1. '...aus dreyerley Arten von Membranen oder Molekülen': George Palmer's legacy

J.D. MOLLON
(Cambridge, UK)

Abstract

It is 200 years since George Palmer suggested that colour vision depended on three classes of molecule or membrane. The present paper gives an introductory review of what is today known of the opsins, of the genes that encode them, and of the alterations to which they are subject. Several theories of anomalous trichromacy are briefly reviewed.

Introduction

We do not know the day on which George Palmer died, but he signed his last will on July 17 1795 and he was buried at Hendon parish church in the Middlesex countryside on August 8 1795[1]. So we can be sure that the 13th Symposium of the IRGCVD fell within a few days of the bicentennial of his death. And it is particularly appropriate that the Research Group should remember this singular man. He was born in Westminster in 1740. By trade he was a prominent dealer in glass, especially coloured glass, and by patrimony he was a member of the Drapers' Company (Mollon, 1993). In 1777 Palmer proposed that the retina contained three types of fibre, corresponding to three physical types of light. His physiological hypothesis, published 25 years before Thomas Young's Bakerian Lecture, is a recognizable version of the modern theory of trichromacy, although his physical theory – of three types of light – is mistaken. Perhaps the most remarkable version of his theory appears in a German popular science magazine of 1781, where the opinions are attributed to a certain 'Giros von Gentilly' and it is suggested that the retina is composed of three kinds of molecule or membrane (Voigt, 1781; Fig. 1). That article contains what is almost certainly the first suggestion that colour blindness arises from an inherited defect of receptor molecules (indeed this is probably the first suggestion that any hereditary defect is due to a molecular error). Palmer

[1]Public Record Office, London, Prob 11/1264, ff 372RH-374LH; Burial register, Parish of St Mary's, Hendon.

C.R. Cavonius (ed.), Colour Vision Deficiencies XIII, pp. 3–20.

gen könnten, und glaubte ſich berechtigt, als einen Grundſatz anzunehmen, daß die Netzhaut aus dreyerley Arten von Membranen oder Molecülen zuſammengeſetzt ſey, davon jede einem von den dreyen urſprünglichen Lichtſtralen entſprechend wäre, und welche die Eigenſchaft hätte, blos von dieſem Einen Bewegungen annehmen zu können, ſo wie etwa die Trommelfelle im Ohr gleichfalls aus Faſern gewebt wären, deren jede einen ihr beſonders entſprechenden Ton verlangte, wenn ſie ihn empfinden ſollte.

Dieſes vorausgeſetzt, laſſen ſich ohne Mühe alle die Unvollkommenheiten des Geſichts durch zwey Urſachen erklären, die einander ſchnurſtracks entgegenſtehen.

Die erſtere iſt die Unwirkſamkeit der Molecülen, und die zweyte deren außerordentliche Beweglichkeit.

Fig. 1. A passage from J. H. Voigt (1781) in which he describes the theory of 'Giros von Gentilly' (*alias* George Palmer, 1740–1795). This passage contains what may be the first explanation of an inherited defect in terms of molecular error.

proposed the existence of two alternative faults that can affect each of the three types of molecule: the first is that it can be inactive, the second is that it can be overactive, constitutively active to use the modern term. If only one of the three types of molecule is affected, then the defect is mild, but if two are affected then the condition is manifestly more severe. In a later monograph (Palmer, 1786), he further suggested that adaptation to saturated colours can transiently produce colour blindness in the normal eye, owing to selective fatigue of the receptors.

It was almost a century before Palmer's idea of light-sensitive molecules was systematically revived by Wilibald Kühne. The subsequent century has seen a slowly growing understanding of the photopigments, an understanding that accelerated with the analysis of the amino acid sequence of rhodopsin (Hargrave *et al.*, 1983) and the sequencing of the genes that code for both the rod and the cone pigments (Nathans and Hogness, 1984; Nathans *et al.*, 1986a, 1986b).

Palmer's suggestion that visual molecules may exhibit constitutive activity was confirmed in 1992, when it was shown that a particularly severe form of dominant retinitis pigmentosa arises from a mutation of rhodopsin that causes the molecule to signal continuously (Robinson *et al.*, 1992). The mutation (Lys-296-Glu) affects the site at which the chromophore normally binds to the protein.

Photopigments and their genes

The structure of opsins

To the left in Figure 2[2] is depicted the outer segment of a cone, with its lipid membranes multiply infolded and packed with molecules of photopigment. Photopigments consist of proteins, typically 348–364 amino acids long, bound to 11-*cis*-retinal, the aldehyde of vitamin A_1. The 11-*cis*-retinal acts as the 'chromophore': it gives the photopigment its colour, its selective spectral absorbance. The protein components of the pigments have come to be called **opsins**. Each consists of seven helices, which span the cell membrane and are linked by loops within and without the membrane (Hargrave and McDowell, 1992). The seven helices form a palisade that surrounds the 11-*cis*-retinal; the latter is bound to a lysine in the seventh helix (Fig. 2). We now know that the palisade is splayed, so that helices are closer together at the cytoplasmic surface, the side within the cell (Baldwin, 1993).

Figure 3 offers a plan view of a photopigment molecule, sketched on the basis of the analysis of Baldwin (1993). To visualize the three-dimensional arrangement of the helices, imagine taking hold of the left-most helix (1) in the cartoon of the molecule at the top, pulling it forwards from the plane of the paper, and bending it round to touch helix 7. If you then viewed the resulting structure from above (the intracellular surface), you would have the plan view shown in the lower part of the figure. The plan view indicates the approximate location of the chromophore, 11-*cis*-retinal, and the probable positions of some of the most important intramembrane amino acids that determine the difference in spectral sensitivity between long- and short-wave pigments.

---

[2]Figures 2, 3 and 4 will be found between pages 20 and 21 of the book.

The opsins are members of the superfamily of G-protein coupled receptors or **heptahelicals**. This family has more than 200 members and includes many of the receptors that mediate communication between our brain cells, and between hormones and neurons. The several forms of the serotonergic, dopaminergic, adrenergic and muscarinic acetylcholine receptor molecules, as well as the large subfamily of olfactory receptors, are all heptahelicals, distant cousins of the opsins (Watson and Arkinstall, 1994). In every case the heptahelical is embedded in the cell membrane, and initiates its signal within the cell by activating a G-protein. In the case of opsins, the G proteins are called **transducins** (Lerea *et al.*, 1989). But whereas photopigments are semi-permanently bound to 11-*cis*-retinal and separate from it after a photon has isomerized the chromophore to its all-*trans* form, other heptahelicals wait with an empty binding pocket and activate their G protein when they bind to a molecule of the right neurotransmitter or hormone or odorant.

The opsin genes

The genes that code for rhodopsin and for the short-wave pigment lie on chromosomes 3 and 7 respectively (Nathans *et al.*, 1986a; Fitzgibbon *et al.*, 1994). Several cases of congenital tritanopia have been found to be associated with point mutations of the gene that codes for the short-wave opsin (Weitz *et al.*, 1992a, 1992b). Thus tritanopia can be produced by a mutation that leads to the substitution of arginine for glycine at position 79 in the amino acid sequence of the opsin, or by mutations that give proline for serine at position 214 or serine for proline at position 264. In the case of the glycine-79-arginine substitution, all homozygotes, but only a minority of heterozygotes, appear to be affected, whereas the two other substitutions show high penetrance in the heterozygous condition (Nathans *et al.*, 1992).

As was predicted from the classical evidence for X-linkage of red–green colour deficiencies (Earle, 1845; Brunner, 1930; Kalmus, 1965), the genes for the long- (L) and middle-wave (M) cone pigments lie on the X-chromosome (Nathans *et al.*, 1986a). The following discussion will concentrate on these genes.

Figure 4 (top) illustrates how the L and M genes lie relatively close together on the q arm of the X-chromosome, although they are further apart than is suggested by traditional diagrams of this kind: the individual genes are separated by some 24 kilobases of non-coding DNA. Typically the X-chromosome carries more than one M gene, although the total number of L and M genes is controversial; Neitz *et al.* (1995) report subjects with as many as nine opsin genes. The complete cluster of L and M genes on the X-chromosome is often referred to as the **opsin gene array**.

It is thought that only one gene in the opsin array is expressed in any given cell. Lying 4 kilobases upstream of the first gene in the array is a locus control region (LCR), which is thought to control which of the genes in the opsin array is expressed in a given cone cell. A man who lacks this LCR will be an S cone monochromat, that is to say, he will express none of the X-chromosome opsin

genes (Nathans *et al.*, 1989; 1993).

How does the LCR exercise its choice in a given cone? Within the nucleus of a cell, the DNA of a given chromosome is not laid out linearly, and in the present case it is thought that the locus control region bends back to couple with one of the promoter regions that lie just upstream of each gene; this coupling makes the favoured gene available for expression. The idea has been advanced that the closer a gene is physically to the LCR the greater its chances of being expressed (Winderickx *et al.*, 1992a). This concept is important in the interpretation of colour deficiency.

Introns and exons

The structure of individual L and M genes is depicted in the lower part of Figure 4. The exons of the opsin genes, i.e. the regions that actually code for the amino acid sequence of the protein, are interrupted by 5 introns – long, non-coding stretches of DNA. In this feature the opsin genes differ from those for many of the heptahelicals, although the dopamine receptor genes are similarly interrupted by introns.

The intronic structure of the opsin genes must be remarkably stable, for introns 2–5 of the L and M genes are homologous in their positions to the introns of the rhodopsin gene, even though the rod and cone genes diverged long before the emergence of the mammals. Introns 2–5 are of equal length in the L and M genes, but in more than 99% of Caucasians and 98% of Japanese, intron 1 of the L gene is 1.9 kilobases longer than that of the M gene. This latter difference has proved useful to molecular biologists in the task of separating two such similar genes, but a curious finding is that 35% of Afro-Americans exhibit L and M genes that are identical in the length of intron 1 (Lund Jørgensen *et al.*, 1990).

The determination of spectral sensitivity

The M and L opsins are extremely similar, differing only by 15 amino acids (Nathans *et al.*, 1986a). This observation has been taken to suggest that the two pigments diverged relatively recently, in evolutionary terms – perhaps 30 million years ago. But are substitutions required at all 15 sites in order to change the spectral sensitivity from middle-wave to long-wave? Neitz *et al.* (1989) showed that much (at the time they suggested *all*) of the spectral difference depended on nucleotide differences in exons 4 and 5. Their inference was based on a protanope whose electroretinographic spectral sensitivity was similar to that of the normal M pigment and who appeared to have only a single X-chromosome opsin gene. The latter was a 'fusion gene': exons 2 and 3 resembled the L sequence of Nathans and colleagues, whereas exons 4 and 5 resembled the M sequence. Since exons 1 and 6 do not differ between the normal L and M genes, the implication was that exons 4 and 5 primarily controlled the spectral

sensitivity of the resulting opsin.

Further detail has come from two sources. In many species of New World monkey the males carry only a single X-chromosome opsin gene but the gene is polymorphic, so that different conspecifics have different photopigments with peak sensitivities in the range 535–565 nm (Mollon *et al.*, 1984; Jacobs, 1993). By correlating genotypes and phenotypes, it was possible to identify the amino acids that are likely to account for most of the variation in spectral sensitivity (Williams *et al.*, 1992). Secondly, the pigments encoded by alternative genes have been measured *in vitro*. Merbs and Nathans (1992a, 1992b, 1993) constructed hybrid genes from cDNA clones for the normal M and L pigments, while Oprian and colleagues (Oprian *et al.*, 1991; Asenjo *et al.*, 1994) chemically synthesized artificial genes. In each case the genes were expressed in cultured mammalian kidney cells. The gene products – the artifical opsins – were then combined with 11-*cis*-retinal, the cell membranes were solubilized in detergent, and the spectral sensitivities of the resulting pigments were estimated from the difference in the absorption at each wavelength before and after bleaching.

The accumulated evidence suggests that the spectral difference between L and M pigments depends mainly on the seven amino acids indicated in Figure 2. In terms of size of spectral shift, the two most important sites are amino acids 277 and 285, which are determined by nucleotide substitutions in exon 5 of the gene. Particular interest has been excited by site 180, which is coded by exon 3 and is polymorphic in both L and M pigments. The substitution of serine for alanine at this site shifts the peak sensitivity (λ_{max}) by a few nanometres to longer wavelengths (Merbs and Nathans, 1992a); and Winderickx *et al.* (1992b) have shown that subjects whose L pigment carries the serine variant tend to require less red in a Rayleigh match than individuals who have alanine at site 180, although the two groups overlap and both groups lie within the normal range. The difference of a single nucleotide in our genomes may therefore mean that different individuals live in subtly different perceptual worlds.

In addition to sites 180, 277 and 285, the following amino acid sites have been reported to influence spectral sensitivity: 65 and 116 (encoded by exon 2), 230 and 233 (encoded by exon 4) and 309 (encoded by exon 5). Only Merbs and Nathans (1993) found a positive effect for site 65 and only Asenjo *et al.* (1994) found a positive effect for site 116; where a positive shift was found for these sites, it was no more than 1 nm.

Other than site 116, which is in the first extracellular loop of the molecule, all the critical sites lie in the transmembrane helices and the spectral shift requires the substitution of an hydroxyl-bearing amino acid for a non-polar amino acid. Individual substitutions are not additive in their effects on spectral sensitivity: the size of the shift may depend on the identity of the amino acids at other critical sites (Merbs and Nathans, 1993; Asenjo *et al.*, 1994). Owing to the number of critical sites and interactions betwen their effects, there appears to be an almost continuous distribution of possible photopigments in the range 530 to 565 nm – a fact that seems salient to the understanding of anomalous trichromacy.

What of the six or seven sites that differ between L and M cones but which do not affect spectral sensitivity (or affect it only minimally)? It is possible that substitutions which do affect spectral sensitivity require other structural adjustments if the molecule is to retain its optimal three-dimensional configuration (Williams *et al.*, 1992). Consider, for example, a gene with exons 5 and 6 drawn from the normal M gene and the remaining exons from the L gene. The opsin encoded by the gene might have a λ_{max} close to that of the normal M pigment but might be compromised in its function: it might transport less readily to the outer segment; it might be unstable in the membrane; it might be reduced in its quantum efficiency; or it might be impaired in its signalling within the cell. In their *in vitro* expression studies, Nathans and colleagues and Oprian and colleagues noted in passing that some of the expressed pigments were unstable or of reduced optical density, although it is difficult to judge whether these observations, in a very different system, are physiologically significant. It is certainly worth considering the possibility that there are forms of daltonism (or of cone dystrophy) that arise from opsins of impaired viability even though the corresponding gene exhibits no gross mutation and appears to code for a functional pigment.

Are the L and M cones labelled?

The midget bipolars and midget ganglion cells of the retina are thought to be the substrate for the phylogenetically recent subsystem of colour vision which compares the quantum catches of the long- and middle-wave cones (Mollon, 1989). In the foveal region, the centre input of a midget ganglion cell is necessarily cone specific, since this input is drawn from a single cone via a single midget bipolar; whether the surround input is cone specific (as suggested by Reid and Shapley, 1992) or is drawn promiscuously from both L and M cones (as suggested by Lennie *et al.*, 1991) is still debated.

The L and M cones of the catarrhine foveola appear to be randomly arranged (Mollon and Bowmaker, 1992). So if the surround input of the midget ganglion cell is specific to either L or M cones, there would seem to be, in principle, only two ways in which specificity could be achieved: either the L and M cones must carry a label that links them during development to their rightful bipolars, or the juvenile retina must undergo a learning process, perhaps reinforcing synapses that carry correlated inputs, and attenuating others, in a Hebbian way.

The same problem is faced, to a greater degree, by olfactory receptor neurons, which owe their stimulus selectivity to many different receptor molecules, all of which are heptahelicals and thus cousins of the opsins. How do such neurons make specific connections to the olfactory glomeruli? Recent evidence suggests that the heptahelical receptor molecules are present not only at the epithelial surface but also in the axons of the olfactory receptor neurones. Singer *et al.* (1995) have identified amino acids in the second extracellular loop of these heptahelicals that are correlated with (and perhaps act as labels for) amino acids

thought to determine the specificity to a particular odorant; the latter amino acids lie in the transmembrane regions, within the binding pocket for the odorant. Might opsins carry an analogous label? Rhodopsin and the short-wave opsin do differ from each other and from the L and M opsins in the latter part of the second extracellular loop (Nathans *et al.*, 1986a), i.e. in the loop that carries the putative label of the olfactory receptors. However, the L and M opsins do not differ from each other in this loop. In their case, the only possible candidates for a label are site 116, in the first extracellular loop, and site 298, in the third. Of these sites, the latter is the more plausible candidate: the codon for site 116 does not differ for all L and M clones (Nathans *et al.*, 1986a) and it is separated by three long introns from the exon-5 codons that primarily determine spectral sensitivity, whereas the codon for site 298 differs for all clones and, being within exon 5, is very tightly linked to the critical codons. Chimpanzees, like man, differ at site 298, but arguing against this site being a label is the fact that gorillas and Old World monkeys have alanine at 298 in both sequences (Dulai *et al.*, 1994).

Ways in which the opsin genes may be compromised

Major deletions

One family is known where colour blindness is associated with a major deletion in an opsin gene (Reichel *et al.*, 1989). The 15-year-old propositus had near-normal acuity and a protan colour deficiency, while older males in the family showed a macular degeneration, with 20/200 vision and a reduced cone electroretinogram. Molecular analysis suggested a 6.5 kilobase deletion between exons 1 and 4 of the L opsin gene.

Point mutations

Some instances of X-linked colour deficiency are likely to arise from point mutations of the L or M opsin genes. One such mutation is the substitution of arginine for cysteine at position 203 in the amino acid sequence of the M gene. Site 203, encoded by exon 4, is in the second extracellular loop of the opsin molecule and the normal cysteine at the site is thought to form a disulphide bond with another cysteine in the first extracellular loop at position 126. Since these two cysteines are highly conserved in evolution (in opsins and in several other heptahelicals) and are thought to be critical for the three-dimensional structure of the opsin, the mutation presumably produces an ineffective molecule. Winderickx *et al.* (1992c) found the cys-203-arg substitution to be associated with extreme deuteranomaly in a man whose opsin array appeared to consist of one L gene and several M genes, the latter all exhibiting the same mutation. The reason why the subject was not a dichromat is a more general problem: some residual discrimination on the anomaloscope has been reported

for several subjects with only a single gene in their opsin array. Winderickx and his colleagues and Nathans *et al.* (1993) found the cys-203-arg mutation occasionally associated with normal colour vision or with simple deuteranomaly, and in these cases the assumption is that the mutated gene lies in the downstream graveyard of the opsin array, distant from the LCR, and therefore not expressed. If the opsin array consists of a single gene exhibiting the cys-203-arg mutation, the result is S-cone monochromatism (Nathans *et al.*, 1993).

It is unlikely that the cys-203-arg substitution is the only opsin gene mutation circulating in man: there must be others that similarly produce a non-viable molecule and still others that impair the quantum efficiency or the stability of the molecule. Despite the large number of daltonians that have been examined by molecular biological techniques, in only a minority of cases have all exons of all their X-linked opsin genes been sequenced, or at least screened for point mutations. Moreover, mutations in the non-coding region immediately upstream of each gene could sometimes lead to failures of expression. If the gross deletion in the family of Reichel *et al.* (1989) produces a slow cone dystrophy, then there may be other daltonians who exhibit a very mild cone dystrophy – as postulated by Regan *et al.* (1994).

For mutations that produce achromatopsia, we should consider molecules that are common to the different cone types but are distinct from their rod counterparts. Thus a candidate gene for rod monochromatism is that for the alpha unit of cone transducin. This gene has been localized to chromosome 1 (Magovcevic *et al.*, 1995).

Unequal crossing-over

A major source of mischief is thought to be unequal crossing-over in the opsin gene array (Nathans *et al.*, 1986b). Corresponding chromosomes align themselves at meiosis, and it is at this stage that crossing-over occurs, stretches of DNA being exchanged between chromosomes. Owing to the juxtaposition and the homology of the L and M genes, they are quite likely to align themselves improperly at meiosis. If now crossing-over occurs and the breakpoint is between genes, one chromosome may lose a gene and the second may gain it. The simplest explanation of dichromacy is that the opsin array has been reduced to one gene by unequal crossing-over, and many examples do fit this hypothesis. The most celebrated of dichromats, John Dalton, appears to have had only a single L gene (Hunt *et al.*, 1995; Mollon *et al.*, 1996). If the breakpoint occurs within the mispaired genes, then hybrid genes will result. Since introns are longer than exons, the chances are that the breakpoint will be within an intron, and so each of the resulting genes will draw some exons from the L sequence and some from the M sequence. Many subjects, both normal and colour-deficient, appear to carry such genes. Hybrid genes, if expressed, will typically produce an opsin that has a λ_{max} intermediate between those of the normal L and M pigments. Some hybrid opsins may be compromised in their viability.

Although the L and M genes typically differ from each other in the size of

intron 1, an interesting and provocative observation is that introns 2 and 4 are almost identical for the L and M genes and less divergent than the coding sequences in between (Shyue *et al.*, 1994): intron 2, which is 1987 base pairs long, differs only at six nucleotides (at least for clones derived from Jeremy Nathans), whereas intron 4, which is 1552 base pairs long, is identical for two of JN's clones (clone JN44 has a gap at position 32). This similarity of introns is curious, since these non-coding sequences might be thought to be free to diverge in the course of evolution – and indeed rapid divergence would have the positive advantage of minimizing the likelihood of unequal crossing-over (Mollon and Jordan, 1988). Either the homology of the introns has a function or it is an indication of how frequently unequal crossing-over occurs.

Theories of anomalous trichromacy

An agreed understanding of anomalous trichromacy has not come as readily as everyone expected in 1986. Any theory must explain how the colour matches of anomals come to be altered, but also faces the long-recognized problem of phenotypic variation: some anomalous trichromats are almost as impaired in their hue discrimination as are dichromats, whereas many are fair, and some have exquisite discrimination, achieving 'superior normal' scores on the Farnsworth-Munsell 100-hue test (Köllner, 1915; Nelson, 1938; McKeon and Wright, 1940; Pokorny *et al.*, 1979; Regan *et al.*, 1994). Moreover, on the Nagel anomaloscope, anomalous trichromats show little correlation between their mid-match points and their matching range (Willis and Farnsworth, 1952; Jameson *et al.*, 1982).

The following list of theories may be useful. It is intended only as a summary and for more detailed discussion the reader is referred to the papers cited below, to Pokorny *et al.* (1979), and to the many contributions to the present volume and its predecessors.

Partial reduction hypothesis

Most current theories of anomalous trichromacy suppose that at least one of the retinal photopigments has been altered in its spectral sensitivity, and this was certainly the assumption of König and Dieterici (1892). But a long-persisting view held that anomalous trichromacy was a partial reduction of normal colour vision: either the normal L or the normal M cones were present in reduced numbers or had reduced sensitivity. That this will not explain simple anomalous trichromacy has often been pointed out (von Kries, 1899; de Vries 1948a; Jameson and Hurvich, 1956; Rushton, 1972) Suppose there were only one L cone left on each side of a foveal matching field: to make a Rayleigh match, the subject must still choose the red–green ratio that produces the same quantum catch in one L cone as the yellow light produces in the single L cone on the other side.

However, there is one phenotype to which the partial reduction hypothesis might well apply, and that is extreme anomaly. By definition, the extreme anomal has a matching range that includes the normal match as well as either the protanomalous or the deuteranomalous match (Pokorny *et al.*, 1979). From the results of Rayleigh matches alone we cannot, therefore, exclude the interesting possibility that some extreme anomals represent partial reductions of normal vision and that their residual colour discrimination depends upon a population of either L or M cones that is reduced in numbers or is in some other way handicapped. If such phenotypes were found, it would be inappropriate to continue to class them with anomalous trichromats. Extreme anomals are a neglected phenotype and it would be valuable to have more information on their colour matching functions for both large and small fields.

The WDW heresy

Wright (1946, pp 353–4) suggested that the 'red' and 'green' cones might be normal in deuteranomaly, but that the 'green' channel might draw its input from both red and green cones; the spectral sensitivity of the channel is thus altered, being shifted to long wavelengths. As an explanation of simple deuteranomaly, this hypothesis faces the same problem as the partial reduction hypothesis (de Vries, 1948a). When an observer makes a colour match in a bipartite field, we believe that he sets the triplet of quantum catches in the three types of cone on the right to be the same as the triplet on the left. The two sides of the field must then match for him whether or not the subsequent neural wiring is normal or abnormal. To accept Wright's position we should have to abandon the precious dogma that colour matches are determined only by the spectral sensitivities of the cones.

Most (but not all) of the remaining hypotheses below are versions of the 'single-pigment shift model' of anomalous trichromacy (Pokorny and Smith, 1977): such theories suppose that the colour vision of the anomalous trichromat depends on two cone types with normal sensitivity and a third cone type with altered spectral sensitivity.

The Schouten hypothesis

De Vries (1948b) attributes to Schouten the idea that there is a single anomalous pigment, intermediate in spectral position between the normal L and M pigments and common to both protanomaly and deuteranomaly. In support of this idea, de Vries cited a dichromat who accepted the matches of both protanomals and deuteranomals. The implication is that the three phenotypes share two pigments, the short-wave pigment and an 'anomalous pigment', and differ only in whether they retain one or other (or neither) of the normal L and M pigments. In other words, de Vries' dichromat is a reduction form of both protanomaly and deuteranomaly. For further discussion of the Schouten hypothesis, see MacLeod and Hayhoe (1974) and Pokorny *et al.* (1975).

The protanolabe/deutanolabe hypothesis

Rushton and his collaborators favoured the view that there were different anomalous pigments in protanomaly and deuteranomaly. They even gave them names: **protanolabe** and **deutanolabe** (Rushton *et al.*, 1973). This position is taken by DeMarco *et al.* (1992), but it does not marry well with contemporary genetics. It would have made sense in the days when there were held to be separate 'protan' and 'deutan' loci: within that framework it is possible to imagine that only certain alterations are possible to the gene at each locus (Pokorny *et al.*, 1979). If, however, we allow that the X-chromosome opsin genes lie in an undifferentiated array and that all that distinguishes them is the spectral sensitivity of the photopigment for which they code, then there is no reason to expect that in protanomaly a normal M gene is always expressed in conjuction with a certain hybrid gene and in deuteranomaly a normal L gene is invariably expressed in conjunction with a different hybrid.

The continuing influence of the two-loci hypothesis has led to an unfortunate contradiction in terminology. DeMarco *et al.* (1992) denote protanolabe by L′ and deutanolabe by M′, reserving L and M for the normal long-wave and middle-wave pigments. Yet their L′ is closer in λ_{max} to the normal middle-wave pigment than to the normal long-wave pigment, while their M′ is closer to the normal long-wave pigment than to the normal middle-wave pigment. Other recent authors have used L and L′ to denote alternative long-wave pigments (e.g. two opsins that differ only at site 180) and M and M′ to denote alternative middle-wave pigments. Although the opsin array may yet have secrets, and the two-loci hypothesis could arise in a new form, current evidence argues against the usage of DeMarco and colleagues.

The Alpern hypothesis

The late Mathew Alpern and his collaborators proposed that there were several normal variants of erythrolabe, the long-wave pigment, and several normal forms of chlorolabe, the middle-wave pigment (Alpern and Moeller, 1977). In short, there was a 'cluster' of erythrolabes with different peak sensitivities and similarly a 'cluster' of chlorolabes. This proposal led to the interesting hypothesis that there are no distinct 'anomalous' pigments. Rather, the protanomal draws two pigments from the chlorolabe cluster, and the further apart their spectral positions, the better we might expect to be his chromatic discrimination to be, *ceteris paribus*. Similarly, the deuteranope draws two pigments from the erythrolabe cluster. Some early evidence for more than one kind of long-wave pigment came from Dartnall *et al.* (1983) who, by means of microspectrophotometry, measured cones from two colour-normal patients with different Π_5 sensitivity; we know now, of course, that the site 180 polymorphism does generate alternative forms of the long-wave pigment.

The colour matches of carriers of anomaly might be seen as supporting Alpern's view that there are no truly anomalous pigments: the mothers of simple

anomalous boys seldom make Rayleigh matches that lie outside the normal range (Jordan and Mollon, 1993). Owing to the phenomenon of X-chromosome inactivation (Lyon, 1972), a subset of cones in the heterozygous retina should express the abnormal X-chromosome which is inherited by an average of half of their sons. If this X-chromosome carries a gene for a true anomalous pigment (i.e. one that is abnormal in its spectral sensitivity) and if this gene is expressed in the anomalous son, then we might expect it also to be expressed in some cones of the mother's retina, and thus to affect her colour matches.

The reduced Alpern hypothesis

A minimalist version of Alpern's hypothesis was advanced by Neitz and Neitz (1992). They suggested that not only are there no anomalous cones but there are only two forms of the normal L cone and two forms of the normal M cones. Protanomaly depends on a comparison of the two possible types of L cone and deuteranomaly on a comparison of the two types of M cone. This hypothesis derived from the idea that only three sites control the spectral sensitivity of the L and M pigments, sites 180, 277 and 285 in the amino acid sequence (Neitz *et al.*, 1991). An implicit assumption is that the latter two sites are so close together that they never become unyoked by unequal crossing-over: they are always tyrosine and threonine (long-wave) or phenylalanine and alanine (middle-wave). Thus, the only source of variation is site 180, which is polymorphic for both L and M genes. A protanomalous observer is one who lacks an effective L gene (through complete deletion or local mutation) but is fortunate enough to have and to express the genes for the two alternative M photopigments. A deuteranomal lacks an M gene, but is fortunate enough to have and to express the genes for the two different L pigments.

Theories that invoke variations in optical density

A photopigment will be changed in its spectral sensitivity if its optical density is reduced from the relatively high value thought to obtain in the normal cone outer segment (Knowles and Dartnall, 1977). When the pigment is present at high density, there is self-screening at wavelengths close to the λ_{max}: as light passes along the cylinder of the outer segment, the available light reaching a given molecule of pigment will be disproportionately attenuated at wavelengths near the λ_{max} but will be less attenuated at the skirts of the absorbance spectrum. Consequently the shape of the pigment's sensitivity curve is broader at high optical densities. We have considered above the possibility that an altered genetic sequence might produce an opsin of virtually unaltered λ_{max} but altered optical density.

Baker (1966) explored the hypothesis that protanomalous vision depends on a normal M pigment in normal density and a normal L pigment in reduced concentration. His experimental test was to measure Rayleigh matches in normals after a far-red bleach: in fact, his subjects did not generate protanoma-

lous matches even during the earliest stages of recovery, when their erythrolable (L-pigment) should have been dilute. Ruddock and Naghshineh (1974) examined the possibility that normal L and M pigments are mixed in one class of cone of the anomalous retina and that optical density is high enough for self-screening, and screening of one pigment by the other, to modify the spectral sensitivity of the components. He and Shevell (1995; see also this volume) showed by computation that variations in optical density can substantially affect the Rayleigh match midpoint and range if the two underlying pigments of anomalous vision are only a few nanometers apart in nominal λ_{max}. These effects arise chiefly from the change in slope of the long-wave limb when self-screening is changed. He and Shevell pointed out that some chromatic discrimination could be achieved if two pigments with the same nominal λ_{max} but different optical densities were present.

An account based on current data

The work of Merbs and Nathans (1992b; 1993) and of Asenjo *et al.* (1994) shows that there is potentially an almost continuous range of photopigments with λ_{max} between, say, 532 nm and 563 nm. Thus in the series of pigments measured *in vitro* by Asenjo and colleagues the largest gap is 4 nm (between 534 and 538 nm). It is true that experimental error might reduce the size of any gap in the series and it is also true that some opsins may not be viable *in vivo*. On the other hand, there are many combinations of substitutions that have not been tried *in vitro* and which might prove to fill in even the small gaps remaining.

The genomic evidence suggests a single array of opsin genes with a single locus control region: there are not, or so it seems, separate L and M loci in the classical sense. In both normal and colour-deficient subjects a variety of genes may be present in the array, some genes having sequences close to the recognized M and L sequences, some having 'hybrid' sequences. Only a subset (perhaps only two) of the genes are thought to be expressed. If we take the *in vitro* and the genomic evidence at its face value, it ought to be possible to find phenotypes with virtually any combination of pigments in the range 530 to 565 nm. Rayleigh match midpoints would depend on the spectral positions of the expressed pigments and chromatic discrimination upon their **spectral proximity**, the distance between their absorbance spectra. Some potential pigments are likely to be reduced in optical density; this would additionally affect matches and discrimination. However, the fact remains that the pool of X-chromosome opsin genes in the human population is dominated by genes that code for λ_{max} values close to 530 and 560 nm. Moreover, the two most potent amino acid sites, 277 and 285, are tightly linked and there has been no report of their being unyoked in a natural human gene (although viable pigments of this kind do occur in platyrrhine monkeys). Although a continuous range of pigments is available in principle there are, therefore, likely to be strong biases in the frequency distribution within this range: at one end there is likely to be a cluster of L and near-L pigments, at the other a cluster of M and near-M

pigments. The typical deuteranomal, we may suppose, draws his two expressed pigments from the first cluster, the typical protanomal expresses two pigments from the second. In both cases, hue discrimination depends primarily upon the spectral proximity of the pigments expressed and secondarily on factors such as optical density. Pigments falling midway between M and L probably exist (they could arise, for example, from the unyoking of sites 277 and 285) but they are likely to be confined to rare phenotypes, phenotypes that would correspond to anomals of the Schouten type. Whether an account of this kind can accommodate the discontinuous distribution of anomals and normals at the anomaloscope (Schmidt, 1953) remains to be seen.

Acknowledgements

I thank C.R. Cavonius, D. Hunt and B.C. Regan for comments on the text.

References

Alpern, M. and Moeller, J. (1977). The red and green cone visual pigments of deuteranomalous trichromacy. J. Physiol. (Lond.), 266: 647–675.

Asenjo, A.B., Rim, J. and Oprian, D.D. (1994). Molecular determinants of human red/green color discrimination. Neuron, 12: 1131–1138.

Baker, H.D. (1966). Single-variable anomaloscope matches during recovery from artificial red blindness. J. Opt. Soc. Am., 56: 686–689.

Baldwin, J.M. (1993). The probable arrangement of the helices in G protein-coupled receptors. EMBO J., 12: 1693–1703.

Brunner, W. (1930). Über den Verebungsmodus der verschiedenen Typen der angeborenen Rotgrünblindheit. Graefes Arch. Ophthalmol., 124: 1–52.

Dartnall, H.J.A., Bowmaker, J.K. and Mollon, J.D. (1983). Human visual pigments: microspectrophotometric results from the eyes of seven persons. Proc. Roy. Soc. Lond., 220: 115–130.

de Vries, Hl. (1948a). The luminosity curve of the eye as determined by measurements with the flickerphotometer. Physica, 14: 319–348.

de Vries, Hl. (1948b). The fundamental response curves of normal and abnormal dichromatic and trichromatic eyes. Physica, 14: 367–380.

DeMarco, P., Pokorny, J. and Smith, V.C. (1992). Full-spectrum cone sensitivity functions for X-chromosome-linked anomalous trichromats. J. Opt. Soc. Am., 9: 1465–1476.

Dulai, K.S., Bowmaker, J.K., Mollon, J.D. and Hunt, D.M. (1994). Sequence divergence, polymorphism and evolution of the middle-wave and long-wave visual pigment genes of great apes and Old World monkeys. Vision Res. 34: 2483–2491.

Earle, P. (1845) On the inability to distinguish colors. Am. J. Med. Sci, n.s. 9: 346–354.

Fitzgibbon, J., Appukuttan, B., Gayther, S., Wells, D., Delhanty, J. and Hunt, D.M. (1994). Localisation of the human blue cone pigment gene to chromosome band 7q31.3-32. Hum. Genet., 93: 79–80.

Hargrave, P.A. and McDowell, J.H. (1992). Rhodopsin and phototransduction. Int. Rev. Cytol., 137: 49–97.

Hargrave, P.A., McDowell, J.H., Curtis, D.R. *et al.* (1983). The structure of bovine rhodopsin. Biophys. Struct. Mech. 9: 235–244.

He, J.C. and Shevell, S.K. (1995). Variation in color matching and discrimination among deuteranomalous trichromats: Theoretical implications of small differences in photopigments. Vision Res. 35: 2579–2588.

Hunt, D.M., Dulai, K.S., Bowmaker, J.K. and Mollon, J.D. (1995). The chemistry of John Dalton's color blindness. Science, 267: 984–988.

Jacobs, G.H. (1993). Mammalian colour vision. Biol. Rev., 67: 413–471.

Jameson, D. and Hurvich, L.M. (1956). Theoretical analysis of anomalous trichromatic color vision. J. Opt. Soc. Am., 46: 1075–1089.

Jameson, D., Hurvich, L.M. and Varner, D. (1982). Discrimination mechanisms in color deficient systems. In: Verriest, G. (ed), Colour Vision Deficiences VI: 295–301. W. Junk, The Hague.

Jordan, G. and Mollon, J.D. (1993). A study of women heterozygous for colour deficiencies. Vision Res. 33: 1495–1508.

Kalmus, H. (1965). Diagnosis and Genetics of Defective Colour Vision. Pergamon, Oxford.

Knowles, A. and Dartnall, H.J.A. (1977). The photobiology of vision. In: Davson, H. (ed.), The Eye, vol 2B: 1-689. Academic Press, New York.

Köllner, H. (1915). Über Übergänge zwischen normalen Farbensinn und angeborene Rotgrünblindheit und über die Möglichkeit ihrer quantitativen Bestimmung. Arch. Augenheilk., 78: 302–335.

König, A. and Dieterici, C. (1892). Die Grundempfindungen in normalen und anomalen Farbensystemen und ihre Intensitätsverteilung im Spektrum. Z. Psychol., 4: 241–347.

Kries, J. von (1899). Ueber die anomalen trichromatischen Farbensysteme. Z. Psychol., 19: 63–69.

Lennie, P., Haake, P.W. and Williams, D.R. (1991). The design of chromatically opponent receptive fields. In: Landy, M.S. and Movshon, J.A. (eds.), Computational Models of Visual Processing: 71–82. MIT Press, Cambridge, MA.

Lerea, C.L., Bunt-Milam, A.H. and Hurley, J.B. (1989). A transducin is present in blue-, green-, and red-sensitive cones photoreceptors in the human retina. Neuron, 3: 367–376.

Lund Jørgensen, A., Deeb, S.S. and Motulsky, A.G. (1990). Molecular genetics of X chromosome-linked color vision among populations of African and Japanese ancestry: high frequency of a shortened red pigment gene among Afro-Americans. Proc. Natl. Acad. Sci. USA, 87: 6512–6516.

Lyon, M.F. (1972). X-chromosome inactivation and developmental patterns in mammals. Biol. Rev., 47: 1–35.

McKeon, K.M. and Wright, W.D. (1940). The characteristics of protanomalous vision. Proc. Phys. Soc., 52: 464–479

MacLeod, D.I.A. and Hayhoe, M.M. (1974). Three pigments in normal and anomalous color vision. J. Opt. Soc. Am., 64: 92–96

Magovcevic, I., Weremowicz, S., Morton, C.C., Fong, S.L., Berson, E.L. and Dryja, T.P. (1995). Mapping of the human cone transducin alpha-subunit (GNAT2) gene to 1p13 and negative mutation analysis in patients with Stargardt disease. Genomics, 25: 288–290.

Merbs, S.L. and Nathans, J. (1992a). Absorption spectra of the human cone pigments. Nature, 356: 433.

Merbs, S.L. and Nathans, J. (1992b). Absorption spectra of the hybrid pigments responsible for anomalous color vision. Science, 258: 464–466.

Merbs, S.L. and Nathans, J. (1993). Role of hydroxyl-bearing amino acids in differentially tuning the absorption spectra of the human red and green cone pigments. Photochem. Photobiol., 58: 706–710.

Mollon, J.D. (1989). "Tho' she kneel'd in that Place where they grew...". J. Exp. Biol., 146: 21–38.

Mollon, J. D. (1993). George Palmer. In: Nicholls, C. S. (ed), The Dictionary of National Biography: 509–510. Oxford University Press, Oxford.

Mollon, J.D. and Bowmaker, J.K. (1992). The spatial arrangement of cones in the primate fovea. Nature, 360: 677–679.

Mollon, J.D. and Jordan, G. (1988). Eine evolutionäre Interpretation des menschlichen Farbensehens. Die Farbe, 35/36: 139–170.

Mollon, J.D., Bowmaker, J.K. and Jacobs, G.H. (1984). Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. Proc. Roy. Soc., Lond., B, 222: 373–399.

Mollon, J.D., Dulai, K. and Hunt, D.M. (1996). Dalton's colour blindness: an essay in molecular biography. In: Dickinson, C., Murray, I. and Carden, D. (eds) John Dalton's Colour Vision Legacy. Taylor and Francis, London, in press.

Nathans, J. and Hogness, D.S. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. Proc. Natl. Acad. Sci. USA, 81: 4851–4855.

Nathans, J., Thomas, D. and Hogness, D.S. (1986a). Molecular genetics of human color vision: The genes encoding blue, green and red pigments. Science, 232: 193–202.

Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B. and Hogness, D.S. (1986b). Molecular genetics of inherited variation in human color vision. Science, 232: 203–210.

Nathans, J., Davenport, I.H., Maumenee, R.A. *et al.* (1989). Molecular genetics of human blue cone monochromacy. Science, 245: 831–838.

Nathans, J., Sung, C.-H., Weitz, C.J., Davenport, C.M., Merbs, S.L. and Wang, Y. (1992). Visual pigments and inherited variation in human vision. In: Corey, D.P. and Roper, S.D. (eds.), Sensory Transduction: 110–131. Rockefeller University Press, New York.

Nathans, J., Maumenee, I.H., Zrenner, E., *et al.* (1993). Genetic heterogeneity among blue-cone monochromats. Am. J. Hum. Genet., 53: 987–1000.

Neitz, J. and Neitz, M. (1992). Do people with anomalous color vision have anomalous cones? Invest. Ophthalmol. Vis. Sci., 33: 754.

Neitz, J., Neitz, M. and Jacobs, G.H. (1989). Analysis of fusion gene and encoded photopigment of colour-blind humans. Nature, 342: 679–682.

Neitz, M., Neitz, J. and Jacobs, G.H. (1991). Spectral tuning of the pigments underlying red-green color vision. Science, 252: 971–974.

Neitz, M., Neitz, J. and Grishok, A. (1995). Polymorphism in the number of genes encoding long-wavelength-sensitive cone pigments among males with normal color vision. Vision Res. 35: 2395–2407.

Nelson, J.H. (1938). Anomalous trichromatism and its relation to normal trichromatism. Proc. Phys. Soc., 50: 661–697.

Oprian, D.D., Asenjo, A.B., Lee, N. and Pelletier, S.L. (1991). Design, chemical synthesis and expression of genes for the three human color vision pigments. Biochemistry, 30: 11367–11372.

Palmer, G. (1777). Theory of Colours and Vision. S. Leacroft, London.

Palmer, G. (1786). Théorie de la Lumière, applicable aux arts, et principalement à la peinture. Hardouin et Gattey, Paris.

Pokorny, J. and Smith, V.C. (1977). Evaluation of single-pigment shift model of anomalous trichromacy. J. Opt. Soc. Am., 67: 1196–1209.

Pokorny, J., Moreland, J.D. and Smith, V.C. (1975). Pigments in anomalous trichromats. J. Opt. Soc. Am., 65: 1522–1524.

Pokorny, J., Smith, V.C., Verriest, G. and Pinckers, A.J.L.G. (1979). Congenital and Acquired Color Vision Defects. Grune and Stratton, New York.

Regan, B.C., Reffin, J.P. and Mollon, J.D. (1994). Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. Vision Res. 34: 1279–1299.

Reichel, E., Bruce, A.M., Sandberg, M.A. and Berson, E.L. (1989). An electroretinographic and molecular genetic study of X-linked cone degeneration. Am. J. Ophthal., 108: 540–547.

Reid, R.C. and Shapley, R.M. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. Nature, 356: 716–718.

Robinson, P.R., Cohen, G.B., Zhukovsky, E.A. and Oprian, D.D. (1992). Constitutively active mutants of rhodopsin. Neuron, 9: 719–725.

Ruddock, K.H. and Naghshineh, S. (1974). Mechanisms of red–green anomalous trichromacy. In: Verriest, G. (ed), Colour Vision Deficiencies II: 210–214. Karger, Basel.

Rushton, W.A.H. (1972). Pigments and signals in colour vision. J. Physiol. (Lond.), 220: 1–31P.

Rushton, W.A.H., Powell, D.P. and White, K.D. (1973). Pigments in anomalous trichromats. Vision Res. 13: 2017–2031.

Schmidt, I. (1953). Some problems related to testing colour vision with the Nagel anomaloscope. J. Opt. Soc. Amer., 45: 514–522.

Shyue, S.-K., Li, L., Chang, B.H.-J. and Li, W.-H. (1994). Intronic gene conversion in the evolution of human X-linked color vision genes. Mol. Biol. Evol., 1: 548–551.

Singer, M.S., Shepherd, G.M. and Greer, C.A. (1995). Olfactory receptors guide axons. Nature, 377: 19–20.

Voigt, J.H. (1781). Das herrn Giros von Gentilly Muthmassungen über die Gesichtsfehler bey Untersuchung der Farben. Magazin für das Neueste aus der Physik und Naturgeschichte (Gotha), 1: 57–61.

Watson, S. and Arkinstall, S. (1994). The G-protein Linked Receptor Facts Book. Academic, London.

Weitz, C.J., Miyake, Y., Shinzato, K. *et al.* (1992a). Human tritanopia associated with two amino acid substitutions in the blue-sensitive opsin. Ann. Hum. Genet., 50: 498–507.

Weitz, C.J., Went, L.N. and Nathans, J. (1992b). Human tritanopia associated with a third amino acid substitution in the blue-sensitive visual pigment. Ann. Hum. Genet., 51: 444–446.

Williams, A.J., Hunt, D.M., Bowmaker, J.K. and Mollon, J.D. (1992). The polymorphic photopigments of the marmoset: spectral tuning and genetic basis. EMBO J., 11: 2039–2045.

Willis, M.P. and Farnsworth, D. (1952). Comparative evaluation of anomaloscopes. Report No. 190, Medical Research Laboratory, Bureau of Medicine and Surgery, US Navy.

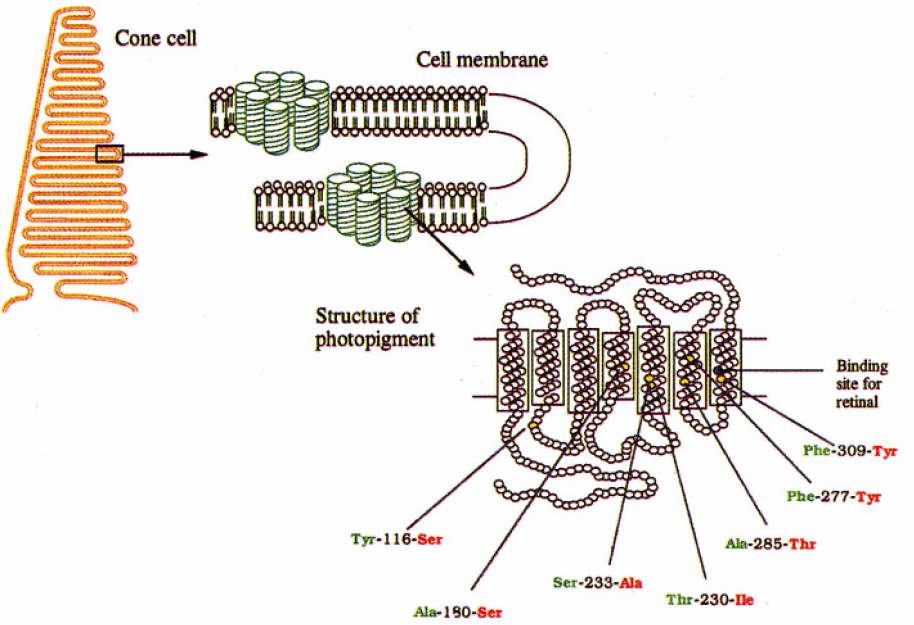
Winderickx, J., Battisti, L., Motulsky, A.G. and Deeb, S.S. (1992a). Selective expression of human X chromosome-linked green opsin genes. Proc. Natl. Acad. Sci. USA, 89: 9710–9714.

Winderickx, J., Lindsay, D.T., Sanocki, E., Teller, D.Y., Motulsky, A.G. and Deeb, S.S. (1992b). Polymorphism in red photopigment underlies variation in colour matching. Nature, 356: 431–433.

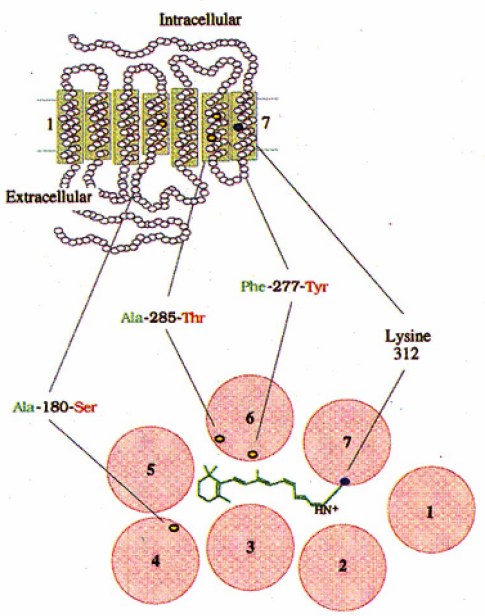
Winderickx, J., Sanocki, E., Lindsey, D.T., Teller, D.Y., Motulsky, A.R. and Deeb, S.S. (1992c). Defective colour vision associated with a missense mutation in the human green visual pigment gene. Nature Genet., 1: 251–255.

Wright, W.D. (1946). Researches on Normal and Defective Colour Vision. Henry Kimpton, London.

Department of Experimental Psychology
University of Cambridge
Downing Street
Cambridge, CB2 3EB
UK

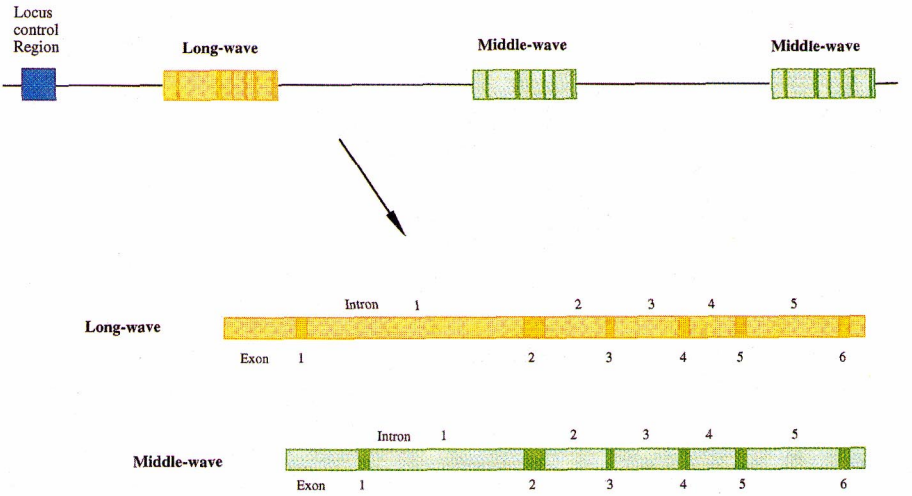


Chapter 1 Fig. 2. The enfolded membrane of the cone outer segment (upper left) is packed with photopigment molecules. Each of these consists of seven helices which span the membrane of the cell and are linked by loops outside the membrane. The seven helices form a palisade surrounding the chromophore, 11-*cis*-retinal. Represented at the bottom right is the sequence of amino acids that make up this heptahelical molecule. Highlighted in yellow are seven amino acid sites that are thought to shift the spectral sensitivity of the opsin from long-wave to middle-wave. The codes below represent the alternative amino acids at these positions: in each case the amino acid indicated in green is the alternative that shifts the peak sensitivity of the molecule to shorter wavelengths, that in red the one that shifts sensitivity to longer wavelengths.



Plan view of opsin from intracellular surface

Chapter 1 Fig. 3. A plan view of the opsin molecule showing the approximate positions of the amino acids that have a major influence on the peak sensitivity of the pigment.



Chapter 1 Fig. 4. Above: the arrangement of L and M genes on the X chromosome. Below: the arrangement of introns and exons for each gene.