of special color confusion charts (e.g., the Dvorine, Ishihara, and Stilling pseudoisochromatic plates), hue discrimination or arrangement tasks (e.g., the Farnsworth–Munsell 100-Hue test, the Farnsworth Panel D-15, the Lanthony Desaturated D-15), and lantern detection tests (e.g., the Edridge–Green, Holmes–Wright), all of which exploit the color deficits of the color blind, have been designed to screen for protan and deutan defects (for an overview of the available clinical tests, see Lakowski, 1969a, 1969b; Birch et al., 1979; Birch, 1993). For more precise characterizations, spectral appearance data (including unique hue loci for anomalous trichromats and spectral neutral points for dichromats), luminous efficiency, and hue discrimination functions may be determined.

Traditionally, however, observers with protan and deutan defects are most efficiently and elegantly characterized by the nature of their Rayleigh matches (Rayleigh, 1881) on a small viewing field (≤2 deg in diameter) anomaloscope (the best known is the Nagel Type I). In this task, the observer is required to match a spectral yellow (ca. 589-nm) primary light to a juxtaposed mixture of spectral red (ca. 679-nm) and green (ca. 544-nm) primary lights. There are two variables: the intensity of the yellow and the relative mixture of the red and green lights. Ideally, the S-cones and rods are excluded so that the match is determined solely by the relative absorptions in the M- and L-cones.

Most trichromats reproducibly choose a unique match between the red/green mixture ratio and the yellow intensity (see Fig. 1.11). Slight differences between their match midpoints are usually attributed to the normal variability in the λ_{max} of the L- and M-pigments, such as caused by the alanine/serine polymorphism at codon 180 in exon 3. Variation can also be attributed to differences in photopigment optical density. Some studies have reported match points as being bimodally (Neitz & Jacobs, 1986; Winderickx et al., 1992b) or quadrimodally (Neitz & Jacobs, 1990) distributed in male trichromats. However, other studies have found only unimodal distributions in both male and female trichromats (Jordan & Mollon, 1988; Lutze et al., 1990). The differences may be specific to the psychophysical methodology: Unimodality may be associated with the method-of-adjustment and multimodality with the forced-choice method (Piantandia & Gille, 1992).

In contrast, individuals with protan and deutan defects have displaced Rayleigh match midpoints (i.e., the mean value of the red–green ratio required to match the yellow primary falls outside the normal range) and/or extended matching ranges (they accept more than one red–green ratio).

Incidences. Incidences of X-chromosome–linked defects vary between human populations of different racial origin (see Table 1.5). The highest rates are found in Europeans and the Brahmins of India; the lowest in the aborigines in Australia, Brazil, the South Pacific Islands, and North America. The differences between populations are well documented (e.g., Garth, 1933; Cruz-Coke, 1964; Iinuma & Handa, 1976, Fletcher & Voke, 1985) and are usually explained by relaxation in selection pressures, which could arise as hunting and gathering cultures evolve toward industrialized societies (e.g., Post, 1962, 1963; however, see Kalmus, 1965). Other possible biological factors include gene flow, the rise of mutant genes, and the migration and mixture of races. Interestingly, there is no evidence for analogous within-species variations in color vision in the best-investigated Old-World monkey, the macaque (for a review, see Jacobs, 1993). An analysis of behavioral results suggests that if any individual variations in color vision among macaques occur, their frequency must be less than 5% (Jacobs & Harwerth, 1989).

Of relevance perhaps to the population differences is the fact that there tends to be fewer opsin genes in

[6]Males who carry genes on the X-chromosome for protan and deutan color vision defects are hemizygous (the mutant gene is present in one copy). In contrast, females can be heterozygous (when a mutant gene is paired with a normal gene), homozygous (when two mutant genes of the same type are paired), or doubly heterozygous (when two mutant genes of different types are paired).

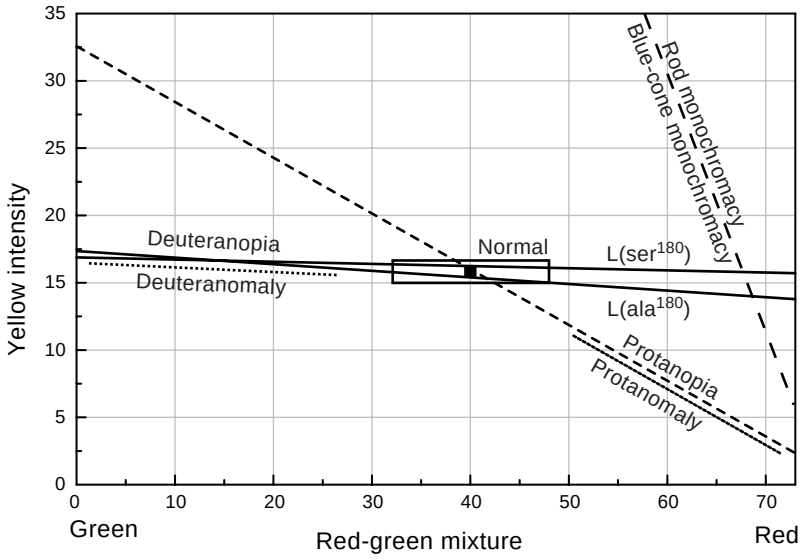

Figure 1.11: Rayleigh matches obtained on a standard Nagel Type I anomaloscope. Characteristic match points or ranges are shown for normal trichromats (the black square indicates the mean, the open rectangle the spread of normal matches), for protanopes with a single L1M2 gene in the X-linked opsin gene array, for deuteranopes with either a single L(ser^{180}) or L(ala^{180}) gene in the array, for an extreme protanomalous trichromat, for an extreme deuteranomalous trichromat, for a blue-cone monochromat, and for a rod monochromat.

the arrays of non-Europeans. Extra M-cone genes are believed to increase the opportunities for intragenic and intergenic recombinations between the L- and M-cone pigment genes, which gives rise to phenotypic color vision defects (see unequal crossing over).

The incidences in females are much lower than those in males because the defects are inherited as recessive traits. Males, who have only one X-chromosome, are hemizygous and will always manifest a color defect if they inherit an aberrant gene at one of the first two positions in the gene array. Females, on the other hand, have two X-chromosomes, one inherited from each parent, so they will not usually show a complete manifestation of the typical colour defect unless they are homozygous. Thus, the incidence in Caucasian females should correspond to the sum of the squared frequencies of singly heterozygous females, which should be the same as the frequencies for protan (p_p) and deutan (p_d) defects in hemizygous males: ($p_p^2 + p_d^2$) or 0.39%. (Table 1.6 provides the incidences of protan and deutan defects, divided according to the dichromatic and anomalous trichromatic forms.)

The slightly higher observed figure (0.42%, see Table 1.6, and 0.50%, see Table 1.5) in Caucasian females, and the even larger discrepancy in Asians, is probably explained by the partial manifestation of color blindness in heterozygotic carriers (see page 38) and by problems of testing methodology and inclusion criteria (for a discussion, see Fletcher & Voke, 1985).

Protanopia and deuteranopia

Protanopia and deuteranopia are the dichromatic or loss forms of protan and deutan defects.

Phenotypes. The classical theory was that protanopes and deuteranopes each lack one of the normal pigments, the L- and M-cone pigments, respectively, and contain the others unchanged (König & Dieterici, 1886). Everything else being equal, they were expected to accept the color matches of color normals; their dichromatic color world being a reduced version of the trichromatic one. We now know that the situation is more complex. Considerable variation occurs in both phenotypes and genotypes.

Although some protanopes and deuteranopes are true reduction dichromats, having only one X-chromosome–linked cone photopigment, which is identical to the normal's M- or L-cone pigment, others are not.

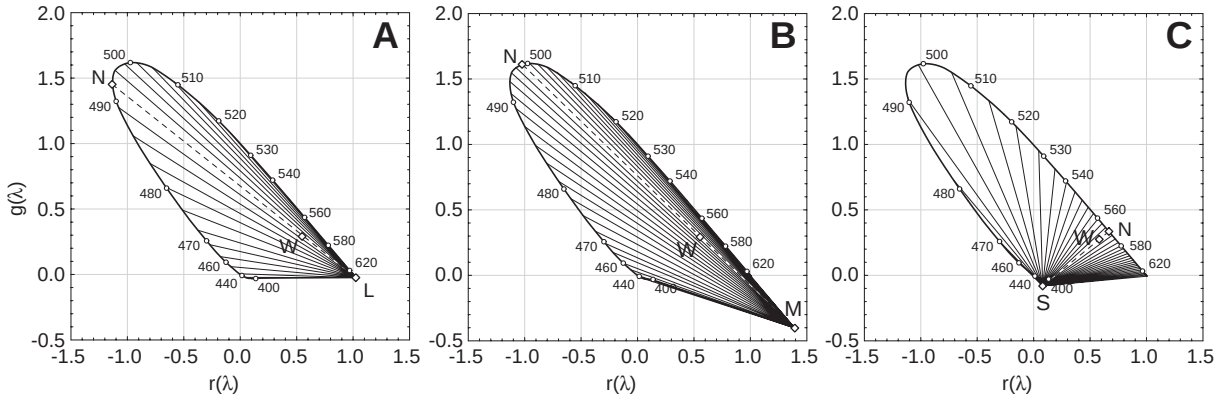

Figure 1.12: Dichromatic lines of constant chromaticity or hue for (A) protanopes, (B) deuteranopes, and (C) tritanopes, drawn in the Stiles and Burch (1955) 2-deg $r(\lambda)$, $g(\lambda)$ chromaticity space (see Chapter 2). The lines have been extrapolated to converge at the confusion or copunctal point, which defines the chromaticity of the missing 2-deg L- (for protanopes), M- (for deuteranopes), or S- (for tritanopes) cone fundamental of Stockman and Sharpe (2000a). The spacing of the lines is based on the spectral wavelength discrimination data of protanopes (Pitt, 1935), deuteranopes (Pitt, 1935), and tritanopes (Fischer et al., 1951; Wright, 1952). The neutral points, N, for the equal-energy white, W, and the corresponding confusion lines (dotted lines) are shown.

Some protanopes have only one X-chromosome–linked cone photopigment, a hybrid, that is not identical to the normal M-cone. Among deuteranopes, the situation is complicated by the L(ser^{180}) and L(ala^{180}) polymorphism. This diversity explains – to a large extent – why systematic individual differences in color discrimination and Rayleigh matches among protanopes (Alpern & Wake, 1977) and deuteranopes have been observed. And, accordingly, why, in so-called classical confrontation experiments, a given protanope did not always accept the color matches made by a normal trichromat (Mitchell & Rushton, 1971; Rushton et al., 1973).

(i) Confusion colors: By definition, dichromats require only two primaries to match all color stimuli. As a result, they confuse or fail to discriminate colors that are easily distinguished by normal trichromats. As first pointed out by Maxwell (1855) and demonstrated by Helmholtz in his *Physiological Optics* (1867), when the colors confused by dichromats are plotted in a chromaticity diagram, the axes of which may be generated from transformations of standard color matching functions or from representations of cone excitations (see Chapter 2), they lie on a series of straight lines called "confusion loci."

The confusion lines of protanopes and deuteranopes are shown in Fig. 1.12. The line spacing is adjusted to correspond to the number of just noticeable wavelength differences within the equal-energy spectrum (i.e., the wavelength represented by one line is just distinguishable from that represented by its neighbor). For protanopes, only about 21 distinct wavelengths can be discriminated; whereas for deuteranopes, only about 31 can be (slightly lower estimates were obtained by Pitt, 1935). In contrast, the normal discriminates about 150 wavelengths[7] in the spectrum (Wright & Pitt, 1934; Boring, 1942).

(ii) Neutral zone and spectral colors: In both protanopes and deuteranopes, the spectrum is dichromic, consisting of just two pure hues (Herschel, 1845). In contrast, normals see at least seven pure hues: red, orange, yellow, green, cyan, blue, and violet. This property was first remarked upon by John Dalton (1798), a deuteranope (Hunt et al., 1995, see below): "I

[7]The total number of discriminable, nonspectral, hues, of course, is many more. Wool graders at the Gobelin Tapestry works, in the 19th century, were known for being able to distinguish at least 20,000 different hues (Chevreul, 1839).

see only two, or at most three, distinctions [in the solar image]. My yellow comprehends the red, orange, yellow, and green of others; and my blue and purple coincide with theirs" (p. 31).

The two spectral regions for protanopes and deuteranopes – one blue, the other yellow – are separated by a neutral or achromatic zone that is indistinguishable from white and corresponds to where the relative absorptions of the two remaining cone classes are the same as for white (see Fig. 1.13). The midpoint of the zone – the neutral point – which, by definition, falls on the confusion line passing through the physiological white point, is relatively easy to specify. Quantitative measures, however, are obscured by the arbitrariness of the physiological white (see Walls & Heath, 1956), individual variability in the density of the ocular and macular pigmentation (which affects the chromaticity of the physiological white), and the failure to take into account phenotypic differences between dichromats of the same class. For protanopes and deuteranopes, representative neutral point values for a white standard light (of color temperature 6774 K) are 492.3 nm (Walls & Mathews, 1952; Walls & Heath, 1956; Sloan & Habel, 1955) and 498.4 nm (Walls & Mathews, 1952; Walls & Heath, 1956; Massof & Bailey, 1976), respectively.

But what are the spectral hues actually seen by the red-green dichromats above and below the neutral zone? Although Dalton and other dichromats often assign conventional hue names to the shorter (blue) and longer (yellow) wavelength regions of the spectrum, it does not follow that their usage coincides with that of the color normal. In fact, the impossibility of passing from which colors color blind observers confuse to which they actually see was recognized almost as soon as color blindness was discovered (Wilson, 1855; see Judd, 1948). The preferred approach to overcoming the ambiguity has been to study the vision of unilateral color blinds, born with one normal eye and one color-deficient eye (the earliest report is from von Hippel, 1880, 1881). Observations, which must be treated cautiously,[8] suggest that, for both protanopes and deuteranopes, below the neutral zone, the blue at 470 nm, and, above the neutral zone, the yellow at 575 nm have the same hue for the two eyes (Judd, 1948 provides a review of historical cases; see also Sloan & Wallach, 1948; Graham & Hsia, 1958; Graham et al., 1961; MacLeod & Lennie, 1976).

(iii) Luminous efficiency: The two types of dichromats also differ in which part of the equal-energy spectrum appears brightest (see Chapter 2 and Wyszecki & Stiles, 1982a). For the normal, the maximum of the luminosity function is on average near 555 nm, for the protanope it is closer to 540 nm (blue-shifted to the λ_{max} of the M-cone pigment), and for the deuteranope it may be closer to 560 nm (red-shifted to the λ_{max} of the L-cone pigment).

(iv) Color perception: Given that red-green dichromats only see two colors in the spectrum, and assuming that the colors are known, it is possible to simulate for the color normal the extent of the color confusions made by the dichromats. The history of such simulations begins with von Goethe (1810). In his *Farbenlehre*, he included a reproduction of a small water color that he painted to demonstrate how the landscape would appear to those lacking the blue sensation (Akyanoblepsie). Since then, many other simulations have been attempted (Holmgren, 1881; Rayleigh, 1890; Ladd-Franklin, 1932; Evans, 1948; Viénot et al., 1995).

Figure 1.14 presents simulations for the normal observer of how a colorful fruit market is perceived by a protanope (Fig. 1.14B) and a deuteranope (Fig. 1.14C). For both, the normally wide gamut of yellowish-green to red colors is dramatically reduced. Clearly, protanopes are not red-blind (because they lack the L- or "red" cones) and deuteranopes green-blind (because they lack the M- or "green" cones); rather, both are red-green blind. They only distinguish such colors on the basis of saturation and lightness variations. The major difference in the images is that

[8]Such cases are extremely rare, and it is never been fully demonstrated that the eyes described as normal and dichromatic are in fact really so (see Bender et al., 1972; MacLeod & Lennie, 1976). Further, the etiology of such a discrepancy between eyes is uncertain. It could be owing to somatic forward or back mutations or, in heterozygotes, to differential lyonization in the two eyes (see heterozygotic carriers of protan and deutan defects).

Racial group	Male		Female	
	No.	Incidence	No.	Incidence
European descent	250,281	7.40	48,080	0.50
Asian	349,185	4.17	231,208	0.58
Africans	3,874	2.61	1,287	0.54
Australian aborigines	4,455	1.98	3,201	0.03
American Indians	1,548	1.94	1,420	0.63
South Pacific Islanders	608	0.82	—	—

Table 1.5: Racial incidence of red-green color vision deficiencies in males and females. The incidences (in percent) correspond to the total number of color defectives found divided by the total number examined in each racial group. The raw numbers were obtained from 67 studies listed in Waaler (1927), Cox (1961), Post (1962), Waardenburg (1963a), Crone (1968), Iinuma and Handa (1976), and Koliopoulos et al. (1976).

Gender	No.	Protan		Deutan	
		Anomaly	Anopia	Anomaly	Anopia
Male	45,989	1.08	1.01	4.63	1.27
Female	30,711	0.03	0.02	0.36	0.01

Table 1.6: Incidences (in percent) for X-chromosome–linked color vision deficiencies in men and women. The raw numbers were taken from populations in Norway (Waaler, 1927), Switzerland (von Planta, 1928), Germany (Schmidt, 1936), Great Britain (Nelson, 1938), France (François et al., 1957), The Netherlands (Crone, 1968), Greece (Koliopoulos et al., 1976), and Iran (Modarres et al., 1996–1997). Many studies included in Table 1.4 could not be included in this table because they did not separate deficiencies according to phenotype.

reds appear relatively darker in the protanopic simulation than in the deuteranopic, owing to the reduced spectral sensitivity of the protanope to long wavelengths. And, protanopes tend to confuse reds, grays, and bluish blue-greens; whereas deuteranopes tend to confuse purples, grays, and greenish blue-greens.

Diagnosis. The Rayleigh match for dichromats, like all color vision tests, exploits the dichromat's color confusions. Because they have only one pigment (or only one functionally distinct pigment), they have one degree of freedom in the Rayleigh equation and are able to fully match the spectral yellow primary to any mixture of the spectral red and green primary lights by merely adjusting the intensity of the yellow, regardless of the red-to-green ratio. Thus, instead of a unique match, they will have a fully extended matching range that encompasses both the red and green primaries. Although the deuteranope will display a normal or near-normal relation in the luminosity of his matches, the protanope will display a luminosity loss for the red primary, requiring less yellow light to match it.

When a regression line is fitted to the matches of either a protanope or deuteranope, its slope and intercept with the yellow intensity axis will depend on the λ_{max} of the underlying photopigment (see Fig. 1.11). In the Nagel Type I instrument, the protanope needs a very dim yellow intensity to match the red primary and a bright intensity to match the green primary (hence a steep slope); whereas the deuteranope uses about the same yellow brightness to match both primaries (hence a flat slope). The intersection of the regression lines of protanopes, who retain normal M-cone pigments, and deuteranopes, who retain normal L-cone pigments, will coincide with the average match of normal trichromats. Regression lines of intermediate slope will be generated by dichromats who have a single X-linked anomalous pigment, such as those encoded by 5'L-3'M hybrid genes. These will not intersect with the average normal match (Jägle, Sharpe, & Nathans, unpublished).

Incidences. Protanopia (1.01%) and deuteranopia

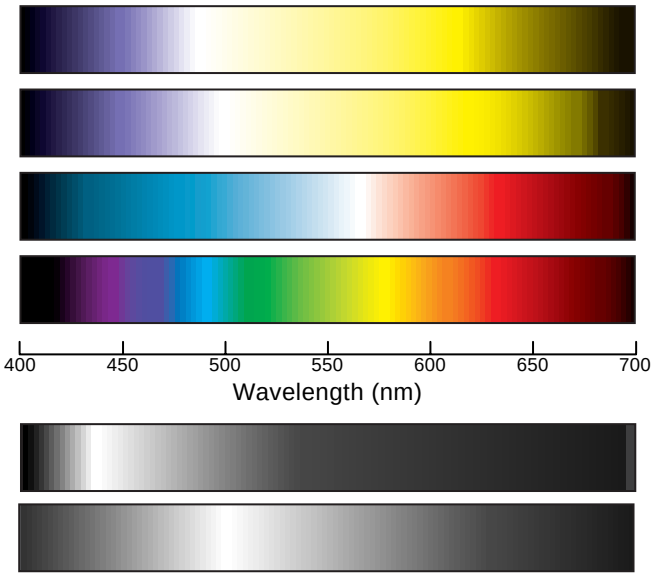

Figure 1.13: The appearance of the visible spectrum for five types of color blindness: protanopia, deuteranopia, tritanopia, blue-cone monochromacy, and rod monochromacy. Neutral points (or zones) are indicated by the white region in the spectrum for the three types of dichromats. These divide the spectrum into two hues. The shorter wavelength hue is perceived below the neutral point, the longer above. The saturation of both increases with distance from the neutral point. The brightest part of the spectrum roughly corresponds to the λ_{max} of the luminosity function (photopically determined, save for the rod monochromat).

(1.28%) are about equally frequent in the European (Caucasian) male population (see Table 1.6). This is in marked contrast to protanomaly and deuteranomaly and is worthy of note.

Molecular genetics. The opsin genes may be compromised in three ways to give rise to red-green dichromacy: by point mutations, major sequence deletions, and by unequal crossing over.

(i) Point mutations: In principle, red-green dichromacy can arise from a point mutation[9]: a single-base alteration in which the normal nucleotide has been substituted by an incorrect one. The point mutation could occur either (i) in the coding sequences of the L- or M-cone pigment gene, which vitiates or destabilizes the opsin molecule or impairs its quantum efficiency, or (ii) in the upstream promoter sequences, which leads to failure of expression of the opsin. One such naturally occurring missense mutation has already been identified (Winderickx et al., 1992c; see also blue-cone monochromacy). It is a thymine-to-cytosine substitution at nucleotide 648 in the coding region, resulting in the replacement of cysteine by arginine at codon 203 ($cys^{203}arg$) in the second extracellular loop of the amino acid sequences in exon 4. The substitution disrupts a highly conserved disulfide bond formed with another cysteine at codon 126, in the first extracellular loop. The rupture is thought to damage the three-dimensional structure of the opsin, rendering it nonviable.

So far, the $cys^{203}arg$ mutation has been found in all of the M-cone pigment genes in one individual with deuteranopia or extreme deuteranomaly (Winderickx et al., 1992c) and in the cone pigment genes of multiple blue-cone monochromats (Nathans et al., 1989, 1993). Although fairly common in the population – estimated frequency of 2% by Winderickx et al. (1992c) and 0.5% by Nathans et al. (1993) – it is not

[9]Point mutations include: (i) synonymous mutations, which do not alter the encoded amino acid residue and, therefore, have no effect on the properties of the opsin; (ii) missense mutations, which alter the encoded amino-acid residue, either with properties similar to (in which case they may be harmless) or different from (in which case they may be disastrous) those of the replaced residue; and (iii) nonsense mutations, which convert an amino-acid–specifying codon to a stop codon, inappropriately signalling the termination of translation and ending the polypeptide chain synthesis of the opsin.

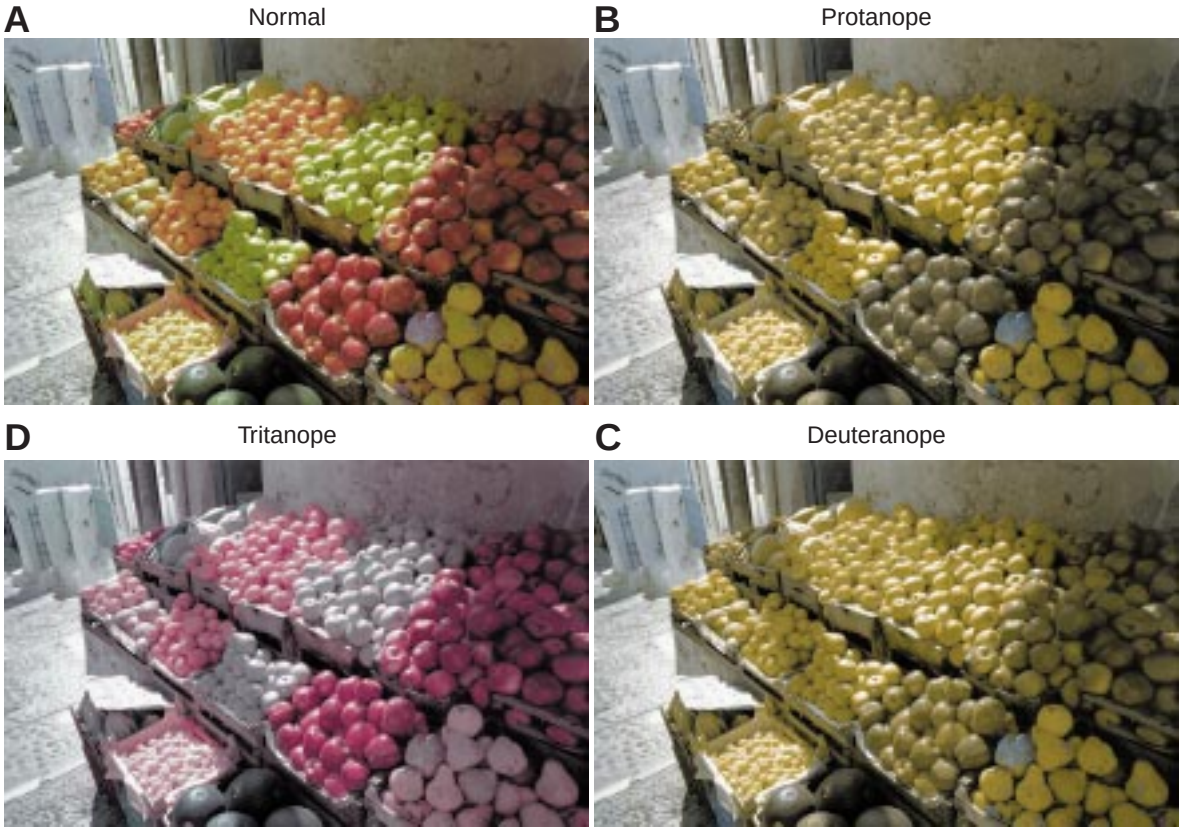

Figure 1.14: A scene from a fruit market as perceived by a normal trichromat (A), a protanope (B), a deuteranope (C), and a tritanope (D). The simulations are based on an algorithm incorporating a colorimetric transformation, which also makes explicit assumptions about the residual sensations experienced by dichromats (see Viénot et al., 1995; Brettel et al., 1997). The transformation replaces the value of each original element by the corresponding value projected onto a reduced color stimulus surface, parallel to the direction of the missing cone fundamental axis in an LMS cone excitation space (see Chapter 2). It is informed by assumptions about the neutral zones of dichromats and the colors perceived by unilateral colorblind observers. The simulations must be treated as approximate, not only because of the limitations of the strictly colorimetric transformation involved and of the printing procedure, but also because the transformed image is viewed by the normal visual system, which will process it differently than the color blind system. (The original scene is reproduced by kind permission of the Minolta Corp.)

always associated with a color vision defect. Its expression may depend on the position in the array of the gene in which it occurs (see gene expression above).

(ii) Major sequence deletions: Red-green dichromacy could also arise from major deletions in the coding sequences or promoter region of one of the opsin genes. A case in which a rearrangement, possibly a deletion, between exons 1 and 4 of the L-cone opsin gene engendered a protan defect with a progressive macular degeneration has been reported (Reichel et al., 1989).

(iii) Unequal crossing over: Dichromacy is most typically the result of unequal homologous crossing over or recombination between the L- and M-cone pigment genes. During meiosis, when the maternal and

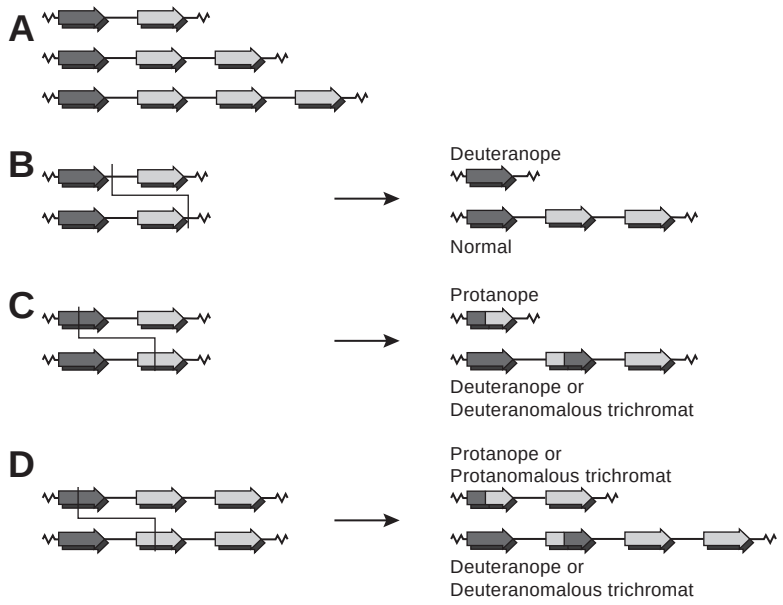

Figure 1.15: Examples of unequal recombination within the tandem array of X-linked genes responsible for the common protan and deutan defects of color vision. Each gene is represented by an arrow: The base corresponds to the 5' end and the tip to the 3' end. Dark gray arrows, L-cone pigment genes; light gray arrows, M-cone pigment genes. Unique flanking DNA is represented by zig-zag lines and homologous intergenic DNA by straight lines. For each recombination event, both products are shown, most of which lead to a color-defective phenotype when inherited in males. (A) Three wild-type (naturally occurring) gene arrays, each containing one L-cone pigment gene and a variable number of M-cone pigment genes. (B) Intergenic recombination events leading to deuteranopia and an enlarged gene array. (C) and (D) Intragenic recombination events leading to protanopia, protanomaly, deuteranopia, or deuteranomaly.

paternal X-chromosomes in a female are aligning to form the X-chromosome of an egg (prophase I), breakage occurs, with an exchange of corresponding sections of DNA sequences and the subsequent rejoining. Normally the exchange is reciprocal, but sometimes, when the chromosomes are misaligned, it is unequal. Such unequal recombination events are facilitated by the juxtaposition of the genes in the head-to-tail tandem array and by their high degree of homology (Nathans, Thomas, & Hogness, 1986; Drummond-Borg et al., 1989).

The mechanisms of recombination inducing dichromacy are summarized in Fig. 1.15 (after Nathans, Thomas, & Hogness, 1986; Nathans, Piantandia, et al., 1986). For simplicity, the common arrangement of the gene array is assumed to be a single L-cone opsin gene followed by one, two, or three M-cone opsin genes (Fig. 1.15A).

Unequal intergenic recombination, in which the crossing-over point occurs in the region between genes, can reduce the gene array to a single L-cone pigment gene (Fig. 1.15B). Males who inherit such an X-chromosome arrangement from their mothers will be single-gene deuteranopes (Nathans, Thomas, & Hogness, 1986; Nathans, Piantanida, et al., 1986; Vollrath et al., 1988; Drummond-Borg et al., 1989). Less than one-half of all deuteranopes display such a pattern (Sharpe et al., 1998). Those who do constitute a true L-cone reduction type of deuteranope (König & Dieterici, 1886), save perhaps for those who carry an M-cone–specific exon 2 embedded within an otherwise L-cone pigment gene, which may slightly change the L-cone photopigment spectrum (see Table 1.3).

Unequal intragenic recombination, in which the crossing-over point typically occurs in the introns between exons, can have several outcomes. It can reduce the array to a single 5'L-3'M hybrid gene (Fig. 1.15C), which will usually be associated with protanopia. Less than one-half of all protanopes display such a pattern (Sharpe et al., 1998); of those, more than one-half possess any one of several M-cone–like hybrid pigments, the spectral sensitivity of which is shifted relative to the normal M-cone pigment (see Table 1.3). Such protanopes can be thought of as anomalous, for

their photopigment spectral sensitivities will be nonrepresentative of those of color normals. The remainder, whose array carries a single 5'L-3'M hybrid gene encoding either an L1M2 or an L2M3 pigment, are of the true reduction type, with a spectral sensitivity resembling that of the normal M-cone pigment (Sharpe et al., 1998).

Alternatively, unequal intragenic recombination can replace the normal L-cone pigment gene with a 5'L-3'M hybrid gene (Fig. 1.15D). The array consists of a 5'L-3'M hybrid gene and one or more normal M-cone pigment genes (Deeb et al., 1992; Sharpe et al., 1998). More than one-half of all protanopes display this pattern (Sharpe et al., 1998). Presumably, they express more than one X-linked visual pigment gene. Protanopia is implied only if the fusion in the hybrid gene occurs before exon 3 (because the gene will encode an M-cone pigment). Otherwise, protanomaly is predicted, because the gene will encode an anomalous pigment that differs in spectral sensitivity from the normal M-cone pigment by 2 to 8 nm, depending on which amino acid residues have been substituted (see Table 1.3).

Finally, unequal intragenic recombination can replace the normal M-cone pigment gene with a 5'M-3'L hybrid gene (Deeb et al., 1992; Sharpe et al., 1998). The array consists of a normal L-cone pigment gene and a 5'M-3'L hybrid gene (Figs. 1.15C and 1.15D). More than one-half of all deuteranopes display this pattern (Sharpe et al., 1998). Deuteranopia is only implied if the fusion occurs before exon 2 (because the gene will encode a pigment with a spectral sensitivity similar to that of the normal L-cone pigment). Otherwise, deuteranomaly is predicted, because the gene will encode an anomalous pigment that differs in spectral sensitivity from the normal L-cone pigment by 4 to 12 nm, depending on which amino acid residues have been substituted (see Table 1.3).

The position in the array of the hybrid pigment gene is likely to be critical. Four to eight percent of Caucasian males with normal color vision have, in addition to normal L- and M-cone pigment genes, 5'M-3'L hybrid genes that apparently are not expressed or are insufficiently expressed to disrupt normal color vision (Drummond-Borg et al., 1989; Deeb et al., 1992). Presumably, in these individuals, the hybrid genes occupy a more distal, 3' position in the gene array (see gene expression, above).

(iv) John Dalton's genotype: Historically, John Dalton is the most famous person known to be color blind. In his honor, the term *daltonism* is used in many languages, including English, French, Spanish, and Russian, to refer to color vision disorders.

Believing his own colorblindness to be caused by his eye humors being blue-tinted, he directed that they be examined upon his death. However, a postmortem analysis conducted on 28 July 1844 failed to confirm his hypothesis (Wilson, 1845; Henry, 1854). Fortunately, sufficient eye tissue was preserved after the autopsy to permit, 150 years later, a molecular genetic analysis (Hunt et al., 1995). PCR amplification of opsin gene fragments from the tissue revealed that only a single L-cone opsin sequence was present, consistent with Dalton being a single-gene deuteranope (see Figs. 1.9A and 1.15B). The genotype is contrary to previous interpretations, including that of Thomas Young (1807), who thought Dalton a protanope. But it accords with a close reading of the historical record and with colorimetric analysis of the color confusions and failures of color constancy reported by Dalton himself (Dalton, 1798; Hunt et al., 1995).

These findings represent, to date, the most impressive example of genotyping of a known historical figure, and they provide a satisfying historical continuity to the field of the molecular genetics of color blindness.

Large-field trichromacy. Nagel (1905, 1907) was the first to observe that many red-green dichromats, himself included, are only completely dichromatic for small viewing fields restricted to the central fovea. With larger viewing fields, they become partially trichromatic and are able to make red-green color discriminations. Nagel's observations have been confirmed by many others relying on small- and large-field Rayleigh or neutral point matches (Jaeger & Kroker, 1952; Smith & Pokorny, 1977; Nagy, 1980; Breton & Cowan, 1981) and color naming (Scheibner &

Boynton, 1968, Nagy & Boynton, 1979; Montag, 1994). It is now widely believed that the majority of dichromats display such behavior.

In cases in which the dichromats have more than one gene in the opsin gene array, the improvement in color discrimination with large fields might be attributed to genotype. Some dichromats could have two separate genes, one normal and one a hybrid, encoding slightly different pigments, the benefit of which only manifests under large-field conditions. In particular, a small minority of deuteranopes may have a normal M-cone pigment gene, distal to the 5'M-3'L hybrid gene (Deeb et al., 1992). If the downstream, normal gene is expressed (in reduced amounts according to its greater distance from the LCR), it might influence color discrimination under large-field conditions. However, 5'M-3'L hybrid genes may only be expressed and only influence phenotype when they occupy the second position of the opsin gene array (Hayashi et al., 1999). A similar explanation is unlikely to apply to protanopes because the presence of downstream L-cone, or, for that matter, 5'M-3'L hybrid pigment genes, has never been demonstrated in these subjects.

However, these explanations fail to account for another observation: some, but not all, single-gene dichromats also display large-field trichromacy (Nathans et al., 1986; Neitz, Neitz, & Jacobs, 1989; Deeb et al., 1992; Crognale et al., 1999). Moreover, at least one individual, who was missing all of the M-cone pigment genes in the X-chromosome array, behaved as an extreme deuteranomalous observer, even under small-field viewing conditions (Nathans et al., 1986). This suggests that Rayleigh matches are not dependent solely on the number of different X-chromosome–linked photopigments. Presumably other factors contribute to the rejection of red-to-green mixture ratios (see also Smith & Pokorny, 1977; Nagy & Boynton, 1979; Nagy, 1980; Breton & Cowan, 1981; Nagy & Purl, 1987; Crognale et al., 1999). These could include: (i) rod intrusion; (ii) S-cone intrusion; (iii) differences in pigment optical density between central and parafoveal cones; (iv) spatial variation in the macular pigment; (v) changes in cone receptor geometry and orientation with eccentricity (e.g., wave-guiding); and (vi) dissimilarities in the morphology and/or function of M- and L-cone photoreceptors, apart from the photopigment that they contain (e.g., placing an L-cone pigment in an M-cone structure might produce a different spectral sensitivity than placing it in an L-cone structure).

Cone mosaic. Dichromats may have fewer cones than normal or some of their cones may lack pigment: the L-cones may be missing or empty in protanopes and the M-cones may be missing or empty in deuteranopes. On the other hand, they may have a full complement of functioning cones: either the missing pigment is entirely replaced by ones of the available type (the M-cones in protanopes and the L-cones in deuteranopes) or the empty cone structures are filled with photopigment of the available type (M-cone pigment in protanopes and L-cone pigment in deuteranopes). A complication here, as alluded to above, is that we do not yet know whether M- and L-cones differ in ways other than the pigment that they contain in their outer segments. It is unlikely that S-cones could replace the missing cones because their numbers are far too small (although it is unknown what ultimately limits their number). Likewise, it is unlikely that the S-cone pigment could replace the M- or L-cone pigment because the S-cone opsin is specified by a gene on a different chromosome.

Surprisingly, the predictions about effects of changes in the cone array on absolute threshold and luminous efficiency have rarely been tested (Abney, 1913; DeVries, 1948a; Hecht & Hsia, 1948; Hecht, 1949; Wald, 1966; Berendschot et al., 1996; Knau & Sharpe, 1998).

In hindsight of the molecular genetics, it is possible to make specific predictions. Dichromats possessing two (or more) pigment genes in their array should not have any missing or nonfunctioning foveal cones (see Fig. 5.2 for a high-resolution image of the living retina of a multigene protanope, subject MM, whose phenotype has been established by Nathans, personal communication). They should have a normal complement of cones with the remaining pigment (either L or M)

plus an extra complement with the de facto pigment encoded by the hybrid gene. This could result in an effective doubling or tripling of the number of L- or M-cones compared with the trichromat eye (assuming a normal ratio of 1.5:1.0; see Fig. 1.1C and Chapter 5).

On the other hand, those dichromats possessing a single L- or 5'L-3'M hybrid pigment gene may or may not have cones that are missing pigment. It would depend on whether all of the non-S-cones are filled with the only available longer wavelength sensitive photopigment.

If the missing pigment is not replaced, then poorer performance on some visual sensitivity and visual acuity tests might be expected (ca. 40% of the cones will be empty). However, an impairment would be difficult to measure because of the limiting effects of the eye's optics and because of the random clustering of the remaining cone photoreceptors (see Fig. 1.1C and Chapter 5) and the constant microsaccadic movements of the eye, which would tend to compensate for any blind areas.

Protanomaly and deuteranomaly

Protanomaly and deuteranomaly are the alteration forms of protan and deutan defects.

Phenotypes. The color vision deficits of anomalous trichromats are usually less severe than those of dichromats, but there is considerable variability among individuals. They can be categorized as simple or extreme, according to their matching behavior on the Rayleigh equation (Franceschetti, 1928). Many simple anomals may be unaware of their color vision deficiency, whereas many extreme anomals may have nearly as poor color discrimination as dichromats.

(i) Color perception: Color terms may be skillfully used by many anomalous trichromats. Unlike dichromats, they do not have a neutral zone and see more than two hues in the spectrum. The saturation of spectral lights may differ from normal (Pokorny & Smith, 1977). But many can assign some hue names, such as blue, cyan, green, yellow, and orange, to unique parts of the spectrum, although the wavelength locations may differ from those of normal trichromats (von Kries, 1919; Rubin, 1961; Hurvich & Jameson, 1964; Smith, Cole, & Isaacs, 1973; Romeskie, 1978). In general, their hue locations are shifted to shorter (protanomals) and longer (deuteranomals) wavelengths. Their spectrum may be conceived as encompassing saturated blues, extremely desaturated bluish-greens and greens, yellows, and relatively desaturated oranges (Pokorny, Smith, & Verriest, 1979).

(ii) Luminous efficiency: Protanomals, like protanopes, have an insensitivity to the far red end of the visible spectrum, with a blue-shifted luminosity function. But deuteranomals, like deuteranopes, may have a normal or slightly red-shifted luminosity function.

Diagnosis. *Rayleigh equation:* The Rayleigh matches of protanomalous and deuteranomalous trichromats are properly characterized by both their displaced match midpoints and their matching ranges. The two measures vary independently (Schmidt, 1955; Hurvich, 1972).

Generally, simple protanomals have a narrow matching range displaced to red (i.e., they require more red than normal in the matching field), with a concomitant luminosity loss for the red primary. Likewise, simple deuteranomals have a narrow matching range displaced to green, with an essentially normal luminosity. Neither tend to accept the matches of normals, and their deviant anomalous ratios are expressed relative to normal ratios by the so-called anomalous quotient (see Pokorny, Smith, & Verriest, 1979).

A very few simple anomalous trichromats accept only one mixture ratio setting (i.e., a unique match), which is outside the normal limits. They are known as minimally affected or *red-green deviants* (Vierling, 1935) and have excellent chromatic discrimination.

On the other hand, extreme anomalous trichromats exhibit a wide matching range, often encompassing the matches of normal trichromats as well as those of anomalous trichromats and sometimes even one (but never both) of the primaries (see Fig. 1.11). The extreme protanomal will match the red-green mixture to the yellow primary except when pure or near pure

green is used; the extreme deuteranomal will do the same except when pure or near pure red is used. The midpoint of the matching range is usually displaced from the center of normal matches.

Incidence. Deuteranomaly (4.61%) is more than four times more frequent than protanomaly (1.07%) in Caucasian males (see Table 1.6). Its greater frequency may be related to a peculiarity of intragenic recombination. Deuteranomaly can arise from unequal crossing over in simple arrays, each having only a single M-cone pigment, whereas protanomaly can only arise if one of the arrays contains two or more normal M-cone pigment genes (see Figs. 1.15C & 1.15D).

Molecular genetics. (*i) Crossing over:* Simple and extreme anomalous trichromacy typically occur as the result of unequal intragenic crossing over between the L- and M-cone pigment genes. Either there is a replacement of the normal L- by a 5'L-3'M hybrid pigment gene or the normal M- by a 5'M-3'L hybrid pigment gene. The position of the hybrid gene in the array is presumably critical (see Figs. 1.10A and 1.10B). For instance, protanomaly results only if a 5'L-3'M hybrid gene replaces the normal L-cone pigment gene in the most proximal position (for examples, see Balding et al., 1998). Likewise, deuteranomaly results only if a 5'M-3'L hybrid gene replaces the normal M-cone pigment gene and is expressed preferentially over the downstream normal M-cone pigment genes (for examples, see Hayashi et al., 1999). This would explain the not infrequent presence of 5'M-3'L hybrid genes in individuals with normal color vision (estimated to be 4–8% in Caucasians and <1% in Japanese).

However, it is still debatable in the case of deuteranomalous trichromats with three or more genes in the array, one of which is a normal M-cone opsin gene, whether the pigment with the anomalous spectral sensitivity encoded by the 5'M-3'L hybrid gene simply replaces the normal M-cone pigment, is present in cones as a mixture with the normal M-cone pigment, or is present in a subset of M-cones.

The first of these possibilities is strongly supported by experiments using long-range PCR amplification. Hayashi et al. (1999) studied 10 deutan males (8 deuteranomalous and two deuteranopic) with three opsin genes on the X-chromosome: an L, an M, and a 5'M-3'L hybrid gene. The 5'M-3'L hybrid gene was always at the second position; the first position being occupied by the L-cone opsin gene. Conversely, in two men with L-cone, M-cone, and 5'M-3'L hybrid opsin genes and normal color vision, the 5'M-3'L hybrid gene occupied the third position. When pigment gene mRNA expression was assessed in postmortem retinae of three men with the L-cone, M-cone and 5'M-3'L hybrid genotype, the 5'M-3'L hybrid gene was expressed only when located in the second position. Thus they conclude that the 5'M-3'L gene will only cause deutan defects when it occupies the second position of the opsin gene array. This is consistent with "two gene only" expression at this locus (see section on gene expression and Fig. 1.10B). An alternative hypothesis for the molecular basis of deuteranomaly (Sjoberg, Neitz, Balding, & Neitz, 1998) – namely, that the M-cone opsin genes are mutated and express a nonfunctional pigment – is unlikely because Hayashi et al. (1999) could not find mutations in the promoter or in the coding sequences of the M-cone opsin genes of the deutan subjects that they investigated (Deeb et al., 1992; Winderickx et al.,1992a; Yamaguchi, Motulsky, & Deeb, 1997). However, it is conceivable that distal gene expression is silenced by some elements in the 3' flanking region of the locus.

(ii) Hybrid pigments: Any comprehensive model of anomalous trichromacy must account for the fact that anomalous trichromats differ greatly in the location and range of their Rayleigh matching points (for review, see Hurvich, 1972; Pokorny, Smith, & Verriest, 1979; Mollon, 1997). The finding of a diverse family of hybrid pigments in single-gene protanopes (Sharpe et al., 1998, 1999) supports a model of anomalous trichromacy in which any one of many M- or L-like anomalous pigments can be paired with either of two major polymorphic versions of the more similar normal pigment (Merbs & Nathans, 1992b; Neitz, Neitz, & Kainz, 1996). As the spectral sensitivities of the normal or the anomalous pigment shift, the midpoint of the Rayleigh match will shift, and as the sep-

aration between the spectral sensitivities of the normal and anomalous pigments increases or decreases, the better or poorer will be the subject's chromatic discrimination (see also Shevell et al., 1998).

Inspection of Table 1.3 reveals a wide range of possible anomalous pigments lying between the normal L- and M-cone pigments. Rather than a continuous distribution, there is a clustering of 5'L-3'M hybrid pigments having λ_{max}'s within about 8 nm of the λ_{max} of the M-cone pigment and a clustering of 5'M-3'L hybrid pigments having λ_{max}'s within about 12 nm of the λ_{max} of the L-cone pigment.

Although the λ_{max} of the anomalous pigment may be the decisive factor in determining the severity of anomalous trichromacy, other factors such as variation in the optical density of the expressed hybrid pigment may also have to be taken into account (He & Shevell, 1995). This could come about if the amino acid substitutions alter the stability and/or activity of the expressed pigment as well as its spectral position. As of yet there is no solid evidence that hybrid pigments have different optical densities than normal M- or L-cone pigments (see sequence variability and opsin viability above).

(iii) Alternative theories: The in vivo identification of the hybrid pigments (see Table 1.3) is inconsistent with the theory that all anomalous trichromats share a single anomalous pigment with a spectral sensitivity at a position intermediate between the normal L- and M-cone pigments (DeVries, 1948a; MacLeod & Hayhoe, 1974; Hayhoe & MacLeod, 1976). Nor does it accord with the theory that there is one anomalous pigment common to all deuteranomalous trichromats and another common to all protanomalous trichromats (Rushton et al., 1973; Pokorny, Moreland, & Smith, 1975; Pokorny & Smith, 1977; DeMarco et al., 1992). Finally, it does not support, in its most extreme form, the theory that there are no distinct anomalous pigments but instead clusters of normal L- and M-cone pigments that differ in their peak sensitivities, and that anomalous trichromats draw their pigments from the same rather than different clusters (Alpern & Moeller, 1977; Alpern & Wake, 1977). While this theory fails to account for the existence of bona fide hybrid pigments, it explains the phenotypes of that subset of anomalous trichromats in which the normal and anomalous pigments are distinguished only by the polymorphic alanine versus serine difference at position 180. It is also prescient in recognizing how the anomalous hybrid pigments cluster near the remaining normal pigment.

(iv) Red-green deviants: Molecular genetic explanations of red-green deviants are not yet forthcoming. Very few have ever been identified and none have been investigated by molecular techniques. One possible explanation is that red-green deviants possess an anomalous pigment, produced by a point mutation, whose λ_{max} lies outside the range defined by the normal M- and L-cone pigments (i.e., either blue-shifted relative to the M-cone pigment or red-shifted relative to the L-cone pigment).

Cone Mosaic. The cone mosaic of anomalous trichromats should resemble those of normal trichromats. The only difference being that the pigment in one of the normal cone classes is replaced by an anomalous pigment.

Heterozygotic carriers of protan and deutan defects

About 15% of women inherit an X-chromosome carrying an abnormal opsin gene array from one parent and an X-chromosome carrying a normal opsin gene array from the other (see Fig. 1.16A). As heterozygous carriers for red-green color vision defects, they partially possess that which they fully bequeath to one-half of their sons. Early in their own embryonic development, by a process of dosage compensation known as X-chromosome inactivation or lyonization (Lyon, 1961, 1972), one of the two X-chromosomes in their somatic cells is transcriptionally silenced (see Fig. 1.16B). The process is normally random, with an equal probability that either the maternal or paternal inherited X-chromosome will be inactivated in the cone precursor cells. However, extremes of asymmetrical inactivation in which the same X-chromosome is inactivated in all, or almost all, of the precursor cells have

been reported in female monozygotic twins and triplets (Zanen & Meunier, 1958a, 1958b; Pickford, 1967; Koulischer et al., 1968; Philip et al., 1969; Yokota et al., 1990; Jørgensen et al., 1992), as have differences in the symmetry between the two eyes (Jaeger, 1972; Feig & Ropers, 1978).

Phenotypes. As a result of X-chromosome inactivation, the retinal mosaic of heterozygotic carriers of dichromacy and anomalous trichromacy will be mottled with patches of color-blind and normal areas (cf., Krill & Beutler, 1965; Born et al., 1976; Grützner et al., 1976; Cohn et al., 1989). Thus, they may partly share in the color vision defects of their sons and fathers (Pickford, 1944, 1947, 1949, 1959). But some carriers of anomalous trichromacy and carriers of anomalous protanopia may actually benefit from the presence of the fourth (anomalous) cone pigment in their retinae. They could be tetrachromatic (DeVries, 1948a), enjoying superior color discriminability and an extra dimension of color vision. The more dissimilar the λ_{max} of the hybrid pigment is to those of the normal M- and L-cone pigments, the greater the potential advantage.

A related issue is to determine the psychophysical consequences, (in the heterozygous human female eye) as the result of X-chromosome inactivation, when the two polymorphic variants of the L-cone opsin gene are both expressed in those women who inherit a gene for the serine variant on one X-chromosome and a gene for the alanine variant on the other (Neitz, Kraft, & Neitz, 1998).

(i) Weak tetrachromacy: A weak form of tetrachromacy arises if the heterozygous carriers have four different types of cones, but lack the postreceptoral capacity to transmit four truly independent color signals. This form has been demonstrated in heterozygotes who accept trichromatic color matches but do not exhibit the stability of matches under the chromatic adaptation that is required when only three types of cones are present in the retina (Nagy et al., 1981). The condition may be analogous to rod participation influencing color matching in the normal trichromatic eye (cf. Bongard et al., 1958; Trezona, 1973).

(ii) Strong tetrachromacy: A strong form of tetrachromacy arises if the heterozygous carriers have four different types of cones plus the capacity to transmit four independent cone signals. Such heterozygotes would not accept trichromatic color matches; rather, they would require four variables to match all colors. Up to now, strong tetrachromacy has never been convincingly demonstrated (Jordan & Mollon, 1993), and it is unclear whether the human visual system is sufficiently labile to allow for it during embryonic and postnatal development of the retina and cortex. A model does exist in New-World monkeys, however, demonstrating how heterozygous females can acquire an extra dimension of color vision from the action of X-chromosome inactivation upgrading dichromacy to trichromacy (Mollon, Bowmaker, & Jacobs, 1984; Mollon, 1987; Mollon & Jordan, 1988; Jordan & Mollon, 1993; see Chapter 6).

Diagnosis. Viewing field size is an important factor in detecting color vision defects in heterozygotic carriers (Möller-Ladekarl, 1934; Walls & Mathews, 1952; Krill & Schneidermann, 1964; Ikeda et al., 1972; Verriest, 1972; De Vries-De Mol et al., 1978; Swanson & Fiedelman, 1997; Miyahara et al., 1998). They are clearly disadvantaged relative to normals, when the target is reduced to a point source (Wald et al., 1974; Born et al., 1976; Grützner et al., 1976; Cohn et al., 1989) or when the time allotted for discriminating small color targets is restricted (Verriest, 1972; Cohn et al., 1989; Jordan & Mollon, 1993). This suggests that eye movements may mitigate against retinal mottle by preventing the target from merely falling on dichromatic areas.

(i) Luminous efficiency (the Schmidt sign): In heterochromatic flicker photometry and other measures of relative spectral sensitivity (see Chapter 2), carriers of protanopia (and protan defects in general) exhibit a more protan-like function (i.e., reduced sensitivity to long-wave lights) – known as the Schmidt sign (Schmidt, 1934, 1955; Walls & Mathews, 1952; Crone, 1959; Krill & Beutler, 1965; Adam, 1969; Ikeda et al., 1972; De Vries-De Mol et al., 1978; Miyahara et al., 1998). On the other hand, carriers of deuter-

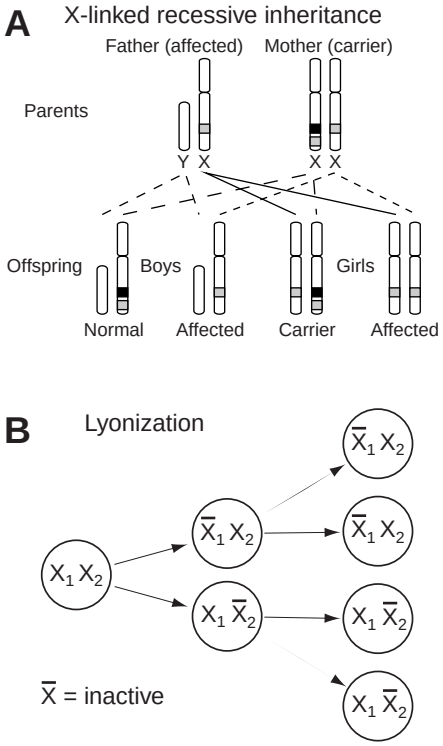

Figure 1.16: Heterozygous defects. (A) X-linked recessive inheritance for protan (or deutan) defects. In the example shown, the father (affected) lacks the L-cone opsin gene on his single X-chromosome, while the mother (carrier) lacks it on one of her X-chromosomes. On average, one-half of the boys will be affected hemizygotes and one-half will be normal; one-half of the girls will be obligate carriers and one-half affected homozygotes. The black and gray squares on the q-arm of the X-chromosome indicate normal L- and M-cone opsin genes, respectively. (B) Schematic representation of X-chromosome inactivation or lyonization. One X-chromosome is normal (X_1), the other is not (X_2). In the early embryonic progenitor cell (left), both X-chromosomes are active. During embryonic development, X-inactivation (middle) is random (50% probability is assumed), so in 50% of the cells X_1 will be inactivated, and in the other 50% X_2 will be inactivated. The descendant lines (right) derived from the original inactivation events retain the same active and inactive chromosomes.

anopia (and deutan defects in general) may exhibit a more deutan-like (shifted to longer wavelengths) function (DeVries, 1948a; Crone, 1959; Adam, 1969; Yasuma et al., 1984).

(ii) Rayleigh equation: Both abnormal and normal Rayleigh matches have been reported in heterozygotic carriers (see Jordan & Mollon, 1993; Miyahara et al., 1998). For most carriers of dichromacy, abnormal matches are predicted only when the matching field size is so small that it falls on a dichromatic region or so that one-half of the field falls on a dichromatic region and the other half on a trichromatic region. For carriers of anomalous trichromacy and for dichromat carriers with a hybrid gene, the presence of a fourth anomalous pigment might disturb the Rayleigh match; introducing the possibility of multiple (3) acceptable matchpoints and/or an extended matching range.

Incidences. Approximately 4.5% of Caucausian females are heterozygous carriers of either protanopia or deuteranopia and about 11% are carriers of anomalous trichromacy (see Table 1.6; Feig & Ropers, 1978; Cohn et al., 1989).[10]

Molecular genetics. The molecular genetic causes of heterozygotic red-green color vision defects are the same as for the hemizygotic defects. A method for the molecular genetic detection of female carriers of protan defects has recently been reported (Kainz, Neitz, & Neitz, 1998).

Cone mosaic. *(i) Carriers of dichromacy:* The cone mosaic of heterozygotic carriers of dichromacy will be composed of mottled patches of color-blind (in which either the L- or M-cones are lacking) and normal areas (Krill & Beutler, 1965; Born et al., 1976; Grützner et al., 1976; Cohn et al., 1989). The relative complement of L- and M-cones in the mosaic can be predicted given certain assumptions, such as a 50% inactivation of both X-chromosomes, a 1.5 ratio of expression of the L- and M-cone pigment genes in the normal opsin gene array, and photopigment replacement operating in the defective opsin gene array. The

[10]The percentage of heterozygotes can be calculated by the equation $2p\,(1 - p)\,(1 - d^2) + 2d\,(1 - d)\,(1 - p^2)$, where p and d are the incidences of protanopia (or protanomaly) and deuteranopia (deuteranomaly), respectively, taken from Table 1.6.

predicted L- to M-cone ratio is 0.43:1.0 for protanopic carriers and 4.0:1.0 for deuteranopic carriers (if it is assumed instead that no replacement is operating, the ratios will be 0.6:1.0 and 3.0:1.0, respectively).

(ii) Carriers of anomalous trichromacy: For carriers of protanomaly and deuteranomaly, the mosaic will be a pattern of retinal areas containing normal L- and M-cones alternating with retinal areas containing either L- or M-cones and hybrid pigment cones containing an anomalous pigment. Assuming 50% inactivation of both X-chromosomes and a normal 1.5:1 L- to M-cone ratio, the predicted ratios will be 0.75:0.75:1.0 for the L-, 5'L-3'M hybrid (M-like), and M-cone pigments in carriers of protanomaly and 3.0:1.0:1.0 for the L-, 5'M-3'L hybrid (L-like), and M-cone pigments in carriers of deuteranomaly.

Blue-cone monochromacy

Blue-cone monochromacy is a rare form of monochromacy or total color blindness caused by loss or rearrangement of the X-linked opsin gene array (Nathans et al., 1989, 1993). It is also known as blue-mono-cone-monochromacy (Blackwell & Blackwell, 1957, 1961), S-cone monochromacy, and X-chromosome–linked incomplete achromatopsia. The S-cones are the only photoreceptors, other than the rods, believed to be functioning (Blackwell & Blackwell, 1957, 1961; Alpern, Lee, & Spivey, 1965; Hess et al., 1989; Hess, Mullen, & Zrenner, 1989).

Phenotype. The disorder is characterized by severely reduced visual acuity (Green, 1972; Zrenner, Magnussen, & Lorenz, 1988; Hess et al., 1989; Hess, Mullen, & Zrenner, 1989); a small, central scotoma (corresponding to the blue-blind central foveola; see Fig. 1.1C); eccentric fixation; infantile nystagmus (which diminishes and/or disappears with age); and nearly normal appearing retinal fundi. There also may be associated myopia (Spivey, 1965; François et al., 1966). A cone-rod break (Kohlrausch kink) in dark adaptation curves occurs, denoting the transition from S-cone to rod function.

In some pedigrees with individuals suspected of having blue-cone monochromacy, considerable intrafamilial variation with respect to residual color discrimination has been reported (Smith et al., 1983; Ayyagari et al., 1998). Thus, phenotypes may vary considerably. Indeed, it may be difficult to distinguish between blue-cone monochromats and some types of incomplete rod monochromats (see section on rod monochromacy).

In some cases, residual L-cone function has been reported (Smith et al., 1983); in others, the presence of cones whose outer segments contain rhodopsin instead of a cone photopigment, replacing the L- and M-cones (Pokorny, Smith, & Swartley, 1970; Alpern et al., 1971), has been inferred. The former reports may be evidence for incomplete manifestation of the disorder; the latter have not been confirmed by other investigators (Daw & Enoch, 1973; Hess et al., 1989) and are unlikely on molecular biological grounds.

(i) Heterozygotic manifestation: Minor abnormalities have been reported in some heterozygotic female carriers of blue-cone monochromacy (Krill, 1964; Spivey et al., 1964; Krill, 1969).

(ii) Color perception: In complete blue-cone monochromats, there is no neutral point; the entire spectrum is colorless (see Fig. 1.13). The photopic luminosity function peaks near 440 nm, the λ_{max} of the S-cone spectral sensitivity function (see Fig. 1.1A). Although blue-cone monochromats are usually considered to be totally color blind (the term monochromacy implies that they can match all spectral colors with one variable), they may have residual dichromatic color perception that arises from interactions between the S-cones and rods at mesopic (twilight) levels where both are active (Alpern et al., 1971; Daw & Enoch, 1973; Pokorny, Smith, & Swartley, 1970; Young & Price, 1985; Hess et al., 1989; Reitner et al., 1991).

Diagnosis. *(i) Specific tests:* The disorder is revealed by its obvious clinical signs and by the accumulation of results on conventional color vision tests. Two tests exist, however, that are specifically designed to detect monochromats: the Sloan Achromatopsia test and the François-Verriest-Seki (FVS; see Verriest &

Seki, 1965) test plates (see Birch, 1993). To distinguish blue-cone monochromats from rod monochromats, a special four-color plate test (Berson et al., 1983) and a two-color filter test (Zrenner, Magnussen, & Lorenz, 1988) exist. The latter test is based on the observation that the blue-cone monochromat's spatial acuity is improved by viewing through blue cut-off filters (Blackwell & Blackwell, 1961; Zrenner, Magnussen, & Lorenz, 1988).

(ii) The Rayleigh equation: Although not optimumly suited for the task, the phenotype can be identified with the Rayleigh equation, in conjunction with other tests. If the luminance levels are such that the rods are functioning, blue-cone monochromats will be able to match the red primary light and some mixtures of the green and red primary lights to the yellow reference light. But they will be unable to match the green primary light or most red-green mixtures to the yellow reference because of their great sensitivity to the green primary light (i.e., there is too little light in the yellow primary). The slope of the regression line fitted to their limited matches corresponds to the λ_{max} of the rods and should be the same as that of rod monochromats (see Fig. 1.11).

Incidence. Like protan and deutan defects, blue-cone monochromacy is transmitted as an X-linked, recessive trait (Spivey et al., 1964; Spivey, 1965; see Fig. 1.13A). The frequency is estimated at about 1:100,000 (exact incidences are unavailable), and it is not associated with or predicted by combined protanopia and deuteranopia inheritances (see Table 1.6). All cases reported so far, beginning with Blackwell and Blackwell (1957, 1961), have been male (see Sharpe & Nordby, 1990a, for a review), which is not surprising because the frequency in females is estimated to be as low as 1:10,000,000,000.

Molecular genetics. The genetic alterations that give rise to blue-cone monochromacy all involve either a loss or a rearrangement of the L- and M-cone pigment gene cluster. They fall into two classes and are characterized by either one-step or two-step mutational pathways (Nathans, Davenport, et al., 1989, Nathans, Maumenee, et al., 1993).

In the one-step pathway, nonhomologous deletion of genomic DNA encompasses a region between 3.1 and 3.7 kb upstream (5') of the L- and M-cone pigment gene array (see Fig. 1.10). The deletions, which range in size from 0.6 to 55 kb, can extend into the gene array itself, which is otherwise unaltered in structure. These deletions occur in an essential segment of DNA, referred to as the locus control region (LCR). Experiments with reporter constructs[11] in transgenic mice show that the LCR is essential for cone-specific gene expression (Wang et al., 1992)

In the two-step pathway, there are multiple causes. Most typically, in one of the steps, unequal homologous recombination reduces the number of genes in the tandem array to a single gene. In the other step, the remaining gene is inactivated by a point mutation. The temporal order of the steps is unknown, but for the most common point mutation unequal homologous recombination is likely to be the last step.

Three point mutations have so far been identified (Nathans, Davenport, et al., 1989, Nathans, Maumenee, et al., 1993): two missense and one nonsense mutation (see Fig. 1.17A). The most frequent missense mutation is in an extracellular loop. It is a thymine-to-cytosine substitution at nucleotide 1,101 (Nathans, Davenport, et al., 1989), with the result that the essential cysteine residue at codon 203 in exon 4 is replaced by arginine ($cys^{203}arg$), thereby disrupting a highly conserved disulfide bridge. The less frequently identified missense mutation is in a membrane loop: a cytidine-to-thymidine substitution at nucleotide 1,414, resulting in a replacement of the proline residue by leucine at codon 307 ($pro^{307}leu$) in exon 5. The nonsense mutation is also in a membrane loop, and it, too, involves a cytidine-to-thymidine substitution at nucleotide 1,233, resulting in the replacement of arginine by a termination codon (TGA; see Table 1.1) at 247 ($arg^{247}ter$) in exon 4.

[11] These are recombinant DNA constructs in which a gene whose phenotypic expression is easy to monitor is attached to a promoter region of interest and then stably introduced into the mouse genome.

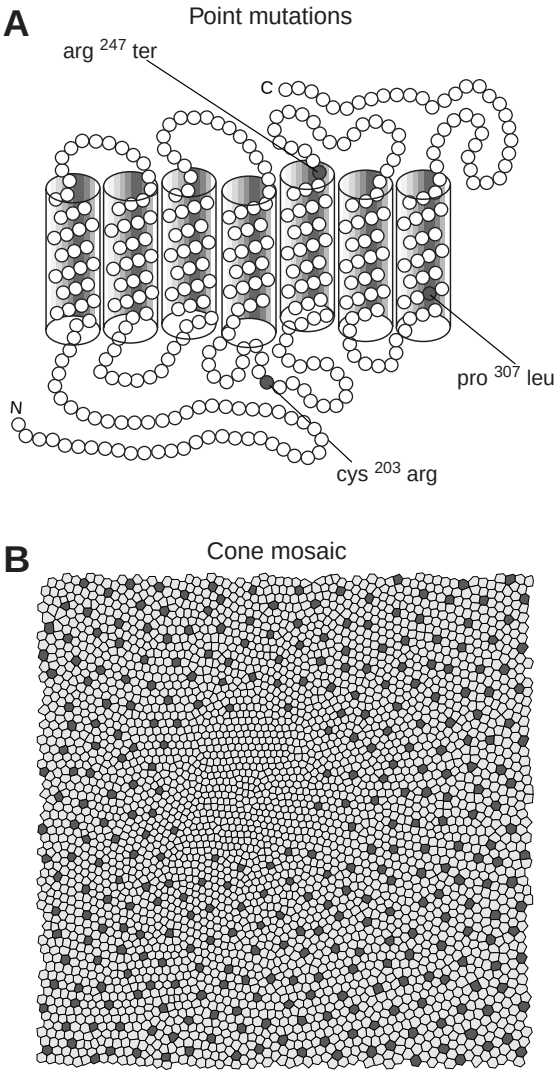

Figure 1.17: Blue-cone monochromacy. (A) Schematic representation of a 5'M-3'L hybrid pigment showing the locations of identified point mutations. The seven α-helices are shown embedded within the membrane. N and C denote amino- and carboxyl-termini, respectively. (B) The cone mosaic of the rod-free inner fovea of an adult human retina at the level of the inner segment; computer altered to simulate the consequences of blue-cone monochromacy. The positions of the S-cones are shown in black; the positions of the M- and L-cones, in white to indicate the lack of a functioning cone pigment. See legend to Fig. 1.1C for other details.

In another variety of the two-step pathway, the gene array is not reduced, but a cys^{203}arg missense mutation in the M-cone pigment gene is transferred to the L-cone pigment gene by gene conversion or intragenic crossing over (Nathans, Maumenee, et al., 1993; Reyniers et al., 1995).

Cone mosaic. Although no histology has been performed on known blue-cone monochromat donor eyes, knowledge about the distribution of S-cones in normal eyes makes it possible to predict the appearance of their cone mosaic (Fig. 1.17B). The drastic effects – the central scotoma, the paucity of pigment-containing cones – arise because the deletion of the LCR or the loss of function of the pigment genes themselves does not permit replacement of the missing L- and M-cone photopigments by another pigment type. Because the gene for the S-cone pigment resides on another chromosome (7), there is no plausible mechanism for placing the S-cone pigment in morphologically intact, but empty L- and M-cones. The same argument applies to the rod pigment (the opsin gene is located on chromosome 3), largely ruling out, on molecular biological grounds, the possibility for the occurrence of rhodopsin-containing cones.

The lack of replacement of the missing M- and L-cone pigments may have immediate consequences for the morphological integrity of the empty cone photoreceptors. The outer segment is relatively fragile, degenerating in response to many genetic perturbations. But there could also be a delayed onset of degeneration. Interestingly, a slowly progressive macular dystrophy involving extensive peripapillary retinal pigment epithelium regression or thinning has been associated with blue-cone monochromacy (Fleischman & O'Donnell, 1981; Nathans, Davenport, et al., 1989).

Cone monochromacy

Besides blue-cone monochromacy, there are other forms of cone monochromacy, known as complete

achromatopsia with normal visual acuity (Pitt, 1944; Weale, 1953). Few cases have ever been described (Pitt, 1944; Weale, 1953, 1959; Fincham, 1953; Crone, 1956; Gibson, 1962; Ikeda & Ripps, 1966; Alpern, 1974; Vajoczki & Pease, 1997; Ross et al., unpublished), and none is fully accepted as authentic.

Phenotypes. Cone monochromats are conventionally assumed to have either L-cones (in which case they are known as L- or red-cone monochromats) or M-cones (in which case they are known as M- or green-cone monochromats), but not both. Although the S-cones are assumed to be totally absent or inactive, they may be partially functioning, contributing to luminance but not color discrimination (Alpern, 1974; Ross et al., unpublished). Evidence of remnant cone function has led to speculation that the defect may be wholly or partially postreceptoral in origin (Weale, 1953, 1959; Fincham, 1953; Gibson, 1962). Weale (1953), for instance, using retinal densitometry in a careful study of three cases, found evidence for normal M- and L-cone pigments. Unlike blue-cone or rod monochromacy, there is no reduced visual acuity, nystagmus, or light aversion.

(i) Color perception: There is no neutral point, and the entire spectrum appears colorless. Although cone monochromats fail to discriminate colors reliably under conventional test situations, they may be able to teach themselves to identify and to discriminate large and saturated colors, slowly and imperfectly, by secondary signs of color, such as brightness and chromatic aberrations that alter the size, sharpness, and apparent depth of objects according to their hue (Ross et al., unpublished).

Diagnosis. The disorder is revealed by the accumulation of results on conventional color vision tests. Cone monochromats make the same color confusions as X-chromosome–linked dichromats in the red-green range.

The Rayleigh matches of cone monochromats are those of a protanope (M-cone monochromat) or a deuteranope (L-cone monochromat). The Moreland equation or adaptation tests (see page 46) can be used, guardedly, to reveal the amount of S-cone function present.

Incidence. The incidence is extremely rare, estimated at 1:1,000,000 (Weale, 1953) or 1:100,000,000 (Pitt, 1944). The latter estimate is based on assumptions about double dichromacy; that is, tritanopia (see below) combined with either deuteranopia or protanopia (see Table 1.6). Revised estimates of the frequency of tritanopia suggest an upper estimate of ca. 1:10,000,000. However, if cone monochromacy is due, in part or whole, to a post receptoral defect, these estimates are inapplicable.

Molecular genetics. Molecular genetic analysis has only been conducted on a single functional M-cone monochromat (Ross et al., unpublished; Nathans, unpublished). Southern blotting/hybridization after gel electrophoresis revealed that the L-cone pigment gene was missing and replaced by a 5'L-3'M hybrid gene, presumably producing an M-like pigment. None of the known mutations in the S-cone pigment gene (see below) were found. The disorder in the S-cone system may be postreceptoral or associated with sequences adjacent to the S-cone pigment gene that direct its expression.

Cone mosaic. The appearance of the cone mosaic in cone monochromats is predicted to be similar to that of a protanope or deuteranope, with possible replacement of the missing L- or M-cone pigment with the remaining one. If the S-cone pigment is rendered partially or totally inviable, no replacement is deemed likely; so that the S-cones will be missing or present but nonfunctioning (see tritan defects below).

Tritan defects

Tritan defects affect the S-cones. They are often referred to as yellow-blue disorders, but the term blue-green disorder is more accurate (see color perception, page 46). As for protan and deutan defects, congenital

tritanopia arises from alterations in the gene encoding the opsin; but, unlike protan and deutan defects, it is autosomal (linked to chromosome 7) in nature.

Phenotypes. Tritan defects affect the ability to discriminate colors in the short- and middle-wave regions of the spectrum. They often go undetected because of their incomplete manifestation (incomplete tritanopia) and because of the nature of the color vision loss involved. From a practical standpoint, even complete tritanopes are not as disadvantaged as many protanomalous and deuteranomalous trichromats because they can distinguish between the environmentally and culturally important red, yellow, and green colors.

On the other hand, the most frequently acquired color vision defects, whether due to aging or to choroidal, pigment epithelial, retinal, or neural disorders, are the Type III acquired blue-yellow defects (see Birch et al., 1979). These are similar, but not identical, to tritan defects. Unlike tritan defects (which are assumed to be stationary), acquired defects are usually progressive and have other related signs, such as associated visual acuity deficits. One acquired disorder – autosomal dominant optic atrophy (DOA) – is commonly described as being predominately tritan in nature (Kjer, 1959) and capable of mimicking congenital tritanopia (Krill et al., 1971; Smith, Cole, & Isaacs, 1973; Went et al., 1974; Miyake et al., 1985). However, it has been linked to chromosome 3q28-qter in many pedigrees (e.g., Eiberg et al., 1994; Bonneau et al., 1995; Lunkes et al., 1995; Jonasdottir et al., 1997; Votruba et al., 1997) and not to the S-cone opsin gene.

(i) Complete and incomplete tritanopia: Tritanopia is the loss form of tritan defects. Like many autosomal dominant disorders, it is complicated by frequent incomplete manifestation (Kalmus, 1955; Henry et al., 1964). The amount of loss of S-cone function ranges from total to minor, but the majority of tritan observers seem to retain some, albeit reduced, S-cone function (Pokorny, Smith, & Went, 1981). Variable penetrance, the degree or frequency with which a gene manifests its effect, has been reported even within families (Sperling, 1960; Cole et al., 1965; Went et al., 1974; Alpern, 1976). However, separate pedigrees have also been documented with complete (Went & Pronk, 1985) and incomplete (Kalmus, 1955; Neuhann et al., 1976) penetrance. Occasionally, tritan defects are observed in combination with deutan defects (van de Merendonk & Went, 1980; van Norren & Went, 1981).

Complete penetrance might only be found in homozygotes and, therefore, would be extremely rare. But it seems more likely that penetrance depends on the location of the mutation or on the influence of other modifier genes.

(ii) Tritanomaly: The classical alteration or anomalous trichromatic form of tritan defects is known as tritanomaly. True cases of tritanomaly, as distinct from partial or incomplete tritanopia, have never been satisfactorily documented. Although the separate existence of tritanopia and tritanomaly, with different modes of inheritance, has been postulated (Engelking, 1925; Oloff, 1935; Kalmus, 1965), it now seems more likely that tritanomaly does not exist, but rather has been mistaken for incomplete tritanopia (Cole et al., 1965) or for acquired disorders such as DOA.

(iii) Tetartan defects: Occasionally, references are found in the literature to a hypothesized second subtype of blue-green disorder, known as tetartan defects (from Greek *tetartos*: fouth). Tetartanopia is the loss and tetartanomaly, the altered form. These are conjectures of the zone (opponent process) theory of Müller (1924), in which tritanopia is considered an outer or retinal (presumably receptoral) defect and tetartanopia, an inner or neural (presumably postreceptoral) defect. No congenital tetartanope or tetartanomalous observer has ever been convincingly demonstrated. However, patients with acquired Type III blue-yellow disorders sometimes exhibit tetartan-like confusions on conventional hue arrangment tests (see Birch et al., 1979). Generally, these disorders involve retinopathies, not disturbances in the S-cone opsin gene.

(iv) Color confusions and neutral point: In the complete tritanope, confused colors should fall along lines radiating from a single tritanopic copunctal point, corresponding to the chromaticity of the S-cones. Calculated tritanopic confusion lines, based on the tritanopic hue discrimination data of Fischer, Bouman, and ten Doesschate (1951) and Wright (1952) are shown in

Fig. 1.12C. They indicate that about 44 distinct spectral wavelengths can be discriminated.

The spectrum is divided by a neutral zone (see Fig. 1.10), which occurs near yellow, ca. 569 nm (6500 K; Walls, 1964; Cole et al., 1965). The violet end of the spectrum may also appear colorless (a complementary[12] neutral zone is predicted to occur in this region, although careful searching often does not reveal it (Cole et al., 1965). But the luminosity is essentially normal, without marked loss of sensitivity at the short wavelengths.

Observations in individuals with unilaterally acquired tritan disorders suggest that, below the neutral zone, a blue-green at 485 nm, and, above the neutral zone, a red at 660 nm have the same hue for the normal and color deficient eyes (Fig. 1.13). The best documented study is Alpern, Kitahara, and Krantz (1983); see also Graham and Hsia (1967).

(v) Color perception: The resulting loss of hue discrimination is in the violet, blue, and blue-green portions of the spectrum (see Fig 1.14D). "Yellow" and "blue" do not occur in the color world of the complete tritanope. They confuse blue with green and yellow with violet and light gray, but not yellow with blue, as the descriptive classification "yellow-blue" defect would seem to imply.

(vi) Large-field trichromacy: Pokorny, Smith, and Went (1981) found that the majority of tritan subjects are capable of establishing trichromatic matches for centrally fixated fields of large diameter (8 deg). The explanation is that S-cones are usually present, albeit in reduced numbers. But the interpretation of trichromatic matching in tritan subjects is complicated by several factors.

Diagnosis. The diagnosis of tritan defects is difficult and the condition often eludes detection.

(i) Plate and arrangement tests: Special pseudoisochromatic plate tests have been designed, including the AO (HRR), Tokyo Medical College, Farnsworth or F2 (reproduced as Plate I in Kalmus, 1965; also see Taylor, 1975), Velhagen, Standard (2nd edition), Stilling, and the Lanthony tritan album. Additional information can be gained from examining chromatic discrimination on pigment arrangement tests, such as the Farnsworth–Munsell 100-Hue test or the Farnsworth Panel D-15 (both have a tritan axis). However, none of these tests, alone or in combination, identify a tritan disturbance unequivocally.

(ii) Blue-green equation: Equations similar in concept to the Rayleigh red-green equation, including the Engelking–Trendelenburg equation (Engelking, 1925; Trendelenburg, 1941) and the Moreland blue-green equation (Moreland & Kerr, 1979; Moreland, 1984; Moreland & Roth, 1987), have been designed to detect tritan defects. The Moreland equation, which is now the most frequently employed, involves matching indigo (436 nm) and green (490 nm) mixture primaries to a cyan standard (480 nm plus 580 nm mixed in a fixed ratio). However, many problems limit the diagnostic ability of this test. A central problem is that the equation applies in a trichromatic region of the spectrum, unlike the Rayleigh equation, which applies to a dichromatic region. Further, the primary wavelengths do not fall on a tritanopic confusion axis (see Fig. 1.12C). Thus, the Moreland equation becomes dichromatic in cases of tritanopia, unlike the Rayleigh equation, which becomes monochromatic in cases of protanopia and deuteranopia. Further, the equation is also affected by individual variations in lens and, to a lesser extent, macular pigment density. Thus, it is not a sufficient test for complete or incomplete tritanopia. Even for normal observers, there is a wide distribution of midmatch points and, frequently, an ambiguous diagnosis.

(iii) Adaptation tests: Two adaptation tests have been designed to isolate any residual S-cone responses: the TNO Tritan test (van Norren & Went, 1981) and the Berkeley Color threshold test (Adams et al., 1987). These tests present a flickering blue target, chosen to favor S-cone detection, on a high-luminance, steady, yellow-orange background chosen to repress the M- and L-cones. Their results, however, are also often indecisive.

(iv) *ERG techniques*: ERG techniques, using chro-

[12]Wavelengths that add together to produce white are complementary.

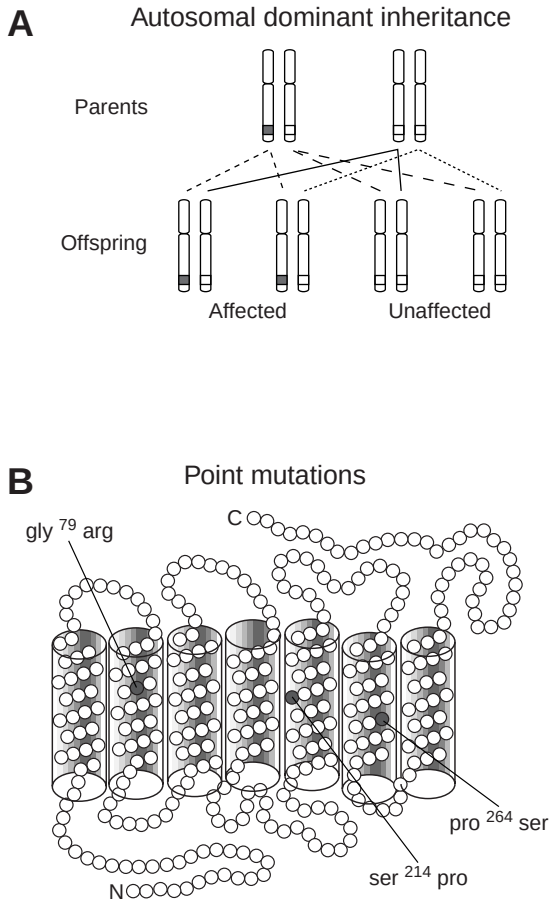

Figure 1.18: Tritan defects. (A) Autosomal dominant inheritance. On average, one-half of the offspring of a heterozygotic parent will inherit the disorder. An open square on the q-arm of chromosome 7 indicates a normal S-cone pigment gene; a filled square, one carrying a missense mutation. (B) Schematic representation of the S-cone pigment showing the locations of three identified amino-acid substitutions.

matic adaptation or silent substitution procedures, have been applied to determining the amount of remaining S-cone function in tritanopes (e.g., Arden et al., 1999). However, isolation of the electrically weak S-cone signal is difficult.

Incidences. Unlike the disorders of the M- and L-cone pigment genes, disorders in the S-cone pigment gene are inherited as autosomal dominant traits (see Fig. 1.18A). Thus their frequencies should be equivalent in males and females, but actual incidences have never been satisfactorily established.

In the United Kingdom the frequency of inherited tritan defects has been estimated as being as low as 1:13,000 to 1:65,000 (Wright, 1952; Kalmus, 1955, 1965), but in The Netherlands it has been estimated as being as high as 1:1,000, (van Heel et al., 1980; van Norren & Went, 1981). The British values pertain only to identified tritanopes and rely on involved calculations, whereas the Dutch samples are too small (480 and 1,023, respectively) and the selection procedures may have been biased.

Molecular genetics. (*i) Missense mutations:* Studies of tritan disorders have established that it is associated with at least three different amino-acid subsitutions that cause missense mutations in the gene encoding the S-cone opsin (see Fig. 1.18B). One substitution, involving a G to A substitution at nucleotide 644 in exon 1, leads to the replacement of glycine by arginine at codon 79 ($gly^{79}arg$; Weitz et al., 1992); another, involving a C to T substitution at nucleotide 1,049 of exon 3, leads to the replacement of serine by proline at codon 214 ($ser^{214}pro$; Weitz et al., 1992); and a third, involving a T to C substitution at nucleotide 1,199 in exon 4, leads to the replacement of proline by serine at codon 264 ($pro^{264}ser$; Weitz, Went, & Nathans, 1992). All three substitutions are in the transmembrane domain of the pigment and are believed to give rise to mutant proteins that perturb the structure or stability of the S-cone pigment and thereby actively interfere with the function or viability of the S-cone photoreceptors.

(ii) Incomplete penetrance: The variable penetrance of the disorder appears to depend, at least in part, on the type of point mutation (Weitz et al., 1992, Weitz, Went, & Nathans, 1992). However, a definitive phenotype–genotype study to examine this question has yet to be performed.

(iii) *Polymorphism and tritanomaly:* The mechanism that permits the frequent manifestation of anomaly in protan and deutan defects – intragenic crossing over – has no analogy in tritan defects: The S-cone opsin gene resides alone on chromosome 7 and has no

neighbor with similar DNA sequences.

Tritanomaly, however, could arise, in principle, from inherited point mutations or polymorphisms that cause shifts in the λ_{max} of the photopigment, but none have been reported so far (see S-cone spectral sensitivity above) and not much variation has been seen in the S-cone pigment gene sequences of the many observers who have been examined. Results from molecular analysis of the S-cone opsin genes in a population of single-gene dichromats and normal subjects revealed only a single common polymorphism that was silent (involving an A to C substitution at nucleotide 775 in exon 2 that leads to the replacement of glycine by alanine at codon 122). No other substitutions were found in the coding sequences and exon–intron junctions (Crognale et al., 1999). In vivo spectral sensitivity (Stockman, Sharpe, & Fach, 1999) and in vitro MSP measurements of the λ_{max} are very limited (see Chapter 2). Moreover, any small variability in λ_{max} would be difficult to disambiguate from individual differences in macular and lens pigmentation in *in vivo* measurements (Stockman et al., 1999) and from the influence of photoproducts produced by photopigment bleaching in *in vitro* experiments. However, some indirect evidence for variation in the λ_{max} of the S-cone pigment, involving factor analysis to disentangle the macular and lens pigment factors, has been reported (Webster & MacLeod, 1988).

Cone mosaic. Expression of the S-cone opsin gene is controlled by adjacent DNA sequences that are located on the same chromosome as the gene itself and that are, therefore, distinct from the sites controlling the expression of the M- and L-cone pigment genes. No LCR seems to be involved. Indirect evidence for the lack of an LCR derives from the finding that, in transgene mice, the ability of a small promoter-proximal sequence to direct expression of the S-cone pigment transgene (Chiu & Nathans, 1994) markedly contrasts with the inactivity of the human L-cone pigment promoter in the absence of an LCR (Wang et al., 1992).

Thus, there is no known mechanism that would allow an M- or L-cone pigment to replace a missing S-cone pigment in an otherwise intact cone structure, if expression of the S-cone pigment gene is invalidated or if the S-cone photoreceptors receive a nonviable gene product. Consequently, empty, nonfunctional, or missing S-cones in tritan defects would be predicted. No histology is presently available.

The lack of replacement should not affect visual resolution, because S-cones are banished from the very central fovea and are too few and widely separated in the periphery to be the limiting factor in visual acuity (see Fig. 1.17B).

Psychophysical and ERG evidence (see Arden et al., 1999) suggests that in some tritanopes the abnormality may lie, in part, in the irregular topographic distribution of the S-cones.

Rod monochromacy

This rare congenital disorder (see Fig. 1.19A) is also referred to as typical, complete achromatopsia, complete achromatopsia with reduced visual acuity, total colorblindness (OMIM 216900), or day blindness (hemeralopia). Its autosomal recessive inheritance distinguishes it from the disorders caused by mutations or structural alterations of the cone opsin genes. However, its manifestation is functionally equivalent to the total loss of all three cone opsin genes. Even if the opsin genes are expressed, they are, nevertheless, never engaged for vision.

Public awareness about the disorder has recently increased, in large part due to the publication of *The Island of the Colorblind* (Sacks, 1997), a description of the Micronesian island of Pingelap and its inhabitants (Brody et al., 1970; Carr et al., 1970) and to the establishment of an Achromatopsia Network on the World Wide Web (http://www.achromat.org/) that is operated for and by achromats (Network facilitator: Frances Futterman, Berkeley, California).

Phenotypes. Rod monochromacy is characterized by photophobia, severely reduced visual acuity (about 20/200), hyperopia, pendular nystagmus, a central scotoma that is associated with the central rod-free

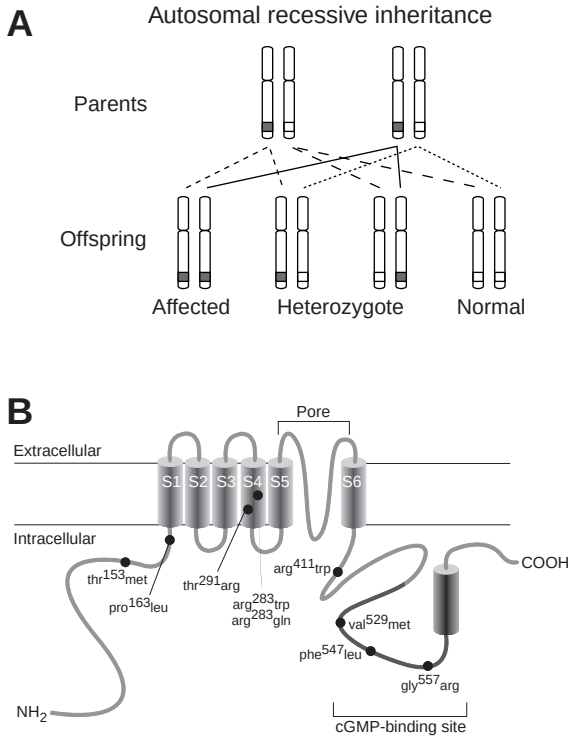

Figure 1.19: Rod monochromacy. (A) Autosomal recessive inheritance for rod monochromacy. On average, one-quarter of the offspring will be affected homozygotes, one-half will be heterozygote carriers, and one-quarter will be normal homozygotes. The light bands indicate a normal gene, the dark bands a gene carrying a mutation for the defect. (B) Putative topology of the human cone cGMP-gated channel α-subunit. The locations of 8 amino-acid substitutions responsible for mutations are indicated by the black dots. The membrane-spanning domains S5 and S6 are thought to line the ion-conducting pore; the cGMP-binding site is located at the intracellular C-terminus (Kohl et al., 1998).

foveola (which is often difficult to demonstrate because of the nystagmus), and the complete inability to discriminate between colors (see Waardenburg, 1963b; Sharpe & Nordby, 1990a).

As first proposed by Galezowski (1868), these symptoms can be explained if the cones are deviant, incompletely developed or completely absent, and the visual functions take place wholly in the rods. This interpretation has since been confirmed by extensive psychophysical measurements and electroretinographic recordings (see Sharpe & Nordby, 1990b, for a review). For instance, there is an absence of the Kohlrausch kink in the dark-adaptation curve. The luminosity function peaks at 507 nm (the λ_{max} of the rod or scotopic visual system), the same as for the dark-adapted normal observer (see Fig. 1.13), instead of 555 nm (the λ_{max} of the cone or photopic visual system). And, in the single-flash ERG, there is a pronounced absence or diminuation of the photopic response.

The Pingelap variant (also known as Pingelapese blindness, total colorblindness with myopia, or achromatopsia with myopia) has been thought to be genetically distinct, mainly because of informal reports of the consistent occurrence of severe myopia (Maumenee, unpublished observations). However, these have not always been confirmed (Wassermann, unpublished observations). In fact, although hyperopia is typically reported in cases of rod monochromacy, myopia is also not infrequent. Moreover, the genetic locus associated with the Pingelapese (Maumenee et al., 1998; Winick et al., 1999) has also been associated with families of European origin (Milunsky et al., 1998; see molecular genetics below).

(i) Incomplete rod monochromacy: An associated disorder is atypical, incomplete achromatopsia (a.k.a. incomplete achromatopsia with reduced visual acuity), in which one or more cone type may be partially spared and functioning with the rods. The symptoms are similar but less severe (see Waardenburg, 1963b; Sharpe & Nordby, 1990a).

(ii) Heterozygotic manifestation: Deviation of color vision, both in the red-green and yellow-blue matching ranges, has been reported in the relatives of rod monochromats (Pickford, 1957; Nordström & Polland, 1980). This has led to speculation that the defect has a tendency toward heterozygotic manifestation (Pickford, 1957; Waardenburg, 1963b; Nordström & Polland, 1980). However, other signs, such as diminuition of the photopic response in the flash ERG, have not been reported.

(iii) Color perception: There is no neutral point and no specific confusion colors (see Fig. 1.13). All colors

are monochromatic, such as the normal observer perceives at night. But many rod monochromats learn to associate certain colors with objects and to recognize some colors by discerning differences in brightness. For instance, deep reds appear dark and blue-greens bright, owing to the blue-shifted luminosity function (for a personal account, see Nordby, 1990).

Diagnosis. Diagnosis of rod monochromacy is accomplished through the accumulation of abnormal results on a battery of color vision, ERG, and visual acuity tests. No specific axis of color confusion is found on the Farnsworth–Munsell 100-Hue test, but an achromat axis is characteristic on the Panel D-15 test.

A color match can be made over the entire range of Rayleigh matches, but a brightness match is only possible at the red end of most anomaloscopes. The matches, mediated by the rods, resemble those made by blue-cone monochromats (Fig. 1.11).

Incidence. The prevalence of rod monochromacy for Caucasians is often given as 1:33,300 for men and 1:50,000 for women (cf., Judd, 1943), based on Köllner's summary of studies by Göthlin (1924) and Peter (1926); however, estimates vary widely between 1:1,000 and 1:100,000. There is no reason to expect a different prevalence for men and women. A recent, but preliminary, survey in Norway yielded an incidence of 1:50,000 for both sexes (see Sharpe & Nordby, 1990a).

Parental consanguinity is a frequent factor in the disorder. Waardenburg (1963b) estimated a minimum of 28.2% parental consanguinity in Caucasians. The factor is much higher in Japan, owing to the high incidence of consanguineous marriages (see Waardenburg, 1963b, for a review).

On the island of Pingelap, the incidence varies between 4 and 10%. The high incidence is attributed to a founder effect (genetic drift observed in a population founded by a small nonrepresentative sample of a larger population) as the result of a typhoon that completely inundated the island in around 1775. Subsequent starvation, together with isolation, reduced the population to about 20, including the hereditary chief. Pedigrees establish that he was a carrier for the disorder (Brody et al., 1970; Carr et al., 1970).

Molecular genetics. The disorder is heterogeneous with loci assigned to chromosomes 2q11 (Arbour et al., 1997; Wissinger et al., 1998), 8 (Maumenee et al., 1998; Milunsky et al., 1998; Winick et al., 1998), and 14 (Pentao et al., 1992). Phenotype variability may depend on genotype.

Mutations are implied in genes that are common to all three cone types. In those individuals with the locus assigned to chromosome 2, the gene has been identified. The disorder is caused by missense mutations (eight have been identified so far) in the gene (*CNGA3*) encoding the α-subunit of the cone photoreceptor cGMP-gated cation channel (Wissinger et al., 1998; Kohl et al., 1998) – a key component of the phototransduction pathways (Fig. 1.19B; see Chapter 3). The mutant proteins may be either improperly folded within the plasma membrane or inherently unstable or unable to be transported and integrated into the plasma membrane. A lack of membrane targeting has also been demonstrated for mutations in the homologous gene in rod photoreceptors that cause autosomal recessive retinitis pigmentosa (Dryja et al., 1995). An inability to bind cGMP may result in a permanent closure of the channel and elimination of the dark current, a situation comparable to a continuous photoreceptor stimulation. The inappropriate functioning of the phototransduction cascade and the continuous activation of photoreceptors has been speculated to be involved in other stationary retinopathies (Dryja et al., 1993). This finding – the first demonstration of a color vision disorder caused by defects other than mutations in the cone opsin genes – implies a common genetic basis of phototransduction in the three cone types (Kohl et al., 1998).

The genetic locus of the disturbance for the Pingelapese has now been identified on chromosome 8q21-q22 (Winick et al., 1999). It is not unique to the Pingelap; for it has also been reported in families of European origin (Milunsky et al., 1998).

The assignment to chromosome 14 relies on a single case of a very rare disorder, uniparental isodisomy, in which the patient inherited two maternal isochromes

14 (Pentao et al., 1992). In light of other data (Arbour et al., 1997; Wissinger et al., 1998), it seems likely that this association occurred by chance or that only a minor fraction of cases with rod monochromacy are caused by a genetic defect on chromosome 14.

Cone mosaic. Unlike for the other forms of color blindness, some histology has been performed on rod monochromat eyes. However, the results are very discrepant, presumably reflecting the heterogeneity of the disorder (see Sharpe & Nordby, 1990a, for a review). In one histological case study, a 29-year-old woman, cones were scarce and malformed in the fovea but were normally distributed and shaped in the periphery (Larsen, 1921); in another, a 19-year-old male, they were imperfectly shaped and markedly reduced in numbers throughout the entire retina (Harrison et al., 1960); in yet another, a 69-year-old woman, they were normally distributed, although abnormally shaped, in the foveal region and scarce, although less often malformed, in the periphery (Falls et al., 1965); and in a fourth, an 85-year-old man (the best documented clinically and psychophysically), they were completely absent in the fovea, abnormally shaped near the fovea, and severely reduced in number (5–10% of normal values) throughout the entire retina (Glickstein & Heath, 1975).

The lack of unanimity in these anatomical findings may reflect different pathologies: Some of the patients may have suffered from stationary, congenital color blindness, while others may have suffered from an early onset, progressive cone degeneration disease. However, it is conceivable that they may also reflect the nature of the different genetic mutations involved. In some disorders, the mutation may be such that the cones remain intact, whereas in others they may degenerate or never develop.

Interestingly, one of the subjects known to have the missense mutation in the *CNGA3* gene, fixated normally and kept his eyes fully open, even in bright light during the first seven to eight months of life. Only thereafter did he exhibit squinting, nystagmus, and light aversion behavior (Nordby, 1990). In contrast, other rod monochromats are known to already exhibit light avoidance symptoms during the first months of life (see Sharpe & Nordby, 1990a).

Conclusions

The discovery by recombinant DNA techniques of the opsin genes responsible for encoding the cone photopigments (Nathans, Thomas, & Hogness, 1986; Nathans et al., 1986) has significantly advanced our understanding of the initiating stages of color vision. Not only has it revealed how common variants in color vision are produced by alterations in the opsin genes, it also informs speculations about the concatenated development of the cone retinal mosaic and associated neuronal network. The logical consequence of changes at the lowest levels of visual processing are changes at the highest, because whatever alters the function of the cone photopigments and their distribution in the cone mosaic will influence the organization of the retinal subsystems and their representation in the cortex. Genetic variation in the opsin genes implies separate perceptual worlds that differ in more than color.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (Bonn) under grants SFB 340 Tp A6 and Sh 23/5-1 and a Hermann-und-Lilly-Schilling-Professur award (LTS), by National Institutes of Health grant EY 10206 (AS), and by the Howard Hughes Medical Institute (JN). We thank Samir Deeb, Frances Futtermann, Karl Gegenfurtner, Jan Kremers, Anne Kurtenbach, Donald MacLeod, Ethan Montag, Hans-Jürgen Schmidt and Bernd Wissinger for comments and Sabine Apitz and Christine Jägle for help and encouragement.