Molecular Genetics of the Middlewave and Longwave Sensitive Opsin Genes of Higher Primates

Kanwaljit Singh Dulai

Institute of Ophthalmology

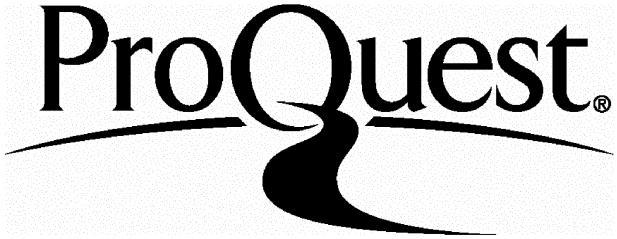
University College, London

University of London

Thesis submitted to the University of London

for the degree of Doctor of Philosophy

1996



Abstract

Over the 10 year period since the cone opsin genes of humans were first isolated and sequenced, a wealth of information has accumulated on the genetics of primate colour vision. However, a number of intriguing questions have remained unanswered. Do other apes and monkeys see with the same basic complement of photopigments as humans do? How are these different opsin genes temporally and spatially regulated? At present a complete understanding of human colour anomalies is lacking.

Based on the analysis of exons 3 to 5 of the X-chromosome opsin genes of two members of the Hominidea (the chimpanzee, *Pan troglodytes* and the Gorilla, *Gorilla gorilla*) and five members of the Cercopithecoidea family of Old World primates (the Diana monkey, *Cercopithecus diana*; lesser spot-nosed guenon, *Cercopithecus petaurista*; African green monkey, *Cercopithecus aethiops*; talapoin monkey, *Miopithecus talapoin*; and patas monkey, *Erythrocebus patas*), it is predicted that the long-wave (LW) and middle-wave (MW) pigments of these primates have similar spectral properties to those of human. Multiple copies of the same opsin gene sequence were identified in the chimpanzee and talapoin. Humans and the other primates show a bunching of polymorphic sites in exon three. The ancestry of the separate LW and MW genes of Old World primates is discussed.

Two anomalous trichromatic human males, each exhibiting the same phenotype, were found to present with differing genotypes. One subject, SSJ, possessed a normal LW and a MW-LW hybrid gene. The other subject, PSJ, had in addition a normal MW gene. It is suggested that the same phenotype may be possible due to selective expression of some of PSJ's opsin genes.

Finally, the phenotype of John Dalton, considered the father of colour vision, has been revised to deuteranopia , following amplification, cloning, and sequencing of his opsin genes from the remains of his eyes.

Acknowledgements

I wish to begin by acknowledging the tremendous support that Dr. David M. Hunt, my supervisor, has provided over the years, especially during the final stages. His faith in my abilities and understanding of my circumstances have not gone unnoticed.

I would like to thank the Medical Research Council for awarding the grant which supported this project.

This paragraph must be the longest due to the wonderful people I have had the pleasure of working alongside. It would only be fair of me to mention them all by name, and in alphabetical order. Alex (proof-reader) Morris, who kindly gave his time to read, and correct, this work; Alex (handball not football) & Alan (of England); Alison (Jet-Setter) Hardcastle; Annette (the Vampire); Binoy (down there somewhere) Appukuttan; Bride Woodcock; Cheryl and Kevin Evans; Chris (Is football on then?) Inglehearn, David (PC) Bessant; Dawn, for corrupting me with her TV; Debbie (spinster) who needs to escape, fast; Donna Kebab; Fabiana (tea) a most sophisticated person; Froggie (one of the best female soccer players around); Hamas Mai; Irish; Jeffrey (GeneReleaser) Keen, who I must thank for making the final results chapter possible; Godzilla, for the bloods; Jill Cowing, who's been there from the beginning; Jim (Boat Crusher) Bellingham, who loves to tinker; Jude Fitzgibbon, always on the ball; Karen (I will come along if I can!) Elder; Leslie Heath (The Fly III); Linford; Mani (The eye is the light of the body); Marcela Excel; Mark PC Hampson; Martin (Specs) Reichel; Naheed Hussain, daughter of Saddam; Neil (pregnant) Ebenezer scrooge; Oragutan; Peter (Gold's PC Gym) Clements; Rachel (:.: ::. ..: ::.:.;); Reshma ("I love America") K-tel, who desperately needs a husband; Richard the Lionheart, (please don't read this); Robin Ali; Rosemary (my neighbour); Sana (613); Sarah (wear that veil); Vectorette Sinthuja; Sue (3 legged); Sue (find me a husband quick); The Protein Lab; Tortilla (wide as a bus); Vinny from Bombay; Zoe (Hot dog lover); Zorro

Zarra; and Martin, who has been inflicted by computer troubles. Thanks everyone for everything. And remember, I love you all.

A huge thank you for all members of my family who have given me the space and time to complete this work. Their patience and wholehearted support has been greatly appreciated. My father, Tarsem Singh Dulai, who has always believed that an education is the greatest gift a parent can bestow on their offspring. I understand the difficulties that he experienced when he sought to educate himself. My mother, Baljit Kaur, for the constant encouragement. Professor Tarsem Singh, who started the "ball rolling", and Jarnail Dosanjh who has always believed in me. To all other members of the family, too numerous to mention, I wish to thank you too.

Finally, the biggest ever thank you must go to my wife, Rani Dosanhj-Dulai, without whom I would never have completed this thesis. She has always supported me, through both the good times and the hard. From the bottom of my heart I thank her very much.

Dedication

For my wife, Rani

the love of my life

Table of Contents

List of Figures

LIST OF TABLES

CHAPTER 1

Introduction

Sight is one of our most valuable senses. Using vision we can see the universe around us and are able to appreciate the visual spectrum, detecting food and dangers. New understandings have developed quickly in the last few years since the techniques of molecular biology have been brought to bear on long pondered problems. How well do we understand the basis of this sense?

Our eyes enable us to perceive the world around us, by capturing the direct and reflected light from objects. Like other organs of the body these structures are also subject to defects or disease.

Greater understanding of all aspects of vision and its defects can be obtained by expanding the analysis to other species which share comparable (physiological and morphological) structures. Naturally, because of evolutionary closeness, the eyes of other primates function in a similar manner to other vertebrate eyes. However, thus far there seems to be a disparity between the presence of colour-blindness in the human species and its lack in other primates.

Normal retinal morphology, biochemistry and genetics will be reviewed, focusing primarily upon human and other primate species. The alternative models of colour vision defects are presented, with particular consideration to possible mechanisms leading to colour vision anomalies. Finally, the gene regulation of opsin genes is discussed.

1.1 Structure of the eye

Animals have an amazingly diverse array of light-sensitive structures capable of detecting light, which range from the very simple to the extremely complex. The eyes of the deep ocean Nautilus consist of a simple spherical invagination which is open to and filled with sea water (Land, 1984). The eyes of birds and many reptiles contain oil

drops which act as cut-off filters (Yoshizawa and Kuwata, 1991). Vertebrate and cephalopods eyes, which are the most complex, function much like cameras: a single lens focuses light images on to a sheet of light receptors (equivalent to the film), and the image is then "developed" by neurone processing in the visual cortex of the brain. Typical "camera eyes" are roughly spherical and are mostly surrounded by a tough, protective sheet of connective tissue called the sclera, and by the choroid, a layer containing blood vessels that carry oxygen and nutrients to the eye and remove waste products (Figure 1.1). Pressurised fluid-filled spaces, the aqueous humour and vitreous humour chambers, help give the eye its shape. The cornea, the tough, transparent portion of the sphere through which light enters, is a specialised epidermal tissue composed mainly of aligned collagen fibres.

1.2 Eye function

When light enters the eye, it passes through the cornea and the aqueous humour. It may then be halted by the iris, a pigmented ring of tissue (blue, brown, green, grey, or intermediate shades in the human eye) that can be opened or closed by radially oriented muscles under reflex control. Or the light may pass through the iris's central opening, the pupil, to the lens, a structure built of highly ordered cells packed with structural proteins called crystallins. During animal evolution, lens proteins may have been selected from stress response proteins and different metabolic enzymes partly because they were suited to the mechanical changes that occur and the transparent crystals they form.

Light rays travelling through the lens and the pupil are focused on to a light-sensing epithelial layer, the neural retina. An important component of the neural retina in some reptiles, birds, and primates is the fovea, a region about 1 mm in diameter where maximal acuity is achieved. In all vertebrate eyes, there is an optic disk, or "blind-spot", where nerve fibres of the neural retina exit from the eye over the optic nerve, which goes to the brain. Finally, surrounding the neural retinal is the pigmented

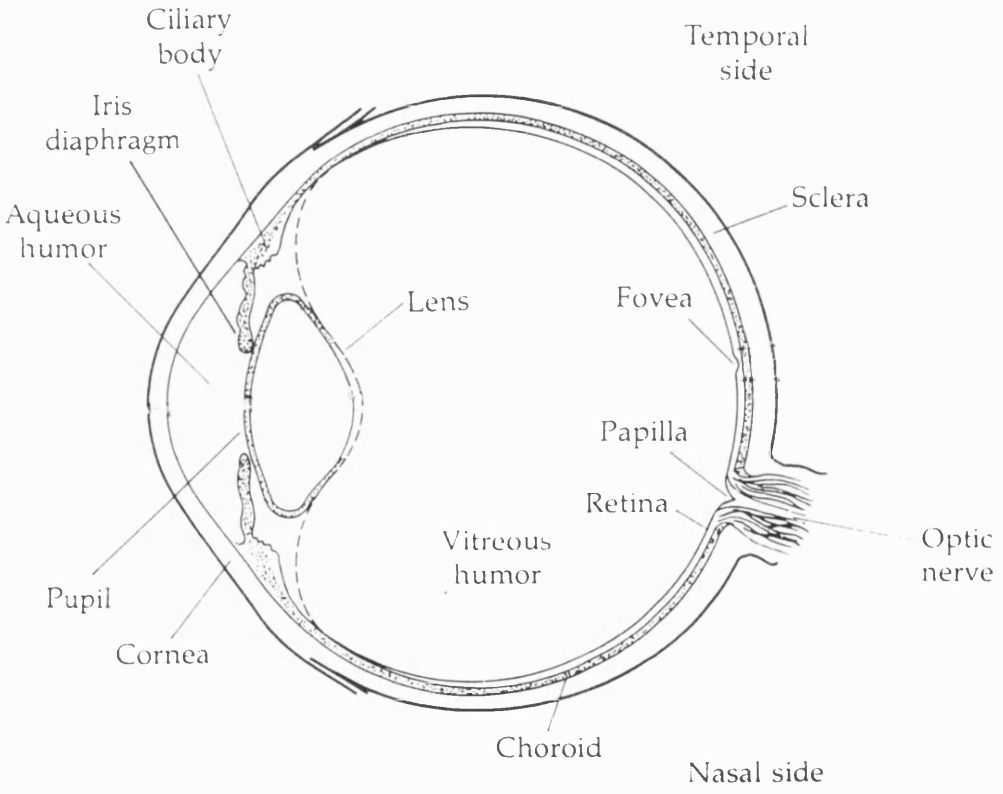


Figure 1.1 Structure of the vertebrate eye (after Hurvich 1981).

retina, a layer that absorbs or reflects light that passes unabsorbed through the neural retina.

When focusing a camera the lens is actually moved toward or away from the film. The eyes of fishes, amphibians, and snakes also focus in this manner. In birds and mammals, however, both cornea and lens participate in the focusing process. First, the curvature of the cornea causes light rays to bend (refract) to varying degrees. This corneal focusing is usually sufficient to provide a sharp image of distant objects. As focusing takes place, the lens is held taut in a somewhat flattened position by the elastic ligament of the cillary body, which supports it. When the object being viewed is near, however, the cornea cannot refract the light enough to focus the image on the retina, so the cillary muscles contract, allowing the lens to round up. This increases the angle of refraction of the light rays, bringing the image into sharp focus. Birds also bend the cornea to help this process. In land vertebrates, this process, called accommodation, goes on constantly as an animal shifts its glaze from one object to another at differing distances. The near-sightedness, myopia, and far-sightedness, hypermetropia, so common in human eyes stems from abnormalities in the shape or functioning of the cornea or lens; abnormalities that can usually be corrected with artificial lenses, or by more invasive treatments.

1.3 The retina

The neural retina, on which an image is focused, is derived from the embryonic neuroectoderm of the optic cup, which is an outgrowth of the primitive forebrain. The highly complex neural retina contains the visual receptor cells, or photoreceptors, as well as a variety of nerve cells.

Light microscopic cross-sections of the combined neural and pigmented retina reveal a striated 0.22 mm thick structure in which ten layers are evident (Figure 1.2). The layer furthest from the incident light, the vascular choroid, supplies the outer third of the retina with nutrients. Bruch's membrane, a thin layer which delimits the retina proper supports the basal surfaces of cuboidal cells of the retinal pigment epithelium

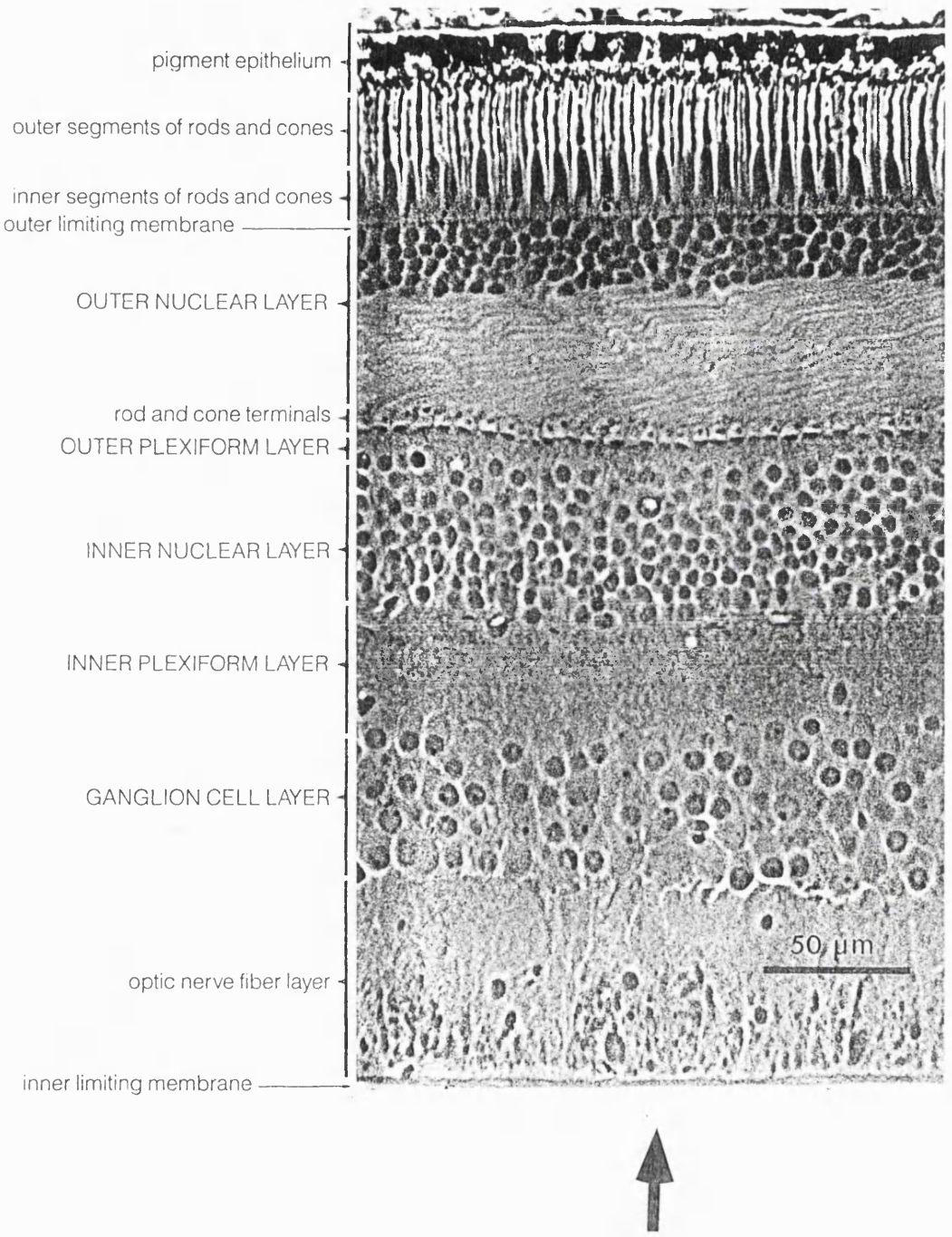


Figure 1.2 Vertical section through a human retina. The micrograph shows an area about 1.25 mm from the center of the fovea. Light impinges on the retina from the direction indicated by the broad arrow. (Adapted from Ali and Klyne, 1985).

(RPE). The space between the RPE cells is closed off by the zonulae occludentes, forming the blood-retina barrier. Adjacent to the RPE are the outer segments of the photoreceptor cells, the space between which is filled with interphotoreceptor matrix. The matrix, which is composed of carbohydrates, such as glycosaminoglycans, and glycoproteins, participates in the cycling of retinoids between the surrounding cells via the interphotoreceptor retinol binding protein(Bunt-Milam and Saari, 1983),and in the anchoring of the photoreceptors and the glial cells to the RPE.

The outer nuclear layer is formed by the photoreceptor perikarya. The synapses between the terminal processes of the photoreceptor cells, the bipolar cells, and the horizontal cells give rise to the outer plexiform layer. The inner nuclear layer consists of the nuclei of the horizontal cells, bipolar cells, amacrine cells, and the Müller cells. The synaptic connections between the bipolar, amacrine, and ganglion cells give rise to the inner plexiform layer. The nuclei of the ganglion cells constitutes the ganglion cell layer. The nerve fibre layer comprises the axons of the ganglion cells as they travel towards the optic disc. The internal and external limiting membranes are not membranes at all. The external limiting membrane is the region in which the photoreceptors and the radial processes of the Müller cells form the zonulae adherentes. Likewise the internal limiting membrane, at the vitreous surface, is the combination of the expanded termination's of the Müller cells and the basement membrane (Dowling, 1987).

The axons of ganglion cells are the sole information channels from eye to brain. This means that the wiring of rods, cones, bipolar cells, and other nerve cells in the retina determines the quality of vision. In the primate fovea there is thought to be a simple one-to-one hook-up from a cone to a bipolar cell to a ganglion cell. This means that the brain receives signals from precisely those cones struck by photons. In most parts of the retina, however, many rods, or cones, synapse with a single bipolar cell, and a number of bipolar cells synapse with each ganglion cell. In the human eye, some 131 million photoreceptors are connected via bipolar cells to only 1 million ganglion cells. The full set of rods and cones joined in this manner to a single ganglion cell make

up the visual field of that ganglion cell: photons hitting any of the photoreceptors connected to the ganglion cell can thus trigger it to generate a nerve impulse.

Two other types of neurones that participate in information processing in the retina are horizontal cells, which interconnect rods or cones with bipolar cells, and amacrine cells, which in turn interconnects sets of bipolar and ganglion cells. This latticework of cross connections within the neural retina permits the visual field pathways of different ganglion cells to interact with each other, providing heightened contrast and thus information about an object's contours. A more detailed account can be found elsewhere (Ali and Klyne, 1985).

In many vertebrates, including humans, the eye contains two types of photoreceptors, rods and cones (Ali and Klyne, 1985), which are shaped as their names imply (Figure 1.3), and which contain different sets of light-absorbing pigments.

1.4 Photoreceptors

All mammalian retinas examined show the presence of both rods and cones (Jacobs, 1993), although the ratio of the two can vary both in total terms (for example, rats have predominantly rod retinas, whilst that of the ground squirrel contains almost entirely cones (von-Schantz *et al.*, 1994; West and Dowling, 1975)), and in different regions of a single retina (the human fovea contains exclusively two classes of cones, the middle-wavelength sensitive and long-wavelength sensitive, no rods or short-wavelength sensitive cones are present (Bowmaker, 1984)). Whereas humans and other animals that are active mostly during the day have both rod and cone retinas, nocturnal species that live in low-light environments, such as deepwater fish(Bowmaker *et al.*, 1994a; Hunt *et al.*, 1996), tend to possess predominantly "rod-eyes". Rods are highly sensitive: they respond even when the amount of light entering the eye is very low (scotopic vision), such as at night or in sheltered environments (e.g. caverns). Cones, on the other hand (which contain less of their respective pigments than do rods) respond to higher light intensities (photopic vision), and so are most active in daytime

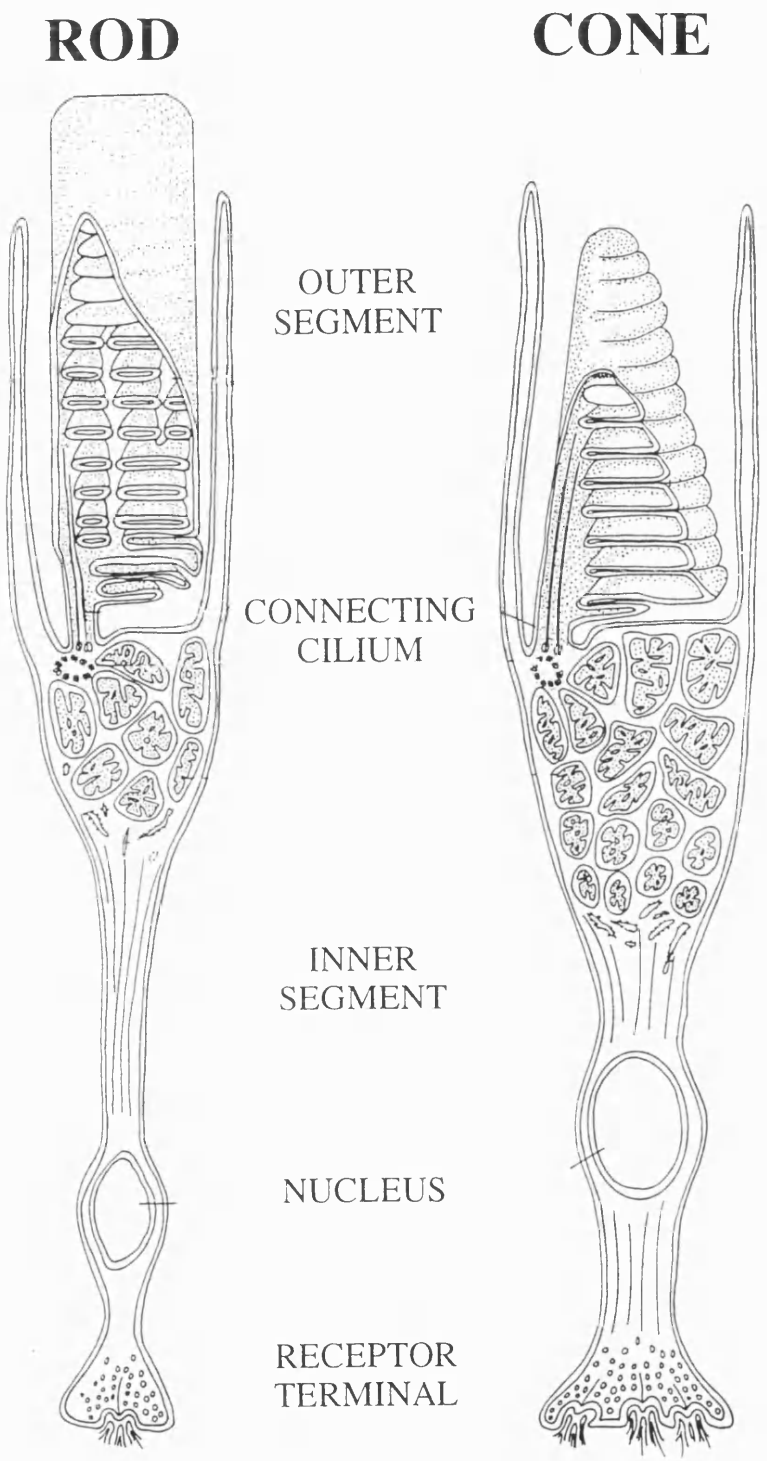


Figure 1.3 Diagrammatic representation of mammalian rod and cone photoreceptors. Adapted from Ali and Klyne (1985).

light levels. Cones also provide high visual acuity and, under appropriate conditions, colour vision (Bowmaker, 1990; Mollon, 1991b).

Rods and cones can be distinguished both morphologically and functionally, indeed their names are derived in large part from their shapes. Each cell type has the same basic plan: an outer segment is linked via a connecting cilium to the inner segment, the inner segment containing the nucleus and finally the synaptic body, which allows transmission of impulses (Figure 1.3). Rods maintain a uniform diameter along the length of the outer segment, whilst cones exhibit a tapering towards the apex (Figure 1.3). Photoreceptor cells are not neurones, but are highly specialised receptor cells, possibly hair cells which evolved to fill this niche.

The outer segments are filled with hundreds of flattened membranous saccules (disks). The membranes of these disks contain millions of molecules of the light absorbing visual pigments. These pigment molecules, which constitute 80% of all protein and 95% of membrane protein of the outer segment, respond to photons of light and initiate the visual process. As Young (1971) has shown, outer segment biosynthesis in a continuous process. His work on Rhesus monkey rods indicated that disks are continuously added from the cilium to the base of the outer segment and ten days later are phagocytosed at the distal end by pigment endothelial cells.

In the dark, the membrane potential of a rod cell is about -30 mV, considerably less than the resting potential (-60 to -90 mV) typical of neurones and other electrically active cells (Ali and Klyne, 1985). As a consequence of this relative depolarisation, rod cells in the dark are constantly secreting neurotransmitters, ~~and the bipolar neurones, with which they synapse, are continually stimulated~~. A pulse of light causes the membrane potential in the outer segment of the rod cell to become slightly hyperpolarised - that is, more negative. The light-induced hyperpolarisation causes a decrease in release of neurotransmitters (Rayer *et al.*, 1990).

The relative depolarised state of the plasma membrane of resting, dark-adapted rod cells is due to the presence of a large number of open sodium (Na+) channels. An increase in Na+ permeability causes the membrane potential to become more positive: that is,

depolarised. The effect of light is to close Na+ channels. The more photons absorbed, the more Na+ channels are closed, the more negative the membrane potential becomes, and the less neurotransmitter is released. Remarkably, a single photon absorbed by a resting rod cell produces a measurable response, a hyperpolarisation of about 1 mV, which lasts for about a second. Humans are able to detect a flash of as few as five photons (Ali and Klyne, 1985). A single photon blocks the inflow of about 10 million Na+ ions due to the closure of hundreds of Na+ channels. Only about 30-50 photons need to be absorbed by a single rod cell in order to cause half-maximal hyperpolarisation (Ali and Klyne, 1985), whereas cone cells require considerably more photon strikes. Rod photoreceptors, like many other types of receptors, exhibit the phenomenon of adaptation. That is, more photons are required to cause hyperpolarisation if the rod cell is continuously exposed to light than if it is kept in the dark (see section 1.6).

1.4.1 Rods

As noted above, rods are responsible for scotopic vision. In a dark-adapted human eye, rods are about ten thousand times more sensitive to photons of light then are cones. Numerous rods connect to a single ganglion cell, thus allowing summation of a response. Unlike cones the photosensitive lamellae of rods are not continuous with their plasma membranes, but are individual discs. The slender cilium does not vary a great deal in length, even between species.

1.4.2 Cones

Photopic vision is the domain of cones. Fewer cones are interconnected in a manner analogous to rods, enabling a finer resolution of detail, especially in the foveal region. The sharpness of an animal's vision depends on the density of cones in the fovea. Each human fovea has approximately 160,000 cones per square millimetre, (compared to roughly 1 million in a similar region in the hawk's eye).

In the outer segments, plasma membrane invagination gives rise to the lamellar body (Figure 1.3). A fluctuation in the length of the cilium has been noted between

organisms. In the fovea, individual cone cells are thought to synapse with bipolar cells in a one to one connection, as opposed to other regions where numerous cone cells connect to a single bipolar cell. See Ali and Klyne (1985) for a full account.

1.5 Photopigments

Packed into the outer segments of photoreceptors are the photopigments, which mediate the light response. The absorption of photons by these light-sensitive pigments triggers a conformational change that eventually results in the generation of a neural output from the photoreceptor cell (see section 1.6).

The visual pigments (opsins) belong to a vast superfamily of structurally similar integral membrane proteins (over 200 members to date; Baldwin, 1993) with an α-helical conformation. They are characterised by seven hydrophobic membrane-spanning segments, that are linked by extramembrane hydrophilic loops: the so called heptahelicals (Mollon, 1991a; Khorana, 1992; Trumpp-Kallmeyer *et al.*, 1992). Receptors from this family have been found in a wide range of organisms and are believed to be involved in the transmission of signals across membranes. These receptor molecules bind a signalling molecule on the extracellular face of the membrane and following activation, engage with a heterotrimeric guanine nucleotide-binding protein (G protein) on the cytosolic face, which in turn initiates a second messenger system of intracellular signalling. The numerous receptors each consist of a single polypeptide whose length can range from 340 to 500 amino acids. The transmembrane segments, which are thought to exist as α-helices of approximately 26 residues each (Baldwin, 1993; Schertler *et al.*, 1993; Alkorta and Du, 1994), constitute 50 % of the molecule, whilst the remaining loops, connecting helices on either side of the membrane, exist as straight chains (Hargrave *et al.*, 1984). The seven helices are believed to assemble to form a palisade structure (Figure 1.4) that encompasses a central extracellular facing ligand-binding site (for reviews see Collins *et al.*, 1989; Mollon, 1991a; Kobilka, 1992; Applebury, 1994).

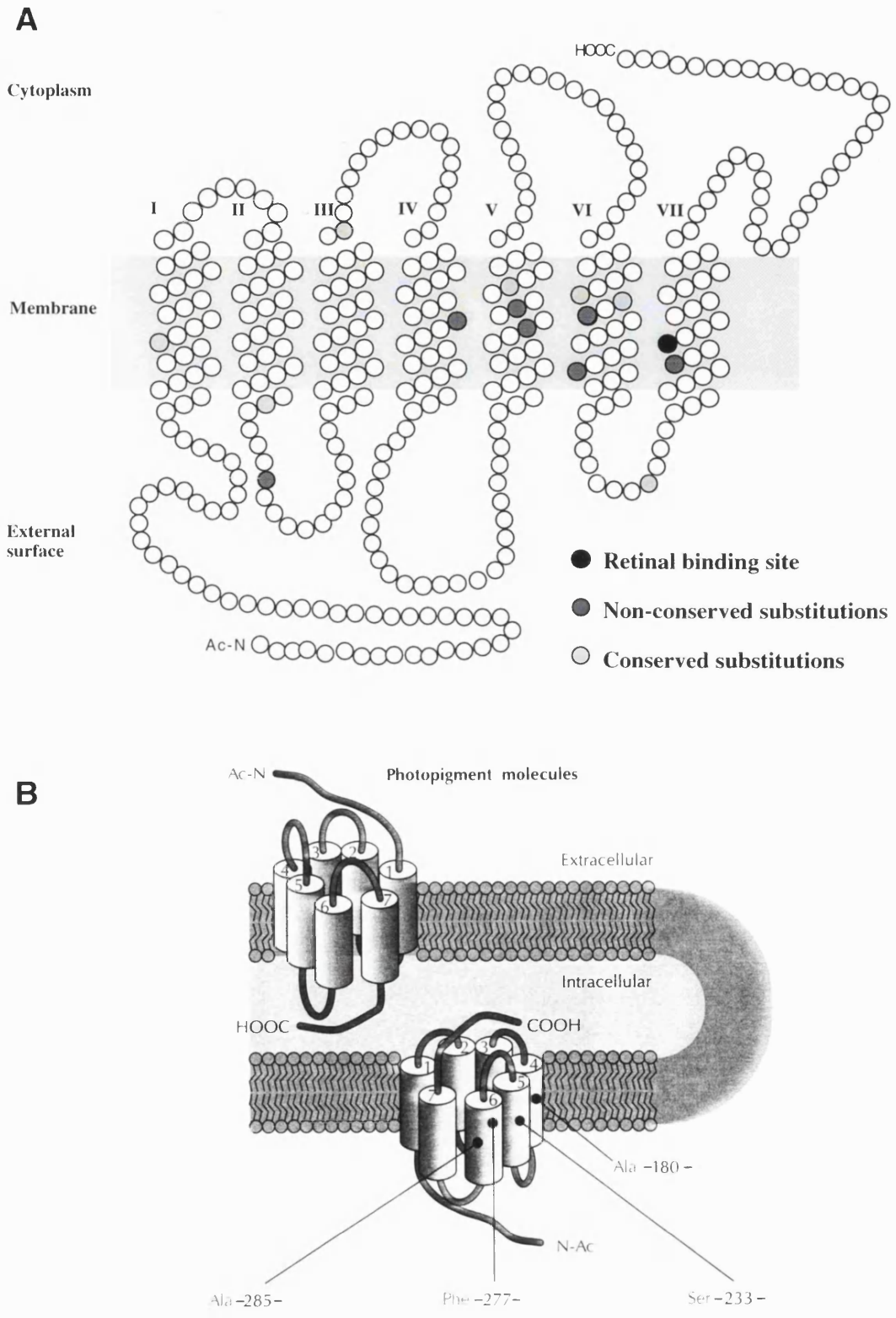


Figure 1.4 **A**. Two dimensional representation of the opsin protein; referred to as a heptahelical since the polypeptide chain transverses the membrane seven times. HOOC and Ac-N denote carboxy and amino terminal ends, respectively. The differences indicated are those between LW and MW opsins. **B**. The "palisade" arrangement, thought to depict the 3-D conformation adopted by the functional opsin molecule. HOOC and Ac-N denote carboxy and amino terminal ends, respectively. (Adapted from Mollon, 1993).

Vertebrate visual pigments are all generated by the addition of a chromophore, either 11-*cis*-retinal, the aldehyde derivative of vitamin A (A1 pigments), or 11-*cis*-3-dehydroretinal (A2 pigments), to an opsin protein encoded by the appropriate genes although some invertebrate pigments employ 11-*cis*-3-hydroxy-retinaldehyde as their chromophore (Vogt and Kirschfeld, 1984). The formation of a covalent Schiff base linkage with a lysine residue in the seventh helix of the opsin molecule sensitises the chromophore to light (discussed further in section 1.6). The absorption of photons by the chromophore is the result of the shape of the π electron shell and double bonds that exist along the length of the molecule (see section 1.6.1).

Each pigment has a characteristic response curve to the wavelength of incident light (Figure 1.5), measured utilising the microspectrophotometry (MSP) technique (Partridge, 1986). The apex of the curve indicates the wavelength to which that particular pigment is the most responsive - this is known as the λ_{max} of the photopigment. However, the effect of absorbed photons on the pigment is independent of their wavelength, thus the area under each curve represents the probability of photon capture for that pigments at that wavelength. This is, the principle of univarance, first proposed by William Rushton (1963). The absorption maxima (λ_{max}) of human cone pigments are 430, 532, and 563 nm for the SW, MW, and LW pigments, respectively (Oprian *et al.*, 1991; Merbs and Nathans, 1992a), as measured by MSP (the values fluctuate slightly between studies).

1.5.1 Rod opsin

The receptor for scotopic stimulation is rhodopsin, which consists of the apoprotein rod opsin and a chromophore, 11-*cis* retinal, which is common to all vertebrate visual pigments. Following the publication of the amino acid sequence of bovine rod opsin (Ovchinnikov, 1982; Hargrave *et al.*, 1983; Ovchinnikov *et al.*, 1983), the genes encoding bovine (Nathans and Hogness, 1983) and human (Nathans and Hogness, 1984) rod opsins were sequenced. The human rod opsin protein is 348 amino acids in length and possesses regions of alternating hydrophobicity (non-polar)

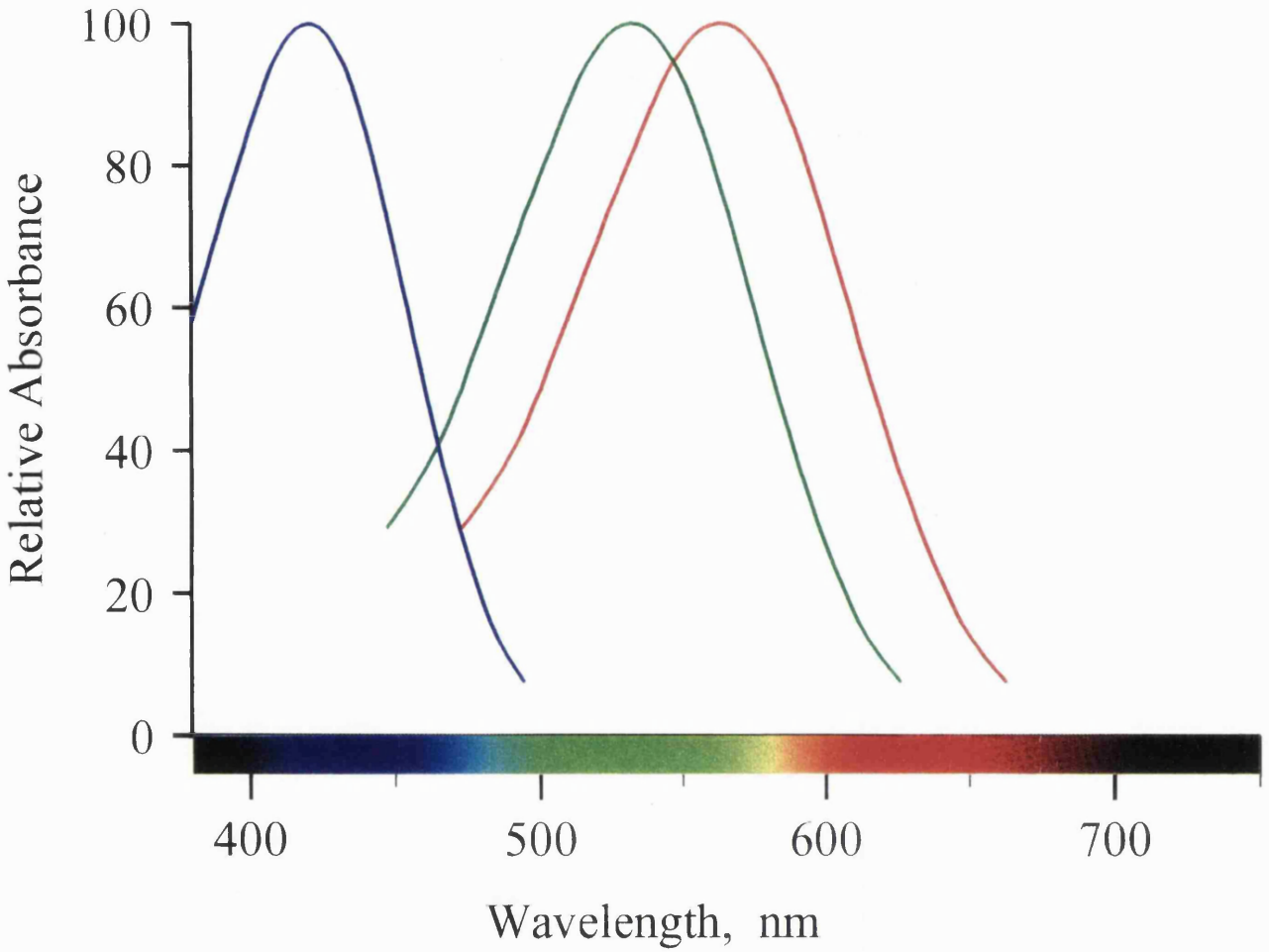


Figure 1.5 The spectral sensitivities of the three types of cone photoreceptor in the normal human eye.

and hydrophilicity (polar), such that *in vivo*, the former regions are embedded in the membrane bilayer, and the latter reside at the aqueous surface of the membrane (Hargrave, 1982; see Figure 1.4). The three dimensional structure is at present not known because of the difficulties in forming crystals. However, predictions based on topographical data suggest an organisation very similar to that of bacteriorhodopsin (Baldwin, 1993; Du and Alkorta, 1994; Alkorta and Du, 1994), and a general model for G-linked heptahelical receptors has been proposed interpreted by Baldwin (1993; Figure 1.6). The rod opsin does not absorb visible light, but when bound to the chromophore, via the Lys 296 residue (numbering of amino acid residues in rod opsin follows that of bovine rod opsin; Wang *et al.*, 1980), it is able to. To stabilise the charge a second site was predicted. The counterion to the Schiff's base linkage was identified by site-directed mutagenesis studies to be Glu113 (Zhukovsky and Oprian, 1989; Nathans, 1990b; Cohen *et al.*, 1992; Zvyaga *et al.*, 1994). In terms of the tertiary structure, these two residues face the retinal binding pocket in the centre of the ring of transmembrane helices (Baldwin, 1993). An account of the 3-D model can be found in Alkorta and Du (1994).

1.5.2 Cone opsins

Human colour vision is based on light sensitive pigments that absorb maximally in different regions of the spectrum (see section 1.7). Structural differences between the proteins govern the spectral properties of these pigments. Pigments can be classed according to the spectral region of maximal absorption (λ_{max}). Various terms are encountered in the literature when reference is made to these different pigment classes. The terms red, green, and blue are often used (Boynton, 1979), however there has been a move to avoid them since they can be misleading if they are taken to mean that the different pigment classes are most sensitive to the regions of the spectrum seen as red, green, and blue, which is not the case. For example, the "red" pigments have peak absorption near 560 nm, a wavelength that appears greenish-yellow, not red, and the blue pigment peaks around 430 nm, a spectral region that appears violet (see Figure

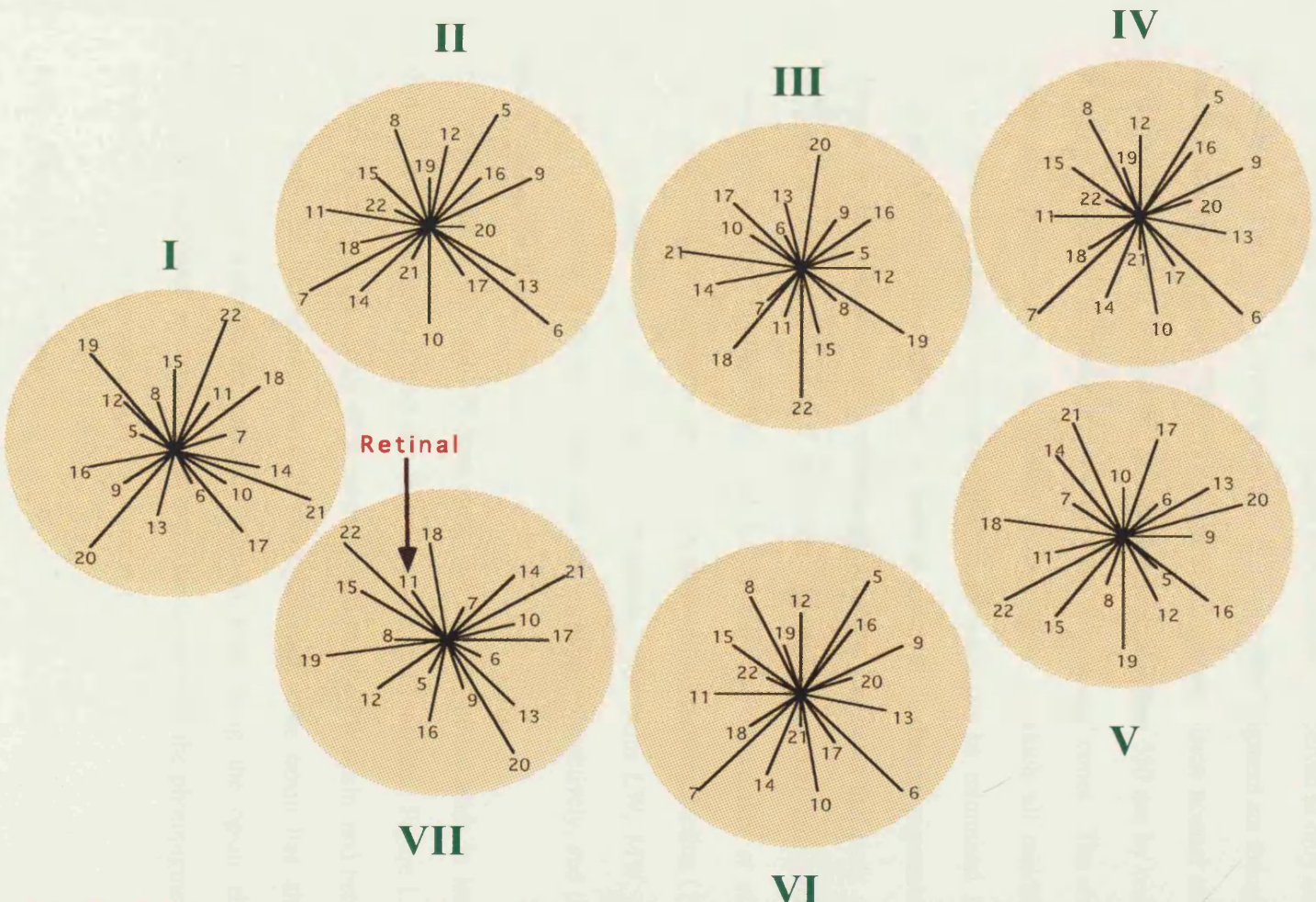


Figure 1.6 The Balwin model representing the probable arrangment of the seven helices, as viewed from above, based on data from a large number of G protein-coupled receptors. Numbers in each circle represent the position of amino acids within that helix. The retinal binding site (312 for the LW and MW cone opsins) is indicated. Adapted from Balwin (1993).

1.5). The sensation of "redness" comes about because of the way output from the receptors is processed by the nervous system, not because humans have a receptor that is most sensitive to the red region of the spectrum.

Persons who lack the "red" pigment are sometimes perceived as only able to see green and blue but not red, while those who lack the "green" pigment are thought to see red and blue but not green. This is not the case. Colour-vision normal individuals determine the colour of objects with a wavelength greater then 550 nm by comparison of the outputs from two receptor types, the "red" and "green" cones. The absence of either cone type results in dichromatic colour vision in which all middle-to-long wavelengths are virtually indistinguishable. Confusion can be minimised if colour terms are avoided when reference is made to pigments. Human pigments can be classed into three categories: long-wavelength sensitive (pigments that peak near 560 nm), middle-wavelength sensitive (pigments that peak near 530 nm), and short-wavelength sensitive pigments (that peak below 430 nm) (Bowmaker *et al.*, 1980; Bowmaker, 1984), which can be conveniently abbreviated following Mollon (1991a) to long-wave, middle-wave, and short-wave pigments. The terms LW, MW, and SW will be used to refer to the red, green, and blue pigments, respectively, and the opsin genes that encode them.

The human MW and LW cone opsins are each 364 amino acids in length and show 96 % identity to each other (i.e. they differ at only 15 residues; Figure 1.7). The degree of identity drops to only 40 % between these and the SW opsin, and between all cone opsins and rod opsin. As with rod opsin each cone opsin has alternating hydrophilic and hydrophobic regions. Certain sites along the opsin chain are recognised as playing an important role in the functionality of the photopigment, most of which parallel the rod opsin molecule (section 1.5.1).

1.6 Phototransduction

Stimulation of the photoreceptors is the beginning of a series of changes that occur in photoreceptor cells, without which the sensation of vision would not be

Figure 1.7 Nucleotide and amino acid alignment of human MW and LW opsin gene coding regions. Nucleotide sequences are labelled with Seq, whilst amino acid sequences are denoted by aa. A (.) indicates identity to the sequence, and a (-) indicates identity to the amino acid above. Single letter amino acid code used. Amino acid differences are illustrated in colour. Total length of the coding region is 1092 bp and 364 amino acids. Data from Nathans *et al.* (1986a).

complete. The term phototransduction is used to describe the high gain and fast kinetic signal transduction cascade which takes place in mature photoreceptor cells, converting quanta of light energy within a particular range of wavelengths into electrochemical changes. This results in membrane hyperpolarisation, reduction of neurotransmitter release, and ultimately in the generation of a nerve impulse which is directed to the brain. Amplification occurs at every stage, therefore feedback and adaptation mechanisms are employed to modulate the cascade.

Over the last 15 to 20 years, there have been major advances in understanding not only the molecular basis of this process, but also aspects of its regulation, and consequently the field has been extensively reviewed (Koutalos and Yau, 1993; Yau, 1994). The biochemical events that occur following the absorption of a photon have been best worked out for the rod photoreceptor cells, but increasing similarities are being discovered in the cone cells.

1.6.1 Phototransduction cascade in rods

Once 11-*cis*-retinal absorbs photons of light the primary photochemical event which takes place is isomerisation of the 11-*cis*-retinal moiety in rhodopsin to all-*trans*-retinal. Thus atomic motion is derived from the energy of light. The intermediate conformation, in which opsin is bound to all-*trans*-retinal is called *meta*-rhodopsin II. Histidine residues of the opsin appear to participate in this controlled conversion (Weitz and Nathans, 1992). This change eliminates the counterion-Schiff base salt bridge and exposes residues necessary for the activation of transducin, the G protein coupled to rhodopsin. Two residues in particular, Glu134 and Arg135 are conserved in the receptor family and are required for transducin activation (Franke *et al.*, 1992). The formation of *meta*-rhodopsin II is both extremely efficient and rapid (10 ms), but this structure is unstable and spontaneously dissociates, releasing opsin and all-*trans*-retinal. All-trans-retinal is isomerised back to 11-cis-retinal by components in the rod-cell membrane, prior to re-forming rod opsin. The biochemical cascade that is initiated in rod outer segments as a result of this change to retinal is illustrated in Figure 1.8.

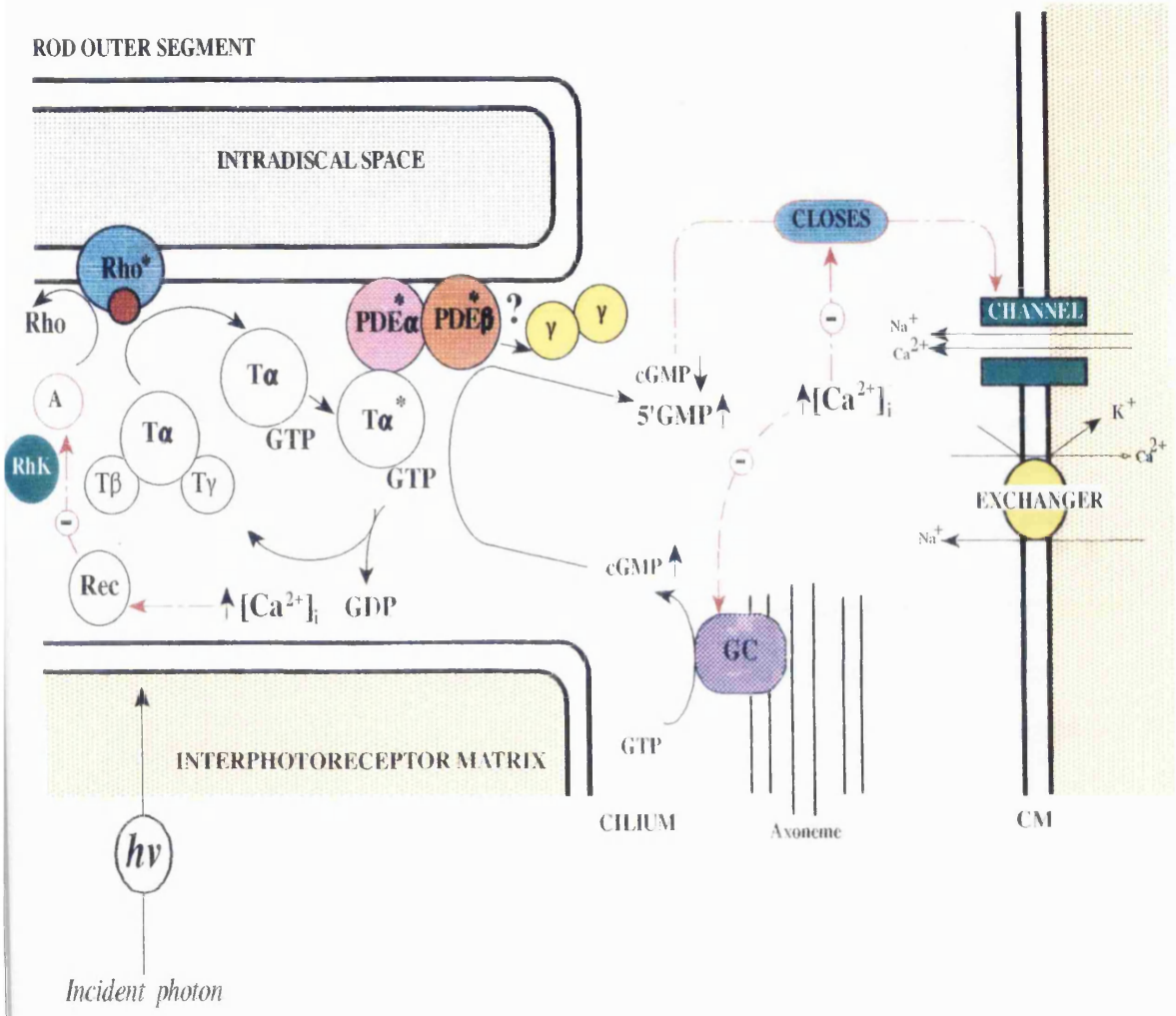
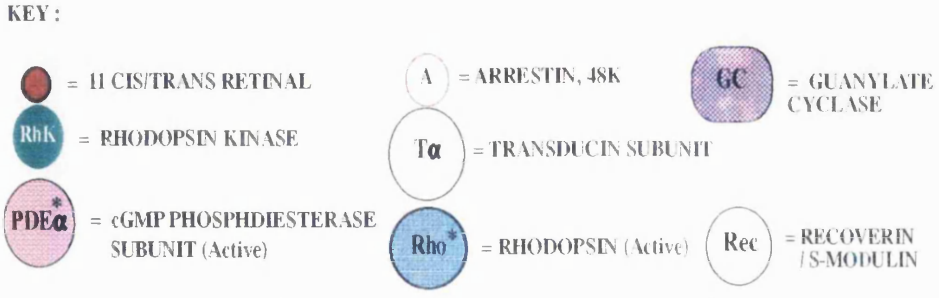


Figure 1.8 Phototransduction cascade in mammalian rod photoreceptor cells (from Clements, 1995).



The molecule 3',5'cyclic GMP (cGMP) is considered the key molecule in transduction. It links the activation of opsin to opening of the Na+ channels, and regulates the adaptation response (Farber, 1995b). The enzyme phosphodiesterase (PDE), which hydrolyses the effector molecule cGMP, is central to the process (Farber, 1995b). It itself is activated by the G protein, transducin.

Transducin is a heterotrimeric GTP binding protein. The three constituent subunits, Tα, Tβ, and Tγ all have separate isoforms in rods versus cones (Lerea *et al.*, 1986; Peng *et al.*, 1992). Post-translational modification of Tα (Fukada *et al.*, 1990) and Tγ (Fukada, 1993) exists, a refinement which has significant functional implications. The phototransduction cascade is controlled and regulated by a number of accessory enzymes and factors. Full accounts are provided by Koutalos and Yau (1993); Yau (1994) and Farber (1995a).

1.6.3 Phototransduction in cones

A similar sequence of events is thought to take place in cones (Cobbs *et al.*, 1985). The main difference lies in the quantity, not the quality of the cascade. Data based on polyclonal antibodies suggest that more than one gene may exist for cone specific α transducin (Lerea *et al.*, 1989). Other experimenters have reported additional cone homologues of rod phototransduction pathway components (full accounts are given in Yau, 1994; Farber, 1995a).

Phototransduction is a complex biochemical pathway. Whilst much of the stimulatory pathway has been described, more intriguing details regarding its regulation are still surfacing.

1.7 Spectral tuning

Nature is quite apt at utilising a single structure and fitting it to a number of solutions. The photopigments are an excellent example. All known photopigments have two components: the opsin protein and the prosthetic group bound to it (see section 1.5). The combination of the two within a photoreceptor determine the wavelength at which the pigment is maximally sensitive. This makes it possible for one

type of opsin to be bound to different prosthetic groups, giving rise to spectrally different pigments. As an example the fresh water eel switches its chromophore during different stages of its life (Archer *et al.*, 1995).

Spectral tuning, also referred to as wavelength regulation, is the modification of the photopigment allowing it to respond maximally to different wavelengths.

1.7.1 Coarse tuning - Schiff's base

A single retinal chromophore covalently links to a lysine amino acid residue of an opsin molecule through a protonated Schiff base linkage (Oseroff and Callender, 1974). The absorption peak of free 11-cis-retinal in aqueous solution is 385 nm (Sakmar *et al.*, 1989), while the protonated Schiff's base has maximum absorbance at 440 nm. This shift towards the longer wavelengths, or red end, of the spectrum is known as the opsin shift, or bathochromic shift. Association with different opsin proteins can alter the wavelength of maximal absorption of the chromophore from 350 nm to 620 nm *in vivo* (Bowmaker, 1991).

Spectral tuning has long been postulated to be influenced by the environment within the opsin binding pocket where the retinal chromophore resides. The formation of the Schiff's base involves the protonation of the nitrogen atom of the lysine residue to which retinal binds. The positive charge thus created has to be stabilised by some mechanism to allow the photopigment to operate in the range of visible light. The inability to obtain photopigments in a crystalline form has prevented the exact molecular interactions from being deduced. However, a number of models have been proposed to account for the molecular interplay which takes place.

One of the first models, the external point charge model, is based on chemical analogue investigations, and postulates the existence of two negative charges to stabilise the positive charge of the protonated Schiff's base (Honig *et al.*, 1976; Honig *et al.*, 1979). Site-directed mutagenesis studies, where all possible acidic residues in the membrane spanning regions of rhodopsin were systematically replaced, suggested the carboxylate anion of glutamate 113 (E113; rod opsin numbering system) side chain was

the negative counterions (Sakmar *et al.*, 1989; Zhukovsky and Oprian, 1989; Nathans, 1990a). Confirmation has been obtained by resonance Raman spectroscopy (Lin *et al.*, 1994).

1.7.2 Fine tuning of photopigments

The finding by Nathans *et al.* (1986a) that for human visual pigments the retinal component of the visual apoproteins remained unchanged, meant that any differences in the λ_{max} must be due to variations in the amino acids of the opsin protein. As stated, comparison of the LW and MW amino acid sequences shows the existence of 15 differences (see Figures 1.7 & 1.9; Merbs and Nathans, 1993). Which of these different amino acids are responsible for the 30 nm or so shift in the λ_{max} between these two pigments has been the subject of much debate!

Eight out of the fifteen amino acid positions are not considered as candidates because they either lie outside the transmembrane regions or involve conservative substitutions (Kosower, 1988). The remaining amino acid sites are all candidates for spectral tuning because they involve swapping of amino acids with dissimilar properties. Most exchanges involve either the gain or loss of hydroxyl groups, or changes in charge, which is thought to influence the stability of the ground state relative to the excited state of the pigment, causing a shift in its λ_{max}.

A comparative study involving opsin genes from humans and two species of South American monkey (Neitz *et al.*, 1991) indicated that just three amino acid substitutions, at positions 180, 277, and 285, could account for the spectral difference between human LW pigments and MW pigments. MW pigments possess alanine at site 180 encoded by exon 3 (helix four of the protein), and phenylalanine at position 277 and alanine at 285, both encoded by exon 5 (helix six), whereas LW opsins encode serine, tyrosine, and threonine at the same three sites, respectively.

The phenylalanine to tyrosine substitution at site 277 (F277Y: amino acid substitutions are referred to by the identity of the pre-exiting residue, abbreviated by using the single-letter amino acid designation, followed by the codon number, followed

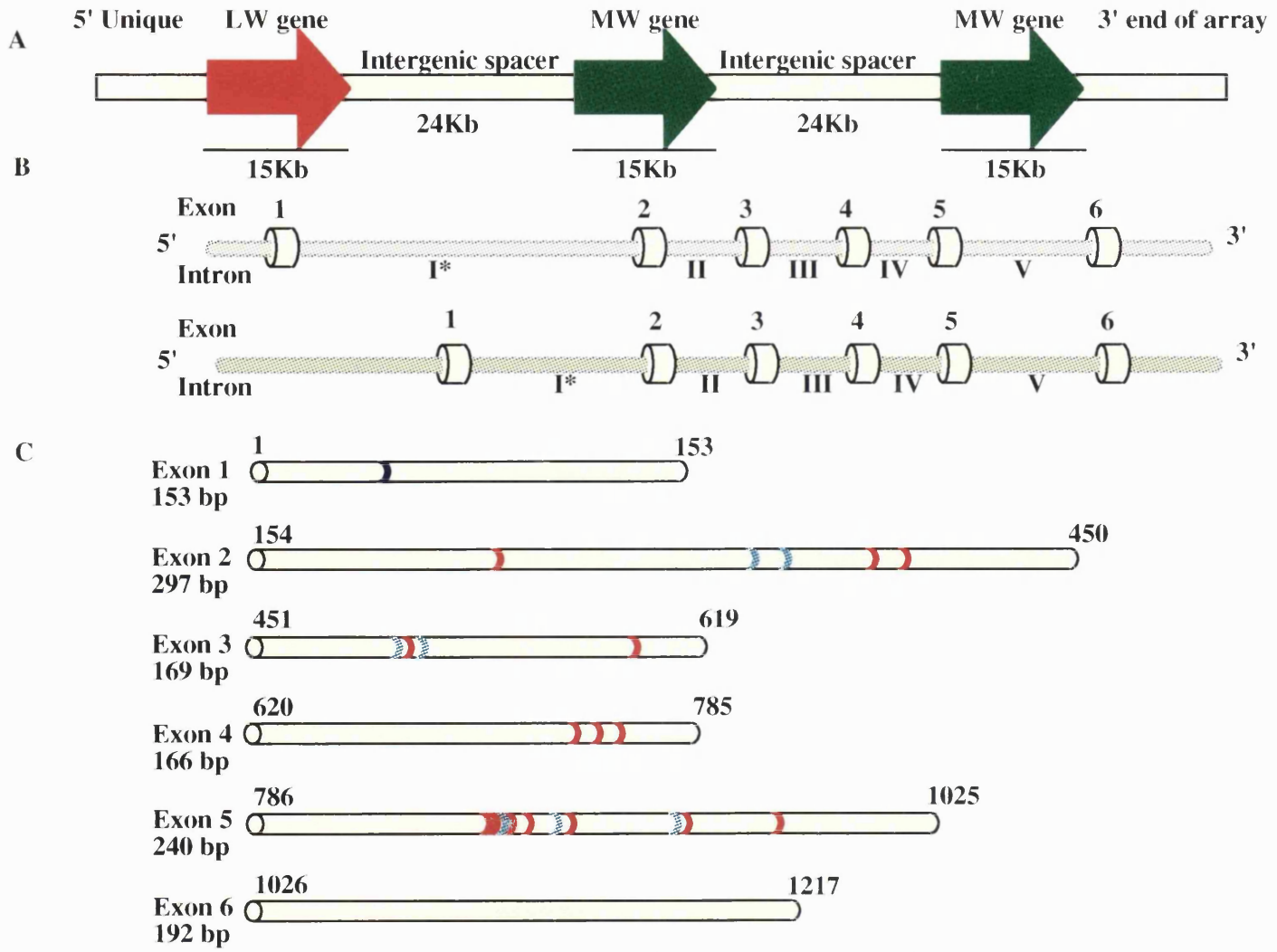


Figure 1.9 Organisation of the human LW and MW opsin loci. **A**. Illustrates the intergenic distances and arrangement of the opsin genes along the X chromosome. **B**. Intron /exon organisation of the respective genes. (* the size of this intron can vary between different human populations). **C**. The length of each exon and the sites of amino acids that differ between the LW and MW opsins (red), transcription start site (blue), and non-coding nucleotide differences (light blue). Numbering system adopted in this figure assumes mRNA start site is +1.

by the introduced residue), was reported to produce a shift of 10 nm, and the alanine to threonine change at site 285 yielded a 15 nm shift (Neitz *et al.*, 1991). Together, these two sites alone have been proposed to determine which class of opsin (LW or MW) is encoded, whilst the 180 position (which causes a 4-6 nm shift) fine tunes the pigment within its class. The importance of the above three amino acid positions have been confirmed by other studies (Chan *et al.*, 1992; Ibbotson *et al.*, 1992; Williams *et al.*, 1992; Merbs and Nathans, 1992a; Merbs and Nathans, 1992b).

However, other findings have suggested that two additional sites (230 and 233) may also influence the spectral positioning of the LW and MW pigments (Ibbotson *et al.*, 1992; Williams *et al.*, 1992; Winderickx *et al.*, 1992c). Both these sites are situated in helix five of the folded protein, and encoded by exon 4 of their respective opsin genes. Each site is believed to alter the λ_{max} by only a few nm.

Whilst all the above investigations relied on comparisons between the sequences of opsin genes from animals for which the absorption spectra of the individual pigments had been recorded by MSP, more recent work has concentrated on the techniques of site-directed mutagenesis and *in vivo* expression of reconstituted pigments to advance our understanding of spectral tuning. A number of studies employing these methods to identify amino acids involved in wavelength regulation have been reported (Sakmar *et al.*, 1989; Zhukovsky and Oprian, 1989; Nathans, 1990b; Nakayama and Khorana, 1991; Chan *et al.*, 1992; Zhukovsky *et al.*, 1992; Merbs and Nathans, 1992b).

The work of Asenjo *et al.* (1994) provided strong evidence, based on an extensive mutagenesis study, that seven amino acid residues are responsible for the 31 nm difference that exists between the MW and LW pigments. They generated, via the method of restriction fragment replacement and subsequent expression, a large set of chimaeric proteins, modelled both on pigments present in the colour-deficient human population and additionally single and multiple point mutants (Figure 1.10 shows a schematic representation of their data). Measurement of the λ_{max}, by absorption spectroscopy, suggested that sites 116, 230, 233, and 309, as well as, the previously implicated sites 180, 277, and 285, were responsible for the complete spectral shift

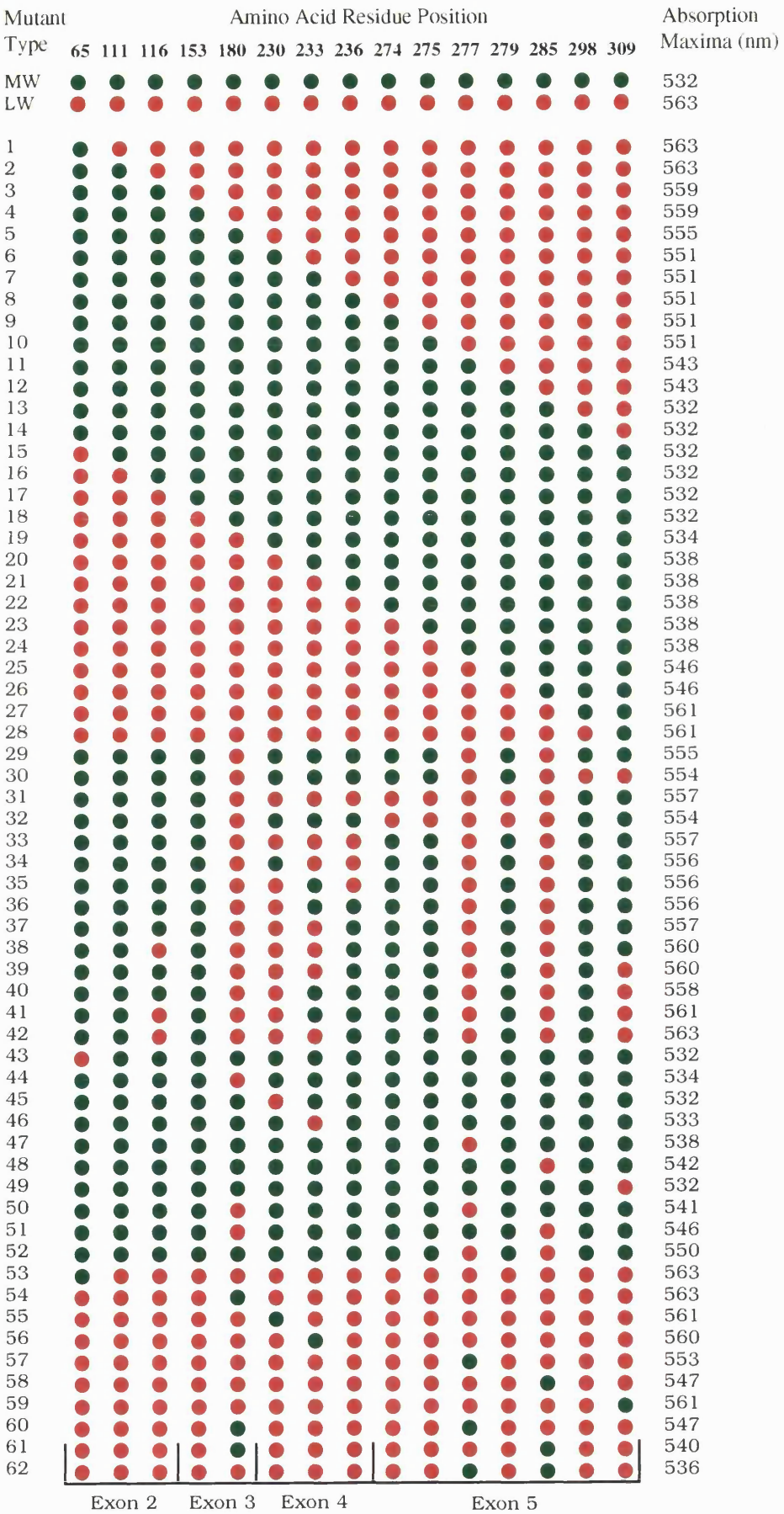


Figure 1.10 Schematic representation of the LW/MW chimaeric receptors tested by Asenjo *et al.* (1994). A red circle indicates that the amino acid is from a LW class pigment; a green circle indicates the amino acid is from a MW class pigment. Based on Asenjo *et al.*(1994).

between the MW and the LW colour vision pigments. Only by changing the amino acids at all seven position from those found in the MW pigment to those found in the LW pigment is it possible to get a complete conversion from a MW pigment to a LW one. The greatest changes were still observed when changes were made to sites 180, 277, and 285 (Figure 1.10).

Chang *et al.* (1995) took a different approach by aligning published opsin protein data from both vertebrates and invertebrates to construct a phylogenetic tree. By limiting their analysis to regions of the opsin molecule thought to interact with the chromophore based on structural and crystalline evidence (Barlow and Thornton, 1988; Stryer, 1988), examination of the tree enabled them to suggest that substitutions for polar or charged amino acids at four sites close to the protonated Schiff base end of the chromophore result in blue shifts of the absorption spectra of the visual pigment. It has been suggested that polar residues in the vicinity of the protonated Schiff's base stabilise the ground state over the excited state of the chromophore resulting in a blue shift (an increase in the energy required for photon absorption). Contrast this to the situation in MW and LW opsins where polar residues, distributed over the backbone of the chromophore towards the ß-ionone ring, stabilise the excited state thus lowering the energy needed for photon absorption, resulting in a red shift..

More insights into spectral positioning are emerging from the ongoing investigation of both fresh-water and deep-sea fish opsin absorption maxima (Bowmaker *et al.*, 1994; Hope *et al.*, 1996; Hunt *et al.*, 1996).

Further work is needed to completely understand the *in vivo* interactions and the mechanism regulating spectral tuning. Indeed, the greatest shift in λ_{max} for the longer-wavelength opsins have been reported as the result of anion binding (Fager and Fager, 1979; Knowles, 1980; Kleinschmidt and Hárosi, 1992; Wang *et al.*, 1993), where ions such as chloride are proposed to bind to the hydrophilic domains of the extracellular face of the opsin causing an increased delocalisation of the π electron shell. Support has come from the findings of Wang *et al.* (1993) who identified two residues (histidine

at position 197 and lysine at site 200: cone opsin numbering), as the most likely candidates for chloride-binding.

The present consensus view is that substitutions at sites 277 and 285, in exon 5, determine the spectral differences that separate X-linked opsin pigments into the LW and MW classes, whilst, polymorphisms at sites situated in exons 2 to 4 (65, 180, 230, 230 and 309) produce the subtypes within each class (Asenjo, 1994; Neitz *et al.*, 1995).

1.8 Colour vision

1.8.1 Definitions

Colour vision is characterised by the ability to distinguish light that differs in wavelength composition. Given a test light of any wavelength composition, humans with normal colour vision can make an identical colour match using just three appropriately chosen colour primaries, e.g., red, green, and blue - either by mixing the right amounts of the three primaries, or by mixing two primaries and adding the third to the test light. This three-dimensionality or trichromacy is the hallmark of normal human colour vision.

Other forms of colour vision can be similarly classified by the number of primaries required in colour matches. An individual who is completely colour-blind, classified in this way, is a monochromat. A monochromat can exactly match the appearance of any light, regardless of its wavelength content, with a single fixed primary light by simply adjusting the intensity of the comparison primary. An example would be when an reference yellow light is matched by adjusting the intensity of a red light.

Dichromats are those individuals who need only two primaries to match the reference light. For instance a yellow reference light would be matched by adjusting the intensity and proportion of red and green lights.

The terms monochromacy, dichromacy and trichromacy strictly refer to the number of primary colours required in colour matches, however these terms are often

used to refer to the number of spectrally different cone pigments an individual has, or is believed to have. While it is true that three cone pigments are the minimum required for trichromatic colour vision and at least two are needed for dichromacy, there may be hazards in equating visual capacities with the number of cone pigment types. For instance an individual may be found to possess say three cone classes yet may still be classed as a monochromat (colour-blind) if the neural processing system has a single non-interacting output.

1.9 Colour vision tests

A range of non-invasive colour vision tests have been developed over the years to detect deficiencies of colour vision and to assign phenotypes. Psychophysical studies rely on the judgement of the subject, whilst, invasive procedures provide direct measurements of photopigments.

1.9.1 Ishihara plates

Dr. Shinobu Ishihara (1968) designed a series of plates which provide a test that gives a quick and accurate assessment of colour vision deficiency of congenital origin. Subjects are asked to view and comment on each plate that they are shown. Their answers are used in conjunction with the appendixed chart, provided with the plates, to ascertain if they are colour defective. The test is non-invasive and can be used to screen a large number of subjects quickly, and is sensitive enough to differentiate the different categories of colour blindness.

1.9.2 Rayleigh colour match

The Rayleigh colour match is named after John William Strutt (better known as Lord Rayleigh (1881)) who introduced the use of this test to diagnose red-green colour vision abnormalities. A full description of the apparatus and test procedure is provided in Piantanida and Gille (1992) and Lutze *et al.* (1990). Basically, the subject is asked to look through an aperture with one eye at a time, where (s)he sees an bipartite circular field. The subject is required to produce a match with a reference yellow (589 nm)

light, in the upper half of the field, by altering the ratio of two primary lights, red and green, which combine to illuminate the lower half of the field. The readings on the apparatus (R/R+G), allow comparisons of red-green colour vision between subjects (see section 1.10).

1.9.3 Microspectrophotometry

MSP is a means of measuring the absorbance spectra of individual photoreceptor cells, which have been dissected away from the retina. This technique, first developed by Caspersson (1940) and since enhanced (Bowmaker, 1984), entails the passage of a fine monochromatic wavelength light through the outer segment of the photoreceptor, whilst a second reference beam is directed through a clear region. Light intensities of the two beams are compared by the instrument, and following certain corrections, the pigment transmission is determined at that wavelength. The procedure is systematically repeated with wavelengths from 600 nm to 350 nm. The characteristic absorbance curves are generated by this method (see Figure 1.5), and the λ_{max} of the photoreceptor can be determined. The technique has been instrumental in collating data on the colour vision of many species. Partridge (1986, 1989, and 1991) fully describes the procedure.

1.10 Colour vision defects

Colour vision defects are known to affect different human populations at different frequencies. For example, a few opsin gene polymorphisms are restricted to certain races (Winderickx *et al.*, 1993; Deeb *et al.*, 1994), resulting in minor differences. Most frequencies quoted below refer to Caucasian populations, unless otherwise stated. During the course of this project, an opportunity arose to reassess the colour vision defect of John Dalton. Dalton was amongst the first to critically examine his own form of colour anomaly, namely red-green colour-blindness. Chapter 6 provides a detailed account.

1.10.1 Monochromacy

Monochromacy, also known as achromatopsia, is a rare condition affecting 1 in 100,000 individuals. Those with this condition have no colour discrimination; in colour discrimination tests they match any one coloured light to another simply by adjustment of the relative intensities of the two (Neitz and Neitz, 1994). The first recorded case in the Western world appeared in 1777 (Huddert, 1777). Two possible mechanisms at the photoreceptor level are responsible for the condition; either the loss of two of the three classes of cone pigments normally present, or the absence or dysfunction of all three. In the former case the remaining cone class functions at photopic (bright light) levels alone. Since there is no other pigment class with which to compare the signal, the information cannot be resolved into wavelength composition, necessary for colour discrimination. When all cone pigments are absent, vision is mediated solely by the rods - the affected individuals characteristically shy away from bright light. Cases of acquired monochromacy that follow from damage to the cortical region of the brain and involve higher order neural processing have also been reported (Damasio *et al.* 1980).

Monochromatic individuals are completely colour-blind, having poor visual acuity and may even show photophobia and nystagmus. Within this group there are recognised three distinct clinical types of congenital monochromacy (Blackwell and Blackwell, 1961); reviewed by Nathans *et al.* (1989). In the first class, known as typical rod monochromacy, all cone function is ablated (congenital achromatopsia). Rods and cones are present but the later are functionally defective. These individuals are highly sensitive to daylight since they depend exclusively on their rods for all vision, which become bleached as light levels rise. In some instances cones give action spectra typical of rod opsin, suggesting that this pigment is present in both rods and cones. The condition is inherited as a non-progressive autosomal recessive. In a second related class, atypical rod monochromacy, cone pigments are present within cones, as are rod opsins within rods, however, these individuals present with the same symptoms as the first class. The visual defect must lie beyond the point of light

absorption. The inheritance of this condition is also autosomal recessive. The third category is blue cone monochromacy, also known as $\pi 1$ monochromacy or X-linked incomplete achromatopsia, which is a rare sex-linked condition characterised by the loss of LW and MW cone pigments, but not SW cones. There are three know genetic mechanisms leading to the presentation of the same phenotype.

For the first mechanism, two independent events are necessary. Initially, an unequal cross-over leads to a reduction in the number of the opsin genes on the X-chromosome to only one. The sole remaining gene undergoes a single point mutation which results in the replacement of an amino acid, cysteine, by an arginine (C203R; Nathans *et al.*, 1989). The cysteine residue is required to form an essential stabilising disulphide bond with another site (C126) on the opsin protein (Karnik and Khorana, 1990). This substitution causes the dysfunction of both the LW and MW cone pigments, with the remaining functional SW cone pigment left to mediate vision. The affected individuals see a colourless world because there are no other functional cone cell types to resolve the wavelength composition. Interestingly, at intermediate light levels there appears to be interaction between the SW cones and rods to produce a simple colour discrimination (Reitner *et al.*, 1991).

The second mechanism causing monochromacy is a deletion of a critical region encompassing 3.1 kb to 3.5 kb upstream of the LW opsin gene (Nathans *et al.*, 1989). This region, now known as the locus control region (LCR), defines a 579-base pair region that is located 4 kb upstream of the red pigment gene and 43 kb upstream of the nearest green pigment gene. This region is thought to be a master switch, directing downstream expression of the opsin gene array (Wang *et al.*, 1992). Lack of the LCR eliminates all expression of the LW and MW opsin genes and has been shown experimentally by using reporter gene constructs (Wang *et al.*, 1992). The third recorded form of genetic defect is the presence in all copies of the LW and MW opsin genes of a single base mutation (C203R; Reyniers *et al.*, 1995).

1.10.2 Dichromacy

Dichromacies, which represent a common form of colour blindness (one Caucasian in every 100), result when a single class of cone pigment is absent. As there are three classes of cone, there are three classes of dichromacy: Protanopes possess a SW and a MW pigment, but lack LW pigments; deuteranopes have a SW and a LW pigment, but lack MW pigments; and tritanopes do not have any SW pigment but do have the LW and MW pigments (Figure 1.11).

1.10.3 Anomalous trichromacy

Anomalous trichromats require three primaries in colour matches, but do not perform with the same consistency or accuracy as normal trichromats. In Rayleigh colour tests their matches either fall outside what is considered the normal range of values, or there is less uniformity from test to test. Again there are three categories: protanomaly, deuteranomaly, and tritanomaly (see Figure 1.11). Protanomalous individuals, who represent 1% of all colour vision defects in Caucasians, require more red light in the Rayleigh test. Deuteranomalous persons, representing the largest single class of colour vision of the Caucasian male population, require a greater amount of green light in the Rayleigh match. Tritanomalous individuals are very rare.

Protanomalous and deuteranomalous individuals display a wide range of colour discrimination phenotypes, from very mild, where it is difficult to distinguish them from normal observers, to very severe, where colour vision is almost as poor as that of dichromats.

1.10.4 Frequency of colour anomalies

Dichromacy and anomalous trichromacy attributable to defects in the LW or MW opsin genes are inherited in a sex-linked manner, which makes males much more susceptible to these conditions. For European Caucasian 8 % of males show some forms of variant LW / MW colour vision. For Africans and Native Americans the frequency is 1-4 %, and for Asians it is 4 to 5 % (Post, 1962). Females exhibit a frequency less than the square of the male value (Francois, 1961; Nathans, 1992). Of

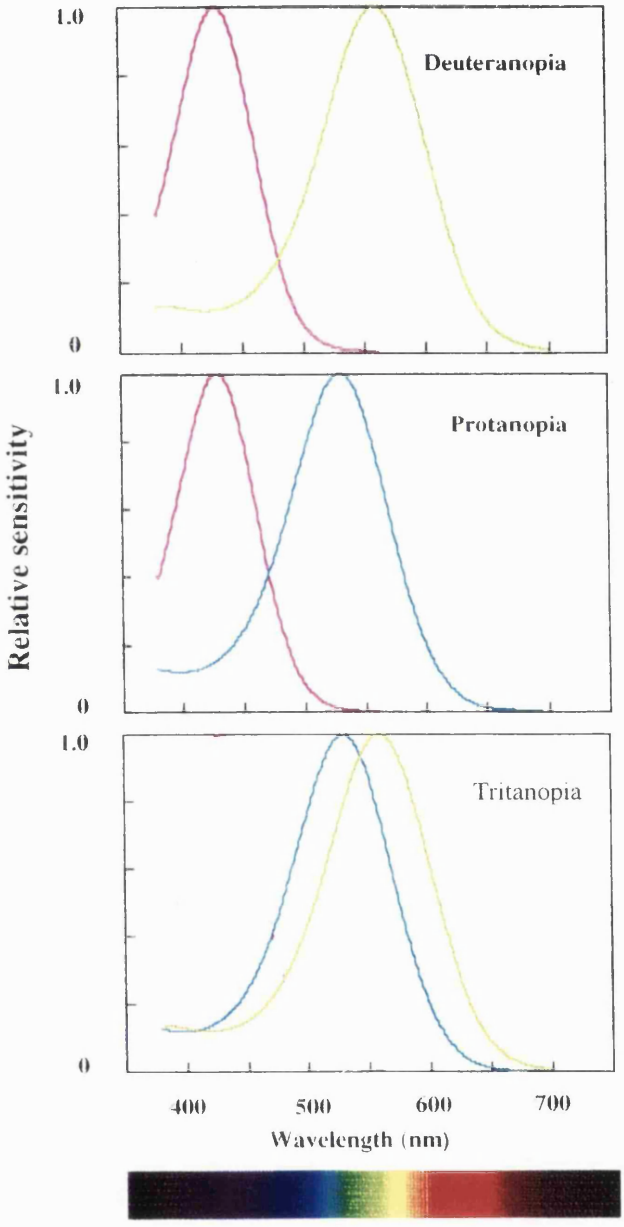


Figure 1.11 Spectral sensitivities of the cone receptors in three types of colour anomaly.

the 8 % of European Caucasians who are colour-blind, 15 % are protanopes, 16 % are deuteranopes, while 12 % are protanomalous and 57 % are deuteranomalous (Francois, 1961; Nathans *et al.*, 1992).

1.11 Molecular genetics of the visual pigments

1.11.1 Localisation of the genes of human cone opsins

Jeremy Nathans and colleagues were the first to isolate the genes coding for human cone opsins (Nathans *et al.*, 1986a, b). They reasoned that if the different visual pigments evolved from a common ancestor, then the genes that encode them may retain a degree of sequence homology to each other. A molecular clone of one gene would allow the isolation of the others by hybridisation under appropriate stringency. Using probes based on the nucleotide sequence of bovine rhodopsin (Nathans and Hogness, 1983), human genomic DNA was screened. At relatively high stringency a single class of clone was detected, which was identified as the human rhodopsin gene. Further screening at relaxed stringency revealed a further three classes of opsins; LW, MW, and SW, which were sequenced and characterised.

Pulsed-field gel electrophoresis studies of Vollrath *et al.* (1988) showed that these genes are arranged in a tandem array with the LW pigment gene 5' of the MW pigment gene(s) as previously hypothesised by Nathans *et al.* (1986a,b). These same studies further indicated that the length of the MW gene unit (from the 5' end of one MW gene to the 5' end of the next) was 39 kb. The LW gene unit is very homologous to the 39-kb MW gene unit: the mapped restriction site patterns are identical except for the 5' part. It was shown that intron 1 in the LW gene is 2 kb larger than the same intron of the MW gene in Caucasian populations, whereas a section of the African American population does not show this disparity (Jorgensen *et al.*, 1990). Furthermore, at 237 bp upstream of the first exon, the LW and the MW gene sequences lose their homology (Nathans *et al.*, 1986a). This divergence in the 5' region between the LW and MW genes probably accounts for the rarity of loss of the 5' region of the LW gene by unequal recombination.

The localisation of the SW opsin gene on chromosome seven was first reported by Nathans *et al.* (1986b) and its locus has since been further refined by fluorescent in-situ hybridisation (Fitzgibbon *et al.*, 1994).

Both the MW and LW opsin proteins are 364 amino acids in length (see Figure 1.7). Their nucleotide sequences are almost identical, differing in only a very limited number of exonic positions. Each is composed of six exons (the protein coding regions) and five introns (regions removed at the RNA processing level by splicing). The length of each intron appears to be conserved to 100 % for introns 2 to 5 (Ibbotson *et al.*, 1992), however there are differences in the length of the first intron (see Figure 1.9). Caucasian populations exhibit a longer exon 1 for the LW opsin gene than for the MW gene (Jorgensen *et al.*, 1990).

1.11.2 Classic theory of human colour vision

Long before the tools of modern molecular biology were brought to bear on the questions about the genetics of colour vision, inheritance patterns of colour vision defects indicated that genes on the X-chromosome encoded the LW and MW opsins and an autosomal gene encoded the SW opsin. Two opsin genes on the X-chromosome were anticipated, one for the LW pigment and one for the MW pigment. Contrary to this, quantitative Southern analysis of genomic DNA revealed that the number of opsin genes on the X-chromosome varies in colour normal males (Nathans *et al.*, 1986b).

The model forwarded by Nathans *et al.* (1986a) proposed that colour vision normal males typically have a SW opsin locus on the seventh autosomal pair, a solitary LW gene, and from one to six MW genes, with a mode of two (Figure 1.12). This model is based on results from densitometric analysis of Southern (DNA; Southern *et al.* 1975) hybridisation blots, and restriction digests of opsin genes. This gene arrangement has since been confirmed by Feil *et al.* (1990).

Genomic DNA from colour-blind and colour deficient males was also subjected to Southern analysis and a model for the molecular genetic basis of colour vision defects was formulated (Nathans *et al.*, 1986b). According to this model, 5'MW-

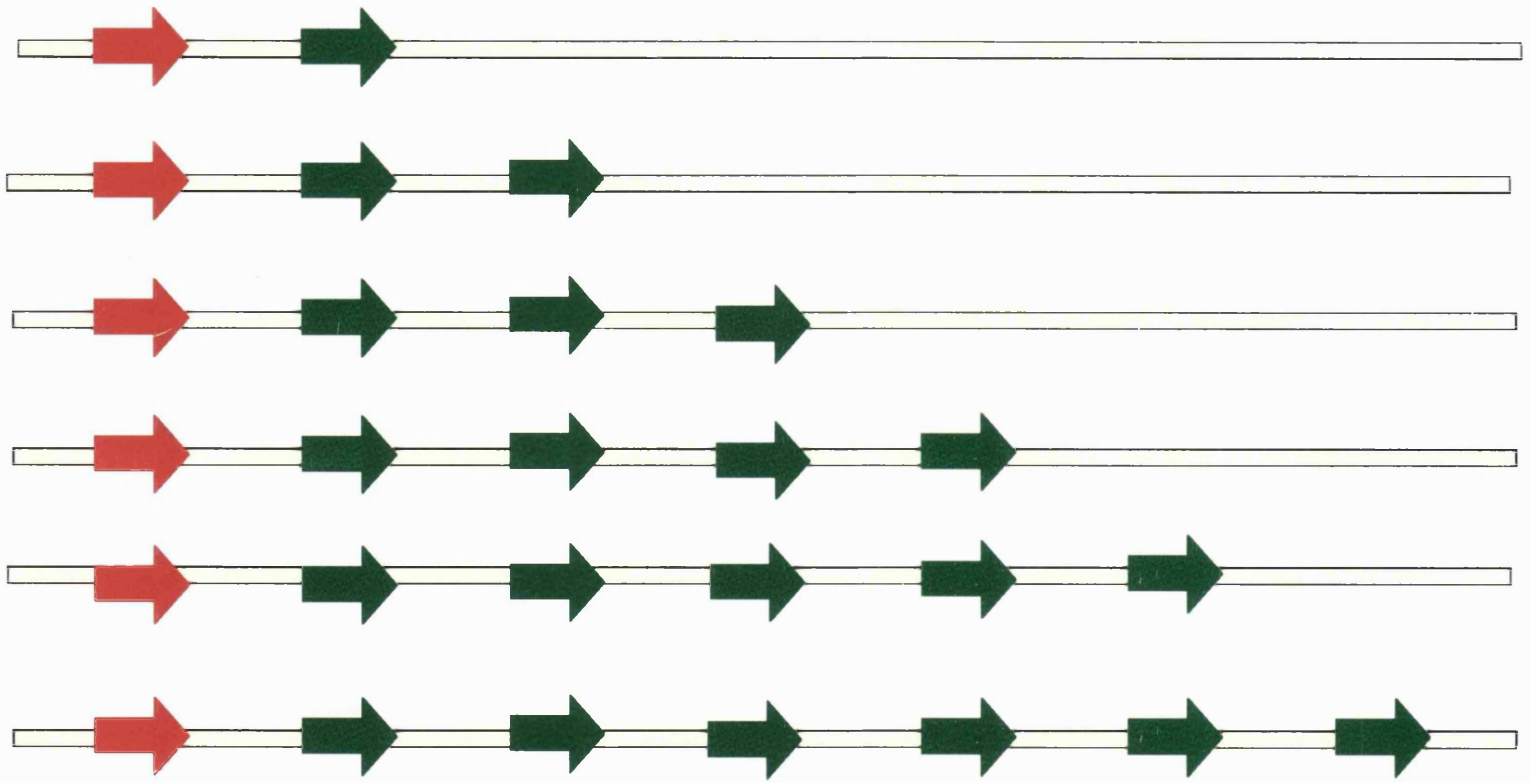


Figure 1.12 Presently accepted model (proposed by Nathans *et al.*, 1986a) of the arrangement of the opsin genes along the X-linked gene arrays of humans. The number of LW (red arrow) genes is limited to one, but the number of MW (green arrow) can vary from to grater than 6.

3'LW hybrid genes, generated by homologous recombination, are proposed to encode abnormal or "anomalous" pigments that are spectrally different from the pigments that underlie normal colour vision (Figure 1.13). The addition of such a hybrid gene to an otherwise normal opsin gene array containing normal MW and LW genes is proposed to cause the colour vision defect (Nathans, 1994).

1.12 Classification of primates

It was the Swedish botanist Karl Linné, better remembered as Linnaeus, who coined the term Primate (from the medieval Latin word meaning "first in rank") in his *Systema Naturae*, (Linnaeus, 1758). Linnaeus recognised that human beings, *Homo sapiens* (Latin meaning "wise man"), share this order with other apes and monkeys based on morphological similarities.

Until recently the order Primates had traditionally been divided into the suborders Prosimii and Anthropoidea, with a separate order, Tarsioidea, for tarsiers, based primarily on anatomical and morphological comparisons (Napier and Napier, 1985). The fossil record is largely lacking during the period of early primate history due in main to the acid soils above which primates lived, the small size of the early primates and possibly the influence of plate tectonics (Martin, 1990). The prosimians and tarsiers constituted the so-called "lower primates", while the remaining "higher primates" had been classified as anthropoids. Anthropoids were divided into platyrrhines (New World monkeys or ceboids) and catarrhines. Catarrhines divided into cercophithecoids (Old World monkeys) and hominoids (apes and humans). Figure 1.14 illustrates the complete classification of living primates as interpreted by Napier and Napier (1985).

Phylogenetic relationships among all primate groups have since been re-examined utilising the power of molecular biology. Comparison of nucleotide sequence data, from regions of identical function across primate species vindicated the rearrangement of the primate order (Figure 1.15; Goodman *et al.*, 1994). An example of a large cross-species data set that has accumulated is the sequences of ε-globin genes

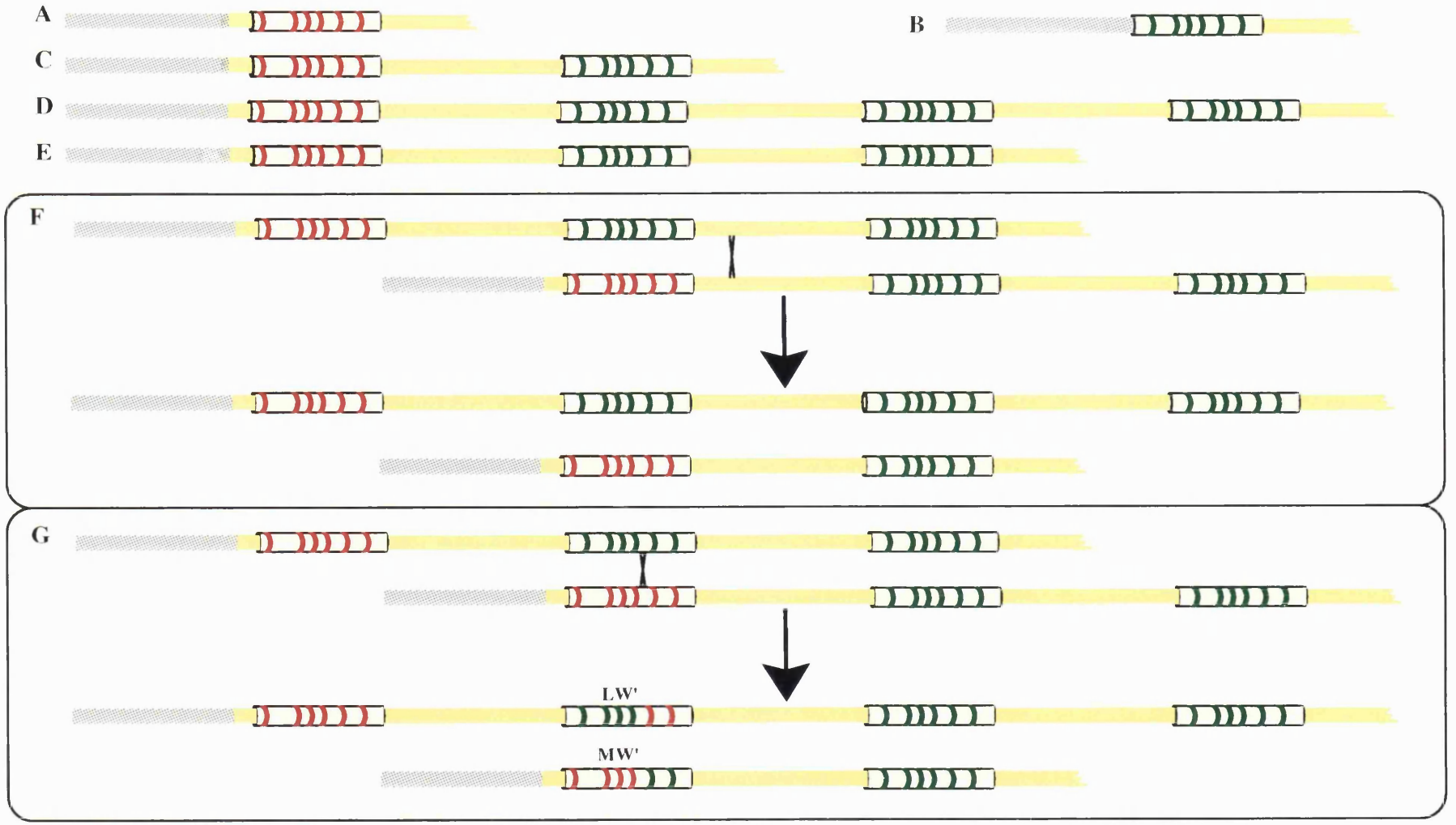


Figure 1.13 Representation of Nathan's hypothesis on the arrangement of the X-chromosome linked opsin gene array, and the probable mechanism by which hybrid genes and changes in the length of the array may occur. Segments **A** to **E** represent the order of different wild-type arrays. **F** shows X chromosomes misaligned at meiosis followed by an intergenic crossing-over event (black cross) that results in the change in the lengths of the arrays. **G** shows a similar situation but with a recombination event occuring between exon 4 and 5. The result is the formation of hybrid genes (LW' and MW'), as well as a change in the array length. Grey lines depict unique 5' flanking regions to the LW opsin, that do not partake in recombination. Bright yellow regions are common intergenic spacers.

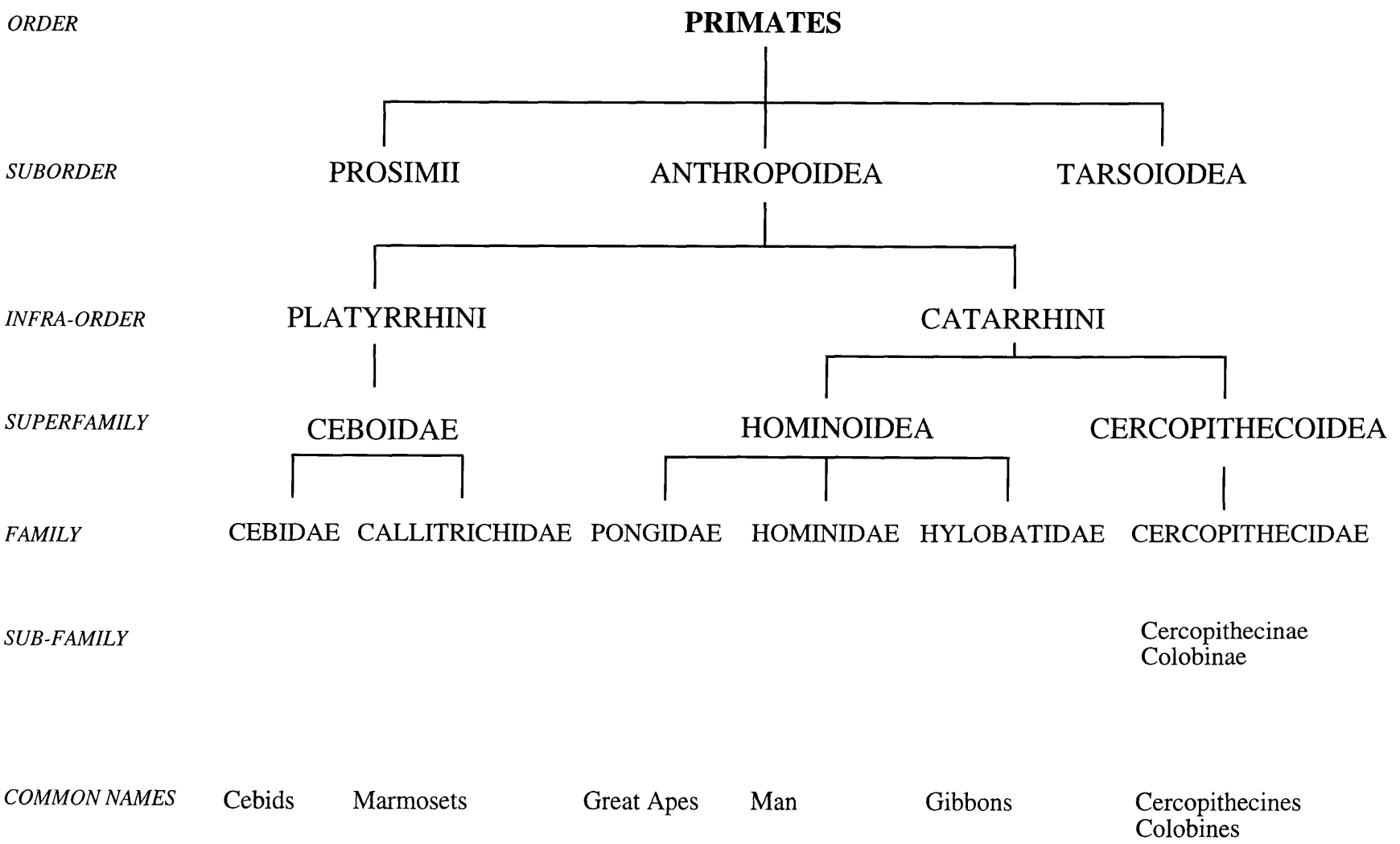


Figure 1. 14 Classification of primates (after Napier and Napier, 1985).

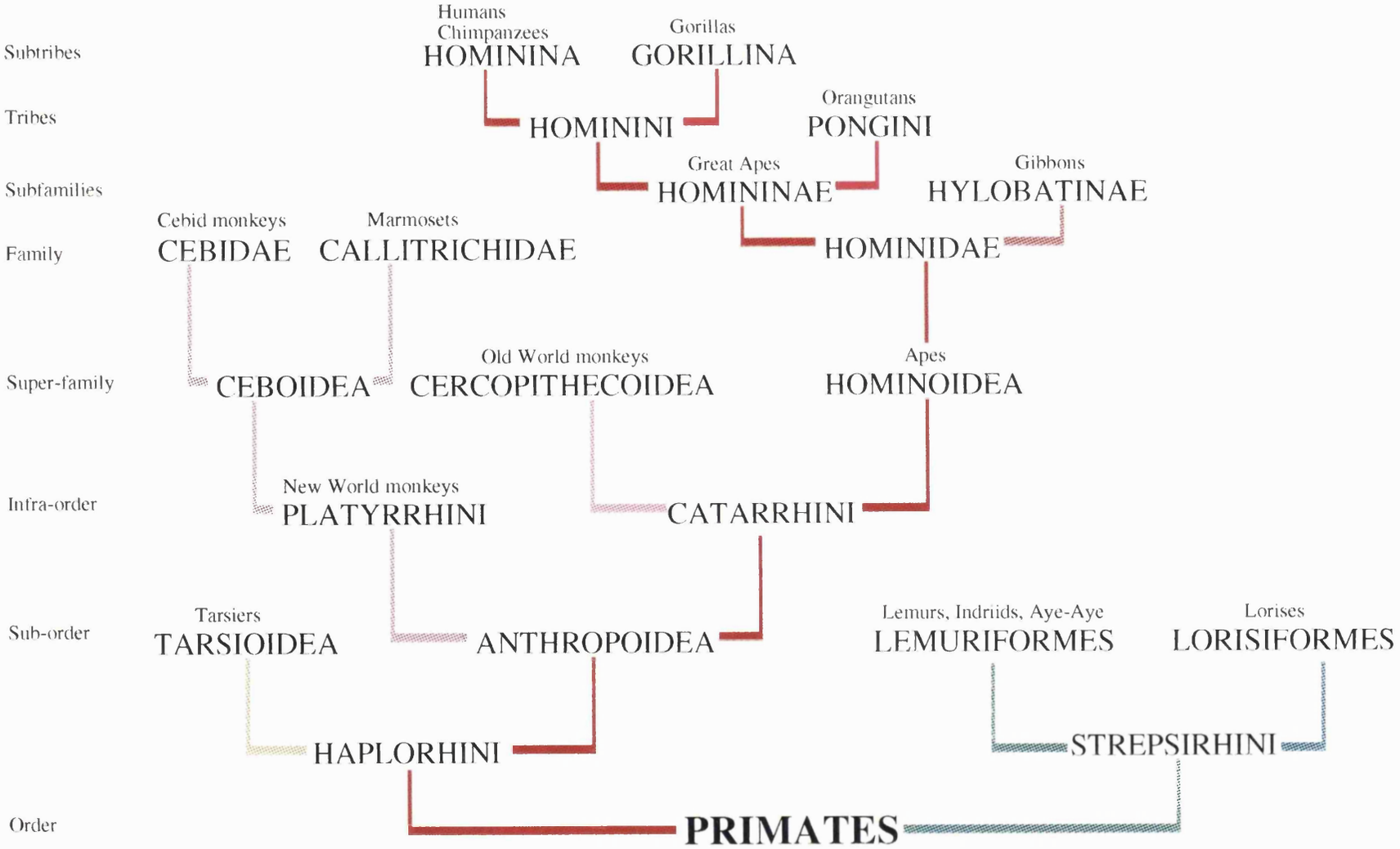


Figure 1.15 Primate phylogeny from molecular evidence based on DNA sequences. Adapted from Goodman *et al.*(1994).

from a range of primates, both higher and lower. This suggested the rearranging of the order to group tarsiers and anthropoids into Haplorhini, and the grouping of lorisoids and lemuroids into Strepsirhini (Porter *et al.*, 1995). Indeed, a time relationship study of primate phylogeny calculated using comparative determinant analysis of 69 proteins (Bauer, 1993) gave ages of the last common ancestors of man and other primates, allowing a more realistic ancestry to be determined (Figure 1.16). The 69 proteins considered in that study constituted 0.1% of the total genetic information (assuming a protein complement of approximately 70,000). From this and other similar investigations there is now general agreement that the last common ancestor of man, chimp and gorilla lived in Africa about eight million years ago (Bauer, 1993; Rogers, 1994), and that chimpanzees are more closely related to man than they are to gorillas (Ruvolo, 1994).

1.13 Colour vision in non-human primates

Evidence has continued to mount showing that there exists a disparity between the colour vision systems of Old World monkeys and that of New World monkeys (Bowmaker *et al.*, 1980; Jacobs and Neitz, 1987; Wikler and Rakic, 1990; Tovée, 1994).*

1.13.1 Colour vision in Catarrhine monkeys

The investigations of Bowmaker (1991) have concluded that both Old World monkeys (Cercopithecoidea) and apes (Hominoidea) possess the same form of colour vision as man. Trichromacy has been adequately demonstrated for a number of catarrhine monkeys (Bowmaker *et al.*, 1991; Ibbotson *et al.*, 1992). In all cases there appears to be at least two loci on the X chromosome plus the SW autosomal locus (Figure 1.17).

---

* For an exception to this rule see chapter 3

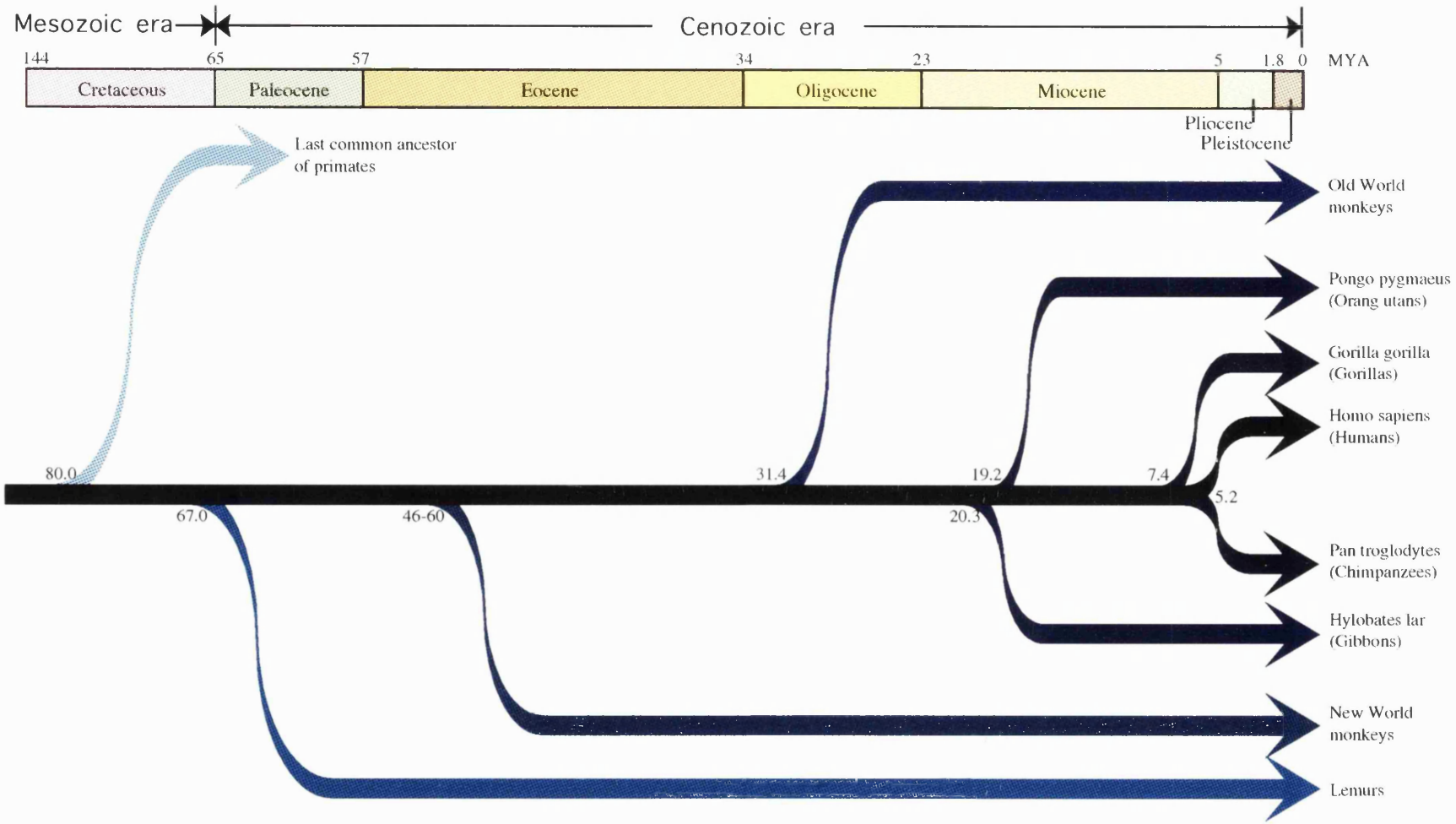


Figure 1.16 Time relations of primate phylogeny. All values stated along the main axis are the ages of the last common ancestor in millions of years before present. (MYA = millions of years ago). Adapted from Bauer (1993).

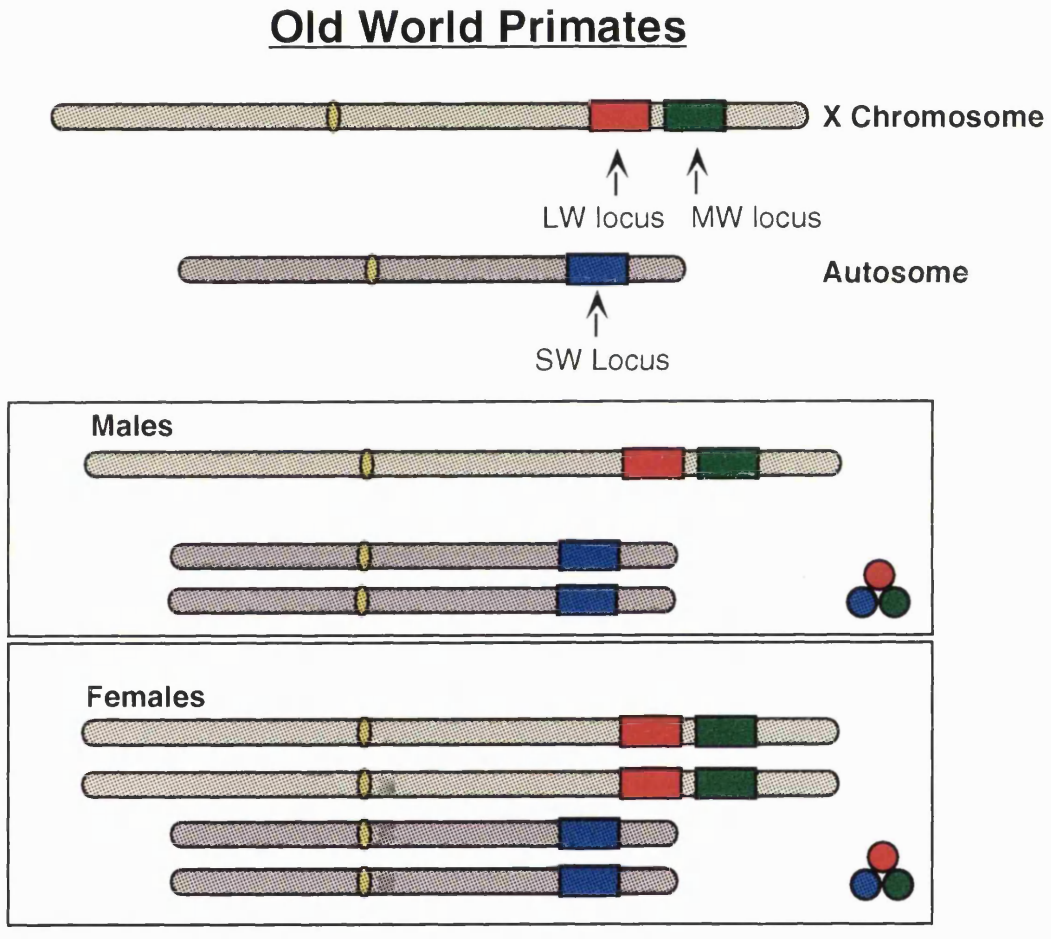


Figure 1.17 Diagramatic representation of chromosomes inherited resulting in normal trichromacy in OW primates. The circles to the right represent the classes of cone photopigments expressed in the whole retina. Note the number of MW loci can vary in the human population from one to six or more (not shown).

1.13.2 Colour vision in Platyrrhine monkeys

The colour vision of New World monkeys was expected to be similar if not identical to that observed in catarrhine monkeys, and it came as a surprise when differences were discovered.

The earliest studies conducted on Cebus monkeys were those of Grether (1939) who found that this species had dichromatic colour vision. Further studies produced variable results (reviewed in Jacobs, 1981). The reasons for these inconsistencies were 1) the relatively few numbers of animals examined, 2) the difficulty of psychophysical testing of non-human primates, and 3) the existence of photopigments with λ_{max} values different from those found in humans.

The present picture of colour vision in New World monkeys arose principally following the work of Jacobs (1977, 1984, 1986; Jacobs and Neitz, 1987) on squirrel monkeys (*Saimiri sciureus*), and subsequently by others on related platyrrhine species (Bowmaker *et al.*, 1987; Tovée *et al.*, 1992; Williams *et al.*, 1992). The current understanding holds that there is probably a single locus on the long arm of the X chromosome of all New World monkeys and that in every species there are three alleles, coding for three spectrally different pigments, able to occupy this locus. Since males have a single X chromosome they would only possess a single class of X-linked cone pigment in the middle- to long-wavelength (M-L) region, together with a SW autosomal pigment (Travis *et al.*, 1988), making every individual a dichromat. Females present with a more complex situation due to the presence of two X chromosomes. Because of the process of Lyonisation (Lyon, 1962), during development one X chromosome in each cone cell has become randomly inactivated. The overall result is that a proportion of cone cells express the gene on one chromosome, while the remainder express the gene on the other. A female inheriting the same allele on each X chromosome is therefore a dichromat, just like the males. However, should the two X chromosomes possess different alleles then two classes of opsin genes will be expressed, resulting in the production of two M-L pigments throughout the retina, and

thus these animals would be phenotypically trichromats. As Figure 1.18 illustrates, three colour classes of males and six classes of females would be present in the population.

1.14 Opsin gene regulation

Regions surrounding the gene, both upstream (5' flanking region) and downstream (3' flanking region), are extremely important in regulating gene expression in many classes of genes. The distance from the gene(s) at which these regulatory sequences are effective is variable: generally there are three classes, the proximal promoter elements, the distal promoter elements (also called enhancers), and the very distant control elements, generally referred to as the locus control regions (LCRs). The proximal promoter elements are generally located within 500 bp or so immediately upstream of the 5' end of most coding genes. They include the well characterised TATA boxes (which bind with TATA box binding proteins (TBP)), Sp1 sites, or G-C boxes (which represent the regions to which Sp1 proteins bind), and CAAT boxes (the sequence to which C/EBP factors bind). These regions promote the accurate initiation of transcription by the ribosomal machinery, by allowing appropriate transcription factors to bind at suitable positions. Experimentation involving alterations (e.g. deletions and/or translocations to expression systems) of these regions confirm their vital roles in gene regulation. There are no hard and fast rules as to which elements will be present at any one gene. However, certain patterns are recognisable for some categories of genes; housekeeping genes (those expressed constitutively in all cells) share many common motifs. Comparisons, involving a large complement of genes, reveal two categories of regulatory elements upstream: those that are found in genes showing distinct patterns of regulation, likely to be involved in basic transcription (housekeeping), and those seen only in genes expressed in a tissue specific, or signal specific, manner. Experiments with opsin genes and associated regulatory regions show that they are of the tissue specific class, since their expression has only been observed in a very limited number of tissues (Lem *et al.*, 1991). The sequencing data

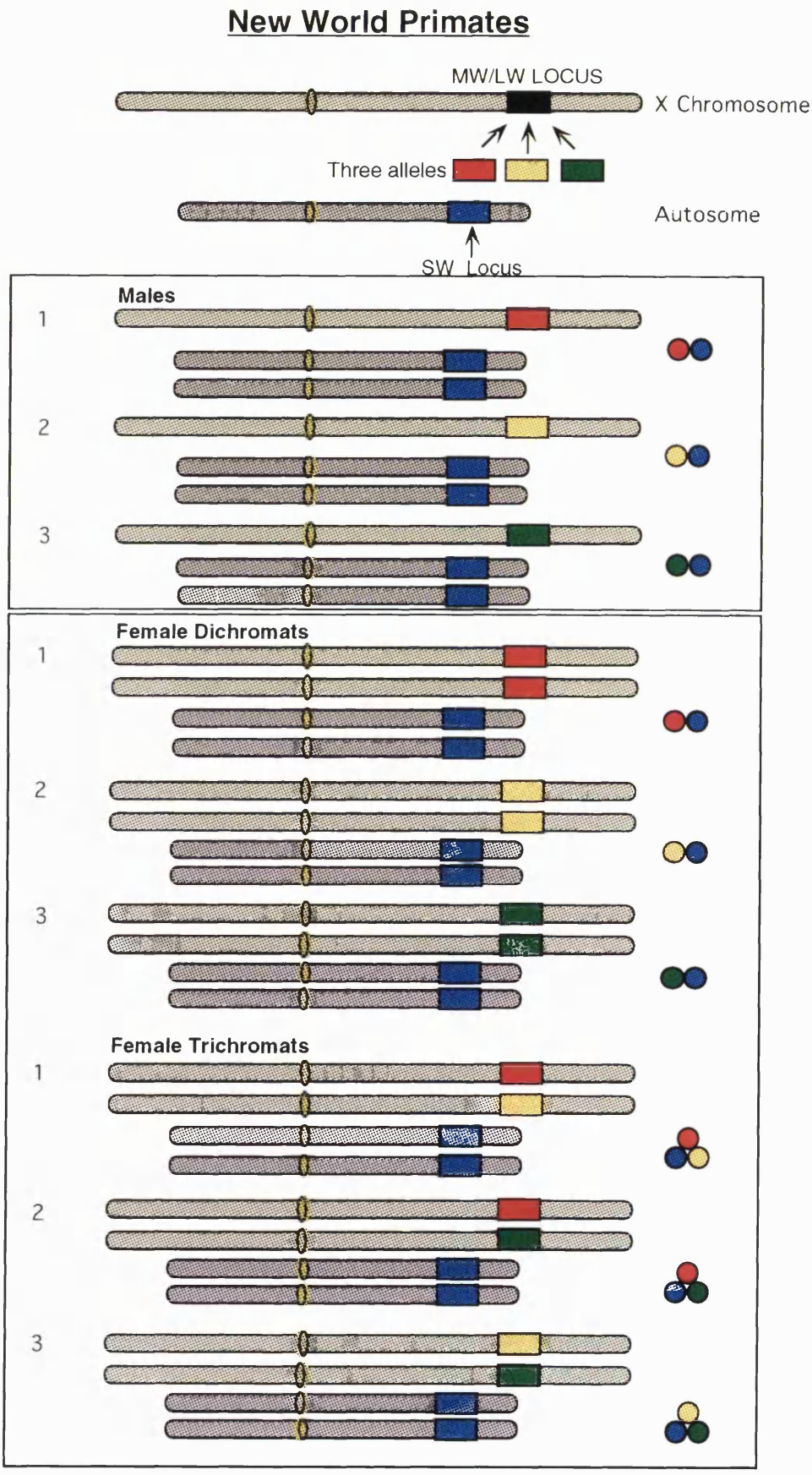


Figure 1.18 Diagramatic representation of chromosomes inherited resulting in dichromacy and trichromacy in NWM. The circles to the right represent cone types expressed in the retina of those animals.

obtained from the MW and LW human opsin clones by Nathans *et al.* (1986a, b) included approximately 450 bp of upstream sequence from each gene. The TATA boxes for both genes were identified.

Wang *et al.* (1993) extended the sequence information to include 6 kb upstream of the LW opsin gene of human, as well as 5 kb upstream of mouse L/MW cone opsin, and two regions upstream of the bovine L/MW cone opsin. By simple sequence alignment they identified a conserved 39 bp stretch, present approximately 2.9 kb upstream of exon 1 of the human LW gene and a comparable distance in the other two species. This they labelled as the LCR core sequence.

In experiments where the LCR core sequence was investigated by placing it immediately upstream of a reporter gene, no transcripts were produced (S. Deeb, personnel communication). This suggested that the relative position of this region to the gene array is important in regulating transcription of the genes, acting as a master switch, due to its influence on gross chromatin structure.

Distal promoter elements are known for a large number of genes (Li and Rosen, 1994). Two functions have been attributed to these sites: tissue specificity and transcription efficiency. Experiments in which these regions are mutated or deleted cause a reduction in the efficiency of transcription (Jones *et al.*, 1988). Chiu and Nathans (1994b) demonstrated that sequences 5.4 kb upstream of the human SW opsin gene, when placed in an appropriate reporter vector, direct expression of the reporter gene only in the SW cones (and a subset of cone bipolar cells) of transgenic mice. A similar set of findings has been reported for the mouse SW cone opsin (Chiu and Nathans, 1994a), suggesting that all the sequence elements necessary for the control of SW cone-specific expression are encoded within the 6.4 kb flanking region. It should be remembered that whereas the human LCR region resides upstream of a opsin gene array that may contain as many as 10 genes, both the mouse and bovine LCR is thought to "overlook" just a single gene. The question then arises as to what function the LCR performs in the latter two species?

1.14.1 Globin gene cluster as a model

The genes of the human α–globin and β–globin loci are expressed in a specific temporal pattern, and in specific haemopoietic tissues during development. The embryonic globins (ζ and ε) are expressed in the yolk sac blood islands until about the fifth week of gestation. At that time, adult α globin (α) and foetal β globin ($^{A}\gamma$ and $^{G}\gamma$) genes begin to be expressed in the liver. The liver is gradually replaced as the major site of haemopoiesis by the spleen and bone marrow, which express the adult globin genes (α, δ and β; Kulozik *et al*, 1988). The switches which ensue are governed by a region upstream of the array, which is analogous to the LCR region identified in the opsin gene array (Tuan *et al*, 1985). The two types of gene array may therefore be regulated in an analogous manner. These regions are discussed fully in chapter 4.

1.14.2 The mouse model

Cone development and topography have been studied in the mouse. Two cone types, as well as rods, have been defined by immunohistochemistry in the mouse retina: one type (hereafter referred to as the M-LW cones) contain visual pigments that react with an antibody, mAb COS-1, and with antibodies raised against human LW and MW visual pigments, and the second type (hereafter referred to as the SW cones) contain a visual pigment that reacts with another antibody, mAb OS-2, and with antibodies raised against the human SW visual pigment (Szél *et al.*, 1992; Wang *et al.*, 1992). The M-LW cones and the SW cones most likely give rise, respectively, to the 510-nm and 360-nm response maxima observed in the mouse electroretinogram under photopic conditions (Jacobs *et al.*, 1991). The mouse retina shows that immunohistochemical staining and morphologic analysis both show that the two cone types are not uniformly distributed throughout the retina (Szél *et al.*, 1992; Wang *et al.*, 1992). Instead, a high concentration of M/LW cones is found in the upper retina and low concentration in the lower retina. The SW cones show a reciprocal pattern, with the result that the overall cone density is approximately uniform. The transition between the upper and lower zones occurs over a narrow equatorial band which is populated by cones which are

labelled by both antibodies (Rohlich *et al.*, 1994). It is apparent that these transitional cones express both classes of opsin, in contract to the generally accepted view of one visual pigment per cone cell (Rohlich *et al.*, 1994). The retina of the rabbit and guinea pig show the same pattern. As yet unidentified factors must be responsible for determining the pattern of cell expression in the retina of these animals. The coexpression of pigments in transitional cones is probably the result of overlapping regulatory factors.

1.14.3 The rat anomaly

Work performed by Szél *et al.* (1994) on the retinae of rodents (rats and gerbils) suggests that, in these species at least, SW photopigments are expressed before MW photopigments. Further, MW cones developed from SW cones, that is, a proportion of SW cones differentiate into MW cones in a temporal manner. This suggests that SW cone development must be the default pathway, with the switch to the MW cones occurring as a result of the action of an unknown factor. Such a temporal switch has not been found in primates.

1.15 Aims of the project

- Determine the sequence and evolutionary relationships of the X-linked opsin genes of apes and OWM.

  Is spectral tuning of other OW primates dependent upon the same amino acid sites as those of man? Do the opsin genes of other primates also exhibit polymorphism? Is it possible to correlate the evolution of the LW and LW opsin genes with the accepted evolutionary emergence of extant simians?

- Determine the regulatory important regions 5' flanking the opsin genes of apes, OWM and NWM.

  In order to elicit the mechanisms responsible for determining the class of opsin gene expressed in any one cone cell, it is necessary to determine which differences exist in the regulatory regions of these genes. Do the proximal promoter regions of the LW and MW opsin gene show differences in sequence, and in the presence of binding sites for known transcription factors? Which 5' flanking region does the corresponding region from the NWM opsin L/MW resemble?

- Investigate the genetic basis for two classes of human red-green colour vision anomalies.

  Is it possible to correlate differences in amino acids at particular sites with the differences in colour perception exhibited by these subjects?

- Examine the genetic basis of the colour vision defect of John Dalton, who died over one hundred and fifty years ago, utilising tissue from his preserved eyes.

  Is it possible to isolate and amplify DNA from tissue 200 years old? Was Dalton an protanope?

CHAPTER 2

Materials & Methods

2. Methods

2.1 Sampling of human subjects

10 ml of venous blood was collected in sealed blood tubes containing sodium EDTA and gently agitated to mix the reagent. The samples were stored at -20°C at the earliest opportunity, awaiting genomic DNA (gDNA) extraction.

Semen samples were provided in screw-top specimen jars. These were frozen at -80°C upon arrival.

2.1.1 Sampling of non-human primates

Tissue samples were stored in a -80°C freezer, after collection for earlier studies (Ibbotson *et al.*, 1992; Williams *et al.*, 1992) Non-human primate DNA samples were kindly furnished by Dr. Helen Stanley, Institute of Zoology, The Zoological Society of London, London, and primate lambda genomic libraries donated by Little John Pharmaceuticals Inc. USA.

2.2 Preparation of nucleic acids from tissue

2.2.1 DNA extraction from whole blood (Nucleon II kit protocol)

Sodium EDTA tubes containing human blood samples (approximately 10 ml) were allowed to thaw at room temperature. 5-10 ml were carefully transferred in to 50 ml Falcon tubes. 40 ml of Reagent A (supplied as part of the Nucleon II DNA Extraction Kit (Scotlab, Scotland)) were added to the Falcon tube prior to rotary mixing for 5 minutes at room temperature. Centrifugation for 10 minutes at 1000 X g was followed by safe disposal of the supernatant. 2 ml of Reagent B was used to resuspend the cell pellet, preceding transfer to a 5 ml screw-capped tube. Addition of 500 µl of

sodium perchlorate (5M) preceded a further rotary mix at room temperature for 15 minutes. The tubes were incubated at 65°C for 25 minutes, with occasional shaking, then placed on ice to cool. 2 ml of chloroform and 300 µl of Nucleon Silica suspension were added, followed by a 10 minute room temperature rotary mix. After a further 5 minute spin at 1400 X g the top layer was carefully aspirated to a fresh Universal tube, to which two volumes of ethanol were added. DNA was precipitated by gentle inversion of the tubes, and subsequently spooled out using a surgical needle, The newly extracted gDNA was re-dissolved in 350 µl of sterile distilled water.

2.2.2 DNA extraction from whole blood (conventional method)

Individual blood samples (approximately 10 ml) were thawed slowly on ice, and their volumes increased to 25 ml with water. Erythrocyte lysis was achieved by incubation with lysis buffer (25 ml) for 30 minutes on ice. Intact lymphocytes were pelleted at 3,500 X g for 15 minutes, and the supernatant reduced to 8 ml by aspiration. A further 42 ml of lysis buffer was added to the supernatant on ice for 10 minutes to ensure complete lysis of the erythrocytes. Intact lymphocytes were again pelleted at 3,500 X g for 15 minutes then resuspended in 3 ml of suspension buffer. Overnight rotary incubation with 10% (w/v) SDS and proteinase K (0.4 mg/ml) at 37 °C was followed by a further 2 hour rotary incubation at 37°C with 0.6 M sodium perchlorate. The suspension was extracted with an equal volume of phenol, phenol:chloroform (1:1, v/v) and twice with chloroform respectively at 5,000 X g. The DNA in the aqueous layer was precipitated with 3M sodium acetate (0.1 volume) and 100 % ethanol (2.5 volumes). The precipitated DNA which formed on tube inversion was removed with a sterile needle and washed in 70% ethanol. The DNA was dissolved in an appropriate volume of TE buffer at 4°C over a period of one to two days. Integrity and yield of gDNA was estimated by gel electrophoresis (Section 2.3.5) in a 1 % agarose gel (FMC, Kodak/IBI).

2.2.3 DNA extraction from solid tissue samples

Frozen tissue samples were recovered from the -80°C freezer and quickly weighed. The following protocol applies to each gram of tissue used. The tissue was ground up using a pestle and mortar whilst immersed in liquid nitrogen. Finely ground tissue was resuspended in 5 ml of NE buffer within an Universal tube. After a spin at 4000 X g for 7 minutes the supernatant was decanted off and a further 5 ml of NE buffer added to the pellet. The process was repeated up to five times or until the supernatant was clear. Following the final spin the pellet was resuspended in 4 ml of NE buffer containing 1.2 mg of Proteinase K and briefly vortexed. 2 ml of 10 % SDS was added before incubation at 50 °C overnight. The next day, the sample was incubated for 20 minutes with an equal volume (6 ml) of equilibrated phenol, then spun at 200 X g for 10 minutes. The upper aqueous layer was removed to a fresh tube. The process was repeated twice with an equal volume of a phenol/chloroform solution, and once with chloroform alone. The DNA in the aqueous layer was precipitated with 5 M NaCl to a final concentration of 0.4 M, and 2 volumes of absolute ethanol, then recovered by either spooling or centrifugation at 3500 X g for 10 minutes. DNA was dissolved in approximately 1 ml of sterile distilled water by keeping at 4 °C overnight. A second precipitation employing 500 µl of 8 M ammonium acetate and six volumes of ethanol was followed by a 70 % ethanol wash. The spooled DNA was transferred to a clean tube where it was left to air dry for 3 hours. Finally, 500 µl of sterile distilled water was used to redissolve the gDNA.

2.2.4 DNA extraction from semen samples

Frozen semen samples were recovered from the -80 °C freezer and allowed to thaw on ice. Each was diluted with an equal volume of NE buffer. For each 5 ml aliquot of diluted semen 500 µg of proteinase K, 50 µl of 10 % SDS, and 50 µl of 60 mg/ml DTT were added. After gentle mixing and overnight incubation at 50 °C, the samples were subsequently processed as per overnight digests of solid tissue samples (section 2.2.3).

2.3 The polymerase chain reaction (PCR)

2.3.1 General considerations

PCR was performed in proprietary buffers and used with the specific *Thermus aquaticus* (*Taq*) polymerase supplied by the manufacturer (NBL, Bioline, Advanced Biotechnologies, and Perkin Elmer). Initially, all 50 µl reactions were carried out with 1-5 units of polymerase per tube. However, experience proved that using 0.05 units or less gave equally good and often better results (i.e. the desired band was more specific and abundant).

Two classes of 10 X buffer were routinely used: initially 100 mM Tris-HCl/ 500 mM KCl based, pH 8.8 at 25 °C (NBL and Bioline) and latterly 750 mM Tris-HCl/ 200 mM $(NH_4)_2SO_4$, pH 9.0 at 25 °C (Advanced Biotechnologies and Pharmacia). These higher pH values are thought to protect strands against depurination and subsequent nicking during thermocycling (Barnes 1994; Cheng *et al.* 1994). Under certain circumstances Perkin Elmer Cetus 10X PCR buffer 1 or buffer system II was used.

Primer design, thermal cycling conditions, dNTP concentration and magnesium concentration were initially based on parameters reviewed by Innis and Gelfand (1990) and Saiki *et al.* (1985). Different sets of primers were individually optimised, based upon their 3' complementarity and primer melting temperatures calculated either by comparing sequences by eye, and using the formula

$$4(G+C)+2(A+T) = \text{dissociation temperature, Td}$$

then reducing the annealing temperature by 2-5 °C (for primers up to 20 bases in length), or more accurately for longer primers by using computer software to test primer pairs (see section 2.8). Where consensus sequence primers were used, the reduction in temperature was initially greater than 2-5 °C.

Cycle length and number varied between 25 and 40 cycles, depending upon the template, particular methodology, and the model of thermocycler used. These will be

noted individually. Once optimised, PCRs were generally performed on the same machine whenever feasible. The following models of thermocycler were available: M J Research; Hybaid Thermal Cycler; Hybaid OmniGene; Perkin Elmer GeneAmp PCR system 2400 and system 9600.

2.3.2 Standard parameters for a typical PCR

Unless otherwise stated, the following constitute a "standard" PCR reaction in a 50 µl volume :-

- 1X concentration of proprietary buffer
- dNTPs at 200 µM each dNTP
- Each primer at 30 pmol
- $MgCl_2$ at 1.5 mM

A typical temperature cycling profile consisted of denaturation at 94°C for 15 seconds, an annealing step for 20 seconds, and extension at 72°C, allowing approximately 30 seconds per kb to be extended. All PCR reactions were begun with an initial 3-5 minute denaturation step.

2.3.2.1 Use of "touchdown" PCR , "hot starts" , dimethyl sulphoxide (DMSO) , and Perfect Match™

Certain modifications of the standard reaction were used in different situations in order to improve the fidelity of the amplification.

"Touchdown" PCR involves setting an initial annealing temperature above the calculated melting temperature of both primers. In each subsequent cycle, the annealing temperature is reduced by 0.5 or 1.0°C until a level is reached which is arbitrarily set slightly below the primer melting temperatures. The remaining cycles are then carried out at this annealing temperature. Theoretically, the first templates to be primed and extended will be those with a perfect match to the primers as the thermal stringency is very high. Over the first few cycles, these products will then form the bulk of the

template for subsequent cycles, so that despite the slightly lower temperature, the kinetics will favour the perfectly matched product.

The technique of "Hot Starting" (Chou *et al.*, 1992) was routinely used. There are a number of ways to achieve this end, utilising wax beads (Perkin Elmer) and, more recently, a neutralising antibody specific to Taq polymerase (Clontech). However, the cheapest and most convenient method used (given that relatively small numbers of tubes were being processed at any one time) was to add the enzyme after the temperature of the cycler had exceeded 80 °C. At this stage any secondary structure in the template or inappropriate primer annealing will have been denatured by the thermal energy. Routinely, the enzyme was added at 94 °C in an extended initial denaturation step with perfectly acceptable results.

DMSO (as well as glycerol) acts to effectively lower the melting and strand separation temperatures (Wong *et al.*, 1991 and Chester and Marshak, 1993). This facilitates the denaturation of the template and primer/template duplexes, thus enhancing the specificity of the reaction. However, since concentrations greater than or equal to 10 % (v/v) can inhibit Taq polymerase activity by 50 % or more (Innis and Gelfand, 1990), DMSO was only used infrequently to eliminate multiple bands. The maximum concentration was 10 %. A beneficial effect was only observed for certain primer/template combinations and was by no means a universal phenomenon.

Perfect Match™ polymerase enhancer (Stratagene, California, USA) is an additive reagent which can increase the specificity of primer extension reactions, e.g. PCR. It has been shown to destabilise many mismatched primer-template complexes that would otherwise result in a heterogeneous pool of amplified molecules. It is especially beneficial where templates larger than 2 kb are used (Mathur, 1991). The use of Perfect Match™ was only included in PCR reactions which gave multiple products on initial attempts. 1 µl was included in the set-up of a 50 µl PCR reaction. However, the protocol was subsequently modified by the supplier requiring the addition of the reagent after the first cycle of PCR.

2.3.3 Visualisation of PCR products: agarose gel electrophoresis

Products of the PCR reactions were routinely size fractionated by electrophoresis through either a 1.8 % (w/v) agarose gel (1% ordinary agarose, 0.8 % low melting temperature (LMT) agarose) in 1 X TAE buffer with Øχ174/*Hae*III as molecular size standards, or a 0.8 % (w/v) agarose gel (ordinary agar only) in 1 X TAE with both Øχ174/*Hae*III and λ/*Hind*III, or a 1 kb ladder marker, as molecular size standards. Bands were stained in ethidium bromide solution (0.5 μg/ml) for 10 minutes and visualised on a UV transilluminator for photography.

2.3.4 Visualisation of PCR products: polyacrylamide gel electrophoresis

In order to accurately resolve small differences in size between fragments of approximately 100 bp in length, non-denaturing 8 % polyacrylamide gels were used. For a 50 ml gel, 16 ml of 25 % acrylamide (19:1 acrylamide:bisacrylamide) concentrate (Sequagel, National Diagnostics) was mixed with 5 ml 10 X TBE buffer and 29 ml of sterile distilled water. Polymerisation was catalysed by the addition of 20 μl TEMED and 600 μl 10% ammonium persulphate. The mixture was quickly poured in to a vertical slab gel unit (Sturdier SE 400, Hoefer Scientific Instruments) and left to polymerise. DNA was loaded with sucrose/orange G loading buffer and run at approximately 100 V overnight. Bands were stained in ethidium bromide solution (0.5 μg/ml) for 10 minutes and visualised on a UV transilluminator for photography.

2.3.5 PCR based techniques

2.3.5.1 Vectorette PCR: a method for genomic walking

This process is also known as "Chemical Genetics" since it does not involve a biological system in which to propagate cloned DNA. Briefly, it involves digestion of gDNA with different restriction endonucleases, followed by ligation of a defined, presynthesized double stranded species of DNA compatible with the ends generated by the cut. This piece of DNA is called a "Vectorette" (ICI Biological Products). The

protocol is summarised in Figure 2.1. Anchored PCR is then performed using a specifically designed sequence-specific primer and a "Vectorette-specific" PCR primer to generate a product which reads in to unknown sequence.

The binding site for the "Vectorette-specific" PCR primer is not actually present initially, since the Vectorette itself contains a "bubble" of mismatched sequence in the middle, which spans the Vectorette PCR primer binding site. The site is only created when the strand complementary to the PCR primer sequence is generated. This strand should arise by DNA synthesis solely from the gene specific primer within the restriction fragment of interest, thus reducing the probability of obtaining unwanted products upon amplification.

2.3.5.2 Construction of Vectorette genomic libraries

2 μg of gDNA was digested overnight with one of the following restriction endonucleases in a volume of 100 μl according to the manufacturer's recommendations:-

- *Bam*HI [1]
- *Hind*III [2]
- *Pvu*II [3]
- *Hae*III [3]
- *Sau*3AI [1]
- *Rsa*I [3]
- *Taq*I [4]

The extent of digestion was monitored by running 0.5 μg of the DNA on a 1% agarose gel (see section 2.4.3). 3 pmol of the appropriate Vectorette ([1] = *Bam*HI compatible, [2] = *Hind*III compatible, [3] = blunt-ended, [4] = *Cla*I compatible,) was ligated to 1 μg of each of the completely digested gDNAs in the presence of the restriction enzyme buffer, 2 mM ATP, 2 mM DTT and 10U of T4 DNA ligase (Pharmacia).

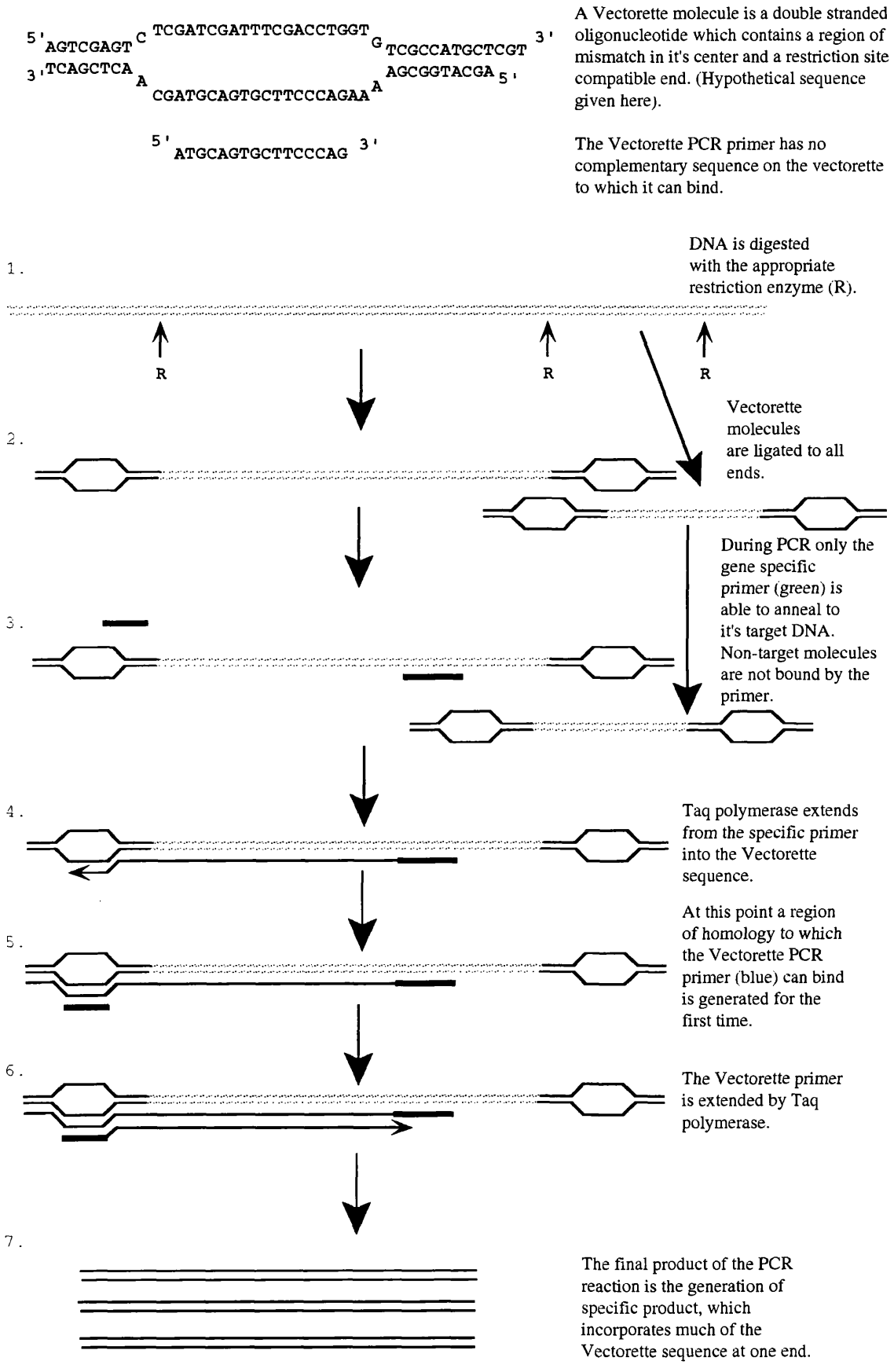


Figure 2.1 Schematic representation of the principle of Vectorette PCR.

Except for the *Sau*3AI digest/*Bam*HI Vectorette ligation reaction, ligation of the Vectorette unit to the end of the gDNA fragment destroys the recognition site. This means that the temperature can be cycled between the temperature optima for the ligase and the restriction endonuclease (but not *Taq*I, whose temperature optimum is at 65 °C) without destroying the ligase activity. This allows ligated gDNA fragments to be recleaved, thus driving the reaction towards attachment of Vectorette units. Therefore, the temperature was cycled between 20 °C, 60 minutes and 37 °C, 30 minutes for 10 cycles. The *Sau*3AI digest was ligated to the *Bam*HI Vectorette at 20 °C for the same duration as the temperature cycling, before denaturing the ligase at 70 °C for 10 minutes.

Following this incubation, 200 µl of sterile distilled water was added to each of the Vectorette genomic libraries, which were then split in to 20 µl aliquots and stored at -20 °C. 2 µl was then used as a template in a first round PCR reaction using gene specific primers in combination with the Vectorette PCR primer.

2.3.5.3 PCR amplification of genomic Vectorette libraries

Two rounds of PCR were found to be required to generate sufficient specific product for further manipulation. The PCRs were standard "hot start" reactions, but significantly different in terms of thermal cycling conditions.

The Vectorette PCR primer supplied has a calculated dissociation temperature of 74 °C using the formula given in section 2.3.1 and even at annealing temperatures as high as 72 °C, gives a ladder of bands formed by spurious annealing. Therefore, the gene specific primer had to be designed with its dissociation temperature very close to 74 °C. This was done using PCR primers portion of the GeneWorks™ software package and is discussed in section 2.8. The PCR reaction thus incorporated a combined annealing/extension step at 72 °C for 2.5 minutes following the 20 second, 95 °C denaturation step.

Additionally, to enhance the specificity of the amplification, the first 5 cycles were carried out solely in the presence of the gene specific primer. The Vectorette PCR primer was then added and a further 35 cycles performed.

A further, semi-nested "hot start" amplification reaction was performed on 1 µl of either

(i) a 1:10,000 dilution of the initial PCR, or

(ii) a LMT agarose purified band of interest (see section 2.4.4.1), to generate quantities of specific product that were easily visible with ethidium bromide staining, and suitable for cloning into a plasmid vector (section 2.4.5).

2.3.5.4 Construction of Vectorette libraries from cosmid clones

Essentially the same as the construction of Vectorette genomic libraries (section 2.3.5.2), except that only a single round of PCR was necessary to produce sufficient product to visualise on a agarose gel and both primers were present at the onset of PCR cycling.

2.3.6 "Walking" PCR

"Walking" PCR was an adaptation of a technique (unpredictably primed PCR) described by Dominguez and Lopez-Larrea (1994). The method employed two successive PCR's, for which four primers were necessary: two sequence specific (SPs) and two universal walking primers (WPs). The specific primers were complementary to known sequence, one being nested to the other. The walking primers were oligonucleotides of defined and artificial sequence, designed following preliminary experimentation.

First round PCR reactions were set up in 50 µl volumes (section 2.3.2) without the specific primer being present. The temperature was varied as follows: 90°C for 60 seconds, 80°C for 30 seconds, 15°C for 2 minutes, 25°C for 10 minutes, 75°C for 30 seconds, and finally held at 90°C whilst the outer specific primer was added to the reaction mix. PCR cycling was then commenced with the following parameters: 30

seconds at 94°C, 30 Seconds at Ta (see below), 30 seconds at 72°C, for 30 cycles. The annealing temperature, Ta, was calculated using the equation:

$$Ta = Tm^{sp} + 4 \sim Tm^{wp}$$

Where the Tm was calculated from the following equation:

$$Tm = 62 + 24\ [(nG + nC)/1] - (360/1)$$

Generally, the Ta was around 52- 56°C. The products of the first round of PCR were diluted 1/1000 in distilled water and one µl of this was included in a second round of PCR utilising both nested specific and walking primers. Any bands visualised upon gel electrophoresis were excised, cloned, and sequenced (sections 2.4.4.1, 2.4.5. and 2.7).

2.4 Manipulation of DNA

2.4.1 Purification of PCR products prior to restriction enzyme digestion

Prior to restriction endonuclease digestion, PCR reactions (100 µl volume) were checked by running 1/10th volume on an agarose mini-gel (See section 2.3.3). Following this, the aqueous layer was separated from the oil overlay by freezing the reactions at -20 °C and pipetting off the oil whilst the aqueous layer was still solid. The volume was adjusted to 100 µl and the PCR product precipitated with 0.1 volumes 3M sodium acetate and 300 µl absolute ethanol.

After 15 minutes at room temperature, the DNA was pelleted at 13,000 rpm for 10 minutes in a benchtop microcentrifuge. The pellet was washed with 70 % ethanol, vacuum dried for 10 minutes and resuspended in 10 µl distilled water. 5 µl of this was used in a digest.

2.4.2 DNA digestion with restriction endonucleases

Genomic DNA, cloned DNA and PCR products were digested with a number of restriction endonucleases according to manufacturers instructions (BRL, NBL, New

England Biolabs, Promega, and Pharmacia). Incubation was generally carried out for a minimum of 16 hours for gDNA but shorter times were used when screening recombinant plasmids and PCR products (See section 2.4.5.5). Digestion reactions were checked for completion on agarose gels prior to Southern blot preparation. Reactions in high salt buffers (e.g. *Eco*R1 digestion's) were often subjected to the addition of spermidine (an enzyme concentrator) to a 5 mM concentration to avoid partial digestion. Once the digestion was complete, reactions were stopped by the addition of EDTA in the loading buffer, or by heating to 85 °C for several minutes.

2.4.3 Electrophoresis of DNA

The method is modified from Sambrook *et al,* 1989. Digested gDNA was separated for analysis by agarose gel electrophoresis. To achieve separation in the size range 0.5 Kb-20 Kb, 0.8 %-1.0 %(w/v) agarose gels (20 cm X 20 cm for Southern blots; smaller gels for plasmid or cosmid digests) were made in 1 X TAE buffer. Digested gDNA (10-15 µg) was loaded in 1 X loading buffer and electrophoresed at 30 V for 16 hours (overnight) in 1 X TAE buffer. Following electrophoresis, gels were stained in ethidium bromide solution (0.5 µg/ml) for 10 minutes and placed on a UV transilluminator for photography.

2.4.4 Preparation of PCR products for cloning and sequencing

Several methods were used for these purposes, the simplest of which was dilution of a very clean PCR so that the ratio of molar ends of insert: vector was approximately 3:1. If a reaction product required separation from excess primer-dimer, phenol: chloroform (1:1)/chloroform extraction and S-400 micro-spin columns (Pharmacia) were used according to the manufacturer's protocol prior to dilution.

2.4.4.1 Elution of target fragments from gels

Following visualisation by electrophoresis the bands of interest were excised from the gel under long-wavelength UV light, with the aid of scalpel blades. The bands were suspended in a small volume of sterile distilled water (40 µl) and left overnight at

room temperature. The DNA diffused out of the gel into the water. A 10 µl aliquot was run out on a 1.8 % agarose gel to both confirm and quantify the eluate.

2.4.4.2 NA-45 paper purification of PCR products

The PCR reaction was run on a 0.8% agarose gel at 100V (section 3.3), then stained with ethidium bromide. Under UV illumination, a cut was made with a scalpel blade just ahead of the band. In to this slit was inserted a pre-cut piece of DEAE paper (NA-45, Schleicher and Schuell) and the gel re-run at 100V for 5 minutes to transfer the DNA to the paper. The NA-45 paper was washed in 0.5 ml of 0.15M NaCl, 0.1 mM EDTA, 20 mM Tris-HCl pH7.5 at room temperature. DNA was eluted from the paper by 2 washes in 1.0M NaCl, 0.1 mM EDTA, 20 mM Tris-HCl at 65 °C for 30 minutes each. The DNA was then precipitated by the addition of 2 volumes of absolute ethanol and incubation at -20 °C for 30 minutes. After pelleting at 13,000 rpm for 10 minutes at 4 °C, and a 70 % ethanol wash, the pellet was vacuum dried for 10 minutes and resuspended in 10 µl of distilled water. This method gives much greater specific purification than spin columns.

2.4.4.3 3MM paper purification of PCR products

As with the DEAE paper method above, PCR products were size fractionated through 0.8 % agarose gels. After ethidium bromide staining and washing in water, the bands in the gel were visualised using a low-power short-wave hand-held UV source (UVP inc., Ca.), so as to minimise DNA damage. Bands of interest were cut from the gel using a scalpel blade, and placed on to small pieces of folded 3MM paper. The paper was cradled in a 1.5 ml microcentrifuge tube which had its bottom sliced off. This tube was placed within another complete microcentrifuge tube. The two tubes were spun at top speed in a Soreaus benchtop centrifuge for 30 seconds. The DNA and water were forced through the paper in to the bottom tube while the gel matrix was trapped on the paper. An aliquot was quantified by means of another gel electrophoresis. The recovered DNA was pure enough for cloning or digestion.

2.4.4.4 Oligonucleotide probing of PCR products

PCR products were separated by 1.5% agarose gel electrophoresis and transferred to a Hybond-N membrane as described in section 2.4.7.2. The oligonucleotide probe was prepared by end-labelling with [γ-^{32}P] dATP, specific activity 3000 Ci/mmol as described in section 2.4.8.2. The filter was prehybridised according to Church and Gilbert (1984) as in section 2.4.8.5 at 55-62°C. The temperature was chosen to be 2-3°C below the calculated dissociation temperature of the oligonucleotide, with an extra arbitrary allowance for mismatches between primer and template. The labelled probe was added directly to the prehybridisation solution and incubated for 2 hours, after which the filter was washed in 2 X SSC, 0.1 % SDS at room temperature and then at 55-62°C for 20 minutes. Further washes at 55-62°C for 5 minutes, down to as low as 0.5 X SSC, 0.1 % SDS were sufficient to reduce background counts. Autoradiography was carried out using Kodak X-OMAT AR film.

2.4.5 Ligation of PCR products to plasmid vectors

2.4.5.1 Vectors and *E. coli* strains used

Plasmid vectors for direct cloning of PCR products were supplied by Invitrogen (pCR™1000, 3 kb; pCR™ Vector, 3 kb; pCR™II, 3.9 kb), Novagen (pT7Blue T-vector, 2.9 kb), and Promega (pGEM-T, 3.0 kb). They are all pUC-derived and contain dT overhangs at the cloning site in order to generate a sticky end which is compatible with the terminal dA overhangs produced on any PCR product due to the template-independent action of certain Taq polymerases.

Each vector carries one or more antibiotic resistance genes enabling the host bacteria to grow in the presence of certain antibiotics, thereby allowing selection of transformed cells. All the pCR vectors encoded the Kanamycin phosphotransferase gene permitting growth in the presence of Kanamycin. All vectors, except pCR™1000, carry the ß-lactamase gene for Ampicillin resistance. SP6 and T7 promoters flank the cloning site in opposite orientations, allowing SP6 and T7 sequencing primers to be

used on all vectors to sequence in to the inserts. The *lacZ* sequence present allows for blue/white selection of recombinants (see section 2.4.5.4).

In all cases, these plasmids were transformed in to *E. coli* DH5α host competent cells (Gibco BRL or Invitrogen), genotype *supE44*, Δ*lac*U169, (Ø80*lacZ*ΔM15), *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1 permitting α-complementation with pUC based plasmids containing the N-terminal portion of ß-galactosidase and allowing blue/white colour selection (section 2.4.5.4).

2.4.5.2 Ligation dependent cloning

Ligase mediated cloning was performed for the majority of PCR products using reagents supplied by Invitrogen (TA cloning kit), Novagen (pT7 Blue Kit), Promega (pGEM-T Vector System), and R & D Systems (The LigATor kit) according to the recommended protocols.

In each case 50 ng of vector was ligated to an appropriate amount of purified PCR product such that the ratio of molar ends of vector to PCR product was 1:1 to 1:3 (See Innis, 1990). Ligations were allowed to proceed for 2 hours to overnight at the recommended temperatures of 12 °C (Invitrogen), 16 °C (Novagen, and R & D systems), and 15 °C (Promega). Reactions were stored at -20 °C following withdrawal of an aliquot for transformation of competent *E. coli* (section below).

2.4.5.3 Transformation of competent *E. coli*

DH5α competent cells were supplied by Invitrogen or Gibco BRL or made in-house, and were transformed exactly according to the manufacturer's recommended protocol with satisfactorily reproducible results. The transformation itself was tested by the inclusion of a closed circular plasmid control and a medium-only negative control. Aliquots of the recovered transformed *E. coli* were plated out and grown at 37 °C overnight.

2.4.5.4 Growth media and selection of recombinants

LB broth plus agar (bacteriological, 15g/l), or MacConkey Agar (see section 2.9.3) plates were used for the growth of all bacterial strains.

Antibiotic selection of transformants was achieved with 50 μg/ml Ampicillin (pCR™ Vector, pCR™II, pGEM-T), 50 μg/ml Kanamycin (pCR1000, pCR™II), or a combination of 50 μg/ml of both (pCR™II). 50 μg/ml Ampicillin and 15 μg/ml Tetracycline were required with pT7Blue T-vector and the pTAg vector. In general, very few or no satellite colonies were observed when Kanamycin was used whether singly, or in combination with Ampicillin.

Blue/white colour selection of transformants/recombinants by the process of α-complementation was achieved by spreading 20 μl of X-gal (40 mg/ml in dimethylformamide) on to the surface of LB agar plates and allowing the solvent to evaporate about an hour before plating out the transformation reaction. IPTG was not required since DH5α cells do not contain a functional *lac*I gene and therefore the *lac* operon does not need to be induced.

Occasionally, growth on MacConkey Agar plates was performed. Selection was achieved by differentiating lightly pigmented colonies, possessing inserts, from darkly pigmented ones.

2.4.5.5 Screening recombinant *E. coli* for inserts

Two methods were used to screen white colonies from the selection plates, (i) restriction endonuclease digestion and (ii) colony PCR.

(i) 2-5μl of plasmid DNA from each clone, depending on the mini-prep yield, was digested with a restriction endonuclease or combination thereof, designed to release the insert from the multiple cloning site. Conditions were according to manufacturer's recommendations (see section 3.3). Reactions were fractionated and visualised on agarose gels (section 2.3.3). DNA from those clones with the desired insert size were then sequenced (see section 2.6).

(ii) In order to screen recombinant *E.coli* for vectors containing a desired insert, single colonies were sampled using a sterile toothpick in to 10µl of distilled water in a 0.5 ml Eppendorf tube. The rest of the colony was subcultured on to a numbered reference plate or inoculated in to LB broth with the appropriate antibiotic. The inoculated distilled water was heated to 95 °C for 5 minutes in a heat block to lyse the cells and spun down at 13,000 X g in a benchtop microcentrifuge for 1 minute to clear any cellular debris. 1 µl of the supernatant was used in a standard PCR. Positive clones were selected for sequencing (section 2.6).

2.4.6 Plasmid and cosmid preparation

2.4.6.1 Plasmid and cosmid mini-preps

5 ml of overnight broth culture was pelleted in a 50 ml Falcon tube at 4,000 rpm for 15 minutes in an MSE Mistral 3000i using a swing-out rotor. The supernatant was discarded and the pellet subjected to a method which is essentially a scaled down version of the midi-prep method (section 2.4.6.2).

The procedure for plasmids was performed initially with user-prepared solutions and latterly with the Magic/Wizard Miniprep kits (Promega). The kit was used exactly according to the manufacturer's instructions. Distilled water was heated to 80 °C before elution of DNA from the spin columns provided.

The non-kit protocol comprised the resuspension, lysis and macromolecule removal steps using solutions I, II and III as for the midi-prep. At this point, the pellet was spun down at 13,000 rpm in a benchtop microcentrifuge for 5 minutes, the supernatant was removed and subjected to a phenol: chloroform (1:1 v/v) extraction followed by an sodium acetate/isopropanol precipitation (section 2.4.6.2). The pellet was washed in 70% ethanol before being dried and resuspended in 25µl TE or distilled water.

The plasmid and cosmid yields were quantitated by agarose gel electrophoresis as in section 2.3.3.

2.4.6.2 Plasmid midi-preps

One technique for preparing plasmid DNA for digestion and sequencing reactions was a mid-range (midi) preparation technique modified from an alkaline-SDS method described by Birnboim and Doly (1979).

A 50 ml overnight culture was centrifuged at 3,000 X g for 10 minutes, and the resultant cell pellet was resuspended in Solution I (4 ml). Chromosomal and other macromolecular components were denatured on ice with 8 ml of Solution II for 5 minutes, and precipitated on ice with high salt Solution III (4 ml) for 15 minutes. The sample was then centrifuged at 3,000 X g for 10 minutes, and the supernatant collected. Plasmid DNA was precipitated from the supernatant with 0.6 volumes of isopropanol and pelleted at 3,000 X g for 10 minutes. The supernatant was removed and the pellet was dissolved in 4 ml of TE buffer and extracted with an equal volume of phenol. DNA was then precipitated with 0.1 volumes of sodium acetate and 2 volumes of absolute alcohol.

The DNA was pelleted at 10,000 X g and resuspended in 400 µl of TE buffer. At this stage contaminating RNA and protein was digested, firstly by incubation with RNAse A (0.5 mg/ml) and 3M sodium acetate (15 µl) at 37 °C for 30 minutes, followed by incubation with proteinase K (100 µg/µl) again at 37 °C for 30 minutes. Following phenol/chloroform purification (section 2.6.3.1), the DNA was pelleted at 14,000 X g for 5 minutes, and washed in 70 % alcohol before air drying and resuspension in an appropriate volume of TE buffer.

Yield was estimated by agarose gel electrophoresis of 2 µl of the preparation along with known quantities of molecular weight standards (λ / *Hind*III).

2.4.6.3 Geneclean II kit

The Geneclean II kit (Bio101, La Jolla, California) was employed essentially as per instructions provided by the supplier. The kit was used to purify DNA from a heterogeneous mixture of proteins, RNA, and other organic materials, usually following plasmid preps (section 2.4.6.1).

2.4.7 Southern blotting

2.4.7.1 DNA denaturation

Electrophoresed DNA was denatured before blotting by immersing the gels in denaturer (see appendix B) for 30 minutes. After rinsing in distilled water the gels were immersed in neutraliser. The pH of the gels was monitored using litmus paper, but two incubation periods each of 15 minutes with a fresh buffer exchange were routinely used. Gels for transfer on to Hybond-N or Hybond-N^+ (Amersham) were then rinsed in 20 X SSC and assembled for transfer using the same buffer.

2.4.7.2 Capillary transfer

Capillary transfer of DNA from an agarose gel to a membrane was first described by Southern (1975). Although based on the same principles, the techniques described below have since been developed to be membrane specific.

Equilibrated gels were placed on a wick of 3MM Whatman paper which was assembled over a reservoir of 20 X SSC. Membranes were cut to exactly the same size as the gel and placed on top. This assembly was then covered with a further two pieces of 3MM Whatman paper soaked in transfer buffer. Care was taken throughout not to introduce bubbles between the various layers. A stack of paper towels were then placed on top to a height of several inches and a uniform weight (500-750 g) was added. Transfer was accomplished overnight, after which the membrane was marked by clipping the top right-hand corner, for orientation purposes. The DNA was fixed to the membrane according to manufacturers instructions by either baking at 80 °C for 2 hours or by UV cross-linking after drying at 80 °C for 15 minutes with an Amersham UV Crosslinker set for 70,000 $\mu J/cm^2$.

2.4.7.3 Alkali transfer

Certain gels were placed without prior treatment on to the Southern transfer apparatus as in 2.4.7.2 but using 0.4 M sodium hydroxide as the transfer agent. The only requirement of this set-up was the need to use Hybond N+ membranes. After

transfer the membrane was washed in 2 X SSC. UV crosslinking with an Amersham UV Crosslinker set for 70,000 μJ/cm^2 was preceded by drying at 80 °C for 15 minutes.

2.4.8 Probe preparation

For Southern blots or library screening, probes used were generated by PCR from gDNA templates or plasmid clones. The desired band was then purified before labelling.

2.4.8.1 Purification of probe template using an LMT agarose gel

The method is modified from Sambrook *et al.* (1989). PCR-generated DNA fragments were electrophoresed through a 1.2% low melting temperature agarose gel (NuSieve; FMC). After staining with ethidium bromide, the insert DNA was visualised under UV light and carefully cut out of the gel in a minimal volume of agarose. This procedure was carried out as rapidly as possible to avoid overexposure of the DNA to UV light. A volume of TE buffer or distilled water was added on the basis of 3 μl/mg of agarose slice and the agarose DNA plug was then incubated at 65 °C for 10 minutes to melt the gel. In order to estimate the concentration of the DNA, aliquots were then electrophoresed in a 1.2% agarose gel alongside known quantities of Øχ174/*Hae*III molecular standards. The stock of template for labelling was then split in to 30-50 μl aliquots and stored at -20 °C.

2.4.8.2 Radioisotopic labelling of probes

The method is based on the "random priming" procedure developed by Feinberg and Vogelstein (1983, 1984) where random hexanucleotides are annealed to denatured template strands and [α-^{32}P]dCTP is incorporated in to the new complementary strands by the action of the Klenow Fragment of *E. coli* DNA polymerase I. The reagents were supplied in kits by Pharmacia, either in aqueous form or in a vitrified state (Ready To Go DNA labelling kit). Briefly, 30-100 ng of DNA template was made up to 25 μl

with distilled water before being heated to 95 °C for 5 minutes, then placed on ice for several minutes. The template was further mixed with 9 µl of reaction buffer containing the random hexanucleotides, 5-10 units of Klenow Fragment, 40-50 µCi of [α-^{32}P]dCTP of specific activity 6000 Ci/mmol to a total volume of 50 µl (for aqueous kit), or by the addition of just the radioisotope (vitrified kit), and incubated at 37 °C for between 15 minutes and an hour.

2.4.8.3 Radioisotopic labelling of λ/ *Hind*III markers

This procedure incorporates [α-^{32}P]dATP (specific activity 3000 Ci/mmol) into fragments of the digest since they all have a 3' overhang which can be filled in using Klenow DNA polymerase. The reaction was performed at room temperature for 10 minutes in *Hind*III buffer with 2 µCi [α-^{32}P]dATP, 1 µg of λ *Hind*III and 2U Klenow DNA polymerase (Pharmacia). A 5 µl aliquot was removed and the remainder stored at -20 °C. Prior to agarose gel electrophoresis the aliquot was denatured in 0.1 M sodium hydroxide for 15 minutes and mixed with 0.1 volumes of 10 X loading buffer.

2.4.8.4 Estimation of incorporation of radioisotope into the synthesised probe

After incubation, 1 µl of the labelling reaction was carefully spotted on to the centre of a glass microfibre filter (Whatman GF/B) and a hand held monitor clamped above it such that the meter was reading approximately 100 counts per second. The filter was then washed with 10 ml of ice cold 5% (v/v) TCA which was drawn through the filter by a vacuum. The filter was then replaced in the same position underneath the monitor and the new meter reading recorded. The specific activity (SA) of the probe could then be estimated, using the following formula :-,

$$SA = \frac{(\mu Ci)(2.2 \times 10^9)(F)}{Di + \left[(4 \times 325)(F)\left(\mu Ci/S\right)\right]}$$

Where μCi = µCi of dCTP in reaction

F = Fraction of input label incorporated in to DNA

Di = Mass of input DNA template (ng)

S = Specific activity of dCTP (Ci/mmol)

325 = Average molar mass of a deoxyribonucleotide

Thus, for 50 ng of template at 50 % incorporation of 50 µCi of dCTP, SA is approximately $1x10^9$ dpm/µg of DNA, whilst for 100 ng, it drops to $5.2x10^8$ dpm/µg. All probes used exceeded 50 % incorporation.

2.4.8.5 Purification of labelled probe

Pharmacia NICK translation columns or S400 micro-spin columns were used exactly according to the manufacturer's protocols to separate the desired probe from the unincorporated nucleotides and hexamers.

2.4.8.6 Nucleic acid hybridisation to Southern blots

The method is modified from Sambrook *et al.* (1989). Hybond-N and -N^+ filters were prehybridised for at least 10 minutes at the hybridisation temperature in 30 ml of a solution containing either, 10% dextran sulphate, 6X SSC, 0.5% (w/v) SDS, 125µg/ml denatured herring sperm DNA and 5X Denhardt's reagent, or Church's solution (50 ml of 0.5M sodium phosphate buffer pH7.2, 7% SDS, 1 mM EDTA (Church and Gilbert, 1984)). The volume was reduced to 20 ml and 50-100 ng of heat denatured oligo-labelled PCR product (or 2-4 ng of oligo-labelled marker (e.g. 1 kb marker ladder)) was added and carefully mixed. Hybridisation was carried out overnight at the hybridisation temperature (generally between 58 °C to 65 °C).

Filters were then rinsed twice in 100 ml 2X SSC immediately following hybridisation. Washes consisted of a single wash at room temperature with 2 X SSC/0.1% SDS, followed by two 20 minute washes at the hybridisation

temperature with 1 X SSC/0.1 % SDS. Stringency was further increased by two 15 minute washes in 0.5 X SSC/0.1 % SDS. Following exposure to XAR-5 film (Kodak), two similar washes with 0.25 X SSC/0.1 % SDS were carried out if required.

All filters were kept moist by sealing them in plastic film and not allowed to dry, which facilitated their stripping for reuse (section 2.4.8.7) if the need arose.

2.4.8.7 Membrane stripping protocols

For Hybond-N and -N^+ membranes several protocols were adopted. i) Filters were washed in 200 ml of 0.4M sodium hydroxide solution at 50 °C. After 30 minutes of constant shaking, the solution was replaced with one containing 0.2M Tris (pH 7.5), 0.1 X SSC, and 0.01 % (w/v) SDS, and a further wash at 50 °C for one hour was performed. ii) pouring boiling 0.5 % (w/v) SDS on to the filter and allowing it to cool to room temperature and repeated several times. iii) Filters were immersed in boiling distilled water for three minutes.

Satisfactory removal of the probe was confirmed by monitoring or autoradiography for an appropriate time period.

2.5 Genomic cosmid and bacteriophage libraries

A human cosmid library (Cachon-Gonzalez, 1991) consisting of approximately $5x10^5$ independent recombinants was obtained from Dr. J. Fitzgibbon, Dept. of Molecular Genetics, Institute of Ophthalmology, London. The Lorist B human cosmid genomic library was generated following a partial *Mbo*I restriction digest of genomic human DNA, as given in Cachon-Gonzalez (1991), with an estimated titre $1.0x10^9$ cfu/ml.

Non-human primate lambda libraries were kindly donated by Upjohn Pharmaceuticals Co., USA.

2.5.1 Screening a human cosmid genomic library

The following protocol was employed to isolate cosmid clones containing the 5' flanking regions of the X-chromosome opsin genes.

2.5.1.1 Plating of a human cosmid genomic library master filter

The human cosmid genomic library was plated out on to a 20 cm x 20 cm charged nylon filter (Hybond-N^+, Amersham) in the following manner.

Megaplates of LB agar were poured, containing 50 µg/ml of Kanamycin, dried, and a filter placed on the surface. A glycerol stock of the library stored at -80 °C was thawed on ice, diluted in 2.5 ml LB/Kanamycin broth and mixed by gentle inversion. Five aliquots, each of 500 µl, were evenly distributed over the surface and gently spread in a single direction at each stroke using a sterile glass spreader, to cover the majority of the filter area. The filters were incubated at 37 °C overnight until single colonies reached approximately 1-2 mm diameter.

2.5.1.2 Making of replica filters

Two replica filters were generated from the master. Megaplates were poured (LB/Kanamycin agar) and set, and Hybond-N^+ filters pre-wetted on their surface. The master filter was removed from its agar bed and placed colony side up on a large sheet of Whatman 3MM paper which had been exposed to UV light for 1 hour in a Microflow microbiological safety cabinet to reduce the possibility of contamination. Ethanol soaked and flamed forceps were used for all manipulations and gloves were worn at all times. The pre-wetted replica filter was carefully placed on top of the master, taking care not to introduce bubbles or ridges between the sheets. A second sheet of UV irradiated 3MM paper was placed over the filters and a glass plate used to exert even manual pressure over the whole area of the filter when transferring colonies. Asymmetric needle holes were then created to orientate the filters, which were gently separated.

The first replica was replaced on its agar plate, whilst the master was used to generate the second replica in exactly the same fashion. The replicas were grown up at 37°C for approximately 8 hours (or overnight at room temperature) whilst the master was restored to 37°C for several hours, then stored at 4°C.

2.5.1.3 Preparation of colony filters for hybridisation

Replica filters were initially floated on denaturing solution (see appendix B) for 5 minutes and then on neutraliser solution, again for 5 minutes. Subsequently, better results were obtained when filters were placed on top of several sheets of Whatman 3MM soaked in denaturer for 7 minutes, blotted on to dry paper to remove the excess fluid, then on to Colony blot neutraliser for two separate 3 minute incubations. This procedure lysed the colonies and allowed the DNA to bind to the filter.

Colony debris was carefully removed by wiping the filter surface with a folded tissue (Kleenex medical wipes) under 2X SSC and the filters air dried. Following a 10 minute incubation at 80 °C, filters were UV crosslinked as in section 2.4.7.2.

2.5.1.4 Nucleic acid hybridisation to colony filters

Replica filters of colonies from the genomic library, and recombinant plasmid clones were prehybridised for 10 minutes at the hybridisation temperature in 50 ml of 0.5 M sodium phosphate buffer pH7.2, 7% SDS, 1 mM EDTA (Church and Gilbert, 1984). 50-100 ng of oligo-labelled purified probe (section 2.4.8.4) was boiled for 5 minutes, quenched on ice and added to 25 ml of the original prehybridisation solution covering the filters (the remainder was discarded). Hybridisation was carried out overnight between 58.5 °C and 65 °C.

Filters were washed exactly as for the Southern hybridisation protocol discussed above in section 2.4.8.5.

2.5.1.5 Isolation of positive colonies on primary screening

Positive signals as revealed by autoradiography which appeared at identical loci on both replica filters were marked, along with orientation pinholes, on an overlying

acetate which was secured on a light box. A few ml of LB broth/antibiotic was placed on top of the acetate to facilitate fine positioning of the master filter which was orientated so as to align the pinholes.

A sterile plastic loop, or the large end of a glass Pasteur pipette, was used to pick all the colonies in a 5 mm radius area centred on the positive signal in to 50 µl of LB broth/Kanamycin. PCR screening and further treatment of the positive pool is described in section 2.5.1.6 below.

2.5.1.6 Genomic library / positive pool PCR

An aliquot of each monkey gDNA Charon library was directly screened for the presence of the opsin genes by diluting 2 µl of the glycerol stock in to 10 µl total volume with sterile distilled water. This was heated to 99 °C in a heat block for 10 minutes before being spun down for 1 minute at 13,000 rpm in a benchtop centrifuge. 10 fractions each of 1 µl were used in a standard "hot start" PCR.

PCR was also utilised in order to test colonies in an area of 2-3 mm radius, centred on each replicated positive signal from a primary cosmid library screen. Colonies were lifted from the master filter with a sterile plastic loop which was inoculated in to 50 µl of LB broth with the appropriate antibiotic and incubated at 37 °C, 225 rpm for 1 hour to allow for recovery and some growth.

The rest of the pool was diluted with 100 µl of LB broth plus antibiotic and stored at 4 °C overnight. For longer term storage, a 5 ml culture was made of the pool which was grown up overnight (37 °C, 225 rpm). A glycerol stock (final glycerol concentration of approximately 15% v/v) was made from this culture and stored at -80 °C.

2.5.1.7 Secondary screening

A 10^3 dilution of each positive pool was made in LB broth/antibiotic and 5 µl, 10 µl and 100 µl were plated out on an 8.5 cm diameter plate of LB agar/antibiotic. The particular volume giving approximately 300 or so colonies of good separation was

diluted in broth to a volume of 250 µl and gently spread on to a circular filter using a sterile glass spreader. This was grown up at 37 °C overnight and treated exactly as for the primary filters to make replicas, lyse colonies, fix DNA to the membrane, hybridise and wash for subsequent autoradiography.

Positive signals were picked as single colonies in to 50 µl of broth/antibiotic and processed in the same way as their primary screen counterparts (see above).

2.5.1.8 Tertiary screening

Tertiary screening was conducted in the same manner as the secondary screen. The density of the colonies was adjusted such that only about 50 grew on each plate, allowing for cultivation of well separated and defined clones.

2.6 Manual sequencing of DNA

2.6.1 Overview of dideoxy chain termination sequencing

Sequencing reactions were carried out using Sequenase version 2.0 (USB), or T7 polymerase (Pharmacia) which are genetically engineered T7 DNA polymerases lacking 3' to 5' exonuclease activity. The dideoxynucleotide protocols described are modified from Sanger *et al.* (1977).

The ddNTPs lack a hydroxyl residue at the 3' position of the deoxyribose, which prevents formation of a phosphodiester bond when incorporated in to the elongating DNA molecule and leads to chain termination. By using four different ddNTPs in four separate reactions, populations of oligonucleotides are generated that terminate at positions occupied by every A, C, G or T in the template strand. Band compressions on gels indicated the use of the nucleotide analogues 7-deaza dGTP and 7-deaza dATP (Pharmacia) to try to alleviate intrastrand secondary structure. This met with limited success and the best method for resolving compressions was to either sequence the opposite strand in this vicinity, or perform automated cycle sequencing (section 2.7). Sequencing reactions were carried out on double stranded plasmid DNA prepared as described in the following section using a selection of SP6, T7, M13

universal forward or reverse primers initially, followed by insert specific oligonucleotides as required.

2.6.2 Dideoxy sequencing with Sequenase(USB) and T7 (Pharmacia) kits

Double stranded plasmid DNA (approximately 1-2 μg) was alkali denatured in sodium hydroxide (2 mM) and EDTA (20 μM) at room temperature for 10 minutes. The DNA was then precipitated in 100 μl of absolute alcohol containing 5 μl of 3M sodium acetate at -20°C for 30 minutes, and recovered by centrifugation at 14,000 X g (13,000 rpm in a microcentrifuge) for 20 minutes at 4°C. The resultant pellet was washed in 70 % alcohol, dried and resuspended in water (10 μl). Manufacturer's instructions were then followed precisely, using [α-^{35}S] dATP as label.

To extend the sequencing reactions in order to read further into the clone of interest, long termination mixes (Pharmacia T7 protocol) were used which are effectively diluted ddNTPs. Some success was achieved in increasing the readable sequence using this protocol but not to a reproducible, practically useful level.

2.6.3 Direct sequencing of PCR products

The sequence of interest was amplified from gDNA or cDNA with specific primers. DNA was extracted with 1 volume of 1:1(v:v) phenol/chloroform followed by 1 volume of chloroform at 14,000 X g for 1 minute (section 2.7.5). A 50 μl aliquot of the aqueous phase was then purified through a Sephacryl S-400 spin column (Pharmacia) following manufacturer's instructions.

2.6.3.1 5'-end labelling of primer

Primer (10 pmoles) was labelled using [γ-^{32}P]dATP in a final volume of 10 μl by adding the following; 1 μl of T4 polynucleotide kinase (5-10 units; Pharmacia), 3 μl of [γ-^{32}P]dATP (3000 Ci/mM) and 1 μl of One-Phor-All buffer (Pharmacia). The reaction mix was incubated at 37 °C for 1 hour.

2.6.3.2 Sequencing using Sequenase 2.0 (USB) and T7 (Pharmacia) kits

An aliquot of DNA (9 µl for Sequenase, 10 µl for T7) was mixed with 2 µl (2.5 ρmol) of labelled primer and heated to 95 °C for 2 minutes then cooled on ice. The following were then added on ice:-

- 1 µl of DTT, 2 µl of reaction buffer (USB) or
- just 2 µl annealing buffer (Pharmacia), then
- 2 µl of Sequenase (1:7 dilution) or T7 polymerase (1:4 dilution).

The reagents were then mixed gently before 3.5-4.0 µl aliquots were removed to be mixed with 2.5 µl aliquots of each dideoxy nucleotide (as described in the manufacturers' protocols). The samples were then incubated at 37 °C for 5 minutes before the reaction was terminated by the addition of 4 µl of stop buffer.

2.6.3.3 Solid phase single-stranded sequencing of biotinylated DNA strands

During the synthesis of certain PCR primers, a biotin group was added to the 5' end at synthesis, in order to generate single stranded DNA for sequencing using streptavidin-magnetic beads (Dynabeads M280, Dynal).

A standard PCR reaction in 100 µl volume was carried out using one biotinylated primer and one normal primer for 35 cycles. The success of the amplification was verified by running 10 µl on a 1% agarose minigel. The remainder of the reaction was fractionated by electrophoresis in a 1% LMT agarose gel made up using 1X TBE as a buffer in both the gel and the tank. The required bands were separated from artefacts, primer dimer, etc., by band excision from the gel under minimal UV exposure, and transferred to 1.5 ml microcentrifuge tubes. 400 µl 1X TBE was added and tubes were incubated at 65 °C to melt the agarose. The streptavidin-magnetic beads were resuspended by vortexing and 16 µl added to the

DNA/diluted agarose. Binding of the biotin to the streptavidin was accomplished over 30 minutes at 65 °C on a rotary platform.

From this point onwards, the anchored biotinylated DNA was separated from the aqueous phase using a magnetic field applied to the microcentrifuge tube (Magnetic Particle Concentrator, Dynal), thus allowing buffers to be easily exchanged with no loss of template. The DNA was washed by resuspending the DNA/beads in 200 µl sterile distilled water, then concentrating the DNA/beads and removing the sterile distilled water. This was repeated before denaturing the template strands in 0.15 M sodium hydroxide on the bench for 5 minutes.

The non-biotinylated strand separated by this treatment was pooled from two identical reactions and precipitated by adding 0.1 volumes 3 M sodium acetate and 2.5 volumes of 100% ethanol, then incubating at -20 °C for 30 minutes. The DNA pellet formed by centrifugation for 15 minutes at 13 K rpm in a benchtop microcentrifuge at 4 °C, was washed with 70% ethanol and resuspended in sterile distilled water. This was then sequenced using a ^{32}P-labelled primer according to the method described in **sections 2.7.3.2 and 2.7.3.3**.

The biotinylated strand bound to the beads was washed twice in sterile distilled water and the sequencing reaction assembled using the reagents from the Sequenase and T7 kits, using [α-^{35}S]dATP following the manufacturers' protocols for double stranded plasmid DNA, taking care to ensure even distribution of the DNA/beads in the solution. The sequencing primer used was either the non-biotinylated one used in the original PCR, or else an internal primer synthesised from previously ascertained sequence.

2.6.4 Polyacrylamide gel electrophoresis of sequenced DNA

Apparatus used for running sequencing gels was purchased from BioRad. The back sequencing plate (holding the buffer reservoir) was silanised before assembly (Sigmacote, Sigma). Acrylamide concentrate (19:1 acrylamide:bisacrylamide) in 8.3 M urea solution (Sequagel, National Diagnostics) was mixed with dilutent (8.3 M Urea)

and buffer (10X TBE in 8.3 M urea) to give a 6% gel solution in 8.3 M Urea and 1X TBE.

For a 50 ml plug, 300 µl of 25%(w/v) ammonium persulphate and 300 µl of TEMED were added and quickly mixed before pouring. For a gel volume of 150 ml, 480 µl of 25%(w/v) ammonium persulphate and 60 µl of TEMED were added and quickly mixed before pouring. Sequencing reactions were heat denatured at 90 °C for 5 minutes before loading on to a pre-warmed (50-55 °C) 6% denaturing polyacrylamide gel and electrophoresed for the desired number of volt hours in 1X TBE.

Samples containing DNA bound to magnetic beads were treated as above, the gel matrix retarding the beads and template at the base of the well. This did not appear to affect the sharpness of the bands on the autoradiograph.

After electrophoresis, the gel was fixed in 10% methanol/10% acetic acid for 5 minutes on the front gel plate. Excess fixative was removed and a sheet of 3 mm Whatman paper was placed on top of the gel (to which it preferentially bound). The gel was covered with cling film and dried on Whatman 3MM paper for 3-4 hours under vacuum at 70 °C before exposure to XAR-5 (Kodak) film overnight. Longer exposures were used if required following initial visualisation of the autoradiograph.

2.7 Automated sequencing DNA

2.7.1 Automated sequencing technology

The main difference between automated sequencing and radioactive sequencing is that all the common formats rely on fluorescent dyes to detect the sequence ladders, and not radionucleotides, which makes them safer to use. Lasers are employed within the design of the apparatus to promote fluorescence, which is detected and translated to a graphic image on a computer. The ABI 373a DNA sequencer was used routinely in this research. Full details can be obtained from the suppliers (ABI cycle sequencer from Perkin Elmer, Foster City, California, and ALF from Pharmacia, Sweden).

2.7.2 Sequencing using the ABI 373a DNA sequencer

A large selection of kits were available from Perkin Elmer for use with the ABI 373a DNA sequencer, and were used exactly as per instructions. Two categories of kit were available; those which utilised fluorescently tagged dideoxy terminators, permitting the use of any primer, and those that incorporated a fluorescently labelled primer, where the template DNA needed to provide a binding site for that primer. Kits from the former category included; ABI PRISM™ dye terminator cycle sequencing kit with AmpliTaq® DNA polymerase, (Part number 401384), and ABI PRISM™ dye terminator cycle sequencing ready reaction kit with AmpliTaq® DNA polymerase FS, (Part number 402078). Kits of the latter category which were selected for use included; Dye primer cycle sequencing ready reaction kit with AmpliTaq® DNA polymerase, CS+, -21M13, (Part number 401843), and M13Rev. (Part number 401941); and ABI PRISM™ dye primer cycle sequencing ready reaction kit with AmpliTaq® DNA polymerase FS, (Part number 402111).

2.7.3 Direct sequencing of PCR products

Direct sequencing of PCR products was permissible using only those kits which contained labelled dideoxyribonucleotides ("terminator" kits). Essentially, 5 µl of each 50 µl PCR reaction were size fractionated on agarose gels to determine the size and purity of product. Solely those PCRs which procreated a solitary target band were processed through Centicon 100 spin columns (Centricon, Princeton, USA) to remove unincorporated primers, oligonucleotides, and dNTPs as follows: 45 µl of the PCR product was combined with 2 ml of sterile distilled water in the upper reservoir of the Centicon 100 column, which was thereupon spun for 15 minutes at 1000 X g. The column was then inverted and the retanate (approximately 50 µl) collected with a slow spin at 300 X g for 5 minutes. A fraction of the DNA was quantified upon a 1.2% agarose gel. These purified PCR products were cycle sequenced as described in section 2.7.4.

Those PCRs which failed to generate a single target band were size fractionated through low melting point agarose gels, objective bands were excised (section 2.4.4.1), and the DNA allowed to elute out overnight in preparation for cycle sequencing (section 2.7.4).

2.7.4 Cycle sequencing

The purified DNA templates, be they from PCR reactions or plasmid preps, were cycle sequenced using either the ABI PRISM™ Taq dideoxy kit, or the newer kits with AmpliTaq® DNA polymerase FS, and the appropriate sequencing primer. Generally all direct sequencing reactions were made up as follows:

OLD KITS		NEW KITS
9.5 µl	Reaction mix	8.0 µl
1.0 µl	sequencing primer (3.2 µM)	1.0 µl
5.0 µl	template (1 µg)	5.0 µl
4.5 µl	distilled water	7.0 µl

The combined reaction volume of 20 µl was then cycled on the Perkin Elmer 9600 or 2400 PCR machines with the following temperature profile

OLD KITS	NEW KITS
ramped to 96°C at 1° per sec	max ramp
96°C for 20 sec.	96°C for 10 sec.
ramped down to 50°C at 1° per sec	max ramp
50°C for 5 sec.	50°C for 5 sec.
ramped to 60°C at 1° per sec	max ramp
60°C for 240 sec.	60°C for 240 sec.

For a total of 25 cycles. The samples were further processed to remove unincorporated dyes (section 2.7.5).

2.7.5 Removal of excess fluorescent dyes

Before loading on to the ABI 373a automated sequencer unincorporated fluorescent dyedeoxy™ terminators had to be removed. Several alternative protocols were suggested by the supplier.

The recommended and best suited approach was to use Centri-Sep spin columns (Centricon, Princeton, USA), in accordance with the manufacturers instructions. These columns permitted the reading of sequence data close to the sequencing primer as well as increasing the reliability of data generated surrounding dye blob migration points. However, the cost per spin column was prohibitive, so only the most critical reactions were processed in this manner.

With the older kits (those employing the original AmpliTaq DNA polymerase) the most cost effective protocol available was the use of phenyl/chloroform/water extraction, succeeded by ethanol precipitation. The later kits (those with a mutant form of Taq DNA polymerases: CS+, or FS) simplified the extraction procedure to just a single ethanol precipitation. The dye content of these kits was reduced as the efficiency of labelled terminator incorporation was improved.

2.7.6 Running of automated sequencer

The 373a DNA sequencer was set up and run in accordance with the supplier's instructions. Sequencing runs involved electrophoresis (of up to 38 samples) for 10 to 13 hours, generally overnight. An Apple Macintosh computer linked to the sequencer was programmed to analyse the data and present it in the from of two files: the text file and an analysis file. The latter file incorporates the electropherogram (scan) data, and constitutes the original sequence data.

2.8 Computational analysis of nucleic acid sequences

2.8.1 Use of localised software for sequence analysis

Analysis of nucleic acid sequences and potential open reading frames was performed using a variety of software developed for the Apple Macintosh computer.

GeneWorks™ (version 4.45) was principally used to collate, store and align sequences of different clones/PCR products to highlight mismatches or nucleotide deletions/insertions and to assemble sequence contigs with full coverage from several clones. In addition, searches for restriction sites, and analysis of sequences for recognised transcription factor binding motifs were performed.

MacVector software (developed by IBI for use on an Apple Macintosh) versions 3.5, 4.0 and 4.1 was used as an alternative package for nucleic and amino acid sequence analysis. Version 2.0 was used, principally for the selection and testing of primer pairs for compatibility prior to their synthesis, and selection of oligonucleotides as probes, as well as accessing sequences from a CD-rom database.

2.8.1.1 Constructing phylogenetic trees

Nucleotide sequences were compared and the synonymous and non-synonymous nucleotide substitution rates calculated according to Li *et al.* (1985). The synonymous rates were used to construct a phylogenetic tree using the neighbour-joining method (Saitou and Nei, 1987).

2.8.2 Use of networked facilities for analysis

The SERC Daresbury laboratory and Human Genome Mapping Group are both host to a suite of programmes, developed by the Genetics Computer Group (GCG) at the University of Wisconsin, which are accessed from a remote site via a computer network service (Devereux *et al.*, 1984). This original VAX system has recently been converted to UNIX. 5' upstream and coding DNA sequences were analysed using this software, as well as, facilities available on the Internet at numerous sites, including the HGMP site.

In order to choose PCR primers from evolutionary conserved regions of the opsin genes, the GAP programme was used to compare the DNA sequences of human and bovine opsins. This program uses the algorithm of Needleman and Wunsch, (1970). Areas of sequence conservation revealed by this comparison enabled consensus oligonucleotides to be synthesised.

Sequences upstream of the opsin genes were compared with entries in the GenBank database (latest version) using the FASTA algorithm (Pearson and Lipman, 1988) to evaluate similarity scores. FASTA was also used to compare all upstream sequence with the Eukaryotic Promoter database in an attempt to identify upstream motifs which were shared with those of other genes. Default settings were used.

Signal Scan was also used to analyse the upstream sequences by searching for known eukaryotic transcription factor binding motifs (Prestridge, 1991), using the mammalian subset of Ghosh's transcription factor database (Ghosh, 1990). This program uses Staden format.

2.9 Reagents, suppliers and materials

2.9.1 General reagents

- TE buffer: 10 mM Tris-HCl pH7.5, 0.1 mM sodium EDTA pH8, autoclaved
- 10X loading buffer: 25%(w/v) ficoll (Mr 400 kD), 0.25%(w/v) Orange G, 0.25 M EDTA in 1X TAE buffer
- 20X SSC: 3 M NaCl, 0.3 M sodium citrate
- 10X TAE: 0.4 M Tris-acetate, 10 mM sodium EDTA, pH8
- 3 M sodium acetate (sodium acetate): 408.1 g/l of the salt dissolved in water and adjusted to pH 4.8 with glacial acetic acid. Autoclaved.

2.9.2 Solutions used for DNA preparation and Southern blotting

- 10X lysis buffer: 50 mM Tris-HCl pH7.5, 25 mM $MgCl_2.6H_20$, 0.6 M sucrose, 5%(v/v) Triton X100
- 10X suspension buffer: 0.1 M Tris-HCl pH7.5, 0.1 M NaCl, 0.1 M sodium EDTA pH8
- Denaturer: 1.5 M NaCl, 0.5 M sodium hydroxide
- Neutraliser: 3 M NaCl, 0.3 M sodium citrate, pH 5.5
- Colony blot neutraliser: 1.5 M NaCl, 0.5M Tris-HCl pH7.2, 1 mM EDTA

2.9.3 Media for microbiological manipulation

- LB (Luria Bertoni) broth: 1%(w/v) Bactotryptone, 0.5%(w/v) yeast extract, 1%(w/v) NaCl, autoclaved
- LB agar: LB broth containing 1.5%(w/v) agar, autoclaved
- Solution I: 50 mM glucose, 25 mM Tris-HCl pH8, 10 mM EDTA, autoclaved
- Solution II: 0.2 M sodium hydroxide, 1%(w/v) SDS, made fresh from autoclaved solution
- Solution III: 10 mM Tris-HCl pH7.5, 0.1 mM sodium EDTA pH8, autoclaved

2.9.4 Gel electrophoresis

- 10X loading buffer: 25%(w/v) ficoll, 0.25%(w/v) OrangeG, 0.2 mg/ml ethidium bromide in DEPC water
- 10X Sucrose loading buffer: 860 mM sucrose, 5.5 mM Orange G in aqueous solution in distilled water
- 10X SSC: 3M NaCl, 0.3 M sodium citrate, DEPC treated and autoclaved

2.9.5 Manual sequencing solutions

- 10X TBE: 90 mM Tris borate, 10 mM sodium EDTA, pH 8.3
- Stop Solution: 0.3% Bromophenol blue, 0.3% Xylene cyanol FF, 10 mM EDTA (pH7.5), 97.5% deionised formamide.
- Sequagel. Sequencing system concentrate (# EC-830) National Diagnostics.

2.9.6 Automated sequencing reagents

- 10X TBE: 90 mM Tris borate, 10 mM sodium EDTA, pH 8.3
- Loading Solution: Deionised Formamide : 0.05 M EDTA (5 : 1 v/v)
- SequaGel^{TM} complete buffer reagent (# EC-841), National Diagnostics
- SequaGel^{TM} ready-to-use 6% sequencing gel solution (# EC-836), National Diagnostics
- 10% Ammonium Persulphate (IBI), freshly prepared.

- 2 M Sodium Acetate
- Phenol/Water/Chloroform (Applied Biosystems; # 400765)

2.9.7 General reagent suppliers

All general laboratory reagents were of "Anala R" grade (BDH) unless stated.

Phenol	Rathburn Chemicals Ltd.
Bovine serum albumin (BSA)	Sigma Chemicals Co. Ltd.
Dimethylforamide (DMF)	Sigma Chemicals Co. Ltd.
Ficoll	Sigma Chemicals Co. Ltd.
Herring testis DNA	Sigma Chemicals Co. Ltd.
Isopropyl-B-D-thiogalactoside (IPTG)	Sigma Chemicals Co. Ltd.
Geneclean Kit	BIO101
XGal	Bethesda Research Laboratories
Lambda DNA	Pharmacia Ltd.
Hyper-film Bmax	Amersham
Hyper-film MP	Amersham
Hybond N	Amersham
Electran 15 Agarose	Sigma
NuSieve GTG Agarose	FMC
Genereleaser™	BioVentures, Inc.
Perfect Match™	Scotlab, UK.

2.9.8 Manual sequencing reagents

Acrylamide	Kochlight Ltd.
Ammonium persulphate	Kochlight Ltd.
Boric acid	Sigma Chemical Co. Ltd.
N,N'-Bis-acrylamide	Kochlight Ltd.
N,N',N',N'-tetramethylethylenediamine (TEMED)	Sigma Chemical Co. Ltd.

Sequenase Version 2.0	Cambridge Bioscience
Urea	Sigma Chemical Co. Ltd.

2.9.9 Suppliers

- Amersham International plc. UK.
- BioRad laboratories Ltd. UK.
- Cambio Ltd. UK.
- Clontech laboratories Inc. USA.
- DuPont Biotechnology Systems, USA.
- Flowgen (suppliers of FMC bioproducts) Ltd. UK.
- Gibco BRL life technologies, UK.
- ICI (Zeneca), Cambridge Biosciences, UK.
- Northumbria Biologicals Ltd.(NBL) UK.
- Pharmacia Ltd. UK.
- Schleicher and Schuell, Germany.
- Scotlab, UK.
- Stratagene Ltd. UK.
- United States Biochemicals (USB), USA.

Chemicals not specified were purchased from:

- BDH Chemicals Ltd.
- Gibco BRL
- Unipath (Oxoid) UK.
- Pharmacia Biotech
- Sigma Chemicals, USA

2.9.10 Radiochemicals

- Amersham

2.9.11 Computer software and access

- SERC Daresbury Laboratory, Daresbury, Warrington, WA4 4AD
- UK Human Genome Mapping Resource Centre, Clinical Research Centre,Watford Road, Harrow, Middlesex, HA1 3UJ.
- GeneWorks, Intelligenetics, Inc., 700 East El Camino Real, Mountain View, CA 94040.
- MacVector, International Biotechnologies Inc., P.O. Box 9558, New Haven, CT 06535.

2.9.12 Oligonucleotides

- Synthesised in house on a MilliGen DNA synthesiser.
- Commercially by Oswel DNA, British Biotechnology, Pharmacia and Cruachem.

2.9.13 Glassware

All glassware was washed thoroughly with Teepol detergent and then rinsed twice with double distilled water or deionised water before use.

2.9.14 Plasticware

Microcentrifuge tubes and tips were used direct from the supply bags and were only sterilised prior to use when employed for subsequent culturing work.

1.5 ml microcentrifuge tubes (MCC) and 0.5 ml microcentrifuge tubes were purchased from Elkay. 1.0 ml blue and 200 µl yellow micropipette tips were from Algen. Sterile 50 ml disposable, screwcap tubes were obtained from Falcon (referred to as Falcon tubes). These items were sterilised by autoclaving at 15 psi for at least 20 minutes.

CHAPTER 3

Catarrhine Primate Opsin Gene Evolution

3.1 Introduction

3.1.1 Primate evolution

The Order Primates includes monkeys, apes, and humans - 15 families, comprising 223 extant species, which are divided into four sub-orders (Nowak, 1991). These species differ from other mammals in having dextrous hands and feet, binocular vision and a well-developed brain. Catarrhina (Old World monkeys & Apes) exhibit narrow internasal septums which bring the nostrils close together, whilst platyrrhina (New World monkeys) possess a broad internasal septum separating their nostrils. Apes do not have tails (Napier and Napier, 1985). All New World monkeys (NWM) possess tails which they use to great effect in navigating through the forest canopy. Humans, along with chimpanzees, gorillas, and other Great Apes are usually referred to as Hominoids, whilst the Old World monkeys (OWM) are known as Cercopithecoids. These species have diverged morphologically from each other, occupying a wide range of ecological niches. In contrast, the primate populations of the New World exhibit a high degree of uniformity; all species looking much like any other, where interestingly all species are arboreal (none is ground-living). Just three families of NWM are recognised (Figure 3.1), Atelidae, Cebidae, and Pitheciidae (Meireles *et al.*, 1995), with a total of 9 subfamilies and 19 extant genera (Porter *et al.*, 1995).

The principal areas of natural primate distribution are Africa, Madagascar, India, south-east Asia, and South and Central America (Figure 3.2). Catarrhine species are confined to Africa,, Madagascar, India, south-east Asia, and many East Indies islands, where both ground-living and arboreal groups are found. Africa is considered the home of the prosimians, Old World monkeys of the two cercopithecid subfamilies, whilst lemurs are exclusive to Madagascar. The Old World monkey species of Africa

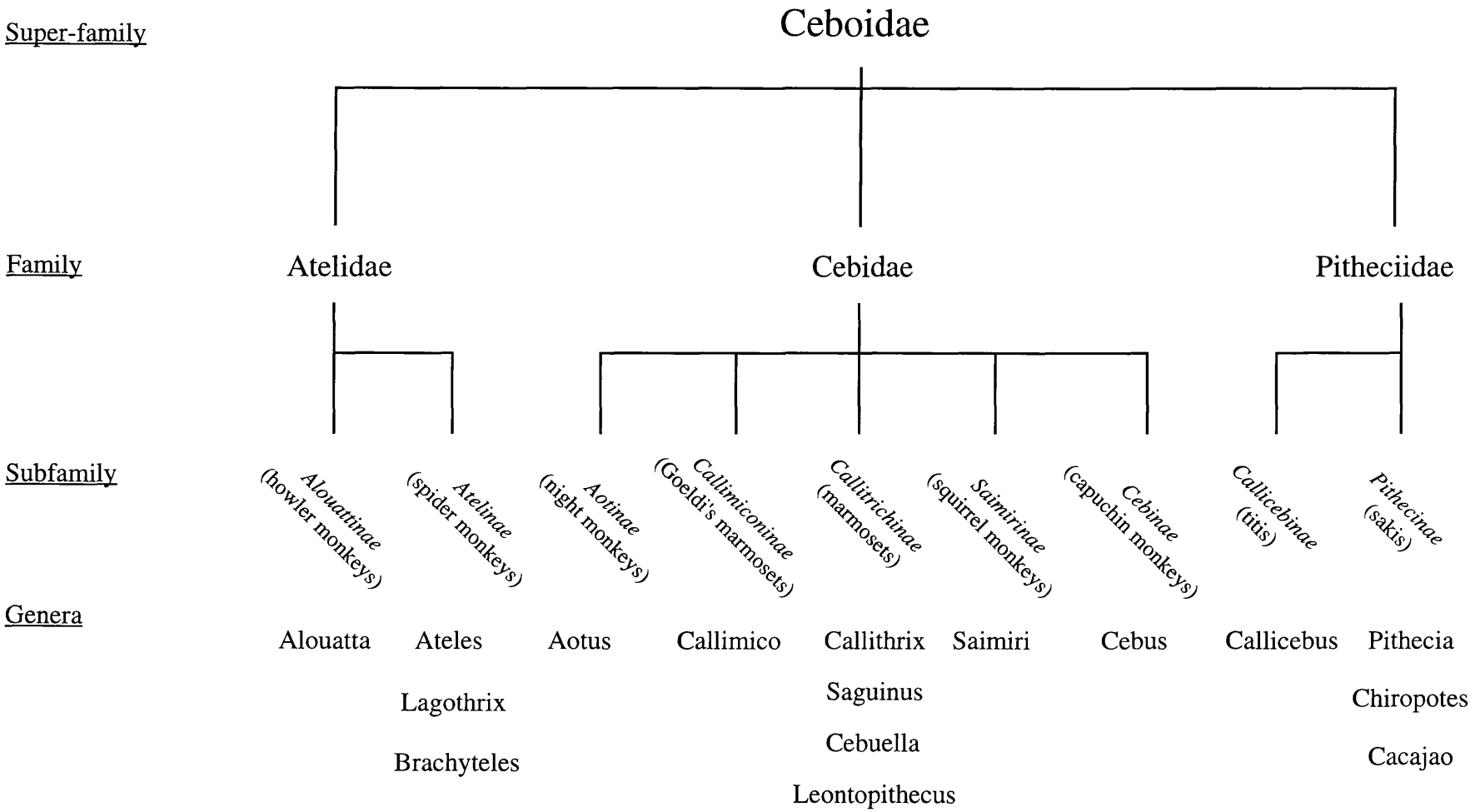
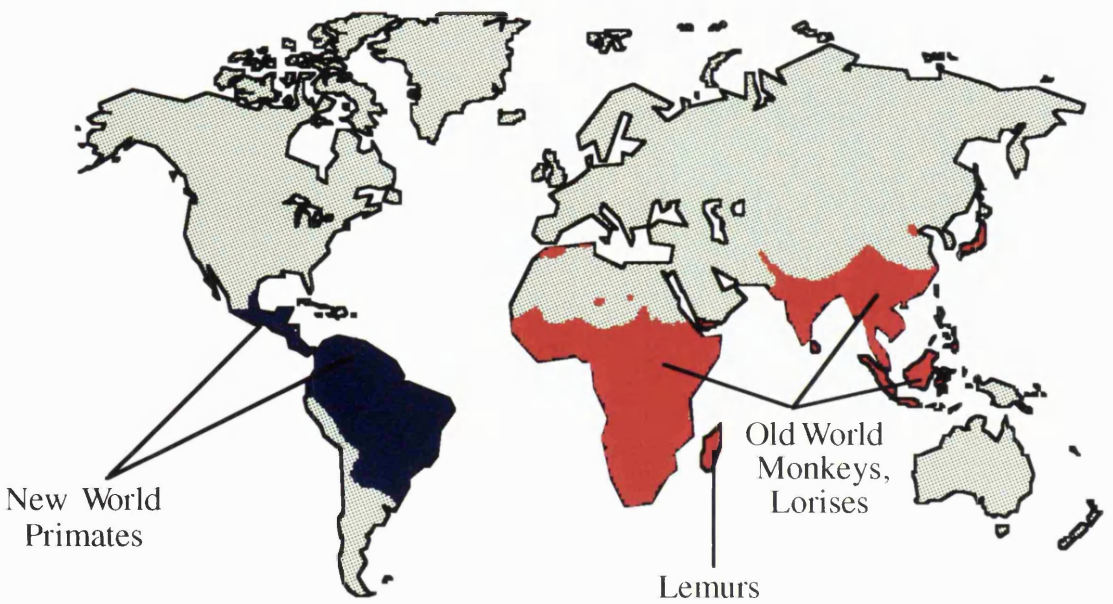


Figure 3.1 New World monkey phylogeny. Adapted from Meireles *et al.* (1995).

Figure 3.2 Geographical distribution of non-human primates. Old World primates inhabit areas depicted in red, and New World primates principally live in areas highlighted in blue.



are clearly distinguishable from the extant species in Asia. Platyrrhina species are exclusively located in a region extending from 24°N in Central America to 30°S in South America. No lemurs, lorises, tarsiers, or apes are found in the New World.

The ideal approach to identifying the true relatedness of extant primates would be to recover the remains of the common ancestors of all the lineages which would lead from the common primate ancestor to the diversity of primate species present today. Important information regarding the evolutionary lineage of the primates has been gathered from examination of the limited fossil finds. These studies rely on the comparison of the fossilised hard tissues, usually bone and teeth, of favourably preserved remains. The critical factor influencing the scarcity of fossils is thought to be the acid soils in which the corpses of the primates decayed: these soils constitute the floor of the tropical rain forests, in which the vast majority of early primates are considered to have evolved. However, these common ancestral remains are not forthcoming to the degree that is necessary to formulate links, thus alternative approaches have to be adopted. An approach utilising molecular data has been considered a good alternative, although by itself would not suffice to replace the fossil approach, since greater time spans are required for the accumulation of molecular changes, from which reliable phylogenetic trees can be constructed (Rogers, 1994). Scientists, aligned to each of the above camps, are accepting that a co-operative analysis involving the two approaches is the best method of understanding the true ancestry of the modern primates.

The ideal molecular approach would be based on whole genome analysis, but considering that there most probably exist 50,000 - 70,000 genes per genome per animal, and a tremendous amount of non-coding DNA too, this is at present out of the question. An acceptable alternative is to compare individual genes or parts thereof, which not only provides information on the evolutionary perspective, but also addresses the functional constraints on that gene.

This study of X-linked opsin genes provides relevant information on the spectral positioning of these genes in primates other than man. Other features, such as the presence of gene polymorphisms, common to humans can also be investigated.

3.1.2 Primate X-linked opsin genes

Among the mammals, only Old World primates have separate MW and LW opsin genes, allowing for trichromatic colour vision (Bowmaker *et al.*, 1991a; Ibbotson *et al.*, 1992), whilst their New World counterparts possess a single X-linked L/MW opsin gene (Williams *et al.*, 1992; Hunt *et al.*, 1993a); section 1.13)*. This implies that the duplication event that gave rise to the trichromatic mode of colour vision in Old World primates arose subsequent to the separation of this lineage from the New World primate lineage, approximately 46 to 60 million years ago (Goodman *et al.*, 1994). The spectral positioning of OW primate pigments is determined by a small number of amino acid substitutions in the transmembrane regions of the opsin pigment molecules (Nathans *et al.*, 1986a; Neitz *et al.*, 1991; Chan *et al.*, 1992; Ibbotson *et al.*, 1992), and the same is true for the spectrally-distinct but allelic forms of the opsin gene in New World monkeys (Neitz *et al.*, 1991; Williams *et al.*, 1992). However, certain consistent differences, in humans, are present and it has been possible to determine that the spectral shift between these pigments is largely achieved by amino acid substitution at only three sites, one in transmembrane 4 (TM4), encoded by exon 3, and the other two in transmembrane region 6 (TM6) of the opsin protein, which is encoded by exon 5 (Figure 3.3; Neitz *et al.*, 1991; Merbs and Nathans, 1993; Williams *et al.*, 1992; Winderickx *et al.*, 1992c). In order to establish whether these substitutions are

---

* Recent findings have shown that for one genus of NWM (*Alouatta* , Howler monkeys) routine trichromatic colour vision, paralleling the situation in their OWM cousins, does exist (Jacobs *et al.*, 1996; Dulai and Hunt, unpublished data). None of nine other NWM genera examined exhibit this feature (Jacobs and Deegan, 1993; Jacobs and Deegan, 1994). It remains to be seen how many other NWM species present with this arrangement.

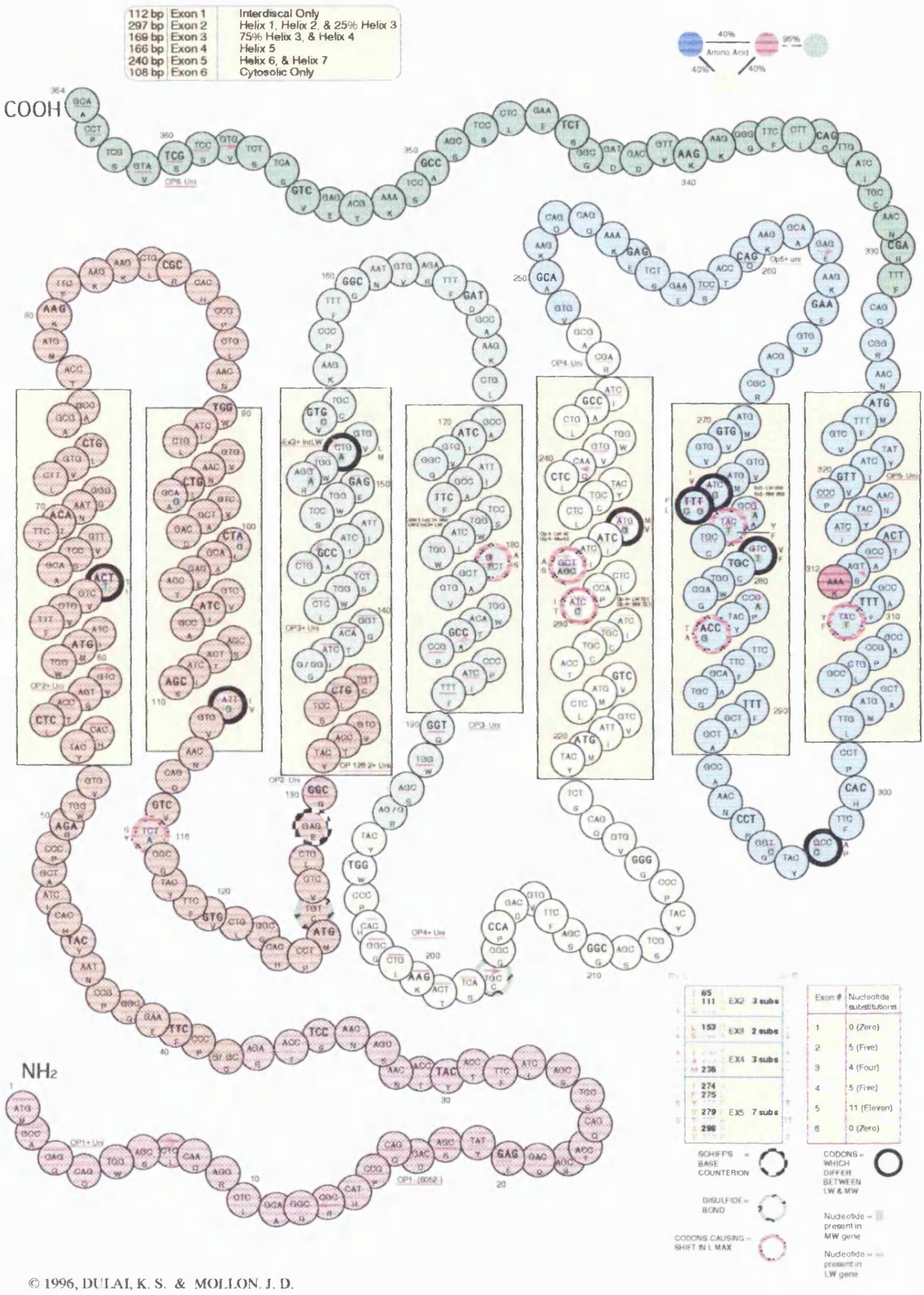


Figure 3.3 2-D representation of the X-linked opsin protein showing the exon boundaries. (Additional information relates to the LW opsin sequence. The corresponding amino acids are indicated).

conserved over other primate species, the MW and LW opsin gene sequences of two additional species of Great Ape, the gorilla, *Gorilla gorilla*, and chimpanzee, *Pan troglodytes*, and five species of Cercopithecoid monkeys, were examined. Except for the gorilla, all of these species are predominantly frugivorous in their diet; and it has been suggested that primate colour vision co-evolved with yellow and orange fruit, which only trichromatic observers can readily distinguish at a distance among dappled foliage (Mollon, 1991a). There is therefore a special interest in the gorilla, a bulk eater, whose diet is largely composed of stalks, vines, leaves, bark, shoots and roots. Although the diet of the western gorilla includes some fruit, the gorilla is a slow-moving forager who does not seek out fruiting trees at a distance in the manner of many monkeys. Has the gorilla experienced a relaxation of the selection pressure for acute colour vision, analogous to the relaxation of selection pressure by which Post (1962) explained the high incidence of colour deficiencies in modern human populations?

In their original report of the sequence of the human MW and LW opsin genes, Nathans *et al.* (1986a) reported the presence of multiple copies of the MW gene, and in a more extensive study (Drummond-Borg *et al.*, 1989), up to five such copies were identified, with a modal value of two per X chromosome. Sequences from within an opsin class exhibited a number of polymorphic sites (Nathans *et al.*, 1986a; Winderickx *et al.*, 1992c), and the same has been reported for the opsin complement of the talapoin monkey, *Miopithecus talapoin* (Ibbotson *et al.*, 1992). The sequences of human also reveal the existence of hybrid LW/MW genes which are thought to underlie certain forms of anomalous trichromacy (Nathans *et al.*, 1986b; Neitz *et al.*, 1991; Deeb *et al.*, 1992), although these have so far not been observed in any other primate species. The sequences of the MW and LW genes of the Great Apes and Cercopithecoid monkeys were examined for the presence of polymorphic sites. Also, using a molecular phylogenetic approach the ancestry of these genes in eight species of Old World primates was examined.

Seven Old World primates species were examined; two Great Apes and five monkeys. The apes were the common chimpanzee (*Pan troglodytes*) and gorilla

(*Gorilla gorilla*). Chimpanzees are found in East, Central and West Africa. They are both arboreal and terrestrial, nesting in trees at night and feeding in fruit trees during the day. They travel on the ground. Their diet is basically frugivorous, however nuts, seeds and leaves are also eaten. Occasionally, termite-fishing and the killing of young baboons and monkeys to supplement the diet has been reported (Goodall, 1986). Gorillas are distributed in Africa in a discontinuous manner, being found mainly in the Congo basin and Cameroon, with further small pockets in the Rwanda and Zaire regions. Gorillas are almost wholly ground-living, travelling in groups of 5-30 individuals. Their diet has already been discussed above. Three of the monkey species examined are from the Cercopithecus genus, belonging to the Guenon group (Martin, 1990): Diana monkey (*Cercopithecus diana*) from Gabon, Cameroon and Congo basin where they inhabit the top canopy of the rain forest; lesser spot-nosed guenons (*Cercopithecus petaurista*) of the Congo; and African green monkey (*Cercopithecus aethiops*) a species which is becoming adapted to a ground-living way of life, that ranges from Senegal to Ethiopia to South Africa. All three species have a substantially omnivorous diet, that includes most tree products, and insects and other invertebrates. *C. aethiops* also feeds on savannah grasses, seeds, gums, thorns, and seeds. The final two species, the talapoin monkey (*Miopithecus talapoin*) from most lowland forest types of western Central Africa, whose diet consists of native crops, leaves, buds and insects, and the patas monkey (*Erythrocebus patas*) whose range extends from Senegal to Sudan and to Tanzania, feeding on grasses, fruits, beans, seeds, and insects, lizards and bird's eggs, are each single members of the Miopithecus and Erythrocebus genera, respectively. All seven species studied provide both a varied habitat and diet. However, MSP measurements on eight species of fruit-eating monkeys from the Old World, inhabiting both Africa and Asia, show similar values of maximum sensitivity (λ_{max}) for both MW and LW photopigments (Bowmaker *et al.*, 1991), values which approximate those of humans (see Table 1).

SPECIES	LW photopigment λ_{max} (nm)	MW photopigment λ_{max} (nm)
Homo sapiens (Human)	532	564
Pan troglodytes (Common chimpanzee)	-	-
Gorilla gorilla (Gorilla)	-	-
Cercopithecus diana (Diana monkey)	565.9 ± 2.4	530.7 ± 3.9
Cercopithecus petaurista (Lesser spot-nosed guenon)	562.7 ± 3.2	533.9 ± 2.8
Cercopithecus aethiops (African green monkey)	566.3 ± 1.9	535.0 ± 2.7
Miopithecus talapoin (Talapoin monkey)	564.0 ± 3.9	533.3 ± 2.8
Erythrocebus patas (Patas monkey)	566.3 ± 1.9	533.0 ± 2.3

Table 1. λ_{max} values obtained by MSP analysis of MW and LW photopigments of human (Oprian *et al.*, 1991) and some Old World monkeys (Bowmaker *et al.*, 1991). No direct data are yet available on the chimpanzee or gorilla.

3.2 Methods

3.2.1 Source of primate genomic DNA

Genomic DNA (gDNA) for this investigation was obtained from three sources:-

1) Chimpanzee and gorilla gDNA was supplied in aqueous solution by Dr. Helen Stanley, Institute of Zoology, Reagents Park, London, (section 2.1).
2) Talapoin and rhesus monkey gDNA was extracted from frozen liver samples (as given in section 2.2.3).
3) Diana monkey, patas monkey, and lesser spot-nosed monkey gDNA, following use in a previous study (Ibbotson *et al.*, 1992).

3.2.2 Sexing of primates gDNAs used in this study

A problem that arose during this study was that information on the sex of the animals was not provided with the DNA samples from the Institute of Zoology, London.

PCR was therefore employed to determine the gender of each primate. Primers, XES1 and XES2 (Poulat, 1991), which were utilised in previous studies to determine the presence or absence of the Y chromosome, were obtained from Dr. David Lovell-Badge at the Medical Research Council, Mill Hill. This primer pair amplify a conserved region of the sex-determining gene on the Y chromosome of all primates, producing a band of approximately 330 bp. For each 50 µl sex-determining PCR reaction, a standard set-up was performed (section 2.3.2) containing 1 µl of each primer at 50 ng/µl stock, and 1.5 mM $MgCl_2$. PCR parameters involved an initial 3 minute extension at 94°C, followed by cycling at 96 °C for 1 minute 10 seconds, 60 °C at 1 minute 10 seconds, 72°C for 1 minute 40 seconds, for 30 cycles. An final extension of 10 minutes at 72 °C was also included. For each tube, a 20 µl aliquot of the PCR amplification plus 3 µl of 10 X loading dye was loaded onto a 2 % agarose gel and electrophoresed.

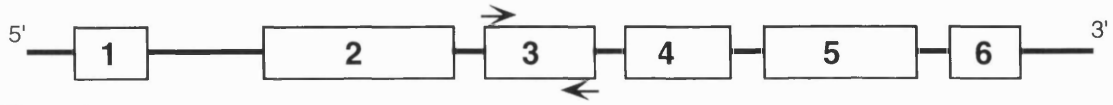
3.2.3 PCR amplification, cloning, and sequencing of exons 3, 4, and 5 of the LW and MW opsin genes

Primer pairs utilised in previous studies (Ibbotson *et al.*, 1992; Williams *et al.*, 1992) were used to amplify DNA fragments from primate gDNAs by the polymerase chain reaction (PCR). Exon 3 primers were calculated to produce fragments of 164 bp, whilst exon 4 primers generate 156 bp fragments, and exon 5 primers 191 bp fragments. The design of these primers was based on the published sequence of the LW opsin gene sequence (Nathans *et al.*, 1986a), and each set would co-amplify both the LW and MW exonic sequences. As Figure 3.4 illustrates these primer sets allow sequence to be generated from regions of exons 3, 4, and 5 that code for approximately half of transmembrane (TM) region 3 and most of TM4, most of TM5, and all of TM6 and half of TM7; regions contributing almost wholly to the spectral absorption characteristics of the opsin The sequence of each primer is given below (the information in brackets is the name of the primer followed by the position of the base along the opsin gene to which the 3' end of the primer hybridises - cone opsin numbering scheme):

Exon 3

5'-ATC ACA GGT CTC TGG TCT CTG-3' (Op3+/base 473)

5'-CTC CAA CCA AAG ATG GGC GG-3' (Op3-/base 597)



Exon 4

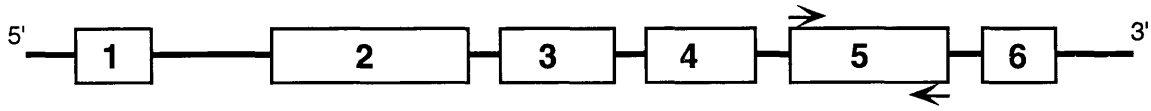
5'-CAC GGC CTG AAG ACT TCA TGC-3' (Op4+/base 650)

5'-CGC TCG GAT GGC CAG CCA CAC-3' (Op4-/base 765)

Exon 5

5'-GAA TCC ACC CAG AAG GCA GAG-3' (Op5+/base 830)

5'-CGG TGA TAG ATG TTG GGG CA-3' (Op5-/base 981)



PCR parameters were: initial denaturation at 94°C for 4 minutes. Denature 94 °C for 45 seconds, anneal at 64°C for 60 seconds, extend at 72°C for 45 seconds, for 35 cycles. 10 minutes at 72°C final extension. Total volume per tube = 50 µl, containing 1.5 mM $MgCl_2$, 2.5 U Taq polymerase, and approximately 200 ng of template DNA.

Following PCR a 20 µl aliquot of each reaction was loaded on to a 1.8 % low melting point agarose gel to check the success, specificity, and yield of the amplification (section 2.4.4).

Target bands were excised and eluted (section 2.5.3.1) and then sub-cloned into pCR1000™ or pCRII™ vectors (section 2.5.4). Following transformation into competent cells (either commercially obtained or produced in-house), plating on appropriate plates, and amplification in culture media (sections 2.4.5.3, 2.4.5.4, and 2.4.5.5), the recombinant plasmids were harvested by utilising the GeneClean™ protocol (section 2.4.6.3). Positive clones (those containing an insert of the correct size) were identified by digesting a small (5 µl) aliquot of the plasmid prep with appropriate restriction enzymes (for vector pCR1000 used *Not*I, and for pCRII used *Eco*RI). Clones containing inserts of the right size were subjected to sequencing (see section 2.6 and 2.7).

To overcome an inherent problem with the PCR amplification process, whereby incorporation of incorrect bases occurs at a low frequency in an enzyme dependent manner (for Taq it is calculated at 10^{-4}; Eckert and Kunkel, 1990), at least three PCR reactions were independently set-up for every exon of each species. By generating

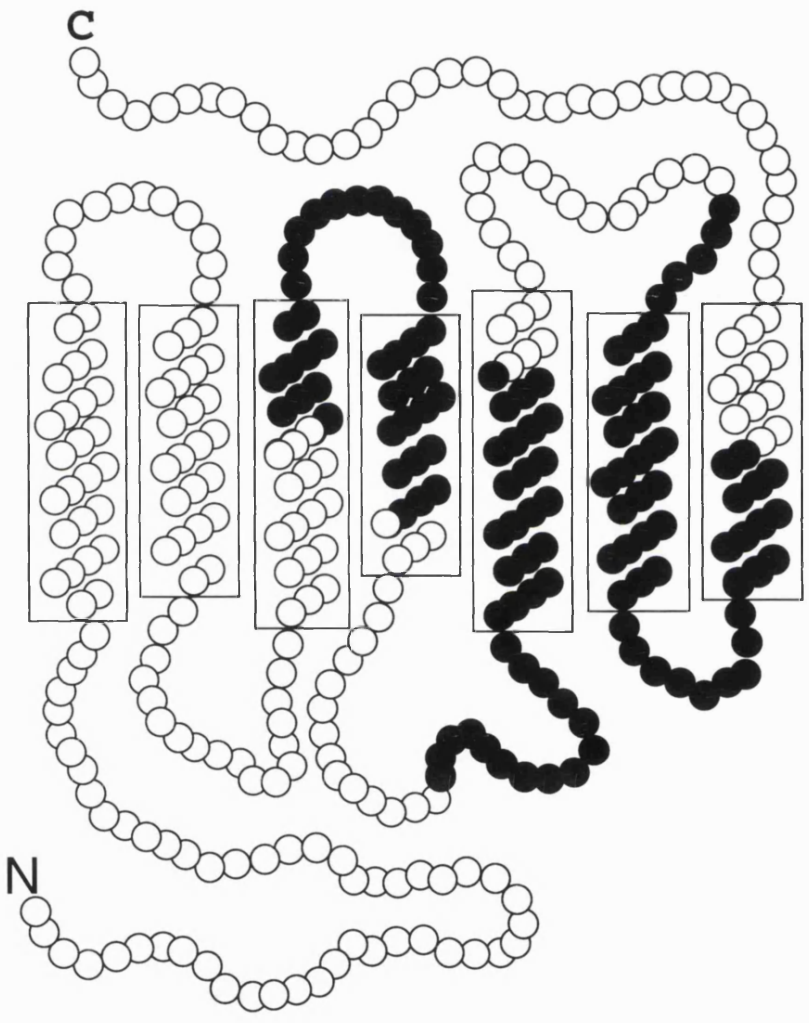


Figure 3.4 Regions (dark cicles) of the LW and MW opsin that were sequenced in the present study

clones from a number of independent PCRs it was possible to identify those sequences which result from incorporation errors and thus eliminate them from the analysis.

Sequencing by the manual dideoxy method using T7 polymerase and ^{35}S-labelled dATP (section 2.6.2), or automated sequencing using the ABI sequencer and Taq Dyedeoxy terminator kits (section 2.7.4), was carried out. For manual sequencing, products of the sequencing reactions were loaded on to 6.0 % polyacrylamide gels, separated at 1500V for about 4.5 hours, and autoradiographed overnight. Sequence data was read directly off the autoradiograph. Automated sequencing generated electropherograms, from which sequence data was directly imported into DNA alignment programmes (GeneWorks, MacVector, or the UW-GCG package - see section 2.8). All clones were sequenced in both directions.

3.2.4 Phylogenetic tree construction

Using the neighbour-joining algorithm of Saitou and Nei (1987) incorporated into the computer programme MEGA (Kumar and Simons, 1993) phylogenetic trees were constructed for the LW and MW opsin genes using the nucleotide sequence and deduced amino acid sequence data generated. The equivalent gene sequences from the marmoset 563 nm allele (Hunt *et al.*, 1993a) was included in this analysis, and the chicken iodopsin (Kuwata *et al.*, 1990) was used as an outgroup to root the tree. The average number of base pair substitutions per nucleotide site (silent and coding) was calculated. To make maximal use of the data, for each species the sequences obtained for exons 3, 4, and 5 were combined, and all substitutions were included.

Individual pair-wise comparisons between all the OW primates were made for each gene, as well as a comparison of each species with the corresponding human, marmoset 563 nm allele, and the chicken iodopsin sequences. The topology and branch lengths of the tree were based on *p* values (where, the *p* value is the frequency of amino acid replacements per site). Support for branches of the tree were estimated by bootstrap analysis with 500 replications (CLUSTAL V; Higgins *et al.*, 1992).

The relatively low frequency of nucleotide differences prompted the calculation of the average divergence of the group from several other closely related genes of other species (Table 2). The sequence from an equivalent region of a single MW/LW sensitive photoreceptor gene (560 nm) of the marmoset (Williams *et al.*, 1992), was included to test the accuracy of the tree. The sequence of the LW photoreceptor gene of the chicken (Kuwata et al., 1990) was used as an outgroup.

3.3 Results

3.3.1 Sex of the primates

As seen in Figure 3.5, a band of the target size, in the corresponding lane indicates that the animal from which the sample of gDNA was obtained was a male; absence of a band of the correct size suggests that that animal lacked a Y chromosome, and thus by inference was a female. DNAs of known sex were included as positive controls. The human female (lane 3), and male (lane 6) did provide the appropriate result. These primers were designed based on the human sequence, therefore they were expected to work only with the male. The results of lanes 1, 2, 7, and 9 show that these primers were able to differentiate correctly between OWM of known sex. Finally, lanes 4, 8, 10, and 11 show the sex of the unknown individuals; orang-utan was a male, as was the chimpanzee, whereas, the gorilla and baboon were females. The experiment was repeated a second time at a later period and gave identical results. The findings are summarised in Table 3

3.3.2 Opsin gene sequences of the primates

For all eight primate samples (includes human control) an intensely staining bands of the expected size (164 bp for exon 3, 156 bp for exon 4, and 191 bp for exon 5) was obtained (data not shown). Sequencing of the clones confirmed the presence of both MW and LW opsin products in the amplified fragments from each PCR. Some clones (a minority) gave sequences that were neither MW or LW opsin fragments, nor did they correspond to any other opsin-like gene. These were considered to be

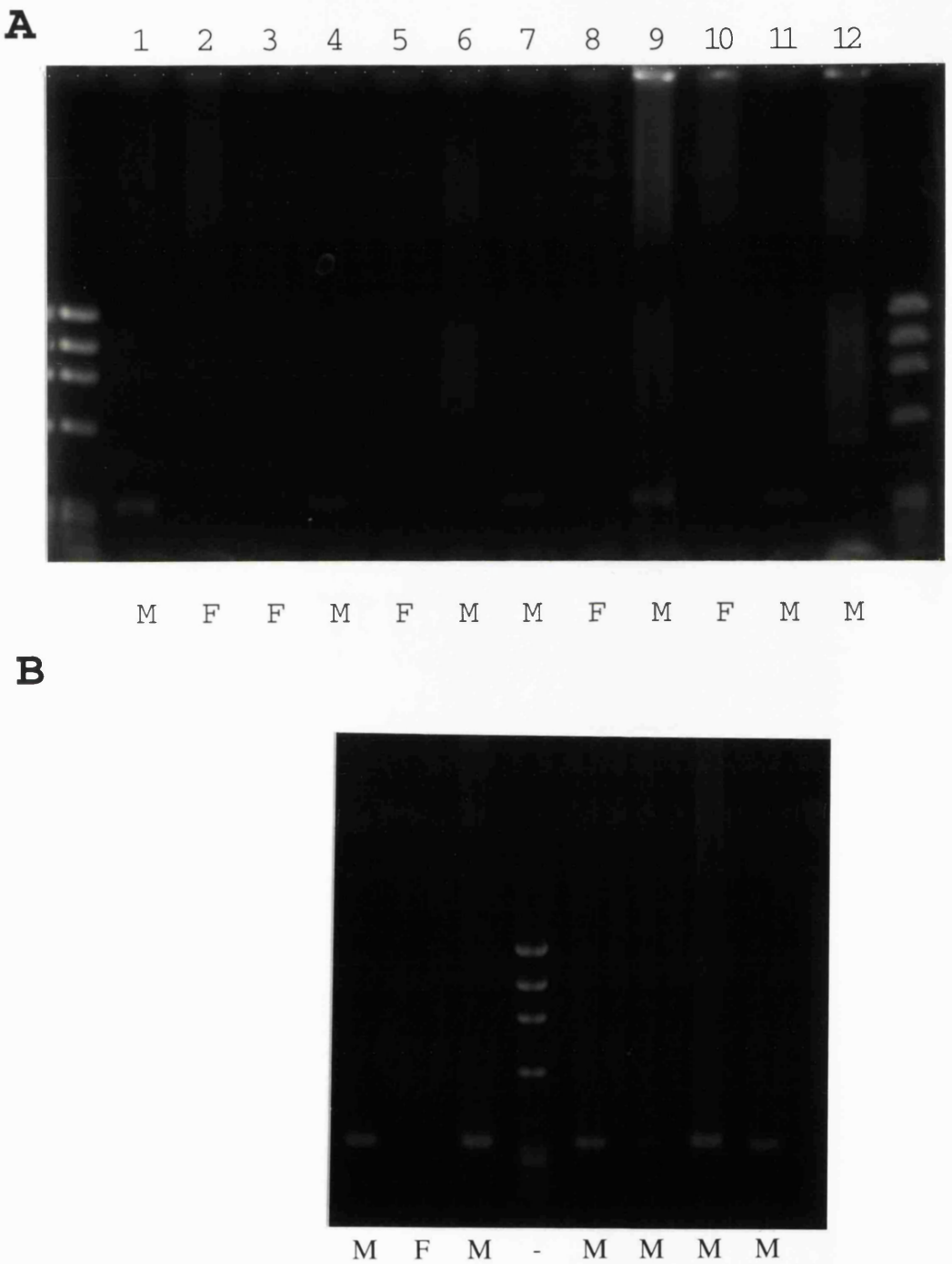


Figure 3.5 Agarose gels showing the amplification of a 330 bp fragment from Y chromosomes, using sex specific primers. Only those DNA samples from male animals produce bands. A). Controls and unknowns tested. The order of loading and the results are given in Table 3. B). Chimpanzee DNAs from animals of known sex PCR'd with the same primers. M= male, F =female. The marker used were ϕX174.

Lane #	Scientific name	Common name	Sex
1	*Cercopithecus petaurista*	Spot-nosed monkey	Male
2	*Cercopithecus diana*	Diana monkey	Female
3	*Homo sapiens*	- Control (human female)	Female
4	*Pongo pygmaeus*	Orang-utan ape	Male
5	*Miopithecus talapoin*	Talapoin monkey- 1	Female
6	*Homo sapiens*	+ Control (human male)	Male
7	*Cercopithecus aethiops*	African green monkey	Male
8	*Gorilla gorilla*	Gorilla ape	Female
9	*Miopithecus talapoin*	Talapoin monkey- 2	Male
10	*Papio papio*	Baboon monkey	Female
11	*Pan troglodytes*	Chimpanzee ape	Male

Table 3 Sex of the primates used in this study

generated by interaction with plasmid or chromosomal DNA from the host bacteria, or a co-amplified non-target region from the original genomic template.

Figures 3.6 to 3.17 show the nucleotide and amino acid sequences of exons 3, 4, and 5 of the MW and LW photopigment genes of the seven species of Old World primates. In each case the sequence has been compared to the equivalent human sequence, derived from published data (Nathans *et al.* 1986a) and from PCR amplified fragments obtained from human DNA using the same primer sets. Although the sequences of exons 4 and 5 of the LW and MW opsin genes of *M. talapoin*, *E. patas*, *C. diana*, and *C. pataurista* monkeys has been reported previously (Ibbotson, 1991) these regions were sequenced again during the course of the present work for the sake of completeness, and as confirmation of the previous work. The sequences are described as MW or LW on the basis of homology with the corresponding human sequences as originally reported by Nathans *et al.* (1986a).

Figure 3.6 LW Exon 3 NUCLEOTIDE SEQUENCE

```
                474-
Human           GCC ATC ATT TCC TGG GAG AGG TGG CTG GTG
P. troglodytes  ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... ...
E. patas        ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ...

Human           GTG TGC AAG CCC TTT GGC AAT GTG AGA TTT
P. troglodytes  ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ..C ... ... ...
C. diana        ..C ... ... ... ... ... ..C ... ... ...
E. patas        ..C ... ... ... ... ... ..C ... ... ...
M. talapoin     ..C ... ... ... ... ... ..C ... ... ...
C. pataurista   ..C ... ... ... ... ... ..C ... ... ...
C. aethiops     ..C ... ... ... ... ... ..C ... ... ...

Human           GAT GCC AAG CTG GCC ATC GTG GGC ATT GCC
P. troglodytes  ... ... ... ... ... ... A.T ... ... ...
G. gorilla      ... ... ... ... ... ... A.T ... ... ...
C. diana        ... ... ... ..C ... ... ... ... ... ...
E. patas        ... ... ... ..C ... ... ... ... ... ...
M. talapoin     ... ... ... ..C ... ... ... ... ... ..T
C. pataurista   ... ... ... ..C ... ... ... ... ... ...
C. aethiops     ... ... ... ..C ... ... ... ... ... ...

                                                    -596
Human           TTC TCC TGG ATC TGG TCT GCT GTG TGG ACA GCC
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... .A.
C. diana        ... ... ... ... ... ... ... ... ... ..G ...
E. patas        ... ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ... ...
```

Figure 3.7 LW Exon 3 deduced AMINO ACID SEQUENCE

	145-									
Human	ala	ile	ile	ser	trp	glu	arg	trp	leu	val
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	val	cys	lys	pro	phe	gly	asn	val	arg	phe
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	asp	ala	lys	leu	ala	ile	val	gly	ile	ala
P. troglodytes	-	-	-	-	-	-	ile	-	-	-
G. gorilla	-	-	-	-	-	-	ile	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

										-185
Human	phe	ser	trp	ile	try	ser	ala	val	try	thr	ala
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	asp
C. diana	-	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-

Figure 3.8 MW Exon 3 NUCLEOTIDE SEQUENCE

```
                474-
Human           GCC ATC ATT TCC TGG GAG AGA TGG ATG GTG
P. troglodytes  ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... ...
E. patas        ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ...

Human           GTC TGC AAG CCC TTT GGC AAT GTG AGA TTT
P. troglodytes  ... ... ... ... ... ... ..C ... ... ...
G. gorilla      ... ... ... ... ... ... ..C ... ... ..C
C. diana        ... ... ... ... ... ... ..C ... ... ...
E. patas        ... ... ... ... ... ... ..C ... ... ...
M. talapoin     ... ... ... ... ... ... ..C ... ... ...
C. pataurista   ... ... ... ... ... ... ..C ... ... ...
C. aethiops     ... ... ... ... ... ... ..C ... ... ...

Human           GAT GCC AAG CTG GCC ATC GTG GGC ATT GCC
P. troglodytes  ... ... ... ... ... ... ... ... ... .T.
G. gorilla      ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ..C ... ... ... ... ... ...
E. patas        ... ... ... ..C ... ... ... ... ... ...
M. talapoin     ... ... ... ..C ... ... ... ... ... .T.
C. pataurista   ... ... ... ..C ... ... ... ... ... ...
C. aethiops     ... ... ... ..C ... ... ... ... ... ...

                                                    -596
Human           TTC TCC TGG ATC TGG GCT GCT GTG TGG ACA GCC
P. troglodytes  ... ... ... G.. ... ... ... ... ... ... ...
G. gorilla      ... ... ... G.. ... ... ... ... ... ... ...
C. diana        ... ... ... G.A ... ... ... ... ... ..G ...
E. patas        ... ... ... G.A ... ... ... ... ... ..G ...
M. talapoin     ... ... ... G.A ... ... ... ... ... ... ...
C. pataurista   ... ... ... G.C ... ... ... ... ... ... ...
C. aethiops     ... ... ... G.C ... ... ... ... ... ... ...
```

Figure 3.9 MW Exon 3 deduced AMINO ACID SEQUENCE

	145-									
Human	ala	ile	ile	ser	trp	glu	arg	trp	met	val
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	val	cys	lys	pro	phe	gly	asn	val	arg	phe
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	asp	ala	lys	leu	ala	ile	val	gly	ile	ala
P. troglodytes	-	-	-	-	-	-	-	-	-	val
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	val
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

										-185
Human	phe	ser	trp	ile	try	ala	ala	val	try	thr	ala
P. troglodytes	-	-	-	val	-	-	-	-	-	-	-
G. gorilla	-	-	-	val	-	-	-	-	-	-	-
C. diana	-	-	-	val	-	-	-	-	-	-	-
E. patas	-	-	-	val	-	-	-	-	-	-	-
M. talapoin	-	-	-	val	-	-	-	-	-	-	-
C. pataurista	-	-	-	val	-	-	-	-	-	-	-
C. aethiops	-	-	-	val	-	-	-	-	-	-	-

Figure 3.10 LW Exon 4 NUCLEOTIDE SEQUENCE

```
                 651-
Human            GGC CCA GAC GTG TTC AGC GGC AGC TCG TAC
P. troglodytes   ... ... ... ... ... ... ... ... ... ...
G. gorilla       ... ... ..T ... ... ... ... ... ... ...
C. diana         ... ... ..T ... ... ..T ... ... ... ...
E. patas         ... ... ..T ... ... ..T ... ... ... ...
M. talapoin      ... ... ..T ... ... ..T ... ... ... ...
C. pataurista    ... ... ..T ... ... ..T ... ... ... ...
C. aethiops      ... ... ..T ... ... ..T ... ... ... ...

Human            CCC GGG GTG CAG TCT TAC ATG ATT GTC CTC
P. troglodytes   ... ... ... ... ... ... ... ... ... ...
G. gorilla       ... ... ... ... ... ... ... ... ... ...
C. diana         ... ... ... ... ... ... ... ... ... ...
E. patas         ... ... ... ... ... ... ... ... ... ...
M. talapoin      ... ... ... ... ... ... ... ... ... ...
C. pataurista    ... ... ... ... ... ... ... ... ... ...
C. aethiops      ... ... ... ... ... ... ... ... ... ...

Human            ATG GTC ACC TGC TGC ATC ATC CCA CTC GCT
P. troglodytes   ... ... ... ... ... ... ... ... ... ...
G. gorilla       ... A.. ... ... ... ... ... ... ... ..C
C. diana         ... ... ... ... ... ... ..T ... ..G ...
E. patas         ... ... ... ... ... ... ..T ... ..G ...
M. talapoin      ... ... ... ... ... ... ..T ... ..G ...
C. pataurista    ... ... ... ... ... ... ..T ... ..G ...
C. aethiops      ... ... ... ... ... ... ..T ... ..G ...

                                                  -764
Human            ATC ATC ATG CTC TGC TAC CTC CAA
P. troglodytes   ... ... ... ... ... ... ... ...
G. gorilla       ... ... G.. ... ... ... ... ...
C. diana         ... ... ... ... ... ... ... ...
E. patas         ... ... ... ... ... ... ... ...
M. talapoin      ... ... ... ... ... ... ... ...
C. pataurista    ... ... ... ... ... ... ... ...
C. aethiops      ... ... ... ... ... ... ... ...
```

Figure 3.11 LW Exon 4 deduced AMINO ACID SEQUENCE

	204-									
Human	gly	pro	asp	val	phe	ser	gly	ser	ser	tyr
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	pro	gly	val	gln	ser	tyr	met	ile	val	leu
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	met	val	thr	cys	cys	ile	ile	pro	leu	ala
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	ile	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

								-241
Human	ile	ile	met	leu	cys	tyr	leu	gln
P. troglodytes	-	-	-	-	-	-	-	-
G. gorilla	-	-	val	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-

Figure 3.12 MW Exon 4 NUCLEOTIDE SEQUENCE

```
                651-
Human           GGC CCA GAC GTG TTC AGC GGC AGC TCG TAC
P. troglodytes  ... ... ..T ... ... ..T ... ... ... ...
G. gorilla      ... ... ..T ... ... ... ... ... ... ...
C. diana        ... ... ..T ... ... ..T ... ... ... ...
E. patas        ... ... ..T ... ... ..T ... ... ... ...
M. talapoin     ... ... ..T ... ... ..T ... ... ... ...
C. pataurista   ... ... ..T ... ... ..T ... ... ... ...
C. aethiops     ... ... ..T ... ... ..T ... ... ... ...

Human           CCC GGG GTG CAG TCT TAC ATG ATT GTC CTC
P. troglodytes  ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... ...
E. patas        ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ...

Human           ATG GTC ACC TGC TGC ATC ACC CCA CTC AGC
P. troglodytes  ... A.. ..A ... ... T.. C.. ... ... ...
G. gorilla      ... A.. ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... .C.
E. patas        ... ... ... ... ... ... ... ... ... .C.
M. talapoin     ... ... ... ... ... ... ... ... ... .C.
C. pataurista   ... ... ... ... ... ... ... ... ... .C.
C. aethiops     ... ... ... ... ... ... ... ... ... .C.

                                                -764
Human           ATC ATC GTG CTC TGC TAC CTC CAA
P. troglodytes  ..T ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ..T ... ... ...
E. patas        ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ...
```

Figure 3.13 MW Exon 4 deduced AMINO ACID SEQUENCE

	204-									
Human	gly	pro	asp	val	phe	ser	gly	ser	ser	tyr
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	pro	gly	val	gln	ser	tyr	met	ile	val	leu
P. troglodytes	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-

Human	met	val	thr	cys	cys	ile	thr	pro	leu	ser
P. troglodytes	-	ile	-	-	-	phe	pro	-	-	-
G. gorilla	-	ile	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	thr
E. patas	-	-	-	-	-	-	-	-	-	thr
M. talapoin	-	-	-	-	-	-	-	-	-	thr
C. pataurista	-	-	-	-	-	-	-	-	-	thr
C. aethiops	-	-	-	-	-	-	-	-	-	thr

								-241
Human	ile	ile	val	leu	cys	tyr	leu	gln
P. troglodytes	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-

Figure 3.14 LW Exon 5 NUCLEOTIDE SEQUENCE

```
                831-
Human           AAG GAA GTG ACG CGC ATG GTG GTG GTG ATG ATC TTT GCG
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... A.. ... ... ... ..C ...
E. patas        ... ... ... ... ... ... ... A.. ... ... ... ..C ...
M. talapoin     ... ... ... ... ... ... ... A.. ... ... ... ..C ..A
C. pataurista   ... ... ... ... ... ... ... A.. ... ... ... ..C ..A
C. aethiops     ... ... ... ... ... ... ... A.. ... ... ... ..C ..A

Human           TAC TGC GTC TGC TGG GGA CCC TAC ACC TTC TTC GCA TGC
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... ... ... ... ...
E. patas        ... ... ... ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... ..T ... ... ... ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ... ... ... ...

Human           TTT GCT GCT GCC AAC CCT GGT TAC GCC TTC CAC CCT TTG
P. troglodytes  ... ... ... ... ... ... ..C ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ..C ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ..C ... ... ... ... ... C..
E. patas        ... ... ... ... ... ... ..C ... ... ... ... ... C..
M. talapoin     ... ... ... ... ... ... ..C ... ... ... ... ... C..
C. pataurista   ... ... ... ... ... ... ..C ... ... ... ... ... C..
C. aethiops     ... ... ... ... ... ... ..C ... ... ... ... ... C..

                                                          -980
Human           ATG GCT GCC CTG CCG GCC TAC TTT GCC AAA AGT
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... T.. ..A ... ... ... ... ... ...
E. patas        ... ... ... T.. ..A ... ... ... ... ... ...
M. talapoin     ... ... ... T.. ..A ... ... ... ... ... ...
C. pataurista   ... ... ... ... ..A ... ... ... ... ... ...
C. aethiops     ... ... ... ... ..A ... ... ... ... ... ...
```

Figure 3.15 LW Exon 5 deduced AMINO ACID SEQUENCE

	264-												
Human	lys	glu	val	thr	arg	met	val	val	val	met	ile	phe	ala
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	met	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	met	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	met	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	met	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	met	-	-	-	-	-

Human	tyr	cys	val	cys	trp	gly	pro	tyr	thr	phe	phe	ala	cys
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-	-	-

Human	phe	ala	ala	ala	asn	pro	gly	tyr	ala	phe	his	pro	leu
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-	-	-

											-313
Human	met	ala	ala	leu	pro	ala	tyr	phe	ala	lys	ser
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-

Figure 3.16 MW Exon 5 NUCLEOTIDE SEQUENCE

```
                831-
Human           AAG GAA GTG ACG CGC ATG GTG GTG GTG ATG GTC CTG GCA
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... ... ... T.. ... ..G
E. patas        ... ... ... ... ... ... ... ... ... ... T.. ... ..G
M. talapoin     ... ... ... ... ... ... ... ... ... ... T.. ... ..G
C. pataurista   ... ... ... ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ... ... ... ...

Human           TTC TGC TTC TGC TGG GGA CCA TAC GCC TTC TTC GCA TGC
P. troglodytes  ... ... ... ... ... ... ..G ..T ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ..C ... ... ... ... ... ...
C. diana        ... ... ... ... ... ... ..T ... ... ... ... ... ...
E. patas        ... ... ... ... ... ... ..T ... ... ... ... ... ...
M. talapoin     ... ... ... ... ... ... ..T ... ... ... ... ... ...
C. pataurista   ... ... ... ... ... ... ..T ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ..T ... ... ... ... ... ...

Human           TTT GCT GCT GCC AAC CCT GGC TAC CCC TTC CAC CCT TTG
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ..G ... ... ... ... G.. ... ... ... ...
C. diana        ... ... ... ... ... ... ... ... G.. ... ... ... C..
E. patas        ... ... ... ... ... ... ... ... G.. ... ... ... C..
M. talapoin     ... ... ... ... ... ... ... ... G.. ... ... ... C..
C. pataurista   ... ... ... ... ... ... ... ... G.. ... ... ... C..
C. aethiops     ... ... ... ... ... ... ... ... G.. ... ... ... C..

                                                          -980
Human           ATG GCT GCC CTG CCG GCC TTC TTT GCC AAA AGT
P. troglodytes  ... ... ... ... ... ... ... ... ... ... ...
G. gorilla      ... ... ... ... ... ... ... ... ... ... ...
C. diana        ... ... ... T.. ..A ... .A. ... ... ... ...
E. patas        ... ... ... ... ... ... ... ... ... ... ...
M. talapoin     ... ... ... T.. ..A ... .A. ... ... ... ...
C. pataurista   ... ... ... ... ... ... ... ... ... ... ...
C. aethiops     ... ... ... ... ... ... ... ... ... ... ...
```

Figure 3.17 MW Exon 5 deduced AMINO ACID SEQUENCE

	264-												
Human	lys	glu	val	thr	arg	met	val	val	val	met	val	leu	ala
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-	phe	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	phe	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-	-	-

Human	phe	cys	phe	cys	trp	gly	pro	tyr	ala	phe	phe	ala	cys
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	-	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	-	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	-	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	-	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	-	-	-	-	-

Human	phe	ala	ala	ala	asn	pro	gly	tyr	pro	phe	his	pro	leu
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	ala	-	-	-	-
C. diana	-	-	-	-	-	-	-	-	ala	-	-	-	-
E. patas	-	-	-	-	-	-	-	-	ala	-	-	-	-
M. talapoin	-	-	-	-	-	-	-	-	ala	-	-	-	-
C. pataurista	-	-	-	-	-	-	-	-	ala	-	-	-	-
C. aethiops	-	-	-	-	-	-	-	-	ala	-	-	-	-

											-313
Human	met	ala	ala	leu	pro	ala	phe	phe	ala	lys	ser
P. troglodytes	-	-	-	-	-	-	-	-	-	-	-
G. gorilla	-	-	-	-	-	-	-	-	-	-	-
C. diana	-	-	-	-	-	-	tyr	-	-	-	-
E. patas	-	-	-	-	-	-	tyr	-	-	-	-
M. talapoin	-	-	-	-	-	-	tyr	-	-	-	-
C. pataurista	-	-	-	-	-	-	tyr	-	-	-	-
C. aethiops	-	-	-	-	-	-	tyr	-	-	-	-

A complication with this assignment is that the human gene is known to be highly polymorphic for exon 3 (Nathans *et al.*, 1986a) and for codon 180 in particular, such that two different MW and two different LW pigments are found in the human population (Winderickx *et al*, 1992; Neitz *et al.*, 1993; and Chapter 5). The association of a particular exon 3 with either a MW or LW exon 4 was not established in this study so the classification of exon 3 sequences as either MW or LW is based solely on homology with the original sequences reported by Nathans *et al.* (1986a). What is clear however is that the two types of exon 3, which code for either serine or alanine at position 180, are present throughout the Old World primates and are not just a feature of the human visual pigment genes.

Amino acid sequence differences among the opsin photopigments of the OWM and apes are shown in Table 4, and silent nucleotide substitutions are shown in Table 5. The coding sequences of the LW and MW opsin genes of the monkeys, chimpanzee, and gorilla are highly homologous to those of humans, which would suggest that their colour vision most probably parallels our own. Initial examination of Tables 4 and 5 reveals several striking features. Firstly, a significant number of both coding (non-synonymous; whereby the change results in the coding of a different amino acid) and silent nucleotide substitutions (synonymous) can be grouped into two categories which demarcates the Hominoids from the Cercopithecoids. Base changes, at positions 155 (the number refers to the amino acid position in the opsin molecule, using the cone opsin numbering system) where OWM possess the same codon in both pigment genes compared to apes, 168 (both opsin genes of the OWM exhibit a third base change in the codon; in apes it is CTG and for OWM it is CTC), 233 (amino acid threonine replaces serine in the MW opsin of OWM), 271 (methionine replaces valine in the LW opsin of OWM), 274 (the phenylalanine present in OWM is replaced by valine for MW opsins), 302 (apes exhibit the codon TTG whilst OWM show CTG for both opsin classes), and 309 (both opsin genes of OWM code for tyrosine) were seen only in the opsin genes of OWM. Also revealed were four differences present in at least two OWM but not across all five members, but not observed with the apes (at codons 184, 276, 306, and 307) .

	Amino acid position																		
	153	171	174	178	180	185	225	229	230	233	236	271	274	275	277	279	285	298	309
Hum LW	leu	val	ala	ile	ser	ala	val	ile	ile	ala	met	val	ile	phe	tyr	val	thr	ala	tyr
P. troglodytes		ile																	
G. gorilla		ile				asp	ile				val								
E. patas												met							
M. talapoin												met							
C. petaurista												met							
C. aethiops												met							
C. diana												met							
Hum MW	met	val	ala	ile	ala	ala	val	ile	thr	ser	val	val	val	leu	phe	phe	ala	pro	phe
P. troglodytes			val	val			ile	phe	pro										
G. gorilla				val			ile											ala	
E. patas				val						thr			phe					ala	tyr
M. talapoin			val	val						thr			phe					ala	tyr
C. petaurista				val						thr			phe					ala	tyr
C. aethiops				val						thr			phe					ala	tyr
C. diana				val						thr			phe					ala	tyr

Table 4. Amino acid differences amongst the opsin photopigments of OWM and apes

Amino Acid #	155	161	164	168	174	184	206	209	226	234	238	276	280	283	284	293	296	302	306	307
Hum LW	GTG	AAT	TTT	CTG	GGC	ACA	GAC	AGC	ACC	ATC	TGC	GCG	TGC	CCC	TAC	GCC	GGT	TTG	CTG	CCG
P. troglodytes																	GGC			
G. gorilla		AAC					GAT										GGC			
E. patas	GTC	AAC		CTC			GAT	AGT									GGC	CTG	TTG	CCA
M. talapoin	GTC	AAC		CTC	GCT		GAT	AGT				GCA	TGT				GGC	CTG	TTG	CCA
C. petaurista	GTC	AAC		CTC			GAT	AGT				GCA					GGC	CTG		CCA
C. aethiops	GTC	AAC		CTC			GAT	AGT				GCA					GGC	CTG		CCA
C. diana	GTC	AAC		CTC		ACG	GAT	AGT									GGC	CTG	TTG	CCA
Hum MW	GTC	AAT	TTT	CTG	GCC	ACA	GAC	AGC	ACC	ATC	TGC	GCA	TGC	CCA	TAC	GCC	GGC	TTG	CTG	CCG
P. troglodytes		AAC			GTC		GAT	AGT	ACA	ATT				CCG	TAT					
G. gorilla		AAC	TTC				GAT							CCC		GCG				
E. patas		AAC		CTC		ACG	GAT	AGT				GCG		CCT				CTG		
M. talapoin		AAC		CTC	GTC		GAT	AGT				GCG		CCT				CTG	TTG	CCA
C. petaurista		AAC		CTC			GAT	AGT						CCT				CTG		
C. aethiops		AAC		CTC			GAT	AGT						CCT				CTG		
C. diana		AAC		CTC		ACG	GAT	AGT			TGT	GCG		CCT				CTG	TTG	CCA

Table 5. Silent nucleotide substitutions amongst the opsin photopigments of the OWM and apes

Secondly, nucleotides unique to human opsins were evident - a base change at codon position 296 of the LW gene and positions 161, 178, 206, and 283 of the MW gene were not shared with any of the non-human primates. Other sporadic base changes were also evident from the data, but these reveal no pattern in their distribution. Comparing the LW opsin gene to the MW opsin gene, with regard to the number of silent site substitutions, it was apparent that the frequency of occurrence, for the primates examined as a group, was approximately the same for each gene.

Certain nucleotide substitutions result in amino acid changes. This in itself is not the critical factor influencing spectral characteristics. The only amino acids that are of interest are those which involve either a change of charge or the loss of an hydroxyl group, and those that occupy strategic positions facing the chromophore binding pocket. Amino acid residues unique to the non-human primates were found at five positions (171, 185, 225, 236, and 271) in the LW, and nine positions (174, 178, 225, 229, 230, 233, 274, 298, and 309) in the MW photopigments. Site 236 of the LW and sites 298 and 309 of the MW were occupied by the reciprocal residue from the other pigment, and therefore can be eliminated from playing a part in any differences in the primates.

The deduced LW opsin amino acid sequences from these primates showed few differences. As Table 4 shows, the MW amino acid sequence was much more variable between the various primates. Some amino acid substitutions were restricted to either the Cercopithecoid or the Hominoid lineages (Table 4). Serine-233, valine-274 and phenylalanine-309 are found only in the Hominoid MW gene, whereas methionine-271 is present only in the LW gene of the Cercopithecoids. Of these sites, only 233 has been implicated in spectral tuning (Williams *et al.*, 1992); the replacement of threonine by serine at this site in Hominoids is considered to be a conservative change (Stryer, 1988), whereby there is no change in the λ_{max}. Work by Merbs and Nathans (1993) indicates that substitution of serine with alanine at this site, as is the case between the two pigment classes, only produces less than a 1 nm spectral shift. The Old World species exhibit no variation in the remaining three amino acids (specified by codons

180, 277 and 285) thought to be critical for spectral tuning (Neitz *et al.*, 1991; Williams *et al.*, 1992; Merbs and Nathans, 1992a). The λ_{max} of the MW (533 nm) and LW (563 nm) pigments of the six species of Cercopithecoid monkeys (Bowmaker *et al.*, 1980; Bowmaker *et al.*, 1991) are known to be very similar to those of man (Bowmaker , 1980) and, given the conserved amino acids at positions 180, 277 and 285 in the Great Apes, we would predict that the MW and LW pigments of gorilla and chimpanzee would have similar spectral characteristics to these Old World monkeys and to man. It would appear unlikely that any change in the spectral characteristics of the MW and LW pigments has arisen from the folivorous life style of the gorilla.

3.3.3 Multiple opsin genes

To address the question of polymorphic genes, numerous clones from each opsin type were sequenced and the data compared. A number of polymorphisms in human MW and LW opsin gene sequences have been reported (Nathans *et al.*, 1986a; Winderickx *et al.*, 1993), and a similar pattern of polymorphisms has been found in non-human primates. The haplotypes of the polymorphic exons are shown in Table 6. The polymorphisms in the talapoin monkey are restricted to exon 3, whereas the chimpanzee and gorilla are polymorphic for both exons 3 and 4, although for both species, the polymorphism in exon 4 is limited to a single site. No polymorphic variation was found in exon 5 of any species examined. Similar findings have been reported for other OW primates (orang-utans (Deeb *et al.*, 1994) and OWM crab-eating macaque (Dulai *et al.*, 1994)). This bunching of polymorphic sites in exon 3 (Figure 3.18) is also seen in humans (Nathans *et al.*, 1986a; Winderickx *et al.*, 1993), with eight sites at nucleotide positions 494, 498, 500, 506, 552, 554, 562 and 573 common to humans and Old World primates. Since the alternative nucleotide present is in all cases identical to that present in the other type of opsin gene, the production of these polymorphisms by intergenic recombination would appear to be the most likely mechanism. Such exchanges must occur therefore much more frequently in exon 3 than in either exon 4 or 5, and since only in a few cases are the positions of these

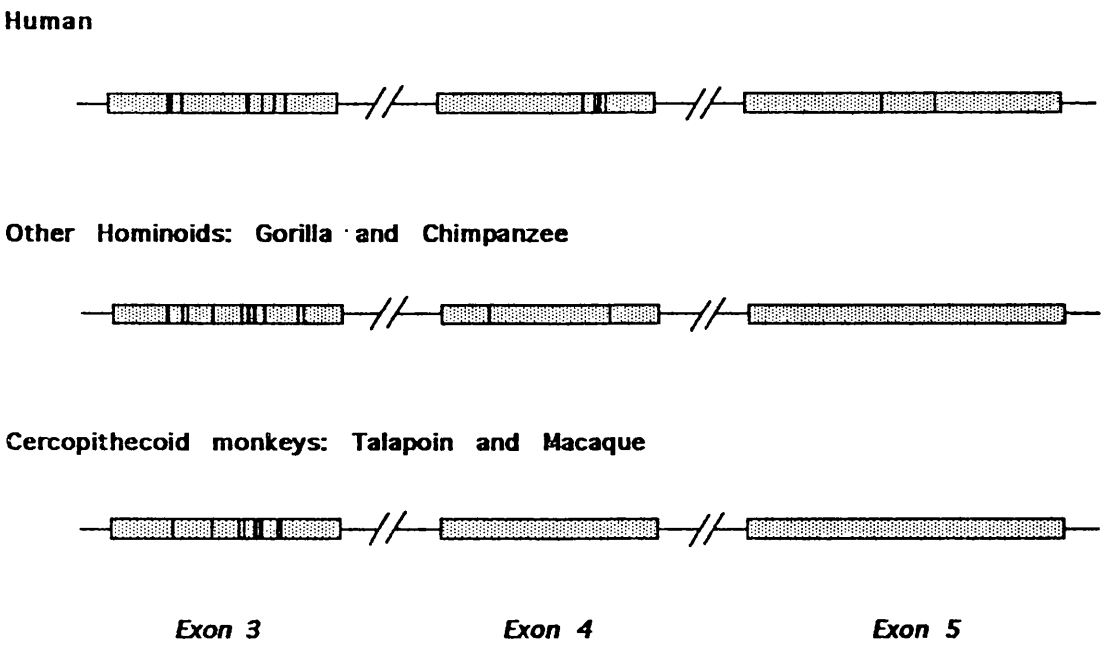


Figure 3.18 Position of polymorphic sites in the MW and LW genes of Old World primates. The human polymorphic sites are taken from previously published data (Nathans *et al.*, 1986a; Winderickx *et al.*, 1993).

polymorphic sites the same in the different species, the production of such exchanges must be an ongoing process. Winderickx *et al.* (1993) have identified a sequence (from 497 to 504) that resembles the hypervariable minisatellite sequence. This sequence has been shown to promote recombination (Wahls *et al.*, 1990) and has been found at certain chromosomal breakpoints (Kenter and Birshtein, 1981; Krowczynska *et al.*, 1990). This region is polymorphic at site 498 but is otherwise conserved across the five species of Old World monkeys examined here. The high rate of intergenic exchanges seen in this region may be the consequence therefore of the presence of this sequence (Winderickx *et al.*, 1993).

The sex of the animals studied is critical to the understanding of these multiple sequences. In the gorilla, the sequence analysis of exon 3 identified two MW class and one LW class sequences, but exon 4 analysis revealed just one MW and two LW sequences (see Table 6). Since the animal studied was a female and not a male, the presence of multiple sequences is less informative because of the presence of two X chromosomes. In this case the results could arise either due to the presence of multiple LW and MW genes on one or both X chromosomes, or because each chromosome carries a different version of these genes. The situation is clearer in the case of the chimpanzee, basically because it is a male. The identification of two LW exon 3 sequences implies the existence of at least two LW genes on the one X chromosome, in addition to the MW gene. Combining this data with that obtained from sequencing of exon 4, suggests that the X chromosome array consists of a minimum of two LW and two MW genes. However, for the reasons discussed above, either of the exon 3 haplotypes identified as LW in the male chimpanzee could be part of a MW gene. These results are consistent therefore with the original suggestion (Nathans *et al.*, 1986a) that only the MW gene can be present in multiple copies. The male talapoin studied, has at least three MW genes and a LW gene. The multiple exon 3 sequences that were found in the chimpanzee and talapoin monkey indicate therefore that these animals carry three and four opsin genes respectively.

			Nucleotide site																		
			494	498	506	524	528	545	551	552	554	557	558	562	563	566	573	575	587	593	659	758
EXON 3	*P. troglogytes*	MW							C	G	G	C	A			C			G	A		
									~	val	~	~	ile			~			~	~		
		LW1							C	A	T	C	A			C			G	A		
									~	ile	~	~	ile			~			~	~		
		LW2							T	G	G	A	G			T			T	G		
									~	val	~	~	val			~			~	~		
	G. gorilla	MW1	A	A	C		G															
			~	met	~		gly															
		MW2	G	C	G		A															
			~	leu	~		arg															
		LW	G	C	G		A															
			~	leu	~		arg															
	M. talapoin	MW1				T		G						C	C		A	C				
						~		~						ala	~		ile	~				
		MW2				T		C						C	T		A	C				
						~		~						ala	~		ile	~				
		MW3				C		C						T	C		G	A				
						~		~						val	~		val	~				
		LW				C		C						C	T		A	C				
						~		~						ala			ile	~				
EXON 4	*P. troglogytes*	MW1																			C	
																					~	
		MW2																			T	
																					~	
		LW																			C	
																					~	
	G. gorilla	MW																				T
																						~
		LW1																				T
																						~
		LW2																				C
																						~

Table 6. Haplotypes of polymorphic exons. ~ indicates synonymous change.

These observations are consistent with the previous identification of multiple opsin genes in humans (Nathans *et al.*, 1986a; Drummond-Borg *et al.*, 1989), and in talapoin monkeys (Ibbotson *et al.*, 1992). In both cases, it was thought that only MW sequences are present in multiple copies, although an alternative hypothesis that multiple LW genes can also be present has now been proposed by Neitz and Neitz (1992, 1995). In either case, the conclusion that multiple opsin genes are present in Cercopithecoid and Hominoid primates is not affected.

The lack of such gene multiplicity in the other monkeys could be due to a number of reasons; no gene multiplicity exists; or one does exist but was not detected by the number of clones sequenced. Another possible reason could be because only one or two individuals from each species was examined. It may be that not every member of the species possesses multiple copies, and those which do may only have a single haplotype across all their different genes. More genes from many individuals of each species need to be examined, before any further conclusions can be drawn.

3.3.4 Evolution of opsin sequences

The reconstruction of phylogenetic trees is a statistical problem (Cavalli-Sforza and Edwards, 1967), whereby the tree that is constructed is just an estimate of the real tree, with a given topology and given branch lengths. Evolutionary biologists often want to determine the history of species: determining when the species split and the divergence times following each split. A species tree is a historical representation of the true lineage.

Figure 3.19 shows the phylogenetic tree constructed from the nucleotide sequence data generated during this study, and the average number of substitutions per site is given in Table 2. The proportion of silent substitutions between the different primate species was considered to be too few for a meaningful analysis to be performed using this data alone. Therefore, all substitutions (silent and coding) was considered.

From examination of the rooted phylogenetic tree (see Figure 3.19) a number of features are apparent: (1) the expected early branching between the chicken red opsin

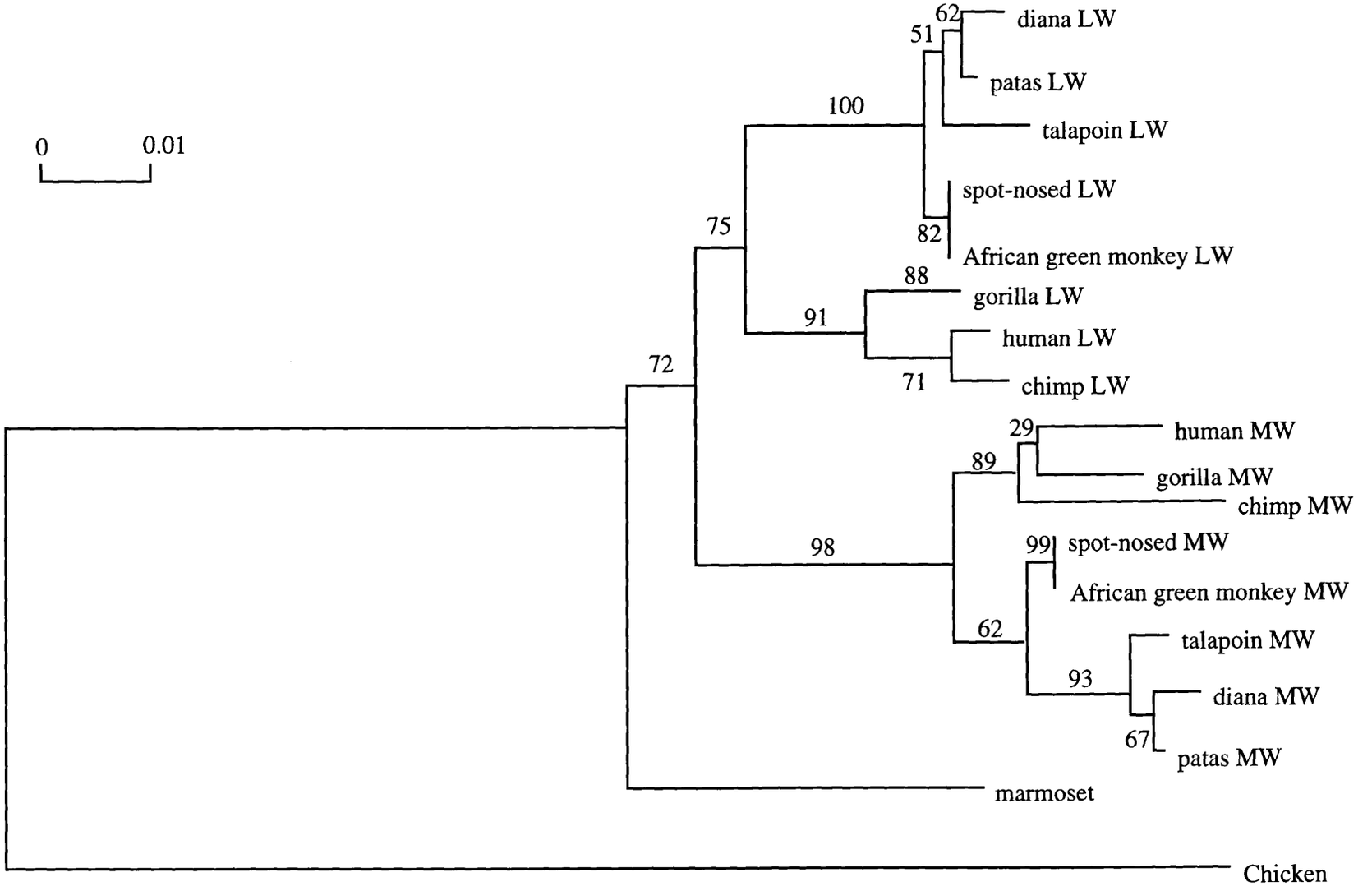


Figure 3.19 Rooted phylogenteic tree generated from sequence divergence of the MW and LW opsin genes of Old World primates, a marmoset gene, and the chicken iodopsin gene. Bootstrap values are given for each branch. The average number of substitutions per site for each branch of the tree can be calculated using the scale bar.

Table 2. Proportion of nucleotide differences among the primates sequenced

OTU Labels

1.. HumanLW
2.. ChimpLW
3.. GORILLALW
4.. DianaLW
5.. PatasLW
6.. TalapoinLW
7.. Spot-NosedLW
8.. AGMLW
9.. HumanMW
10.. ChimpMW
11.. GorillaMW
12.. DianaMW
13.. PatasMW
14.. TalapoinMW
15.. Spot-nosedMW
16.. AGMMW
17.. MARMO
18.. IODOP

Distance: Proportion of nucleotide differences. No. of Codons in Subset: 129 of 129. Codon Position(s) Used: 1 2 3 . Gap Sites and Missing Information Data: All such sites were removed from the subset data

Distances in the upper-right matrix. Standard Errors in lower-left matrix

'*' indicates an invalid distance value

OTUs	1	2	3	4	5	6	7	8
1		0.0078	0.0233	0.0362	0.0336	0.0413	0.0336	0.0336
2	0.0045		0.0155	0.0388	0.0362	0.0439	0.0362	0.0362
3	0.0077	0.0063		0.0439	0.0413	0.0491	0.0413	0.0413
4	0.0095	0.0098	0.0104		0.0026	0.0103	0.0078	0.0078
5	0.0092	0.0095	0.0101	0.0026		0.0078	0.0052	0.0052
6	0.0101	0.0104	0.0110	0.0051	0.0045		0.0078	0.0078
7	0.0092	0.0095	0.0101	0.0045	0.0036	0.0045		0.0000
8	0.0092	0.0095	0.0101	0.0045	0.0036	0.0045	0.0000	
9	0.0113	0.0115	0.0120	0.0134	0.0132	0.0134	0.0127	0.0127
10	0.0138	0.0140	0.0132	0.0144	0.0142	0.0144	0.0138	0.0138
11	0.0123	0.0125	0.0115	0.0134	0.0132	0.0134	0.0127	0.0127
12	0.0129	0.0132	0.0127	0.0113	0.0115	0.0123	0.0120	0.0120
13	0.0125	0.0127	0.0123	0.0113	0.0115	0.0123	0.0120	0.0120
14	0.0127	0.0129	0.0125	0.0115	0.0113	0.0120	0.0118	0.0118
15	0.0125	0.0127	0.0123	0.0123	0.0120	0.0123	0.0115	0.0115
16	0.0125	0.0127	0.0123	0.0123	0.0120	0.0123	0.0115	0.0115
17	0.0127	0.0129	0.0120	0.0118	0.0120	0.0127	0.0120	0.0120
18	0.0203	0.0204	0.0204	0.0206	0.0206	0.0208	0.0205	0.0205

OTUs	9	10	11	12	13	14	15	16
1	0.0517	0.0801	0.0620	0.0698	0.0646	0.0672	0.0646	0.0646
2	0.0543	0.0827	0.0646	0.0724	0.0672	0.0698	0.0672	0.0672
3	0.0594	0.0724	0.0543	0.0672	0.0620	0.0646	0.0620	0.0620
4	0.0749	0.0879	0.0749	0.0517	0.0517	0.0543	0.0620	0.0620
5	0.0724	0.0853	0.0724	0.0543	0.0543	0.0517	0.0594	0.0594
6	0.0749	0.0879	0.0749	0.0620	0.0620	0.0594	0.0620	0.0620
7	0.0672	0.0801	0.0672	0.0594	0.0594	0.0568	0.0543	0.0543
8	0.0672	0.0801	0.0672	0.0594	0.0594	0.0568	0.0543	0.0543
9		0.0310	0.0207	0.0439	0.0388	0.0413	0.0258	0.0258
10	0.0088		0.0284	0.0517	0.0465	0.0439	0.0336	0.0336
11	0.0072	0.0084		0.0413	0.0362	0.0388	0.0233	0.0233
12	0.0104	0.0113	0.0101		0.0052	0.0078	0.0181	0.0181
13	0.0098	0.0107	0.0095	0.0036		0.0078	0.0129	0.0129
14	0.0101	0.0104	0.0098	0.0045	0.0045		0.0155	0.0155
15	0.0081	0.0092	0.0077	0.0068	0.0057	0.0063		0.0000
16	0.0081	0.0092	0.0077	0.0068	0.0057	0.0063	0.0000	
17	0.0142	0.0136	0.0138	0.0136	0.0136	0.0138	0.0140	0.0140
18	0.0211	0.0211	0.0209	0.0210	0.0209	0.0208	0.0209	0.0209

OTUs	17	18
1	0.0672	0.1990
2	0.0698	0.2016
3	0.0594	0.2016
4	0.0568	0.2067
5	0.0594	0.2067
6	0.0672	0.2119
7	0.0594	0.2041
8	0.0594	0.2041
9	0.0853	0.2222
10	0.0775	0.2222
11	0.0801	0.2145
12	0.0775	0.2171
13	0.0775	0.2145
14	0.0801	0.2119
15	0.0827	0.2145
16	0.0827	0.2145
17		0.2041
18	0.0205	

(outgroup) and the primate opsin genes; (2) the separation of the NWM (marmoset) opsin sequence from those of the OW primates; (3) the distinction made between the LW and MW opsin sequences of the OW primates; (4) the clustering of related OW species in both opsin classes, with the exception of the human, chimpanzee, and gorilla clade, where data from the tree places the human and chimpanzee LW sequences together but places the human and gorilla MW sequences together. However, considering the bootstrap support value for the MW branch point in particular (29 %), it can be seen that this order in very tentative.

The tree can be considered as representing a true relationship based on present knowledge of the evolution of vertebrates and of opsin genes. Opsins represent a class of protein with an evolution that predates the separation of the vertebrates and invertebrates (Yokoyama, 1995). This is indicated by the similarities in sequence between all the opsins sequenced to date (Yokoyama, 1995). The chicken is a member of the bird (avian) family, while the primates are members of the mammalian group. The chicken iodopsin (red) sequence roots the tree for this reason. The OW primate LW and MW opsin genes arose due to a duplication event sometime after the separation of the primates into OW and NW families. This feature is illustrated in the tree by the branching of the marmoset sequence prior to the separation of the LW and MW genes of OW primates. The duplicated LW and MW genes passed to the present day species from a common ancestor who possessed this duplication. The genes have diverged slightly from each other during this period. This is also shown by the branching points on the tree. The tree can be considered as true because it maintains the grouping of the three great apes (human, chimpanzee, and gorilla), as suggested strongly by other lines of evidence (Nowak, 1991). The tree also shows that the different clades maintain the grouping of patas, diana, and talapoin monkeys as well as the spot-nosed and the African green monkeys, for both the LW and MW branches.

3.5 Discussion

3.5.1 Sequence comparisons

A high degree of homology is observed in the coding sequences of the LW and MW opsin genes between the OW primates examined. The first feature of note is that the visual pigments of the great apes and OW monkeys would be expected to have λ_{max} values very similar to those of humans based on the conservation of the amino acids at sites 277 and 285. These sites have been strongly implicated for the majority of the (30 nm) difference that exists between the λ_{max} values of the MW and LW photopigments (Asenjo *et al.*, 1994; Neitz *et al.*, 1991; Neitz and Neitz, 1995).

The separate MW and LW sequences of Old World primates appear only after the divergence of New and Old World primates. Each gene then shows a separate lineage into the Cercopithecoid and Hominoid branches. This is consistent with the notion that the duplication event that gave rise to these separate genes occurred early in the evolution of Old World primates, before the separation of the Cercopithecoid and Hominoid lineages.

At present it is unclear why all NWM, with the exception of the howler monkeys (*Alouatta sp.*), still enjoy colour vision based on one X-linked opsin locus in addition to an autosomal one. Amongst mammals just the primates are trichromatic (Jacobs, 1993), thus a pathway leading from a single X-chromosome pigment gene, with a single allele, to the multiple opsin genes of the OWM must have evolved. Considering the high sequence homologies of the X-chromosome linked opsin genes, a gene duplication event is indicated as an intermediale step in the pathway (Nathans *et al.*, 1986a). The discovery of widespread polymorphic colour vision among platyrrhines suggests that the evolution of trichromatic colour vision may have occurred via the polymorphic colour vision system (Bowmaker *et al.*, 1987), which avoids the need for convergent evolution. An alternative hypothesis is that the polymorphic colour vision of NWM may be a degenerative form of trichromacy (Mollon, 1991b). The lines of evidence at present suggests that the former mechanism may be the true route by

which trichromacy has been attained. Considering the problem from a phylogenetic point of view, the data suggests that an ancestral trichromacy would have to have been independently lost in at least three lineages while the evolution of trichromacy from a polymorphic base would have required change in only one line. The molecular analysis of the howler monkey's opsin genes is equivocal at best. Jacobs *et al.* (1996) provide support for these monkeys acquiring their trichromacy from a polymorphic past, based on the sequences of exon 5. However, the molecular analysis of a larger region of these genes by Dulai and Hunt (unpublished data) challenges this finding based on silent site and coding substitutions. Further, the marmoset (New World) gene is represented by three different allelic variants that specify pigments with λ_{max} of 563, 556 and 543 nm respectively, and the phylogenetic analysis indicates that these variants arose some time after the establishment of the New World lineage (Williams *et al.*, 1992; Hunt *et al.*, 1993b; Shyue *et al.*, 1995). Since Old and New World primates depend on a common set of amino acid substitutions to achieve the spectral shifts of the MW (543 nm pigment in marmoset) and LW (563 nm pigment in marmoset) visual pigments (Neitz *et al.*, 1991; Ibbotson *et al.*, 1992; Williams *et al.*, 1992; Winderickx *et al.*, 1992c), this would imply that the mechanism of spectral tuning in Old and New World primates arose separately in the two lineages; the use of a common set of amino acid sites in the spectral tuning process must be the result therefore of convergent evolution (Shyue *et al.*, 1995). The phylogenetic analysis may underestimate however the antiquity of the three marmoset sequences. Since these are allelic variants of a single polymorphic gene, the process of recombination will tend to limit sequence divergence, whereas the separate MW and LW genes of Old World primates will undergo exchange only by the considerably less frequent process of gene conversion (Balding *et al.*, 1992; Winderickx *et al.*, 1993). It is possible therefore that the appearance of this allelic variation pre-dates the separation of Old and New World lineages. In this case, different spectral forms of the opsin gene would have been present in the ancestral primate, and an unequal exchange that placed two different

allelic forms of this gene on to a single X chromosome would have resulted in the retention of the same key amino acid substitutions in both lineages.

Do other opsin genes from distantly related organisms support the view that convergent evolution is operating to spectrally tune the various photopigments? In combination with the above primate data and the analysis of fish LW opsins (Yokoyama and Yokoyama, 1990) it is apparent that sites 180, 277, and 285 have been independently used to achieve the same spectral positioning of the corresponding pigments. The work of Hunt *et al.* (1996) and Morris *et al.* (1993) further implicates site 277 in the spectral tuning of the rod opsin of fresh water fish and cephalopods, respectively.

CHAPTER 4

Sequences Upstream of the X-Linked Opsin Genes of Higher Primates

4.1 Introduction

What mechanism is responsible for determining the opsin class that is expressed by each photoreceptor cell? The mature retina consists of a diverse set of differentiated cell types (see section 1.3). Each type has its place within this highly ordered system, performing its own unique function. The human fovea is occupied exclusively by MW and LW cones, whereas, in the peripheral retina SW cones and rod cells are present also (Bowmaker *et al.*, 1980). During development, most of the different cell types of the retina, including photoreceptor, bipolar, amacrine and Müller cells appear to be derived from common progenitor cells (Turner and Cepco, 1987). It has been suggested that the pattern of retinal cell differentiation is regulated by cell-cell interactions and local environmental factors (Turner *et al.*, 1990). Rod photoreceptors are among the last neuronal cell types to form in the mammalian retina. In rats and mice, the major period of rod photoreceptor generation begins at about embryonic day 19 (E19), reaches a peak at postnatal day 1 (PN1) to PN2 and is over by PN8. Lineage studies using retroviral markers have shown that rod photoreceptors originate from the same neuroepithelial progenitors as other retinal neurones and glia, and that the decision to become a rod is not made until or just after the final mitosis. One of the earliest cell type-specific genes expressed by differentiating rod photoreceptors is opsin. Rod opsin can be detected immunocytochemically in the plasma membranes of immature rods within approximately 48 hr after the cells become postmitotic (Barnstable, 1987; Watanabe and Raff, 1990).

The general principle of molecular biology stipulates that DNA produces RNA which in turn produces protein. The transcription process, whereby an RNA message is produced from the DNA, is a major control point for regulating the expression of genes

in particular cell types or in response to a particular signal. Although mechanisms exist for post-transcriptional regulation (e.g. RNA splicing or RNA message stability) the major control point lies at the level of transcription. Transcription of many eukaryotic genes is regulated by the combination of multiple sequence-specific DNA-protein(s) interactions. In most cases the DNA elements involved in these interactions are located upstream of the genes. Some of these *cis* -acting sequences are common to many genes, whereas others have a more limited distribution (Maniatis *et al.*, 1987).

Promoters are located near the site where RNA polymerase binds to initiate the transcription of the gene. Their position relative to the start site is relatively fixed. The TATA and CAAT boxes are promoter elements. This region abutting the coding region of the gene is generally known as the proximal promoter.

Enhancers are DNA sites that may be far removed from the promoter, but which increase the rate of initiation. They are generally situated in the distal 5' flanking region of the gene, and are orientation-independent. Enhancers can stimulate any promoter placed in their vicinity, thus many are regarded as generic in architecture, any one may be identified in many locations. Regulatory factors, such as hormone receptors or transcription factors, bind and thereby facilitate transcription of the gene. For example, when a steroid hormone binds to its receptor, the complex then binds to enhancers of specific genes.

The distinction between a promoter and an enhancer is generally based on the distance from the site of initiation. Other than this enhancers can be regarded as a clustering of promoter elements, with the ability to function at increased distances from the site of message initiation, and in either orientation.

Transcription factors are the protein components which interact with each other and the sequence elements on the DNA, promoting (or hindering) gene expression. Eukaryotic transcription factors fall into three main structural classes; zinc-finger proteins, leucine zipper proteins, and helix-loop-helix proteins. Zinc-finger proteins possess motifs which, in conjunction with a zinc atom, result in the formation of a secondary structure (a finger) that interacts directly with DNA. The number of "fingers"

in different proteins can vary. Examples include the Kruppel factor found in the fruit fly, *Drosophila* (Miller and Bieker, 1993), steroid hormone receptors (Collins *et al.*, 1989), and Sp1 factors (Pugh and Tijian, 1990). All leucine zipper proteins exhibit a domain consisting of a stretch of amino acids with leucine in every seventh position. To act as transcription factors the proteins must dimerise via the leucine zipper, either as homodimers or as heterodimers. Examples include the CAAT/enhancer-binding proteins (C/EBP), Fos, and Jun. The Ap1 factors, mentioned latter, consist of a heterodimer of Fos and Jun units. Helix-loop-helix proteins share a common type of sequence motif: a stretch of 40-50 amino acids containing two amphipathic α-helices separated by a linker region of varying length. These proteins must also dimerise prior to interaction with DNA. Examples include the E12/E47, that is expressed ubiquitously, and the MyoD protein, which is expressed in a tissue-specific manner.

Exactly how particular combinations of transcription factors are used to produce strikingly distinct temporal and spatial patterns of gene expression in the co-ordination of retinal cell differentiation, is poorly understood.

A common feature amongst multicellular organisms is the differentiation of cells allowing them to function differently from their neighbours. Most of these cells still possess a full complement of genes, but are able to express unique, as well as common, subsets of proteins. The mechanisms underlying gene regulation are being deciphered.

In the vast majority of situations where a protein product is generated in a particular tissue or set of tissues, or in a temporal manner, it is the result of control processes acting on the corresponding gene, allowing transcription to proceed only in that tissue at the appropriate time (for review see Darnell, 1982; Latchman, 1995). Genes regulated in this manner are referred to as "tissue specific", as opposed to those which are constitutively expressed in all tissues at all times; the so called "house-keeping" genes.

What are the molecular events that give each of these cells its distinctive identity, that is what makes a cell a cone cell and not a rod? To approach this question at a genetic level, several groups have begun to dissect the cis-acting regulatory sequences

involved in cell-type-specific gene expression in the retina. Sequences controlling expression of bovine and murine rod opsin in rods, the human red and green cone pigment genes in cones, and retinal bipolar cells have been partially defined using reporter constructs in transgenic mice.

Using chimaeric genes consisting of 4.4 kb of mouse rod opsin 5' flanking sequence fused to a diphtheria toxin gene and 4.4 kb or 500 bp of rod opsin 5' flanking sequence fused to the *E.coli lacZ* gene, Lem *et al.* (1991) demonstrated the targeting of photoreceptor cells. Whilst the 4.4 kb construct expression followed temporal and spatial gradients of expression that approximated opsin expression, the 500 bp fragment was weaker and non-uniform, suggesting that the more distal elements were necessary for enhanced and uniform spatial expression. Similarly, a transgene generated by Chen *et al.* (1994) containing either 3.8 or 1.1 kb of the human SW opsin 5' flanking region, directed expression to both SW photoreceptors and to bipolar cells when tested in transgenic mice.

Wang *et al.* (1992) undertook a equivalent series of experiments whereby transgenic mice carrying sequences upstream of the human LW pigment gene fused to a beta-galactosidase reporter gene were generated, to test whether these sequences were required for transcription of the adjacent visual pigment genes in cone photoreceptors. The patterns of transgene expression indicated that a region between 3.1 kb and 3.7 kb 5' of the red pigment gene transcription initiation site, but 43 kb from the nearest MW gene, is essential for expression.

4.1.2 Sequence motifs upstream of retinal genes

The richness of human, bovine, and mouse retinas in rod opsin has meant that the majority of the work in the past has concentrated on the analysis of the regulation of this gene and the associated phototransduction pathway genes. Thus although most of the transcription elements and factors are unique to the rod opsin system, several motifs are shared amongst a large number of retinal genes, not only in mammals but also in invertebrates such as *Drosophila.*

Electrophoretic gel mobility shift and DNase I protection experiments have identified a *cis*-acting sequence, Ret-I, in the rat rod opsin 5' flanking region, 131 bp from the transcription start site (Morabito *et al.*, 1991). This sequence, 5'-GGAAGCCAATTAA-3', is thought to confer rod cell specificity. It has also been reported in the proximal promoter region of the bovine opsin gene (DesJardin and Hauswirth, 1996). Interestingly, DesJardin *et al.* (1996) further report the existence of elements likely to bind negative transcriptional regulatory factors. These elements, C-rich elements (CRS's), exhibited an inverse relationship to rhodopsin gene expression in protection assays, localised to regions upstream of the CAAT box. In the same report another element, bovine A/T-rich sequence (BAT) was identified. These elements are considered to be important in the developmental regulation of rod opsins.

Analysis of the region upstream of the mouse α-transducin gene has identified a cis-acting element, Tα-1, which shows homology to the Ret-1 sequence. This element was able to direct tissue-specific expression, and binds a retina-specific nuclear factor (Ahmad *et al.*, 1994). Also the work on β-PDE genes (Farber, 1995) has resulted in the detection of potential proximal regulatory elements (HS-like, PCE 1-like, and ATF -like), as well as, the presence of c-myc and Sp1 elements.

A comparison of the sequences immediately flanking the human LW and MW opsin genes shows almost complete identity from the transcription start site (at base +1) to base -195 (7 % of bases differ), beyond which any homology is lost (Nathans *et al.*, 1986a). Within this region there are present a number of identifiable sequence motifs. The TATA box was identified by Nathans *et al.* (1986A) at position -29 to -24. A CAAT box is also apparent at site -103 to -99. The high degree of homology between these two genes makes it difficult to identify other sequence motifs.

The TATA box, an AT-rich sequence (consensus TATAA/TAA/T), is found about 30 base pairs upstream of the transcriptional start site of most eukaryotic genes, but is absent in some genes, notably housekeeping genes expressed in all tissues and in some tissue-specific genes (Latchman, 1995). In promoters that contain a TATA box

and in those that lack it, the low activity of the promoter itself is significantly increased by upstream promoter elements (UPE). UPEs are found in a wide variety of genes with different patterns of expression, and play a role in stimulating the constitutive activity of promoters. Two well characterised UPE are the Sp1 box, which has the consensus sequence GGGCGG, and the CCAAT box, both are typically found upstream of the promoter in many genes both with and without TATA boxes (Dynan and Tjian, 1985). The Sp1 protein is present in all cell types. The DNA-binding activity of the protein encoded by the Sp1 gene is dependent upon a region at the C-terminus which contains three zinc finger motifs. The ability of the protein to activate transcription has been mapped to two glutamine-rich regions of the protein which act by interacting with protein subunits associated with the TATA binding protein (Courey and Tjian, 1988; Hoey *et al.*, 1993). The CCAAT box binds a number of different proteins, some of which are expressed in all tissues whilst others are expressed in a tissue-specific manner (Latchman 1995). Some genes may have both of these elements, whereas others have single or multiple copies of one or the other. These elements are essential for transcription of the genes, and their elimination by deletion or mutation abolishes transcription (McKnight and Kingsbury, 1982; McKnight *et al.*, 1984). Hence these UPEs play an essential role in efficient transcription of the gene.

Several known transcription factor binding motifs have been identified which are conserved in photoreceptor upstream regions of both cone and rod cells. Sp1 control sequences have been shown to footprint in mammalian promoters although the site was first identified in the simian virus 40 (SV40) promoter (Dynan and Tjian, 1983). The prominence of the GC box in the related sequences suggested an important role in binding, which was confirmed by dimethyl sulphate protection (Gidoni *et al.*, 1984). The sites, present in the herpes simplex virus "immediate-early" gene, function in an orientation-independent manner in activating transcription mediated via a TATA box (Jones and Tjian, 1985). The factor itself has been isolated and is composed of two polypeptides, 105 and 95 kD in size, activating transcription via one or more sequence elements in the promoter (Briggs *et al.*, 1986).

The Sp1 transcription factor itself is ubiquitous (Faisst and Meyer, 1992). Its recognition site in DNA is required to drive the accurate transcriptional initiation of the HPRT gene (Dynan, 1986). Such promoters bind a transcriptional complex which contains TATA binding protein and its associated factors (Pugh and Tjian, 1991). The proximity of the site to the start of the opsin transcript suggests a role for the Sp1 site in recruiting the transcriptional machinery.

Ap2 sites may bind a dimerised 52 kD protein which activates transcription in SV40 and human metallothionein IIa promoters (Mitchell *et al.,* 1987). It's overlap with the Sp1 site may indicate some competition between the two proteins for their binding motifs. Ap2 is expressed most abundantly in cells of the neural crest lineage and nephric tissues and is developmentally regulated during embryogenesis. Ap2 levels peak at mid-term pregnancy in the mouse and decline thereafter (Mitchell *et al.,* 1991). The possible effect of Ap2 on transcription in neural and kidney tissue of an adult is not known. Ap2 activates transcription in transduction pathways stimulated by phorbol ester/diacyl glycerol/protein kinase C pathway and protein kinase A/cAMP elevation (Imagawa *et al.,* 1987) which may indicate an inducible property of the opsin promoters.

The AP-1 site represents the core sequence for the group of basic helix-loop-helix transacting factors. Examples of these include the muscle-specific MyoD (Blackwell and Weintraub, 1990), *c-Myc* (Blackwell *et al.,* 1990) and two genes involved in neural development, Mammalian *achaete-scute* homologue-1 (Mash-1) (Guillemot *et al.,* 1993) and the microphthalmia (*mi*) gene (Hodgkinson *et al.,* 1993). The conserved site in the opsin promoter corresponds closely to that of MyoD but very slight alterations could drastically alter the specificity to match that of *c-Myc*. The sequence of the *mi* gene product binding site is not known, but this type of site may be important in the differentiation and perhaps maintenance of specific cell types, notably muscle and neural tissue.

The importance of a region a considerable distance upstream of the LW opsin gene in regulating the X-linked opsins was implicated by observations that deletions far

upstream resulted in a condition known as blue cone monochromacy, a disorder in which both red and green cone functions are absent (see section 1.10.1). Sequences within this region are highly conserved among humans, mice, and cattle, even though the latter two species have only a single visual pigment gene at this locus (Wang *et al.*, 1992). This region has since been referred to as the locus control region (LCR), in line with the findings of similar sequences upstream of globin (Kielman *et al.*, 1994) and the immunoglobulin heavy-chain genes (Madisen and Groudine, 1994). The locus control region of the human beta-globin cluster consists of four major DNase I hypersensitive sites (HS) (Chang *et al.*, 1992). When linked to globin genes, the locus control region conferred a high level of erythroid- specific expression of these genes in transgenic mice or transfected erythroid cell lines. In particular a 36-bp LCR core fragment, containing the NFE-2/AP-1 binding consensus, increased the overall level of expression by several fold (Chang *et al.*, 1992). A similar core fragment, a 37-mer, has also been reported to exist in the LCR regions from human, mouse, and bovine (Wang *et al.*, 1992).

What is the function of the LCR region? Numerous reports pertaining to the functionality of the globin and other LCRs may shed light on the mechanism by which the opsin LCR affects transcription of the LW and MW genes. The human β-globin LCR has two important activities. First, it opens a 200 kb chromosomal domain containing the human epsilon-, gamma- and beta-globin genes and, secondly, it functions as a powerful enhancer of globin gene expression (Caterina *et al.*, 1994a). DNase I hypersensitive site 2 (HS 2) of the human beta-globin LCR directs high level expression of the beta-globin gene located 50 kb downstream. Experiments in cultured cells and in transgenic mice demonstrate that duplicated AP1-like sites in HS 2 are required for this powerful enhancer activity. A cDNA clone encoding a basic leucine-zipper protein that binds to these sites was isolated and designated Locus Control Region-Factor 1 (LCR-F1). These results suggested that LCR-F1 may be a critical factor involved in LCR-mediated, human globin gene expression (Caterina *et al.*, 1994b). A position-independent enhancer effect has been proposed for the LCR,

whereby the inclusion of the human β-globin LCR region was demonstrated to dampen chromosomal position effects and promote high-level expression of a reporter construct inserted into mouse genomes (Yu *et al.*, 1994).

The purpose of this work is aimed at defining cis-acting regulatory sequences that control the cell type-specific expression of opsin genes. By cloning and sequencing the 5' flanking region of the human MW gene it should allow several objectives to be fulfilled; it will allow the comparison of this region to that upstream of the LW opsin gene, for identification of possible conserved motifs which may be involved in the regulation of these cell types. It will allow comparison with homologous regions upstream of the OW primate opsin genes, and from regions flanking the NW primate X-chromosome opsin gene locus, enabling a broader comparison to be made. And it may also enable the ancestral X-linked primate opsin gene to be determined by comparing the region upstream of the NW monkey locus to that of the OW regions.

4.2 Methods

4.2.1 Genomic Vectorette PCR

Vectorette PCR (known officially as Chemical Genetics; see section 2.3.5.1) kits were supplied by Cambridge Research Biochemicals Ltd. The Vectorette method was chosen because it appeared to offer the most robust approach with the least genetic manipulation.

4.2.2 Human gDNA template for Vectorette PCR

Human gDNA (KSD) was used as a positive control template on which to test the viability and efficiency of the Vectorette method. Approximately 1 ug of gDNA from a human male (KSD) was digested separately with the restriction enzymes *Bam*HI, *Eco*RI, *Hind*III, and *Rsa*I, each in a total volume of 50 µl, as detailed in section 2.3.5.2, using 20 U of enzyme. 15 µl aliquots of these digests were size

fractionated on a 0.8 % agarose gel; the resultant smear indicated that the DNA had digested properly (data not shown).

Vectorette libraries were constructed as stated in section 2.3.5.2, using appropriate Vectorette linkers and the restriction enzymes, *Bam*HI, *Eco*RI, *Hind*III and *Rsa*I, which cut within known sequence upstream of exon 1 of the LW opsin gene (Figure 4.1).

Aliquots from both the human *Bam*HI and *Hind*III Vectorette libraries were amplified using the Vectorette PCR 1 primer and primer LW5'Fla-L(-555), which primes to LW 5' flanking regions only (Figure 4.2a).

LW5'FLa-S(-553)	5'-AGA CTG ACT GCG CAC ATC C-3'
LW5'FLa-L(-555)	5'-GCA GAC TGA CTG CGC ACA TCC TG-3'

The gene specific primer was designed with thermal characteristics which matched the Vectorette PCR 1 primer as closely as possible to allow a very stringent annealing/extension temperature of 72 °C to be used, and was chosen to selectively amplify a region upstream of the LW opsin gene for two important reasons; 1) the distance from the priming site of this primer to the known restriction sites for *Eco*RI and *Hind*III is easily within the capacity of the PCR reaction (approximately 1.4 and 0.5 kb, respectively), and 2) to deter the formation of complex banding patterns should a priming site common to both the LW and MW opsin genes be selected (since multiple MW genes can exist). Following first round Vectorette PCR, products were separated on a 1 % agarose gel.

Any discrete bands were subjected to 3MM cleaning (section 2.4.4.3), prior to second round nested Vectorette PCR using the Vectorette sequencing primer and the nested specific primer LW5'Fla-S(-553), which primers to a site within the sequence of primer LW5'Fla-L(-555) (see Figure 4.2a), at the following cycling parameters : denature at 95°C for 5 minutes, then 40 cycles of 94°C / 30 seconds, 66°C / 60 seconds, and 72°C / 180 seconds. A final extension of 10 minutes at 72°C was included. Any target bands were excised, placed in 100 µl distilled water, to elute (section 2.4.4.1),

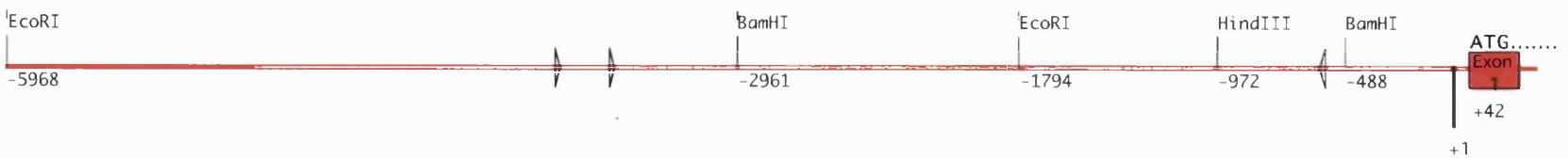


Figure 4.1 Restriction map of the 5' flanking region of the human LW opsin gene. The numbers refer to the base upstream of the transcription start site +1. The left facing arrow identifies the site to which the specific primer employed in the Vectorette PCR primes. The two right facing arrows delimit the LCR region. (Restriction site data from Nathans *et al*, 1986a).

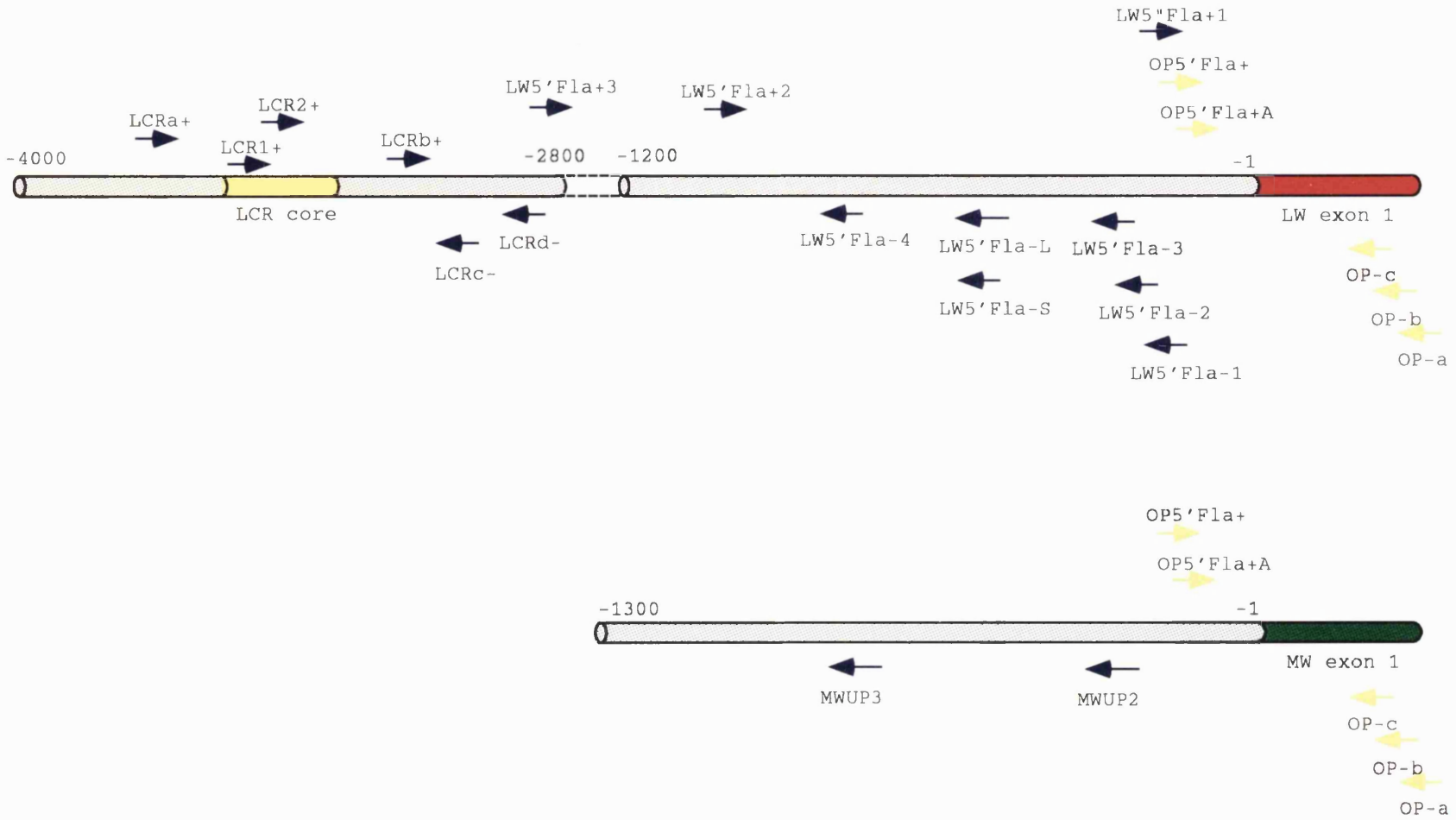


Figure 4.2 Priming sites of the various 5' flanking primers used in the study. The arrows in blue represent primers that are specific to either the LW or MW opsin, whilst those in yellow can co-amplify both opsin sequences.

ligated into pCRII™ cloning vector (section 2.4.5), and subsequently sequenced manually (section 2.6.3.3).

An additional *Cla*1 library was constructed (section 2.3.5.2) and subjected to amplification under the same conditions as the previous libraries.

4.2.3 Generation and amplification of non-human primate Vectorette libraries

Vectorette libraries were generated with gDNA from several OW primates (talapoin, diana, and patas monkeys), and two NW individuals (marmoset 15 and marmoset 16). Genomic DNAs from these animals was digested with *Bam*HI, *Eco*RI, and *Hind*III and ligated to appropriate Vectorette linkers, as per section 4.2.2 above. All possible Vectorette linkers (*Bam*HI, *Cla*I, *Eco*RI, *Hind*III, and blunt ended) were tried. DNAs from a number of species of OW primate and NW primate were re-digested and ligated to Vectorette units a number of times. All PCR products were visualised on agarose gels.

Modifications, such as changing the concentration of various components and using different cycling profiles, employed by other laboratories (Dr. Oglivie; personnel communication), were attempted too.

Trials with inverse PCR were conducted. The method outlined in Stemmer and Morris (1992) was followed, which required that the gDNA be fully digested and upon ligation, for which the dilution step is most critical, that individual molecules circularise, and as with the Vectorette method, that the distance over which amplification is desired is not outside the scope of the PCR reaction.

4.2.4 Screening a human cosmid genomic library for MW upstream

A human cosmid genomic library in vector Lorist B (sections 2.5.1.1), was obtained from Dr. J. Fitzgibbon. The goal was to screen this library with a probe, generated from an opsin gene coding region, in order to isolate individual cosmids containing the human MW opsin gene along with a substantial portion of its 5' upstream flanking region. The strategy entailed the use of an exon 3 probe, that would

co-select both the MW and LW opsin genes, during the primary screen (exon 1 primers were not available at the time). Cosmids containing MW gene fragments would then be differentiated from those containing LW inserts via a PCR screen. Individual positive MW clones would be segregated by secondary and tertiary hybridisation screening. PCR analysis would determine whether or not they contained exon 1 of opsin genes and therefore, probably the desired upstream sequences.

4.2.5 Titrating the cosmid library

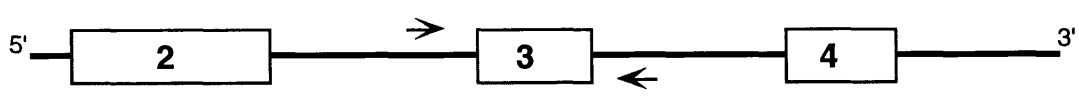
It is general practice to titre libraries prior to their screening, to establish the quantity needed to obtain a satisfactory plating density. However, on this occasion, since the library was being actively used in the laboratory at the time these experiments commenced, the amount needed for a specific density of colonies was already known. The titrating step was therefore omitted. Plating 90 µl of this cosmid library was expected to produce approximately 100,000 clones per 23 cm^2 plate.

4.2.6 PCR screen for LW or MW exon 3

Primers NaOp3+/NaOp3- which had previously been reported to amplify a specific band from genomic DNA (Nathans *et al.*, 1993; hybridising to gDNA 50 bp upstream of exon 3, in intron 2, and 50 bp into intron 3, amplifying a fragment of 269 bp),

NaOp3+ 5'-GTC TAA GCA GGA CAG TGG GAA GCT TTG CTT-3'

NaOp3- 5'-ACG TAA AAG CTT TAA GGT CAC AGA GTC TGA CC-3'



were used to test aliquots of the library (section 2.5.1.6). Ten aliquots, each of 1 µl, were subjected to a standard PCR reaction (section 2.3.2), products of which were size fractionated on an 1.8 % agarose gel and visualised.

4.2.7 Primary screening by hybridisation

Using the library titre as an estimate, approximately 200,000 colonies were plated out on two large 20 x 20 cm filters, A and B (section 2.5.1.1), i.e. 100,000 per filter giving a probability of detection of the desired sequence of over 95 %. Conditions were as described in section 2.5.1. Replica filters were generated from each master and probed with exon 1 regions (see section 2.5.1.1). The filters were washed down to 0.5 X SSC; 0.1 % SDS at 65°C. Following overnight exposure at -80°C, with intensifying screens, the autoradiographs were developed and aligned.

For each instance where a strong duplicated signal was produced on the replica filters, an area corresponding to the 2 mm radius (diameter of a Pasteur pipette) was excised from the master plate and placed into 200 µl of LB broth and grown overnight in a 37°C shaker. Before further analysis a series of PCR amplifications were conducted to confirm the presence of opsin sequences in these positive clones (see below).

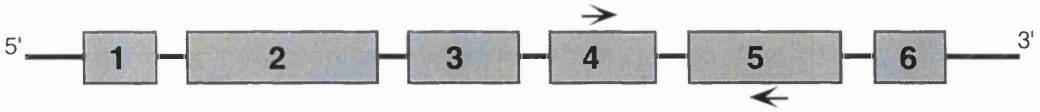
4.2.8 PCR confirmation and classification of the primary lift colonies

Each primary colony pool was subjected to a PCR screen using two sets of primers. The first set (NaOp3+/NaOp3-) was the same as that used for the pre-plating PCR screen. These primers co-amplify both the MW and LW opsin genes, thus they would only indicate the presence of X-linked opsin sequences in the clones.

Since the aim of the experiment was to amplify regions in and around the MW opsin gene, it was necessary to differentiate between those clones carrying the LW opsin sequence from those with the MW sequence. It was decided that the most informative set of primers to use at this stage would be OpMW4+/OpMW5-, since they were readily available.

OpMW4+ 5'-GCT GCA TCA CCC CAC TCA G-3'

OpMW5- 5'-GAA GCA GAA TGC CAG GAC C-3'



This primer pair were designed following their use in another study (Winderickx *et al.*, 1992c) to specifically amplify the MW opsin gene region from the centre of exon 4, across intron 4, to the centre of exon 5. These would confirm the presence of MW opsin gene inserts in the cosmids. Once again the products of the PCR reactions on each colony pool were resolved on an 1.2 % agarose gel.

4.2.9 Further screening by hybridisation and PCR

Following titering of positive pools to determine their density, the desired dilutions were plated out and processed in the same fashion as the primary screen master filters. Only filters giving a subset of clear, replicated positive colonies were processed further, by transferring the positive colonies to 200 µl of LB broth and growing overnight at 37°C.

As a final selection step to ensure that a single cosmid was obtained and that any contaminating non-target colonies were selected out, a tertiary plating and probing of positive colonies was performed, in the same manner as the secondary screen. Independent colonies were grown in 10 ml of LB broth. A aliquot was saved as a glyerol stock for future work, the rest was processed using the Wizard protocol (section 2.4.6).

Positive clones, isolated through the tertiary screen, were PCR screened for MW opsin exon 2 using primers Op2+/Op2- (see section 2.5.1.6). Although these primers were not specific to MW opsin, they did indicate that the tertiary screen did select cosmids containing exon 2 of an opsin gene. PCR products were visualised on 1.8 % agarose gels following electrophoresis.

4.2.10 Cosmid Vectorette construction and PCR

Having isolated cosmids containing the MW opsin gene, it was necessary to determine the extent to which the 5' flanking region was included. One possible

method to accomplish this would be to PCR amplify using a cosmid arm specific primer in conjunction with an opsin exon 1 primer. This was ruled out in favour of cosmid Vectorette construction, because it would not have been possible to differentiate a negative PCR result from one where the distance exceeded the PCR's limits of extension. The Vectorette approach was selected even though the genomic Vectorette method had proved troublesome. The complexity of all the parameters for cosmid Vectorette construction are much reduced compared to genomic Vectorettes.

10 ml aliquots of LB broth containing Kanamycin were inoculated with 1 µl of the isolated tertiary cosmid clones and grown overnight (section 2.4.5.4). Following harvesting via Wizard preps (section 2.4.6.1) the cosmids was size fractionated on a 0.6 % agarose gel, which allowed both sizing and quantification.

1 ug of each purified cosmid was digested with the restriction enzyme *Eco*R1 (section 2.4.2), then ligated to *Eco*R1 Vectorette 1 linkers (section 2.3.5.4). Aliquots of this library were kept at -20°C until needed for Vectorette PCR amplification.

Using the Vectorette 1 PCR primer in conjunction with the Op1- primer a Vectorette PCR was performed with an extension temperature of 72°C for 3 minutes. The PCR set-up was as given in section 2.3.5.3. 2 µl of the CosMW1 Vectorette library was used as template. No information on the expected size of the amplified product was available except from the restriction map published by Nathans (1986a), which suggested that a product of about 2 kb could be expected (see Figure 4.1). Automated direct sequencing of the resultant fragment was performed, with the OP1-primer and the dyedeoxy terminator kit from ABI (section 2.7.3).

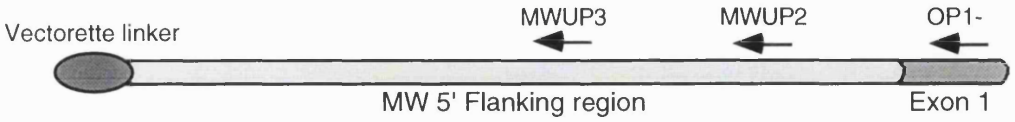
4.2.11 Walk along the MW 5' flanking region

A nested primer was synthesised based on the sequence data obtained from the initial sequencing run. A further 5 µl aliquot of the eluted 2.3 kb fragment was again direct sequenced by employing the primer designated MWUP2.

MWUP2 5'-TGG GCA CAG TGG CAC CAG C-3'

A third walk followed in the same manner, utilising another primer, MWUP3, based on the sequence obtained from the previous walk.

MWUP3 5'-TCG AGG CTT CCG TGA TCG CAC C-3'



4.4.12 Direct PCR Amplification of LW Opsin 5' Flanking Regions

Scrutiny of the manuscript by Wang *et al.* (1992) revealed the sequence of a conserved 37 bp region common to a region upstream of the human LW opsin gene and also upstream of the bovine and mouse X-chromosome opsin gene. This region is regarded as a key *cis* acting site much like the LCR upstream of the ß-globin gene cluster (Li *et al.*, 1990; Caterina *et al.*, 1991). The relevance of the ß-globin upstream enhancer sequence is illustrated by the effects of deletions and substitutions to it on the levels of expression of the gene cluster to which it is connected (Beris *et al.*, 1992; Perichon *et al.*, 1993). A full range of primers to encompass the region from the LCR to intron 1 of the LW opsin gene were designed (see Figure 4.2).

The sequence of the LCR region primers, and otherswhich prime between the LCR and opsin intron 1, is given in Table 7. All these primers were designed based on the published sequence upstream of the human LW gene (Wang *et al.*, 1992).

A fragment of approximately 4 kb should be produced with the primer pair LCRa+/Op1-a. A preliminary test with these, and other primers (see Table 7) from this region, was performed on two human male gDNA templates (KSD & NE). A 50 µl standard PCR reaction (section 2.3.2) cycled with an annealing temperature of 68°C and an extension time of 240 seconds was performed. The products were resolved on a 1.2 % agarose gel.

The same experiment was attempted on the gDNAs of rhesus monkey, diana monkey, patas monkey, as well as on three talapoin monkeys, along with human

LCR1+ (-3484)	5'-GAC TTG ATC TTC TGT TAG CCC TAA TC-3'
LCR2+ (-3474)	5'-TAG CCC TAA TCA TCA ATT AGC-3'
LCRa+ (-3696)	5'-gca tga agc ttC CGC CAC TGC TCC CGC TCCT-3'
LCRb+ (-3375)	5'-GTG GGG GCT GGC ACA CGT GT-3'
LCRc- (-3354)	5'-AAT GAC CCA GCT CAG TGC AA-3'
LCRd- (-3054)	5'-GGA TGG CTG GCC AGC CTT GT-3'
Op1-a(+118)	5'-TGG AGT TGC TGT TGG TGT AGG TGA AGA TGC-3'
Op1-b(+94)	5'-TGC TGG ACT GGG TGC TGT CCT CAT AGC-3'
Op1-c(+78)	5'-TCA TAG CTG TCC TGC GGA TGG CG-3'
LW5'FLa-S(-553)	5'-AGA CTG ACT GCG CAC ATC C-3'
LW5'FLa-L(-555)	5'-GCA GAC TGA CTG CGC ACA TCC TG-3'
LW5'FLa+1 (-186)	5'-GTC TGA GTT TGG TTC CCA GC-3'
LW5'FLa+2(-1039)	5'-TAA GTG TGA GCC AGG AAT GC-3'
LW5'FLa+3(-2946)	5'-GAT CCT CTG ACT CTG GTG GG-3'
OP5'FLa+A (-70)	5'-GGT AGT GTA GGG TTT GGG AG-3'
LW5'FLa-1 (-206)	5'-CTG GGA ACC AAA CTC AGA CG-3'
LW5'FLa-2 (-276)	5'-GCC CTC TCT GTG TTT ATG TG-3'
LW5'FLa-3 (-306)	5'-CTT CTA CTG TCC TGC GTC C-3'
LW5'FLa-4 (-976)	5'-CCT TTT GGG TTT ATT CCT TAA GC-3'

Table 7. PCR primers utilised in the amplification of regions upstream of the opsin genes. (see Figure 4.2 for a diagrammatic illustration of the priming position & selectivity of each primer). Sequence given in lower case represents a 5' restriction site linker.

controls, using the same sets of upstream primers. The PCR parameters were: denaturation of 20 seconds at 94°C, annealing at 69°C for 20 seconds, and extension at 72°C for 4 minutes, for 35 cycles. A final extension of 72°C for 10 minutes was included. Each of these bands was excised and eluted for cloning (section 2.4.4.1). These same primer pairs were utilised in PCR reactions using New World monkey (marmoset) gDNA

Further sets of PCR reactions were set-up with the primer pair LCRa+/LCRc-. The amplification of this region which constitutes the LCR region upstream of the LW opsin gene would allow a quick and simple sequence comparison to be made between the higher primates. The size of the product was expected to be 381 bp, much smaller than that produced with the previous primer sets.

4.2.13 Screening of primate λ genomic libraries for upstream regions

Primate lambda genomic libraries (seven in total), representing both OW and NW species, were supplied as a kind gift by Dr. Jerry Slighton of The Upjohn Company, Kalamazoo, Michigan, USA. Each library was constructed in a slightly different lambda vector, and had been previously re-amplified more than once, which is known to bias the representation of certain sequences at the expense of others. A cautionary note to that effect was enclosed with the samples. Following a PCR screen, utilising 5 µl of each bacteriophage library and the primer pair Op5+/Op5- (section 2.5.1.6), it was determined that the best candidates for full library screening were the Baboon (lambda 1059 - *Mbo*I partial digest - twice plate amplified), Squirrel monkey (Charon 32 - *Eco*RI partial digest - once plate amplified), Spider monkey (Charon 35 - *Sau*3AI partial digest - twice plate amplified), and both Rhesus monkey libraries (Charon 32 - *Eco*RI partial digest - twice plate amplified, and Charon 40 - *Eco*RI partial digest - once plate amplified). The other Spider monkey (Charon 40 - *Bam*HI partial - once plate amplified) and Cebus monkey (Charon 4A - *Eco*RI partial - twice plate amplified) libraries were rejected based on their failure to generate an exon 5 product

(data not shown). Exon 5 primers were employed at this stage because primers from other regions of the opsin genes were not available at the time.

The Rhesus (Charon 40) monkey library was determined as having a titre of 3 x 10^5 plaque forming units (pfu) / ml (section 2.5.1; it was supplied as a library with a titre given as 1 x 10^8 pfu / ml). The Rhesus library was used to transduce DH5α bacterial cells, and then plated out on two large square plastic library plates (section 2.5.1.1). Following incubation, at 30°C overnight, a good dense plaquing pattern was seen on both plates. Replica filters from each master plate were produced using Hybond N nylon paper (section 2.5.1.2). These filters were hybridised with radio-labelled opsin exon 1 probes (generated using PCR primer pair NaOp1+/NaOp1-) at 65°C overnight in Church's solution. The filters were washed at 65°C, as detailed in section 2.5.1.5, using 2 x SSC / 0.1% SDS. The moist filters were sealed in plastic envelopes prior to autoradiography at -80°C in cassettes with intensifying screens and developed three days hence. The filters corresponding to each original plate were aligned. Plaques which gave a positive screen result were excised and placed into individual tubes containing 1 ml of SM buffer.

Secondary and tertiary screens were performed, as well as, PCR screens using exon 1, 3 or 5 primers. Several other libraries were also screened in the same manner.

4.2.14 Walking PCR amplification of regions upstream of the opsin genes

As an alternative to gene libraries, a gene walking method, modified by Dr. J. Bellingham, Dept. of Molecular Genetics, Institute of Ophthalmology, London, from Dominguez and Lopez-Larrea (1994) was used. The method, known as unpredictably primed PCR (or as "Walking PCR"), was employed as detailed in section 2.3.6.

Genomic DNA from human (KSD), two representatives of the OW primates (diana and patas monkeys), and two animals from the NW lineage (marmoset no.15 and marmoset no.16), was subjected to the first round of "walking PCR". Walking primers UNI 33 and UNI 28, combined together, were annealed to the templates,

before PCR cycling with primers LCR1+ or OP1-b, at an annealing temperature of 56°C.

UNI 335'-TTTTTTTTTTTTTTGTTTGTTGTGGGGGGGTT

UNI 285'-TTTTTTTTTTTGTTTGTTGTGGGGGTGT

UNI 17 5'-TTTTTGTTTGTTGTGGG

Aliquots of the first round products were visualised on a 1.25 % agarose gel. 1 µl of the primary PCR reaction products were diluted in 1 ml of sterile water in preparation for second round nested PCR reaction. 2 µl of these dilutions were included in 2nd round PCR reactions, which employed the nested walking primer UNI17. Annealing took place at 58°C. Upon completion of temperature cycling a 1.25 % low melting temperature agarose gel was loaded with 17 µl of 2nd round PCR products. All bands which were seen were excised, ligated into pGEM-T vector, Wizard 'prepped', prior to automated sequencing on the ABI (sections 2.4.5 and 2.7).

4.2.15 Search for sequence motifs in the 5' flanking regions

The upstream genomic sequence information, from the human MW locus was analysed to determine if there were any recognisable eukaryotic transcription factor binding site motifs, whether or not they were also conserved in the human LW opsin upstream region, and if any other areas of sequence were highly conserved, which may suggest some functional significance.

MacVector, GeneWorks, Signal Scan (section 2.8), and the on-line TESS query programmes were used to analyse the upstream region plus the first intron of MW opsin coding sequence for known transcription factor motifs. The mammalian transcription factor recognition motifs present were mapped in the human MW and LW opsin genes, as well as, in the 5' flanking regions of other primates

4.3 Results

4.3.1 Vectorette approach

The results of the trials on human gDNA proved positive. As Figure 4.3 shows, following first round PCR amplification, a number of fragments were generated. Bands corresponding to the expected size (calculated based on primer sites and restriction cutting points) were excised from appropriate lanes (lane 7 - *Hind*III fragment at 72°C; lane 11 - *Rsa*I fragment generated at 66°C; lane 13 - *Bam*HI fragment again at 66°C; lane 15 - another *Hind*III product, this time amplified at 66°C).

Figure 4.3 shows the products generated following nested PCR of the template recovered from the first round of PCR amplification, and in Figure 4.4 the partial sequence obtained from this clone has been aligned to published LW upstream sequence of Wang *et al.,* (1992). The nucleotide sequence information obtained from three clones extends from the *Hind*III restriction enzyme half site ligated to the Vectorette, upstream of the LW opsin gene towards the 5' end of the gene. This *Hind*III site is conserved in the human LW opsin upstream region (see Figure 4.1). The two sequences do not agree entirely along the 200 bp presented. The differences could be due either to the existence of polymorphisms in the human population, to the introduction of nucleotide substitutions during amplification or cloning, or to sequencing errors.

The sequence alignment confirmed that the Vectorette protocol was a quick and viable method to employ in order to obtain unknown sequences flanking a known sequence. However, PCR of the human *Cla*1 Vectorette library generated a smear, with a broad size distribution (data not shown). The Vectorette-only primer control reaction generated bands (without a second specific primer present). This fragment continued to appear even when the temperature of annealing was raised to 76°C. No products should have been generated since the Vectorette amplification process does not provide a template to which the Vectorette primer can bind (see Figure 2.1), unless it happens by chance to be present in the genomic template. Smearing during the first

Fig. 4.3

Figure 4.4 Sequence of the cloned DNA amplified via Vectorette PCR. The Human LW 5' flanking sequence was obtained by Wang *et al.* (1992). The *Hind*III site used in the construction of the Vectorette libraries is shown . Regions of mismatch are indicated by red overscoring. The numbers indicate the distance upstream of the message start site. The data is in agreement with the presence of the*Hind*III site given in Figure 4.1

round of amplification was not considered unusual, however, after second round amplification a single solitary band was expected, but this never materialised. Following most second round PCR attempts smearing in the agarose gel lanes appeared again, however on several occasions no products were detected (data not shown). This failure of the Vectorette system was subsequently attributed to the loss of activity of the *Cla*I restriction enzyme.

Non-human primate gDNAs which had been similarly digested and ligated to Vectorette linkers, were also subjected to Vectorette PCR utilising the same primers. Every attempt to generate single bands, representing opsin 5' flanking regions, failed. All possible Vectorette linkers (*Bam*HI, *Cla*I, *Eco*RI, *Hind*III, and blunt ended) were tried. DNAs from a number of species of OW and NW primates were re-digested and ligated to Vectorette units a number of times, all without success. The Vectorette method was eventually abandoned, without yielding further results.

The limited trials with the inverse PCR method were also unsuccessful. No amplification of any bands was forthcoming, even after minor adjustments to the protocol. In fairness this method was quickly abandoned after a short period in favour of other protocols. Several reports show that numerous conditions have to be satisfied before successful results are obtained.

4.3.2 MW 5' flanking region from a human cosmid library

A PCR screen of the Lorist B human cosmid library did reveal that a band of the correct size was generated in eight of the ten test aliquots (data not shown). Thus, the technique of PCR screening was shown to be very sensitive, indicating that the library did contain clones with the DNA of interest present in 80% of the dilutions generated from the neat library stock (each containing approximately 50,000 cfu based on the titre obtained from the work of Dr. J. Fitzgibbon).

The primary library screen resulted in ten regions (which gave the most intense replicated positive signals) that were selected from the two primary filters for further characterisation.

As Figure 4.5a shows, 9 out of the 10 primary colony lifts did contain opsin exon 3 regions and, as Figure 4.4b shows 3 of these 9 contained MW exon 4 to 5. It was not possible to tell at this stage how much upstream region, if any, was present in each cosmid. Subsequent manipulations used these 3 positive pools (labelled MW1a, MW2a, and MW3a), the others, which were assumed to be cosmids of LW origin, were stored at 4 °C.

Following titration of positive pools MW1a, MW2a, and MW3a to determine their density, the desired dilutions were plated out and processed in the same manner as the primary screen master filters. Only filters giving a subset of clear, replicated positive colonies were processed further. On each of 3 filters approximately 20 positives were obtained, representing about 40 % of all colonies. Two discrete positives were picked from each of these secondary filters (numbered MW1a1, MW1a2, etc.), and again placed into 200 µl of LB broth and grown overnight.

As a final selection step to ensure that a single cosmid was obtained and that any contaminating non-target colonies were selected out, a tertiary plating of positive colonies was performed, in the same manner as the secondary screen. Probing of these filters showed that almost every colony present "lit-up", indicating that the secondary screen was very efficient.

The three positive clones (CosMW1, CosMW2, and CosMW3) isolated through the tertiary screen were PCR screened for MW opsin exon 2 using primers Op2+/Op2- (see section 2.5.1.6). Although these primers were not specific to MW opsin, they did indicate that the tertiary screen did select cosmids containing exon 2 of an opsin gene. All clones produced a positive result (data not shown).

4.3.3 Vectorette library construction from MW cosmids

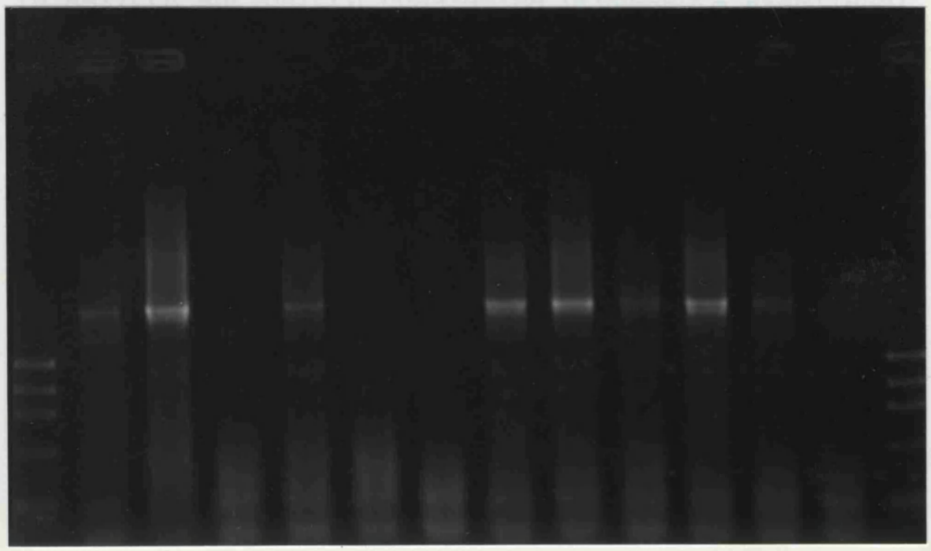
Cosmid clones CosMW1 and CosMW2 were amplified, purified, and size fractionated on an agarose gel. The estimated size of each cosmid was calculated at approximately 45 kb. A smear seen at the bottom of each lane represented undigested RNA, which was removed by further digestion with RNase H (data not shown). Each

Fig. 4.5

of these cosmids was digested with *Eco*RI to completion, Vectorette libraries were constructed, and PCR amplification was performed. No information on the expected size of the amplified product was available except from the restriction map published by Nathans *et al.* (1986a), which suggested that a product of about 2 kb could be expected (see Figure 4.1). As can be seen from Figure 4.6, an amplified band of approximately 2.3 kb was generated. Automated direct sequencing of this fragment, and the subsequent alignment with the published sequence (Nathans *et al.*, 1986a; Wang *et al.*, 1992) confirmed that this fragment was indeed the region upstream of the MW opsin gene. Following two gene walks a total of 1.4 kb of the 2.3 kb fragment was sequenced before direct sequencing of this method failed due to the presence of a repetitive region (the exact cause was latter deduced when the sequence was analysed for sequence motifs - the design of the fourth walking primer was unfortunately based on the sequence of a repetitive region of one of the AluI elements. Thus all direct sequencing attempts would have presented multiple priming sites for this primer, resulting in the generation of multiple sequences). Attempts to use the Vectorette sequencing primer, from the opposite end of the fragment, in a direct sequencing protocol also failed (not an uncommon feature of the Vectorette template, as other researchers have reported similar problems). The full sequence obtained upstream of the human MW opsin gene is given in Figure 4.7.

4.3.4 Direct PCR Amplification of LW Opsin 5' Flanking Regions

PCR amplifications using the LCR region primer, LCRa+, in combination with a series of primers that prime to exon 1 did produce bands (Figure 4.8). The actual size of the bands correlates well with the predicted size. The expected size of the bands in lane 2 and lane 8 was calculated to be 2,761 bp (primer LCRa+ binds to position -3696 and primer LW5'FLa-4 primes to position -975), whilst the primer pairs LCRa+/Op1-c (lanes 3 & 9), LCRa+/Op1-b (lanes 4 & 10), and LCRa+/Op1-a (lanes 5 & 11) were expected to produce bands of 3,920 bp, 3,941 bp, and 3,968 bp, respectively. Lanes 7 & 13 were expected to generate bands of 3,141 bp (LCRa+/LW5'FLa-L). Each of the



Each lane represents the amplified products from a different aliquot of the libraries. In some cases amplification was successful while in others it was not.

Figure 4.6 PCR products generated from aliquots of the two human MW cosmid EcoRI Vectorette libraries, utilising the vectorette sequencing primer and the opsin specific primer Op1-b. The 2.3 kb fragments were eluted from the gel and partially direct sequenced giving the sequence in Figure 4.7.
Markers = ϕX174/*Hae*III

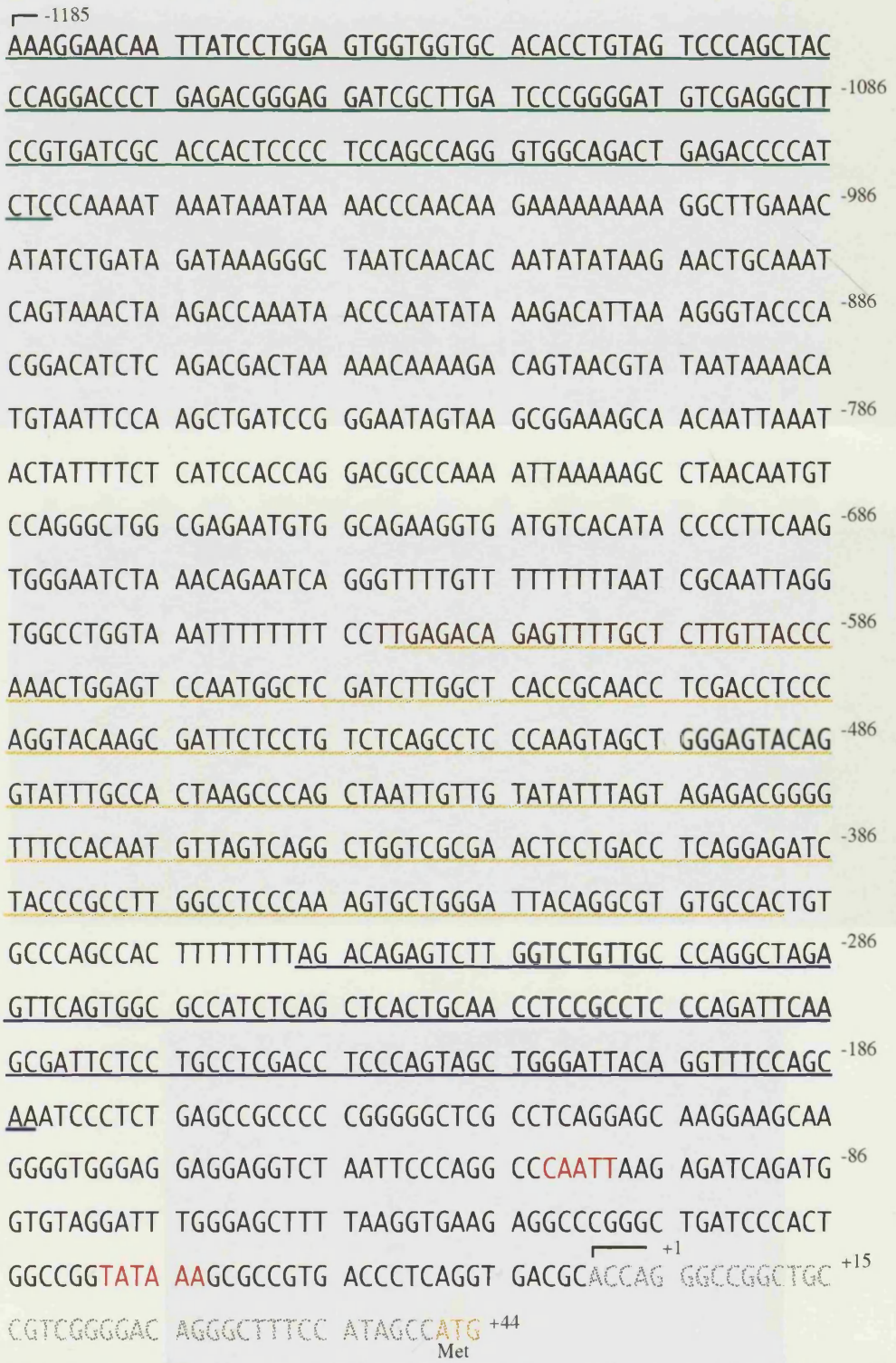


Figure 4.7 Sequence flanking the 5' region of a human MW opsin gene. The TATA and CAAT boxes are highlighted (red type). Underlined sequences represent Alu repeats. Yellow underscoring indicates a full length Alu I in reverse orientation, whilst the region underlined in blue is an Alu 1 half-site. At the beginning, underscored in green, is the 3' end of a full length Alu1 sequence.

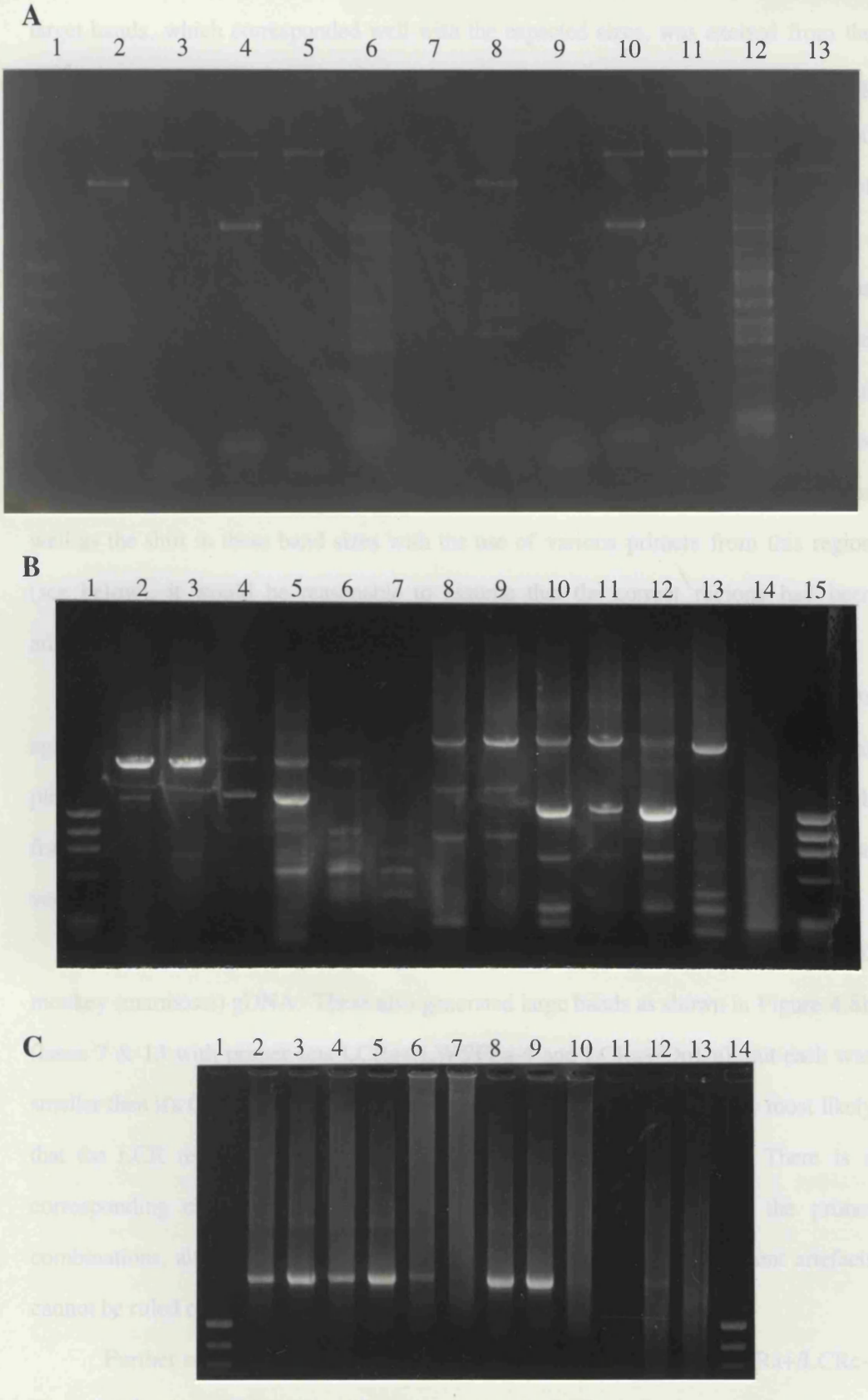


Figure 4.8 PCR amplification products obtained upon utilising primer pairs selected from those listed in section 4.4. **A**. Human DNA controls. **B** and **C**. Various primate templates (see text for loading order). The sites to which these primers bind is illustrated by Figure 4.2.

target bands, which corresponded well with the expected sizes, was excised from the gel and allowed to elute in 40 µl of sterile water (section 2.4.4.1). Ligation and cloning into several different vector systems, to allow sequencing of these fragments, proved ineffective. No positive colonies with the target inserts were detected. The identity of these bands was not confirmed by other means.

The same experiment was attempted on the gDNAs of Rhesus monkey, Diana monkey, Patas monkey, as well as on three Talapoin monkeys, along with human controls, using the same sets of upstream primers. As Figure 4.8b shows successful amplification was obtained in most reactions. The identity of these bands was not established, but considering the agreement between the estimated and the actual size, as well as the shift in these band sizes with the use of various primers from this region (see below), it would be reasonable to assume that the correct regions had been amplified.

Numerous attempts to isolate clones containing plasmids with these inserts again proved unsuccessful. It is possible that the efficiency of the cloning process into plasmids with T overhangs is inversely proportional to the size of the insert; the 3.9 kb fragments may be too large therefore to clone with reasonable efficiency into these vectors.

These same primer pairs were utilised in PCR reactions using New World monkey (marmoset) gDNA. These also generated large bands as shown in Figure 4.8b (lanes 7 & 13 with primer sets LCRa+/LW5'FLa-4 and LCRa+/Op1-a), but each was smaller then it's Old World counterpart. The reason for this discrepancy is most likely that the LCR region may be closer to its opsin gene in marmosets. There is a corresponding change in fragment length which correlates well with the primer combinations, although once more the possibility that these bands represent artefacts cannot be ruled out.

Further sets of PCR reactions were set-up with the primer pair LCRa+/LCRc-. The amplification of this region which constitutes the LCR region upstream of the LW opsin gene would allow a quick and simple sequence comparison to be made between

the higher primates. The size of the product was expected to be 381 bp, much smaller than that produced with the previous primer sets. The alignment of the patas monkey sequence to that of the human (Figure 4.9) reveals that the core LCR region is well preservedand that a high degree of overall homology (91%) exits across the entire region. When the respective bovine and mouse LCR regions are aligned, the identity of the region surrounding the LCR core declines, but within this core region all 37 bp are absolutely conserved (Figure 4.10). This implies that there is an important function performed by this element, as has been shown indirectly when deletions encompass this area (Wang et al., 1992). This finding is not unexpected since the LCR is considered to act at the chromatin level rather that at a more localised one (see earlier; Caterina *et al.*, 1994a).

4.3.5 Screening of primate λ genomic libraries for upstream regions

The Rhesus (Charon 40) monkey library titred at 3 x 10^5 plaque forming units (pfu) / ml (section 2.5.1; it was supplied as a library with a titre given as 1 x 10^8 pfu / ml). Between 4 and 12 replica signals were obtained from this primary screen. The alignment of filters and plates allowed the pinpointing of the corresponding plaques, and all 3 were excised from the plates.

The results of the secondary and tertiary screens were just as promising, with good clear replicates showing up each time. However, PCR screening did not generate fragments corresponding to either exons 1, 3, or 5, when used with the appropriate primer pairs. Several other libraries were screened, but with the same outcome.

4.3.6 Walking PCR amplification of primate opsin gene flanking regions

Several reactions produced bands of approximately 700 bp using the primer combination UNI33 / UNI28 / LCR1+, and about 400 bp with UNI33 / UNI28 / Op1 b. Second round PCR fragments of about 280 bp were amplified in lanes utilising the primer Op1-b; no products were obtained from those tubes containing the LCR1+ primer. Three of the sequences were each confirmed as MW upstream from human,

-3815

```
Human   CCGCCACTGC  TCCCGCTCCT  CTCCCCCCAT  CCCACCCCCT  CACCCCCTCG
Patas                           ----------  ----------  ----------

Human   TTCTTCATAT  CCTTCTCTAG  TGCTCCCTCC  ACTTTCATCC  ACCCTTCTGC
Patas   ----------  ----------  CA--------  -------C--  ----------

Human   AAGAGTGTGG  GACCACAAAT  GAGTTTTCAC  CTGGCCTGGG  GACACACGTG
Patas   ----------  -------G--  ----------  -C--------  ----A--A--

Human   CCCCCACAGG  TGCTGAGTGA  CTTTCTAGGA  CAGTAATCTG  CTTTAGGCTA
Patas   ----------  -----G----  -----C----  ----------  ----------

Human   AAATGGGACT  TGATCTTCTG  TTAGCCCTAA  TCATCAATTA  GCAGAGCCGG
Patas   ----------  ----------  ----------  ----------  --------A-

Human   TGAAGGTGCA  GAACCTACCG  CCTTTCCAGG  CCTCCTCCCA  CCTCTGCCAC
Patas   ----------  ---------A  -----T----  -----A----  ----------

Human   CTCCACTCTC  CTTCCTGGGA  TGTGGGGGCT  GGCACACGTG  TGGCCCAGGG
Patas   ----------  ----------  -A--------  ----------  ----G-----
```

-3335

```
Human   CATTGGTGGG  ATTGCACTGA  GCTGGGTCAT  T
Patas   ----------  -
```

Figure 4.9 Alignment of human and patas LCR regions showing over 95.5% homology over the 341 bases. The region in red type is the LCR core sequence identified by Wang *et al.* (1992) which is fully conserved. Bases in green identify the Ret1/PCE1 element, and type in italics corresponds to the PCR primers.

Figure 4.10. Sequence comparisons of the LCR regions of man, patas, bovine & mouse. A high degree of homology is seen in and around the LCR core sequence (red). The same Ret1/PCE1 element shown in Figure 4.9 is still conserved in these distantly related mammals. Underscored sequences represent identity between the four sequences. (Bovine and mouse LCR sequence data from Wang *et al.*, 1992).

patas and diana monkeys by comparison with published human sequence (Figure 4.11). Sequences generated from other templates were not of opsin origin.

4.3.7 Sequence motifs recognised

Signal scan and TESS software specifically compiled to identify transcription elements was used to analyse the opsin upstream regions. These results map the position of transcriptional start sites and reveal that the promoter region for the human MW opsin gene, like that of its LW counterpart possess certain characteristics of classical promoters.

Sequence analysis of the region extending from site +1 to -1280 upstream of the human MW opsin gene revealed the presence of a large number of motifs. The first feature of note was the unusually high number of Alu I motifs present (see Figure 4.7). A dot matrix (dot plot) analysis was instrumental in the detection of both the full length Alu repeat and the three Alu half-sites (Figure 4.12). Indeed, homology between the human MW and LW genes at the proximal promoter region is interrupted by the presence of one of these Alu repeats. However, if the Alu element is ignored homology between these regions does not return. A number of *cis*-acting motifs were also identified. The TATA box has previously been noted as residing at site -29 to -24 (Nathans *et al.*, 1986a). A CAAT box was identified at site -103 to -99. 11 Sp1 sites, dispersed along the region, were present. Both Ap1 and Ap2 sites were also seen, as well as, 3 sites to which the c-Jun and c-Fos factors bind. All of the above information, including the mapping of the transcriptional start sites has been integrated into Figure 4.7. Interestingly, none of the elements reported to exist upstream of the human rod opsin gene (such as the Ret-I and BAT-I motifs) as well as those reported as being common to retinal genes, are present in this region.

4.3.8 Comparison of 5' flanking regions

An alignment of the human LW and MW upstream sequences in this region was performed initially using the dot matrix comparison of MacVector software (section 2.8), then using GeneWorks and CLUSTAL V (Higgins *et al.*, 1992) software (section

Fig 4.12

2.8), to highlight particular regions of nucleotide sequence conservation which may be responsible for binding transcription factors. Several regions of homology were detected using dot matrix comparisons (Figure 4.13). The first was the region immediately flanking these genes to base - 195, which represents the proximal promoter region, where the TATA and CAAT boxes reside. All other regions representing extensive homology were found to be Alu repeats (Figure 4.14). These results suggest that all the *cis*-acting motifs necessary for proper gene regulation may either lie in this proximal region or beyond the point (-1185) to which the MW upstream region was sequenced.

The limited sequences obtained from OW primate (chimpanzee, diana, and patas) MW opsin 5' flanking regions were aligned with analogous regions from the human MW and LW genes. These show a great deal of homology to the human sequences (both MW and LW), but since the limit of the 5' flanking region generated was confined to the first 200 bp upstream, a region where the human MW and LW sequences are almost identical, this is not surprising (Figure 4.15). Once again the TATA and CAAT boxes are present, as are the Ap1, Ap2, and Sp1 sites. In fact most of the motifs identified in the human opsin promoter proximal flanking regions are present in these primates (see Figure 4.15). The evolutionary closeness of these four species has not allowed enough time for a divergence in this region to have occurred, thus sites required for transcription cannot be resolved from others which are not.

The NW primate opsin upstream regions, from two marmosets, were also aligned to that of the OW primates. Once again, the amount of sequence obtained was limited to about 200 bp, so an assignment as to which OW opsin gene this region most closely resembles was not possible. Of the comparison performed this sequence alternated between the human MW and the LW opsin sequences in this region, as shown in see Figure 4.16. The sequences from the two marmosets were identical.

It was possible to deduce from dot plot comparisons of the 5' flanking regions of mouse, bovine, and marmoset X-linked opsin and the equivalent regions of the LW and MW genes of human (data not shown) that the first 250 bp upstream from the

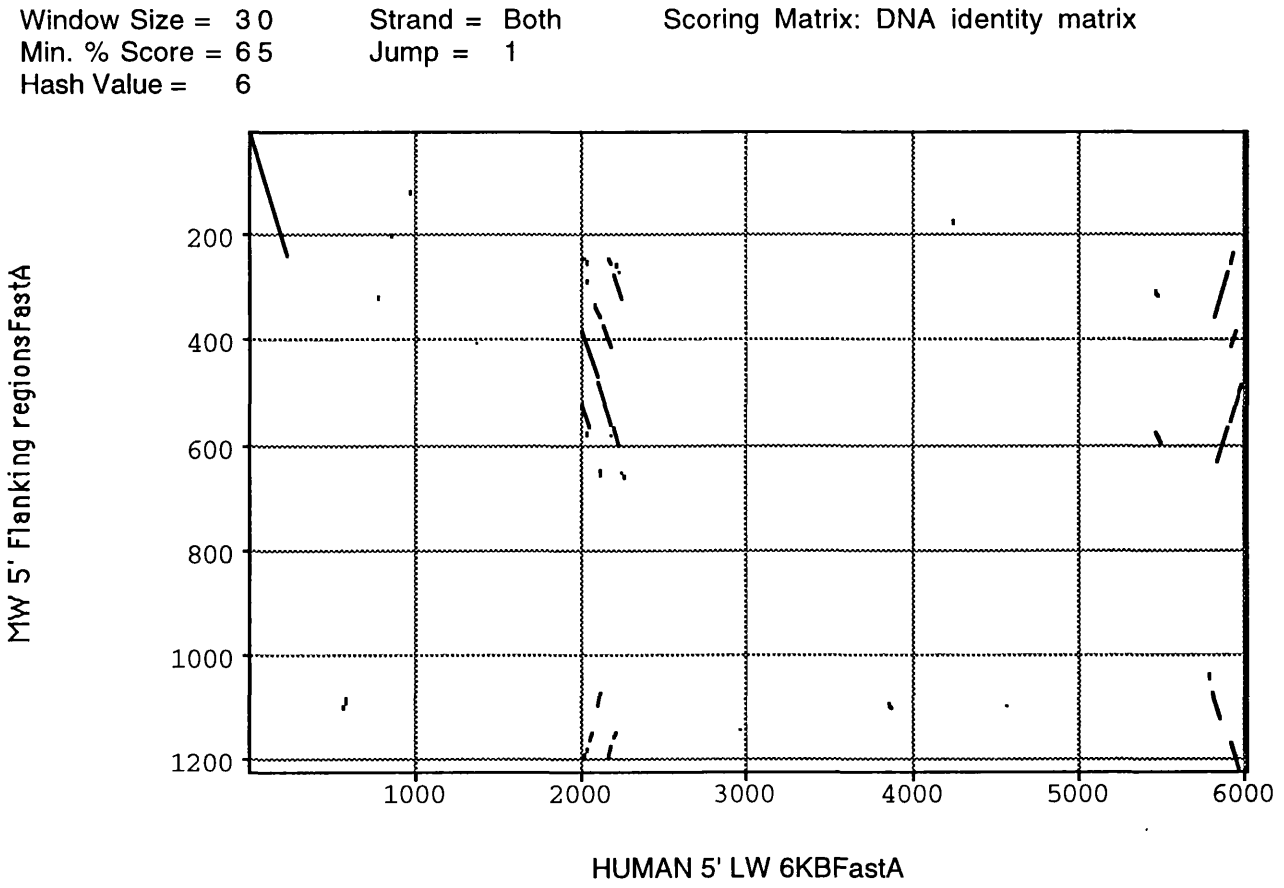


Figure 4.13 Dot matrix analysis of the regions 5' flanking the human MW and LW opsin genes. The data for the MW gene was generated in this study and that for the LW gene was published by Wang *et al.* 1992. The line in the top left corner is the homologous region from the promoter proximal region of the two genes. All other areas correspond to the alignment of Alu repeats.

```
GGCTGGGCGT GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGGCCGA     50
GGTGGGTGGA TCACCTGAGG TCAGGAGTTC AAGACCAGCC TGGCCAACAT    100
GGTGAAACCC CGTCTCTACT AAAAATACAA AAATTAGCCG GGCGTGGTGG    150
CGCGCGCCTG TAATCCCAGC TACTCGGGAG GCTGAGGCAG GAGAATCGCT    200
TGAACCCGGG AGGTGGAGGT TGCAGTGAGC CGAGATCGCG CCACTGCACT    250
CCAGCCTGGG CGACAGAGCG AGACTCCGTC TCA                      283
```

Figure 4.14 Nucleotide sequence of a common element of the human Alu1 family. These sequences are present upstream of all the photopigment opsin genes.

Fig. 4.15

translation start site is relatively preserved across all these species. Naturally, the degree of homology declines as the evolutionary distance between the species increases.

4.4 Discussion

Analysis of the 1.3 kb region immediately flanking the 5' end of a human MW opsin gene revealed a number of prominent features: 1) a region extending from site +1 to -195 showed a great homology to human LW opsin, and to regions upstream of other OW primates and also to a NW species. 2) within this common region the basic elements for gene regulation were present. These included the TATA and CAAT boxes, an AP-1 site, an AP-2 site, an AP-4 site, and two Sp1 sites. 3) Alu elements were identified further upstream, residing in both orientations. In fact one of these is responsible for abruptly ending the homology between the MW and LW genes. 4) Further upstream, in the region where the MW and LW opsin regions show very little homology, a number of other elements which may be functionally important were noted. One of these was the Kruppel site, as well as a "multi-site" for a large number of leucine zipper factors. The significance of these sites is not known at present.

The presence of multiple *Alu* elements in the MW opsin 5' flanking region is not an unusual feature; a similar clustering of *Alu* sites has been reported for the human rod opsin upstream region (Bennett *et al.* 1995), although the closest such element to the transcription site is about 2.4 kb in the latter case. An homology search of the sequence database (GenBank R81.0) by Bennett *et al.* (1995) revealed that the rod opsin and the *Alu* element 5.8 kb upstream of the human LW opsin gene were members of the same subfamily, and therefore may be duplications of each other. Although *Alu* sequence elements are dispersed throughout the human genome and have been implicated in disease processes in only a few cases, they may be capable of modulating transcription if they are near the transcription site of genes (Kim *et al.*, 1989; Brini *et al.*, 1993). It remains to be seen if the Alu repeats identified upstream of the MW and LW human opsin genes play any role in regulation opsin gene expression.

Fig. 4.16

The Ret-1 consensus sequence was identified as part of the LCR core sequence in humans and other mammals (see Figure 4.9 and 4.10). Also, there is an Ap1 site just upstream of this Ret-1 site. Other elements noted in this region were Sp1 binding motifs, the G-C boxes. The presence of all these sites in this region may have a significance since a similar arrangement is found in the proximal 5' flanking region of the rod opsin gene (see figure 2 in Bennett *et al.*, 1995).

A comparison of visual pigment genes from birds, fish, and mammals reveals that the duplication events that generated the rod opsin, SW and LW/MW photopigment genes occurred prior to the vertebrate radiation (Nathans and Hogness, 1983; Takao *et al.*, 1988; Kuwata *et al.*, 1990, Tokunaga *et al.*, 1990; Yokoyama and Yokoyama, 1990). Construction of a phylogenetic tree using this data (Yokoyama, 1995) suggests that the cone opsins gave rise to the rod opsins. If this is the case then the cis-acting upstream sequences controlling expression of the ancient cone system may have evolved to give the present 5' flanking region of the rod opsin gene, but the antiquity of the lineage would suggest that enough evolutionary time has elapsed for the 5' flanking regions to have diverged to such a degree that it may not be possible to detect any common motifs.

The presence of retinal conserved elements in and around the LCR upstream of the LW opsin, rather than in the proximal promoter regions of this gene or the MW gene, may be significant. It can be proposed that whereas other single loci retinal genes are regulated by retinal specific factors which act directly at the proximal promoter level, with the multiple X-linked opsin loci present in the human and presumably other OW primates, this interaction may take place with the LCR region before other mechanisms decide which opsin gene to activate in a particular cone cell (Wang *et al.*, 1992). However, without direct evidence for the involvement of the LCR in determining gene expression, it is only reasonable to assume that the LCR participates in the "opening" of the chromosomal region to transcription factors, so that gene transcription of the LW and/or MW genes can occur. This view satisfies the criteria that the LCR is instrumental in determining expression from these loci; deletions, as those reported in

blue cone monochromats (Nathans *et al.*, 1993) which remove the LCR, do not allow the proper conformation of the chromatin to be adopted, thus denying access to the appropriate transcription factors. If the LCR was functioning in this way it would still require an appropriate interaction with specific factors so that transcription could be regulated in a tissue specific and temporal manner. The presence of the Ret-I motif, in combination with other as yet unrecognised elements, in the LCR may be the sites which determine this specificity.

Further work needs to be performed at ascertaining the exact role the LCR plays in controlling opsin gene regulation.

4.4.1 The proximal promoter architecture of X-chromosome opsin genes

The positioning of a transcriptional start site is the first step in defining the structure of a promoter region. The presence of multiple transcriptional start sites has never been reported for the human opsin genes (Nathans *et al.*, 1986a).

4.4.2 Alu sequences in the 5' flanking regions of human opsin genes

Analysis of the 5' flanking regions of the LW and MW human opsin genes reveal the existence of Alu elements in these regions, although they are at different locations. The MW regions contains one complete Alu plus an Alu half-site (see Figure 4.7). The region 5' to the human SW opsin gene also shows a complete Alu element (B. Appukuttan, personal communication, 1996), with further Alu repeats appearing upwards of 3 kb from the rhodopsin gene. Alu elements are a family of short (150 - 300 bp) interspersed DNA repeats which exist in several hundred thousand copies throughout the human genome (Willard *et al.*, 1987; Batzer and Deininger, 1991; Shen *et al.*, 1991) and have mobilised throughout primate genomes by retroposition from a few "master" genes. Similar sequences, but in far fewer numbers, have been reported in mouse and hamster genomic DNAs (Kominami *et al.*, 1983). Many human Alu sequences have been identified and sequenced; no two copies are identical in sequence, but an 80 % conservation is seen within species, which drops to 50 - 60 % between

species. Most Alu's are full length, however, many partial Alu-like sequences have been found too, mostly scattered between genes and within introns.

What function do these sequences serve? Alu elements exhibit a remarkable sequence similarity with 7S RNA, a 294 base cytoplasmic RNA. This small cellular RNA is part of a cytoplasmic ribonucleoprotein particle, the signal recognition particle, that aids in the secretion of newly formed polypeptides through the membranes of the endoplasmic reticulum. One can hypothesize that since opsins are integral membrane proteins the association between them and the Alu sequences just upstream may not be entirely coincidental. A role in RNA polymerase III transcription units has also been forwarded (Hess *et al.*, 1985). The significance of this is not evident at present.

Alu sequences have been demonstrated in primates other than humans. Koop *et al.* mapped the region surrounding the ß-globin gene cluster where they identified seven Alu repeat sequences. They compared these and other orang-utan sequences to orthologously related human sequences: Alu sequences in human and orang-utan were calculated as having diverged by 3.7%. No evidence for either the addition or deletion of Alu sequences from the beta globin gene cluster nor any evidence for recent concerted evolution among the Alu sequences examined was reported. Both phylogenetic and phenetic distance analyses suggest that Alu sequences within the α- and β- globin gene clusters arose close to the time of simian and prosimian primate divergence (about 50- 60 MYA). Alu sequences have been evolving at the rate typical of noncoding DNA for the majority of primate history.

Young Alu elements which only exist in the human genome have been reported (Batzer *et al.*, 1994); these are members of the human-specific subfamily , as well as a variety of Alu subfamilies which suggest that they were founded independently in chimpanzee, gorilla and human lineages (Zietkiewicz *et al.*, 1994).

The data as a whole suggest that cone opsin gene transcription factor binding sites, defining the promoter proximal region, are located in the 250 bp immediately upstream of each gene, and that a LCR region located several kb away is required as an enhancer element. The homology of marmoset to human MW and LW 5' flanking

region means that differential control of MW and LW genes cannot reside in this promoter proximal region

4.5 Summary

The elucidation of the X-linked opsin upstream genomic sequences and the mapping of transcriptional start sites has defined the proximal promoter elements at the nucleotide level. The promoter appears to possess both a TATA and a CCAAT box, features which are shared by many genes expressed in a temporal and spatial manner. There are numerous transcription factor binding sites conserved between the human and non-human primate genes and areas of conserved sequence which are candidate sites for novel DNA binding proteins. The basic requirements have been met for functional studies of protein binding to extend this work.

CHAPTER 5

Human Colour Vision Anomalies

PART ONE:
The Genetic Basis of Variation in Normal Colour Vision

5.1 Introduction

Colour matching tests (Rayleigh matches, psychophysical studies, etc.) have long revealed that the normal human observer has three cone pigment types. For a time it was assumed that all individuals with normal colour vision possessed an identical complement of cone pigments. Recently, however this view has been challenged by evidence, indicating the existence of significant individual variation in human cone pigment spectra (Dartnall *et al.*, 1983; Neitz, 1986; Neitz and Jacobs, 1986; Neitz and Neitz, 1995; Neitz *et al.*, 1995). The first reports to suggest that the situation may be more complex than initially thought came from the extensive analysis conducted by Stiles and Burch (1959), measuring colour matches in a large group of subjects having normal colour vision. Their results suggested that the spectral sensitivities of the group were broader than one could attribute to experimental error alone. Evidence of variation has continued to mount, with additional studies localising most of this variation to the X-chromosome linked cone pigments (Dartnall *et al.*, 1983; Neitz, 1986). Now the accepted view is that colour normal subjects do indeed possess tremendous variation in the organisation of their opsin gene arrays (Winderickx *et al.*, 1992a, c; Deeb *et al.*, 1994; Nathans, 1994; Neitz and Neitz, 1995), with gross departures from these arrangements leading to colour defects. One important implication of these findings is that colour normal observers may not have exactly the same complement of visual pigments as each other, thus their views of the world may not be as homogeneous as was thought once.

5.2 Methods

5.2.1 Material for analysis

Psychophysical experiments performed at the University of Cambridge, by Dr. John Mollon's group, identified a number of individuals whose colour matches fell on the outer limits of the acceptable range for colour normal observers. These subjects exhibited normal trichromatic colour vision as suggested by non-invasive tests. Genomic DNA extracted from venous blood of two such male individuals (identified as DT and YT) who represented the two extremes of the normal distribution of Rayleigh match midpoints, was obtained. The labelling of the tubes did not identify the subjects or their phenotype - in effect this was a blind study.

5.2.2 Integrity and yield of gDNA

The integrity and yield of the gDNA samples from DT and YT was assayed by agarose gel electrophoresis (section 2.4.3). The yield of DNA was estimated by comparison with the upper band of 1 μg of the marker λ digested with *Hind*III.

5.2.3 PCR amplification of exons 2, 3, 4, and 5 of LW and MW opsins

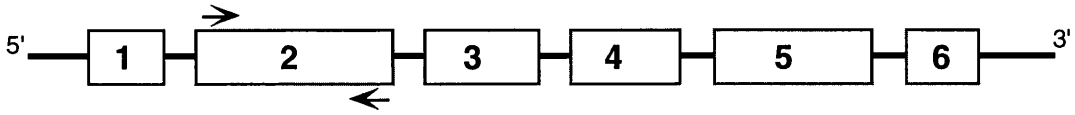
On the premise that exons 2 to 5 of the LW and MW opsin genes are responsible for encoding all the fifteen amino acids that differentiate the two cone opsins classes (section 1.11), the analysis was restricted to these regions.

Independent PCR reactions were set-up to amplify exons 2, 3, 4, and 5 of the LW and MW opsin genes from gDNAs of subjects DT and YT. The primer pairs OpEx3+, OpEx3-; OpEx4+, OpEx4-; and OpEx5+, OpEx5- were used as detailed in section 3.2.2, along with 200 ng of gDNA as template and 5 units of *Taq* polymerase. All PCR parameters were as given in section 3.2.2. Primer pair OpEx2+ and OpEx2-, whose sequence is given below,

Exon 2

5'-GTG TAC CAC CTC ACC AGT GTC-3' (Op2+/b215)

5'-CA CAG GGA GAC GGT GTA GCC-3' (Op2-/b428)



were utilised in an identical PCR set-up of 50 µl volume (section 2.3.2), as the other primer sets. The temperature of annealing was adjusted to 60°C, for PCR cycling. All these primer pairs were designed to co-amplify both the MW and LW photopigments genes.

Products of the first round of PCR were size fractionated on 1.8 % agarose gels as described in section 2.3.3

5.2.4 Cloning and sequencing of PCR fragments

Two strategies were employed for the cloning and sequencing of the PCR amplified products.

The first approach involved each of the target bands being eluted from the agarose gels (section 2.4.4.1). 7 µl of the eluate was combined with 2 µl of pCR™1000 vector or pCR II vector in a ligation reaction (section 2.4.5). Concurrently, 0.5 µl of each neat PCR reaction was added to separate ligation mixtures, along with 2 µl of pCR™1000 vector. All ligations were incubated at 12°C for four hours, following which 2 µl of each was used to transform DH5α competent cells (section 2.4.5.3). Positive clones were selected on LB agar / Ampicillin plates (section 2.4.5.4), mini-prepped (section 2.4.6.1) in 10 ml of LBA broth, and then purified using the Geneclean II protocol (section 2.4.6.3). 3 µl aliquots of each clone were digested with the restriction enzyme *Not*1 (section 2.4.5.1) releasing the insert from the vector. These digests were size fractionated on 1.5% agarose gels (data not shown). Clones containing the correct insert size were manually sequenced in both directions, using the Sequenase Kit (USB) (section 2.6.3.3). Autoradiographs were developed after overnight exposure.

Alternatively, following the size fractionation of exon 5 products, DNA was eluted from excised target gel fragments as before. 7 µl of the eluate was used in the ligation reaction employing pTAg vector. Following transformation, positive clones were selected upon LB / Kanamycin / Tetracycline / IPTG / X-Gal plates. A small portion of each colony was placed in a PCR reaction of 20 µl volume containing the recommended concentration of reagents. PCR products were separated on a 1.5 % low melting temperature agarose gel. Fragments of the predicted size were excised into 50 µl distilled water and eluted overnight at room temperature. 5 µl of the eluate was used in the automated sequencer cycle sequencing reaction (section 2.7.3) prior to loading on to the ABI 373a automated sequencer. The effectiveness of the procedure allowed the throughput of a large number of clones.

To allow for potential errors in sequence data obtained from cloned fragments of PCR products, each set of experiments was repeated at least three times. The reasons for this are discussed in section 3.2.3

5.2.5 Exon 4 to 5 hybrid gene analysis

There are a number of nucleotide differences between the LW and MW opsin genes that can be incorporated into primers that will allow discrimination between these opsin classes. As Figures 1.7, 1.9 & 3.3 reveal the greatest clustering of nucleotide differences is in the centre of exon 4 and in exon 5. Therefore, four primers, Op4+LWint, Op4+MWint, Op5-LWint, and Op5-MWint, were synthesised, that hybridise to these regions. The primers Op4+LWint and Op4+MWint both prime to an identical site in exon 4 (Figure 5.1), the difference between them being that Op4+LWint only hybridises to LW exon 4 templates while Op4+MWint selects MW templates. The remaining two primers both hybridise to a region in exon 5. Once again Op5-LWint primes to LW sequences and Op5-MWint to MW templates. When primer pairs Op4+LWint / Op5-LWint and Op4+MWint /Op5-MWint are used to amplify DNA, fragments encompassing a region from base 721 in exon 4 across intron 4 and approximately half of exon 5, to base 878 (a total length of about 1.7 kb) from LW and

Fig. 51

MW templates are amplified, respectively. When these primers are used in other combinations (Op4+LWint / Op5-MWint or Op4+MWint / Op5-LWint) they generate fragments from any hybrid genes that contain a MW exon 4 and LW exon 5 or *vice versa.* At the temperature of annealing selected these primers are thought not to prime to templates which contain nucleotides other than those corresponding to LW or MW sequence.

For each subject tested with these primer pairs, four PCR reactions were set-up.

Tube a	Op4+LWint / Op5-LWint
Tube b	Op4+LWint / Op5-MWint
Tube c	Op4+MWint / Op5-MWint
Tube d	Op4+MWint / Op5-LWint

Each 50 µl PCR was prepared as stated in section 2.3.2

The following thermal profile was employed

94°C for 180 sec denaturing step

94°C for 20 sec
62°C for 30 sec
72°C for 120 sec
} for 35 cycles

final step of 72°C for 600 sec.

The products of each of these PCR reactions were size fractionated in adjacent lanes upon 1.2 % agarose gels (section 2.3.3). Following staining and visualisation of the gel, photographs were taken.

5.3 Results

Colour vision testing on the Nagel anomaloscope revealed that DT required more green in a red-green mixture to match a yellow than did YT. The settings on the Nagel anomaloscope were 35.6 for DT and 43.75 for YT, where the mean setting for a population of young males is approximately 39. The instrument has a linear scale from

0 (= 100 % green) to 73 (= 100 % red). The mixture primaries are 546 nm (green) and 669 nm (red), while the yellow primary is 589 nm.

5.3.1 Integrity and yield of gDNA from DT and YT

Each 5 ml vial of non-clotted venous blood from subjects DT and YT yielded approximately 2 - 5 mg of high molecular weight gDNA (data not shown). It should be noted that the blood samples were received many days after they were collected. They had not been refrigerated, and were transported at room temperature. The integrity and yield of these DNAs was comparable to that obtained from freshly collected blood, or blood that had been stored at -20°C immediately following collection.

5.3.2 Sequences of opsin exons 2 to 5 of DT and YT

DNA from both DT and YT yielded solitary intensely-staining bands of the expected sizes (212 bp for exon 2, 164 bp for exon 3, 156 bp for exon 4, and 191 bp for exon 5; data not shown). For both subjects, all four exons were successfully amplified, cloned, and sequenced. The identity of the clones was determined by comparisons with the LW and MW sequences reported by Nathans *et al.* (1986a).

For DT, just 4 LW and 7 MW exon 2 sequences were generated, whilst with DNA from YT, 6 LW and 21 MW clones were sequenced. These clones were all identical to the previously published LW and MW sequences (Figure 5.2). It therefore follows that the amino acids encoded by these sequences would also be identical to the published LW and MW proteins (Figure 5.3). Exon 2 possesses just one amino acid site (at position 116) that is implicated in spectral tuning (Asenjo *et al.*, 1994), plus two other positions (65 and 111) which differ between the LW and MW opsin genes. These regions were used to discriminate between the opsin pigments.

In total, 13 opsin exon 3 clones from DT and 34 from YT were sequenced. Sequence data generated from these did reveal several differences between the two templates (Figure 5.4). Firstly, sequences obtained from subject YT consistently gave bases GCT for site 579 to 581, which translates into alanine at amino acid site 180, whilst sequences from subject DT always gave bases TCT for the same region,

Figure 5.2 Comparison of Exon 2 Nucleotide Sequences

```
                  216-
Human LW          TGG ATG ATC TTT GTG GTC ACT GCA TCC GTT TTC ACA
DT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ... ... ... ... ... .T. ... ... ... ... ...
YT  MW            ... ... ... ... ... ... .T. ... ... ... ... ...
Human MW          ... ... ... ... ... ... .T. ... ... ... ... ...

Human LW          AAT GGG CTT GTG CTG GCG GCC ACC ATG AAG TTC AAG
DT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  MW            ... ... ... ... ... ... ... ... ... ... ... ...
Human MW          ... ... ... ... ... ... ... ... ... ... ... ...

Human LW          AAG CTG CGC CAC CCG CTG AAC TGG ATC CTG GTG AAC
DT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  MW            ... ... ... ... ... ... ... ... ... ... ... ...
Human MW          ... ... ... ... ... ... ... ... ... ... ... ...

Human LW          CTG GCA GTC GCT GAC CTA GCA GAG ACC GTC ATC GCC
DT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ..G ... ... ... ..G ... ... ... ... ... ...
YT  MW            ... ..G ... ... ... ..G ... ... ... ... ... ...
Human MW          ... ..G ... ... ... ..G ... ... ... ... ... ...

Human LW          AGC ACT ATC AGC ATT GTG AAC CAG GTC TCT GGC TAC
DT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ... ... ... G.. ... ... ... ... .A. ... ...
YT  MW            ... ... ... ... G.. ... ... ... ... .A. ... ...
Human MW          ... ... ... ... G.. ... ... ... ... .A. ... ...

                                                          -428
Human LW          TTG GTG CTG GGC CAC CCT ATG TGT GTC CTG GAG
DT  LW            ... ... ... ... ... ... ... ... ... ... ...
YT  LW            ... ... ... ... ... ... ... ... ... ... ...
DT  MW            ... ... ... ... ... ... ... ... ... ... ...
YT  MW            ... ... ... ... ... ... ... ... ... ... ...
Human MW          ... ... ... ... ... ... ... ... ... ... ...
```

Figure 5.3 Comparison of Exon 2 Deduced Amino Acid Sequences

	59-											
Human LW	trp	met	ile	phe	val	val	thr	ala	ser	val	phe	thr
DT LW	-	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	-	-	-	ile	-	-	-	-	-
YT MW	-	-	-	-	-	-	ile	-	-	-	-	-
Human MW	-	-	-	-	-	-	ile	-	-	-	-	-

Human LW	asn	gly	leu	val	leu	ala	ala	thr	met	lys	phe	lys
DT LW	-	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	-	-	-	-	-	-	-	-	-
YT MW	-	-	-	-	-	-	-	-	-	-	-	-
Human MW	-	-	-	-	-	-	-	-	-	-	-	-

Human LW	lys	leu	arg	his	pro	leu	asn	trp	ile	leu	val	asn
DT LW	-	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	-	-	-	-	-	-	-	-	-
YT MW	-	-	-	-	-	-	-	-	-	-	-	-
Human MW	-	-	-	-	-	-	-	-	-	-	-	-

Human LW	leu	ala	val	ala	asp	leu	ala	glu	thr	val	ile	ala
DT LW	-	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	-	-	-	-	-	-	-	-	-
YT MW	-	-	-	-	-	-	-	-	-	-	-	-
Human MW	-	-	-	-	-	-	-	-	-	-	-	-

Human LW	ser	ile	ser	ile	val	asn	gln	val	ser	gly	tyr	phe
DT LW	-	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	val	-	-	-	-	tyr	-	-	-
YT MW	-	-	-	val	-	-	-	-	tyr	-	-	-
Human MW	-	-	-	val	-	-	-	-	tyr	-	-	-

											-130
Human LW	val	leu	gly	his	pro	met	cys	val	leu	glu	gly
DT LW	-	-	-	-	-	-	-	-	-	-	-
YT LW	-	-	-	-	-	-	-	-	-	-	-
DT MW	-	-	-	-	-	-	-	-	-	-	-
YT MW	-	-	-	-	-	-	-	-	-	-	-
Human MW	-	-	-	-	-	-	-	-	-	-	-

Figure 5.4 Comparison of Exon 3 Nucleotide Sequences

```
                474-
Human LW        GCC ATC ATT TCC TGG GAG AGG TGG CTG GTG
Subject DT a    ... ... ... ... ... ... ..G ... C.. ...
Subject DT b    ... ... ... ... ... ... ..A ... A.. ...
Subject YT      ... ... ... ... ... ... ..A ... A.. ...
Human MW        ... ... ... ... ... ... ..A ... A.. ...

Human LW        GTG TGC AAG CCC TTT GGC AAT GTG AGA TTT
Subject DT a    ..G ... ... ... ... ... ... ... ... ...
Subject DT b    ..C ... ... ... ... ... ... ... ... ...
Subject YT      ..C ... ... ... ... ... ... ... ... ...
Human MW        ..C ... ... ... ... ... ... ... ... ...

Human LW        GAT GCC AAG CTG GCC ATC GTG GGC ATT GCC
Subject DT a    ... ... ... ... ... ... ... ... ... ...
Subject DT b    ... ... ... ... ... ... ... ... ... ...
Subject YT      ... ... ... ... ... ... ... ... ... ...
Human MW        ... ... ... ... ... ... ... ... ... ...

                                                   -596
Human LW        TTC TCC TGG ATC TGG TCT GCT GTG TGG ACA GCC
Subject DT a    ... ... ... ... ... T.. ... ... ... ... ...
Subject DT b    ... ... ... ... ... T.. ... ... ... ... ...
Subject YT      ... ... ... ... ... G.. ... ... ... ... ...
Human MW        ... ... ... ... ... G.. ... ... ... ... ...
```

Figure 5.5 Comparison of Exon 3 Deduced Amino Acid Sequences

```
                145-
Human LW           ala ile ile ser trp glu arg trp leu val
Subject DT a       --- --- --- --- --- --- --- --- --- ---
Subject DT b       --- --- --- --- --- --- --- --- met ---
Subject YT         --- --- --- --- --- --- --- --- met ---
Human MW           --- --- --- --- --- --- --- --- met ---

Human LW           val cys lys pro phe gly asn val arg phe
Subject DT a       --- --- --- --- --- --- --- --- --- ---
Subject DT b       --- --- --- --- --- --- --- --- --- ---
Subject YT         --- --- --- --- --- --- --- --- --- ---
Human MW           --- --- --- --- --- --- --- --- --- ---

Human LW           asp ala lys leu ala ile val gly ile ala
Subject DT a       --- --- --- --- --- --- --- --- --- ---
Subject DT b       --- --- --- --- --- --- --- --- --- ---
Subject YT         --- --- --- --- --- --- --- --- --- ---
Human MW           --- --- --- --- --- --- --- --- --- ---

                                                      -185
Human LW           phe ser trp ile try ser ala val try thr ala
Subject DT a       --- --- --- --- --- --- --- --- --- --- ---
Subject DT b       --- --- --- --- --- --- --- --- --- --- ---
Subject YT         --- --- --- --- --- ala --- --- --- --- ---
Human MW           --- --- --- --- --- ala --- --- --- --- ---
```

Figure 5.6 Comparison of Exon 4 Nucleotide Sequences

```
                651-
Human LW           GGC CCA GAC GTG TTC AGC GGC AGC TCG TAC
DT LW              ... ... ... ... ... ... ... ... ... ...
YT LW              ... ... ... ... ... ... ... ... ... ...
YT Hybrid          ... ... ... ... ... ... ... ... ... ...
DT MW              ... ... ... ... ... ... ... ... ... ...
YT MW              ... ... ... ... ... ... ... ... ... ...
Human MW           ... ... ... ... ... ... ... ... ... ...

Human LW           CCC GGG GTG CAG TCT TAC ATG ATT GTC CTC
DT LW              ... ... ... ... ... ... ... ... ... ...
YT LW              ... ... ... ... ... ... ... ... ... ...
YT Hybrid          ... ... ... ... ... ... ... ... ... ...
DT MW              ... ... ... ... ... ... ... ... ... ...
YT MW              ... ... ... ... ... ... ... ... ... ...
Human MW

Human LW           ATG GTC ACC TGC TGC ATC ATC CCA CTC GCT
DT LW              ... ... ... ... ... ... ... ... ... ...
YT LW              ... ... ... ... ... ... ... ... ... ...
YT Hybrid          ... ... ... ... ... ... ... ... ... ...
DT MW              ... ... ... ... ... ... .C. ... ... AGC
YT MW              ... ... ... ... ... ... .C. ... ... AGC
Human MW           ... ... ... ... ... ... .C. ... ... AGC

                                                   -764
Human LW           ATC ATC ATG CTC TGC TAC CTC CAA
DT LW              ... ... ... ... ... ... ... ...
YT LW              ... ... ... ... ... ... ... ...
YT Hybrid          ... ... G.. ... ... ... ... ...
DT MW              ... ... G.. ... ... ... ... ...
YT MW              ... ... G.. ... ... ... ... ...
Human MW           ... ... G.. ... ... ... ... ...
```

Figure 5.7 Comparison of Exon 4 Deduced Amino Acid Sequences

```
                204-
Human LW            gly pro asp val phe ser gly ser ser tyr
DT LW               --- --- --- --- --- --- --- --- --- ---
YT LW               --- --- --- --- --- --- --- --- --- ---
YT Hybrid           --- --- --- --- --- --- --- --- --- ---
DT MW               --- --- --- --- --- --- --- --- --- ---
YT MW               --- --- --- --- --- --- --- --- --- ---
Human MW            --- --- --- --- --- --- --- --- --- ---

Human LW            pro gly val gln ser tyr met ile val leu
DT LW               --- --- --- --- --- --- --- --- --- ---
YT LW               --- --- --- --- --- --- --- --- --- ---
YT Hybrid           --- --- --- --- --- --- --- --- --- ---
DT MW               --- --- --- --- --- --- --- --- --- ---
YT MW               --- --- --- --- --- --- --- --- --- ---
Human MW            --- --- --- --- --- --- --- --- --- ---

Human LW            met val thr cys cys ile ile pro leu ala
DT LW               --- --- --- --- --- --- --- --- --- ---
YT LW               --- --- --- --- --- --- --- --- --- ---
YT Hybrid           --- --- --- --- --- --- --- --- --- ---
DT MW               --- --- --- --- --- --- thr --- --- ser
YT MW               --- --- --- --- --- --- thr --- --- ser
Human MW            --- --- --- --- --- --- thr --- --- ser

                                                   -241
Human LW            ile ile met leu cys tyr leu gln
DT LW               --- --- --- --- --- --- --- ---
YT LW               --- --- --- --- --- --- --- ---
YT Hybrid           --- --- val --- --- --- --- ---
DT MW               --- --- val --- --- --- --- ---
YT MW               --- --- val --- --- --- --- ---
Human MW            --- --- val --- --- --- --- ---
```

translating into serine at position 180. Also, all clones from YT were of one type, namely identical to the published MW opsin sequence (Nathans *et al.*, 1986a), whereas clones isolated from DT fell into two classes with approximately equal numbers; those which were identical to the published LW opsin sequence (sequence DT a in Figure 5.4) and those that were the same as MW opsin sequence at the 5' end but encoded serine at site 180, i.e. they differed at site 153 (sequence DT b in Figure 5.4).

Examination of opsin exon 4 clones (Figure 5.6) revealed that DT possessed both LW (28 clones sequenced) and MW sequences (21 clones sequenced), while YT possessed in addition to the LW (12 clones) and MW (20 clones) sequences, a single class of hybrid sequence (12 clones). All the hybrid clones were 5' LW and 3' MW, with a cross-over between amino acid 233 and 236 (Figure 5.7). In the region of exon 4 amplified there are three amino acid differences between LW and MW opsins. At site 236 methionine is replaced by valine in the transition from LW to MW, however, this is considered a conservative change, and does not contribute to the spectral positioning of the photopigment. Site 230, which codes for isoleucine in LW and threonine in MW opsins, and site 233, which encodes alanine in LW and serine in MW pigments, do contribute to the λ_{max} of the pigment and are therefore important in spectral tuning. In the hybrid sequence these two tuning sites are LW in nature, therefore, placing these sequences in the LW class. Tables 8 and 9 show the data.

34 exon 5 sequences were obtained from the DNA of subject DT (19 LW and 15 MW). All DT's sequences were identical to either the LW or MW sequences previously published by Nathans *et al.* (1986a; see Figures 3.14 to 3.17). Thus it was possible to conclude that, with the exception of an unlikely situation where the protocol failed to amplify a particular opsin class, the colour phenotype of DT was not the result of changes in exon 5.

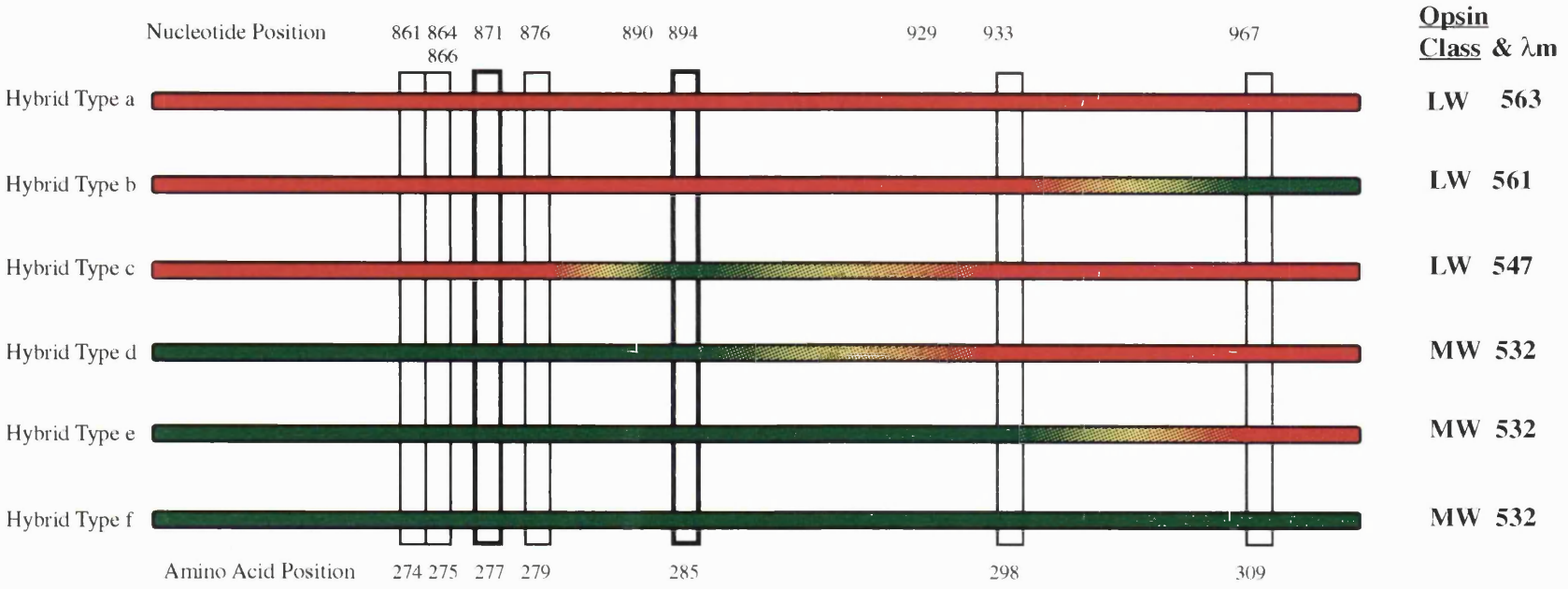
Table 8 shows the composition of all the exon 5 sequences obtained from the cloning of YT's DNA. No sequences agreed absolutely with the published LW or MW opsins. All 97 sequences possessed a C at position 890, and a C at base 929, regardless of which opsin class they encoded. Opsins from the LW class should have

exhibited a T at base 929, while MW clones should have encoded an A at base 890. These nucleotide bases at these positions were never observed following sequence analysis of opsin genes from other human subjects (a total of 9 males were examined, inclusive of controls, during the course of these investigations). Each base is at the third position of an codon and does not cause an amino acid change. Sequences of exon 5 type *a* represent LW exon 5, encoding all seven amino acids expected of a LW class opsin (Table 9). 18.5 % of the clones were of this type. Likewise, exon 5 sequences of type *f* possess all seven amino acids present in a MW opsin (representing 57.7 % of the clones). The remaining exon 5 sequences obtained (*b* to *e*) represent sequences whereby one end is of one opsin class and the other end is of the other class, with the cross-over point varying along the segment (Figure 5.8). These sequences are referred to as being derived from hybrid or chimaeric exons. For example, exon 5 type *c* is LW in nature from the 5' end across amino acids 274 to valine at site 279, but from the alanine at site 285 to the 3' end resembles a MW class opsin.

The spectral positioning of the opsins possessing these exon 5 sequences would be shifted from the values attributed to either the LW or the MW classes, to intermediate values. The work conducted by Asenjo *et al.* (1994), whereby each of the 15 amino acid residues which differentiate LW from MW opsins were each sequentially altered and the λ_{max} of the expressed protein measured, suggests that, with the assumption that exons 2 to 4 are not hybrids and are of the same class as the 5' end of exon 5, then, fragments of exon 5 type *a* would exhibit a λ_{max} of 563 nm and those of type *f* would have a λ_{max} of 532 nm. Hybrids of exon 5 type *b* would absorb maximally at 561 nm, while interestingly, hybrids of exon 5 type *c* would show a maximum absorbance at 547 nm. These values are based on the findings of Asenjo *et al.* (1994); see Figure 1.10.

Could the sequences exhibiting a hybrid nature be artefacts of the amplification process? A considerable proportion (approximately 25 %) of the clones revealed a hybrid sequence. Hybrid exons were encountered infrequently (on three separate occasions, representing less than 1 % of all clones sequenced) in the analysis of opsin

Figure 5.8 Diagramatic representation of Exon 5 hybrids sequenced from the DNA of YT. The red regions correspond to LW sequence, while green areas represent MW sequence. Areas in yellow depict the region across which a cross-over event occured.



gene sequences from other individuals, and it was assumed that they represented artefacts of the PCR and cloning process. The frequency at which they were recovered from the DNA of YT suggest that this explanation alone cannot account for the observations here. YT's sequences must represent amplification from genuine hybrid template(s), since five independent PCR and cloning experiments generated the data presented; in each instance, hybrid clones were produced. Do individuals with this number of hybrid opsin genes actually exist? The answer is yes if one considers the model proposed by Neitz and Neitz (1995a, b), where the analysis of a large number of human subjects suggests that hybrid opsin genes are frequently present in the arrays of colour vision normal individuals. It is probable that YT is such an individual with both LW and MW opsin gene classes and a wide range of exon 5 hybrids.

5.3.3 Disparity in the number of clones generated from each class

As shown in Table 10 there is considerable disparity in the numbers of clones sequenced from each opsin class. Does this reflect in any way the relationship between the numbers of each opsin class present at the genomic level? It is known for humans that the ratio between the LW and MW opsin genes can fluctuate from a ratio of 1:1 to 1:6 or more (Nathans, 1986a, b), with the additional presence of hybrid genes complicating the issue. A single PCR reaction when performed with a universal set of primers, and under the correct circumstances, will tend to maintain these ratios during the amplification process. Thus a sampling of the PCR reaction products should provide fragments with a ratio similar to that present on the template. This approach has been successfully adopted by the Perkin Elmer company in its heterozygote detection kits.

An aliquot of the amplified PCR product was ligated in the cloning vector, transformed into competent cells, and plated out. Individual positive colonies were randomly selected to re-PCR and then cycle sequenced. Each step of this multi-step process may have influenced the ratios of the two different types of opsin sequences represented amongst the sequenced clones. A statistical examination of the number of

clones of each type may shed light on whether the final clone count reflects the genomic complement of these genes.

The χ^2 statistical test was applied to test the hypothesis that the ratio of the LW and MW clones is 1:1. If the test involved the comparison of each individual exon then an inaccurate result would be generated since too few clones would be involved. Also, if the test considered all four exons then the data from exon 3 would provide a bias (since it is known that the difference in this exon is the fact that differentiates DT and YT). Therefore it was decided that the analysis should encompass data from exons 2, 4, and 5 only.

Testing the number of clones of DT first;

Class	Observed (O)	Expected (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
MW	42	45	9	0.2
LW	48	45	9	0.2
			$\chi^2 =$	0.4

degrees of freedom (df) = (2-1) = 1

From a table of χ^2 values a p value of approximately 0.5 is obtained, indicating that a deviation at least as great as that observed is expected at least 50 % of the time from chance alone.

Applying the same test to clones of YT;

Class	Observed (O)	Expected (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
MW	72	59.5	156.25	2.63
LW	47	59.5	156.25	2.63
			$\chi^2 =$	5.26

degrees of freedom (df) = (2-1) = 1

From a table of χ^2 values a *p* value of approximately 0.02 is obtained, indicating that a deviation at least as great as that observed would be expected in only 2 % of repeat experiments from chance alone. From this analysis it is obvious that the frequency with which the clones of YT were obtained does not correspond to a 1:1 ratio (*p* value less than 5 %). A further analysis tested with the null hypothesis that the ratio of the MW to LW clones is 2:1 follows.

Class	Observed (O)	Expected (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
MW	72	79.34	53.88	0.679
LW	47	39.67	53.88	0.679
			$\chi^2 =$	1.358

degrees of freedom (df) = (2-1) = 1

From a table of χ^2 values a *p* value between 0.5 and 0.25 is obtained, indicating that a deviation at least as great as that observed is expected at least 25 to 50 % of the time.

Finally, a ratio of 3:1 was tested;

Class	Observed (O)	Expected (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
MW	72	89.25	297.56	3.334
LW	47	29.75	297.56	3.334
			$\chi^2 =$	6.668

degrees of freedom (df) = (2-1) = 1

A *p* value between 0.01 and 0.005 is obtained, indicating that a deviation at least as great as that observed is expected only less than 1 % of the time. Thus the null hypothesis that the data represent a 3:1 ratio has to be rejected.

The above analysis suggests that for DT the ratio of MW to LW opsin genes is 1:1, and that for YT this ratio is 2:1, assuming that the sampling of clones represents a true reflection of the number of genes on the chromosome.

5.3.4 Exon 4 to exon 5 hybrid gene analysis

The agarose gels used to resolve the PCR products of the sequence specific primers revealed that both DT and YT possess a sequence that is MW exon 4 to MW exon 5, as well as another sequence that is LW exon 4 to LW exon 5. No bands of the target size were seen in the lanes when MW exon 4 was used with LW exon 5 primers or *vice versa* (Figure 5.9). Therefore, it is possible to conclude from these simple diagnostic PCR reactions that the DNAs of DT and YT do not contain any exon 4 to exon 5 hybrid genes. Considering the variability of exon 5 of YT (section 5.3.2), it is surprising not to have detected at least a single hybrid involving intron 4, which is over 10 times longer.

5.4 Discussion

As shown by the sequence data the genotypes of the two subjects are different and these probably account for the differing phenotypes. The opsin genes of subject DT show fewer departures from those of the normal observer (i.e. someone with two opsin genes; one LW and one MW), than do those of subject YT, who possesses a more complex arrangement. The X-linked array of DT contains two different exon 2 sequences (one LW and one MW), two different exon 3 sequences (both with serine, at site 180, but with different amino acids at site 153 - no sequences with alanine 180 were encountered), two different exon 4 sequences (a LW and a MW) and two different exon 5 sequences (LW and MW). Subject YT has two different exon 2 sequences (one LW and one MW), a single exon 3 sequence (MW based on the presence of methionine at site 153, and nucleotide differences), three different exon 4 sequences (a LW, a MW, and a LW-MW hybrid), and six different exon 5 sequences (LW, MW, and four hybrid types).

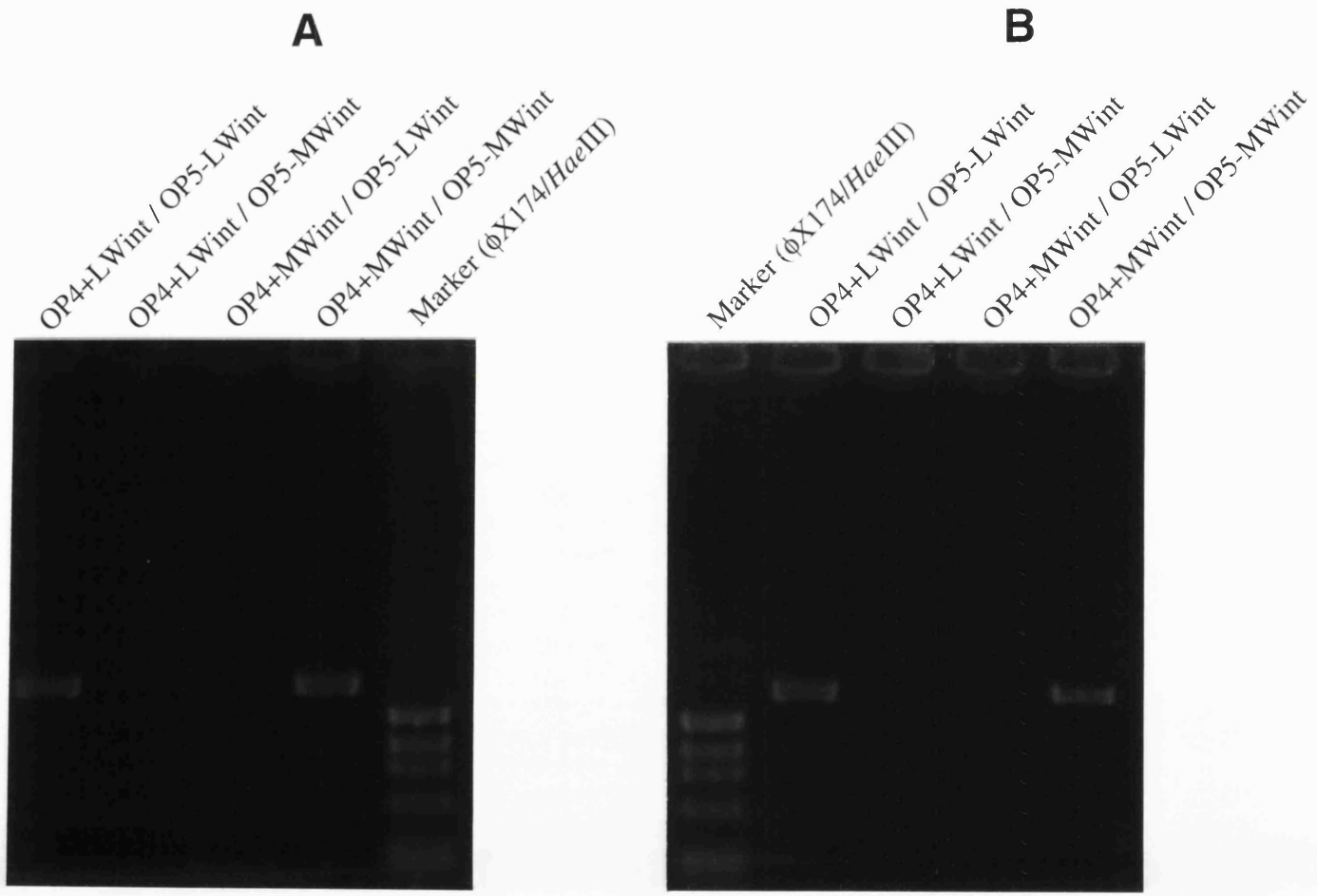


Figure 5.9 Diagnostic agarose gels used to detect the presence of exon 4 to exon 5 opsin gene fragments. The target bands are 1.7 kb in length. (The sequence of the gene specific primers and their specificity are given in the text). **A**. Subject YT. **B**. Subject DT.

All exon 3 sequences generated from subject DT have serine at site 180, whilst YT has alanine at this site. Interestingly, DT shows the presence of two types of exon 3; one with methionine at position 153, and the other with leucine at position 153, which were not observed with YT. Amino acid site 153 is not considered important in spectral tuning (Asenjo *et al.*, 1994).

The work of Winderickx *et al.* (1992) involving an extensive study of 50 young Caucasian males with normal colour vision who were tested using a combination of techniques (Rayleigh colour matches, SSCP analysis, and PCR amplification) also supports the presence of a single amino acid at site 180 (either alanine or serine) in all the LW and MW genes of individuals who produce Rayleigh matches with values from the extremes of the normal distribution. The amino acid change at position 180 of the LW opsin was found to be polymorphic in the sample, with 62 % of the subjects having serine (TCT), while 38 % had alanine (GCT). Further analysis suggested that a similar situation was present for the MW opsin gene, with 84 % having alanine and 16 % with serine. No other anomalies were reported as contributing to the phenotype of the subjects tested in that study (site 153 was not commented upon).

An important distinction between the finding of Winderickx *et al.* (1992) and those reproted here is that suject DT not only has the serine 180 allele of the LW pigment (which alone would place him in the lower half of the distribution of matches) but also the relatively rare (8%) of individuals with the serine 180 allele of the MW pigment (which ought to move his colour match a little further out to the extremes).

Amino acid residue 180 of the opsin apoproteins is now established as being polymorphic in the human population, for both the LW and MW genes (Nathans, 1994; Neitz and Neitz, 1995; Neitz *et al.*, 1995). There should exist sub-populations of each opsin class, each producing a photopigment with a slightly shifted absorbance spectrum. According to Asenjo *et al.* (1994) a prototype LW opsin, corresponding to that published by Nathans *et al.* (1986a) (see Figure 1.10) with Ala at site 180, has an absorbance maximum of 556 nm. The same sequence but with the substitution of Ser at site 180 shifts to the absorbance maxima to 563 nm. Similarly a prototype MW opsin

(Figure 1.10) with Ala 180 absorbs at 532 nm and with Ser 180 its λ_{max} shifts slightly to 534 nm.

The identification of six different types of exon 5 sequence from subject YT must mean that his X-linked gene array contains at least six exon 5's, and because the existence of partial opsin genes have never been reported, it must be interpreted that a minimum of six photopigment genes are present. Assuming the results of the χ^2 test to be a true reflection as to the 1:2 ratio amongst the LW and MW genes, it is possible to extend this argument one step further, by speculating that YT possesses 2 LW opsin genes and 4 MW genes. Both the model of Nathans *et al.* (1986a) and that of Neitz *et al.* (1994, 1995) suggest that multiple opsin genes are present in the arrays of many human observers, and indeed the work of these and other groups (Winderickx *et al.*, 1992) have shown this to be the case.

DT needed less red in his matches, suggesting increased sensitivity to long wavelengths. This correlates with the existence of serine at amino acid site 180 of all his opsin genes. Serine introduces an hydroxyl group into the binding pocket of the opsin photopigments resulting in the delocalisation of the charge on the retinal and resulting in a bathychromatic shift (see section 1.7.1). Thus, the data obtained here suggests the presence of two photopigments; one with a λ_{max} of approximately 563 nm (representing the LW exons 2, 4, and 5 associated with either exon 3 sequence), and the other with a λ_{max} of about 534 nm (representing the MW exons 2,4, and 5 associated with either exon 3 sequence). This narrowing of the spectral distance between the LW and MW pigments is responsible for DT matching at one extreme of the normal distribution of the Rayleigh test.

Subject YT requires more red light to match the standard yellow light. The presence of alanine at site 180 goes a long way to account for this result, but according to Dr. Mollon, YT's phenotype is more severe than can be explained in this way. The presence of the exon 5 hybrid genes will provide YT with a large complement of opsin genes able to code for photopigments spanning the range from 532 nm to 563 nm. At

present it is not known how many opsin genes in an array are expressed and whether the order of these genes is important. Winderickx *et al.*, (1993) have suggested that gene order is relevant in determining which opsin gene is expressed. It is probable that YT expresses a MW photopigment of the 532 nm class, as well as, another that has a λ_{max} blue shifted relative to the classical 563 nm type. However, the 547 nm pigment type is not expected to be expressed, since that would result in the situation seen with anomalous observers.

Base position	861	864	866	869	871	876	890	894	929	933	967	No.
Human LW	**A**	**T**	**T**	**G**	**A**	**G**	**C**	**A**	**T**	**G**	**A**	**0**
Exon 5 type *a*	a	t	t	g	a	g	c	a	c	g	a	18
Exon 5 type *b*	a	t	t	g	a	g	c	a	c	g	t	4
Exon 5 type *c*	a	t	t	g	a	g	c	g	c	g	a	6
Exon 5 type *d*	g	c	g	a	t	t	c	g	c	g	a	5
Exon 5 type *e*	g	c	g	a	t	t	c	g	c	c	a	8
Exon 5 type *f*	g	c	g	a	t	t	c	g	c	c	t	56
Human MW	**G**	**C**	**G**	**A**	**T**	**T**	**A**	**G**	**C**	**C**	**T**	**0**

Table 8. Exon 5 sequences from clones of subject YT

Deduced amino acid	274	275	277	279	285	298	309	No.
Human LW	**ile**	**phe**	**tyr**	**val**	**thr**	**ala**	**tyr**	**0**
Exon 5 type *a*	ile	phe	tyr	val	thr	ala	tyr	18
Exon 5 type *b*	ile	phe	tyr	val	thr	ala	phe	4
Exon 5 type *c*	ile	phe	tyr	val	ala	ala	tyr	6
Exon 5 type *d*	val	leu	phe	phe	ala	ala	tyr	5
Exon 5 type *e*	val	leu	phe	phe	ala	pro	tyr	8
Exon 5 type *f*	val	leu	phe	phe	ala	pro	phe	56
Human MW	**val**	**leu**	**phe**	**phe**	**ala**	**pro**	**phe**	**0**

Table 9. Deduced amino acids at sites that differ in Exon 5 of YT

	Subject DT		**Subject YT**	
	LW	MW	LW	MW
Exon 2	4	7	6	21
Exon 3	0	13	34	0
Exon 4	28	28	18	37
Exon 5	16	7	23	14

Table 10. Number of clones of each opsin class sequenced.

PART TWO

The Genetic Basis of Human Anomalous Trichromacy in Two Observers

5.5 History of anomalous trichromacy

As mentioned in the introduction anomalous trichromats were first identified by Rayleigh (1881) who found that these subjects required abnormal proportions of monochromatic red and green light to make a Rayleigh match. For almost any physical or psychological property, some variability within the normal population exists and, not surprisingly, this is true for colour perception. Consequently, anomalous trichromats have chromatic systems dramatically shifted relative to the normal observer (their Rayleigh matches are many standard deviations away from the normal match). For most practical purposes these differences in colour discrimination are of such a magnitude as to produce a very different colour experience, relative to normal.

The "shifted" wavelength sensitivities of anomalous trichromats are likely to be a consequence of slight changes in their cone pigments. Neitz *et al.* (1987) undertook a study of the role that variation in the number of MW cone pigment genes might have in determining colour preception. However, from their results it appeared that anomalous trichromacy did not correlate with MW opsin gene multiplicity. With the discovery of the hybrid genes by Nathans *et al.* (1986b) it was proposed that these are responsible for the shift in the wavelengths of maximum sensitivity. Therefore, it is of great interest to investigate the genes which code for the cone pigments of such individuals.

5.5.1 Previous findings

The LW and MW opsin loci of subject PSJ have been investigated in a previous study (Ibbotson, 1991). Briefly, he is a young male exhibiting the rare phenotype referred to as "minimalanomale Trichromaten". Just 1 to 5 % (dependent upon the population examined) of all anomalous trichromats are thought to be of this type (Dr. J.

M. Mollon, personal communication). Testing with the Ishihara plates shows him to be normal, whereas with the Farnsworth-Munsell 100-hue test and the Mollon-Reffin computer test of colour discrimination, he shows a superior normal hue discrimination. PSJ's colour anomaly is only revealed when his colour matches are compared to those of normal subjects. On the Rayleigh test (section 1.9.2) he makes a deuteranomalous match, requiring too much green light in the setting. Subjects who present with anomalous trichromacy are generally thought to possess either a LW or MW opsin pigment with a shifted λ_{max}, such that the two absorbance spectra are closer together than normal. However, for PSJ, who has excellent colour discrimination, this explanation did not seem appropriate. The main findings were the presence of multiple MW opsin genes, one of which was reported to carry a single bp deletion in exon 3; hybrid MW and LW opsin genes; and the presence of a normal LW opsin gene (Ibbotson, 1991). Presumably, the gene that carried the point deletion would encode a truncated non-functional photopigment. The opsin genes of subject SSJ had previously not been subjected to a molecular analysis.

The objective of this investigation was to confirm the existence of the deletion in exon 3 of one of the opsin genes from PSJ, and to attempt to determine the molecular basis of the visual anomaly exhibited by another male subject, SSJ, whose presented with the same phenotype as PSJ. It was reasonable to assume that a mechanism similar to that operating in PSJ could explain the visual phenotype of SSJ, since the two are very much alike.

5.6 Methods

5.6.1 Material for investigation

DNA prepared from semen samples provided by PSJ was already available in the laboratory following its isolation for a previous study (Ibbotson, 1991). Further gDNA was isolated from the blood of PSJ and subject SSJ as detailed in section 2.2.1.

5.6.2 Detection of a point mutation in exon 3 of one of PSJ's opsin genes

The primary goal of this investigation was to reproduce sequence data confirming the presence of the point deletion. If verified then the analysis would have been directed towards determining if the mutant protein was expressed in cone cells and thus deduce its role in the manifestation of the phenotype. If the deletion was not detected then an alternative explanation for PSJ's phenotype would be sought.

5.6.3 Amplification and sequencing of exon 3 of PSJ

Exon 3 of the X-linked opsin genes of subject PSJ were PCR amplified exactly as stated in section 3.2.2, using Op3+ and Op3- primers. The products were cloned into either pCR 1000 or pTAg cloning vectors (section 2.4.5) prior to manual (section 2.6) and automated (Section 2.7) sequencing.

5.6.4 Amplification of exon 3 using deletion point specific primers

By carefully controlling the annealing temperature during the PCR cycling process it is possible to generate products based on the sequence of the primers. The primers 3' terminal bases are the most critical. Should they not be exactly complementary to the bases on the template then the DNA polymerase is generally unable to add bases and thus fails to promote primer extension. By performing PCR reactions over a series of annealing temperatures one can calculate the optimal temperature with which to perform the cycling reaction.

The three primers shown below were synthesised, all hybridising from base 552 to either base 579 or 580.

Op3+n/LW 5'-GTGGGCATTGCCTTCTCCTGGATCTGGT-3'

Op3+n/MW 5'-GTGGGCATTGCCTTCTCCTGGATCTGGG-3'

PSJ-deleted 5'-GTGGGCATTGCCTTCTCCTGGATCTGGC-3'

Each was designed to anneal to a specific opsin template; Op3+n/LW will only hybridise to LW opsin exon 3, Op3+n/MW just to MW exon 3, and primer PSJ-deleted

should just hybridise to any exon 3 with a point deletion at base 579. Standard 50 µl PCR reactions were set up using one of the above specific primers along with Op3- as the reverse PCR primer. Products of 62 bp in length were expected. The annealing temperature that allowed discrimination between the different classes of opsin genes, was found to be 65°C, which was obtained experimentally by testing each primer pair on plasmid clones known to possess opsin exon 3 inserts (data not shown). PCR products were electrophoresed through a 8 % non-denaturing acrylamide gel, and then visualised, to detect products.

5.6.5 Direct sequencing of exon 3 of PSJ

Following co-amplification of exon 3 of the MW and LW opsin genes from genomic DNA of PSJ (see section 5.6.3), direct sequencing of these products was performed. This approach has been shown to detect with great accuracy the presence of insertions or deletions in genes with multiple loci. Essentially, during the cycle sequencing process the sequencing primer clamps down on all the templates within the reaction mix. Upon extension deoxyribonucleotides are incorporated until by chance a dideoxyribonucleotide terminates the extension. When these labelled products are electrophoresed on the automated sequencer the electropherogram trace from all the extension products stays in register until a deletion or insertion is encountered. Thereafter, multiple peaks at most positions are encountered, representing more than one sequence, which is analogous to double banding seen on manual sequencing autoradiographs.

The protocol involved the excision of target bands from agarose gels and their elution into water (section 2.4.4.1). 5 µl of this eluate was mixed with automated sequencing components (section 2.7.3; using Taq Dyedeoxy terminator kits) prior to cycle sequencing.

5.6.6 Amplification and sequencing of exons 2 to 5

Exons 3, 4, and 5 of SSJ and exons 4 and 5 of PSJ were PCR amplified exactly as stated in section 3.2.3. The products of the PCR were eluted from agarose gels and subsequently cloned and sequenced via the automated approach (section 2.7) Exon 2 from both SSJ and PSJ was amplified, cloned , and sequenced as given in section 5.2.3 above.

5.6.7 Amplification of opsin exon 4 to exon 5

The gDNA from PSJ and SSJ were each subjected to PCR reactions with the four combinations of transintronic primers (see section 5.2.5). 50 µl standard reactions were set up as given in section 2.3.2. PCR cycling parameters were: initial denaturation at 94°C for 5 minutes, cycling for 35 cycles at 94°C for 30 seconds, 64°C for 30 seconds, and 72°C for 2 minutes; a final extension for 5 minutes was also included. Products of the reaction were resolved on 1.2 % agarose gels.

5.7 Results

5.7.1 Mutation detection in exon 3 of PSJ

72 sequences were obtained from positive clones, of which 31 were identical to published LW opsin sequence (Nathans *et al.,* 1986a), and the remainder (41) were identical to MW sequence. No sequences exhibiting a point deletion were seen. This number of clones was considered more than adequate to have detected the mutation. Other experimental procedures were adopted to reinforce this finding.

Figure 5.10 shows the region of a direct-sequenced electropherogram where the deletion was originally found. The sequence remains in register across this region, indicating that a deleted template was not present.

In addition both primer pairs Op3+n/LW / Op3- and Op3+n/MW / Op3- did produce fragments of the correct size. However, the primer combination, PSJ-deleted and Op3-, did not generate any detectable product (data not shown). The identity of the positive bands was not confirmed. This result also suggested that there may not be a deleted opsin gene in the genome of PSJ.

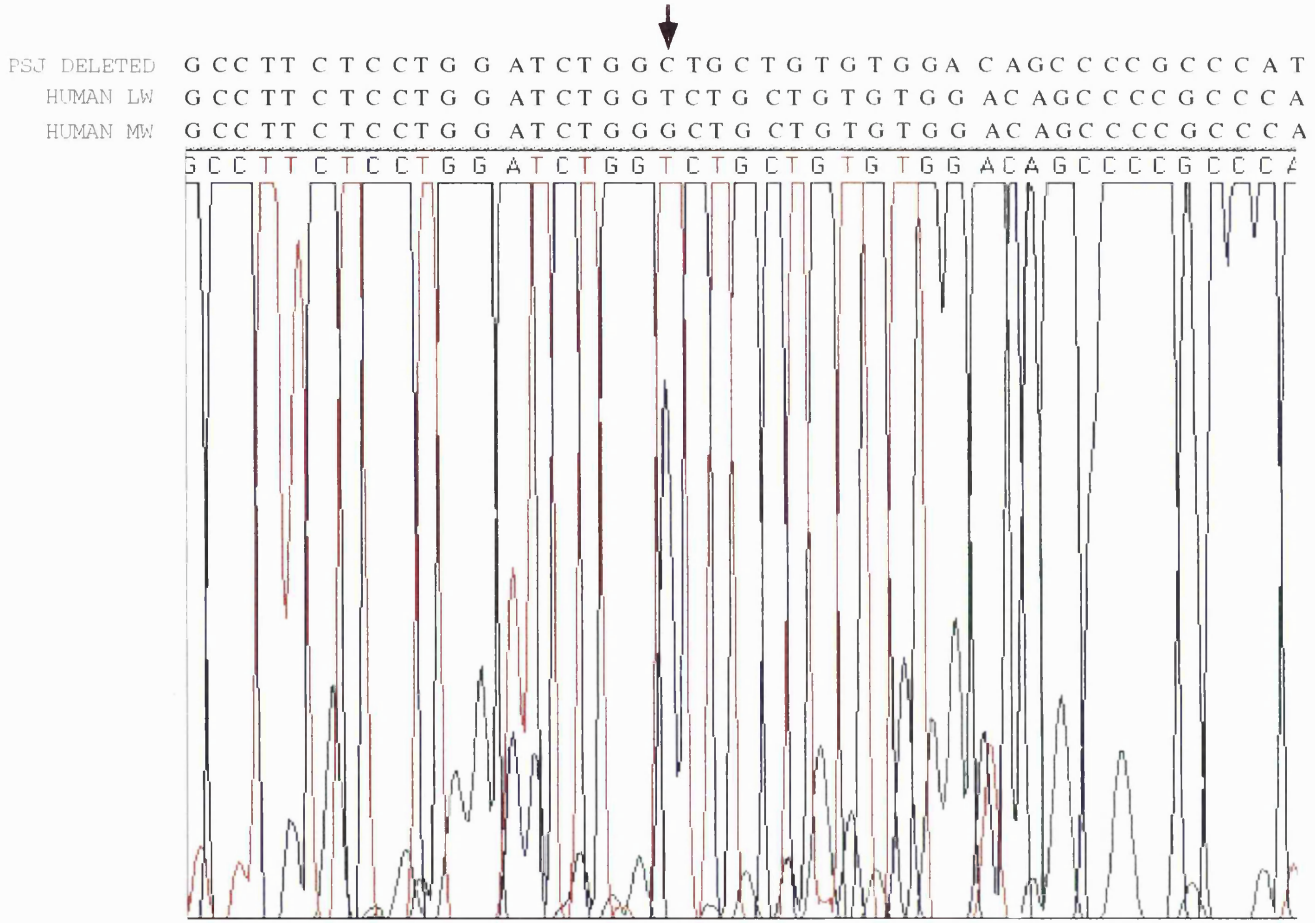


Figure 5.10. Portion of an electropherogram generated upon direct sequencing exon 3 PCR products amplified from PSJ, using primer pair Op3+/Op3-. These primers will co-amplify all X-linked opsin genes to which they hybridise. The arrow indicates the position at which the deleted nucleotide was reported. The sequence given directly above the graph is that generated from the template. The sequences at the top are those expected from the various opsin classes. It is clear that the deleted sequence was not represented amongst the sequenced products.

5.7.2 Cloning and sequencing of exon 3 of SSJ

A total of 26 exon 3 clones were analysed, of which 18 were MW in nature and the reminder were LW. All sequences conformed exactly to previously published sequences (Nathans *et al.*, 1986a).

5.7.3 Sequences of exons 2, 4, and 5 of PSJ and SSJ

From the DNA of subject PSJ a total of 43 exon 5 clones were recovered. Of these 29 were identical to published MW sequence, and 14 were of the LW type.

For subject SSJ a large number of exon 5 clones were sequenced (62). Interestingly, all the clones were LW in spectral class. No MW clones were ever recovered. Direct sequencing of this region was undertaken. Once again the sequence was entirely consistent with the presence of only a LW exon 5 opsin sequence (data not shown).

Exon 4 opsin gene sequences were obtained from both subjects. All sequences conformed to the previously published sequences. All clones were either wholly LW or MW (data not shown).

Clones of opsin exon 2 sequences were amplified too. 8 clones from PSJ (5 MW and 3 LW), and 11 from SSJ (5 MW and 6 LW) were obtained. Once again no departure from the published LW or MW sequences was observed.

5.7.4 Exon 4 to exon 5 opsin gene analysis

Figure 5.11 shows the results of the PCR amplification of the relevant regions. It is apparent that the genomes of the two subjects are not alike in their complement of X-linked opsin genes. DNA from PSJ generated fragments which represented genes which carry exon 4 MW to exon 5 MW, exon 4 LW to exon 5 LW, as well as a product generated with the exon 4 MW and exon 5 LW primer pair (Figure 5.11a lane 3). Therefore, from this data alone it is possible to deduce that subject PSJ possesses at least three opsin genes, one of which is a 5'MW-3'LW, with a cross-over in intron 4.

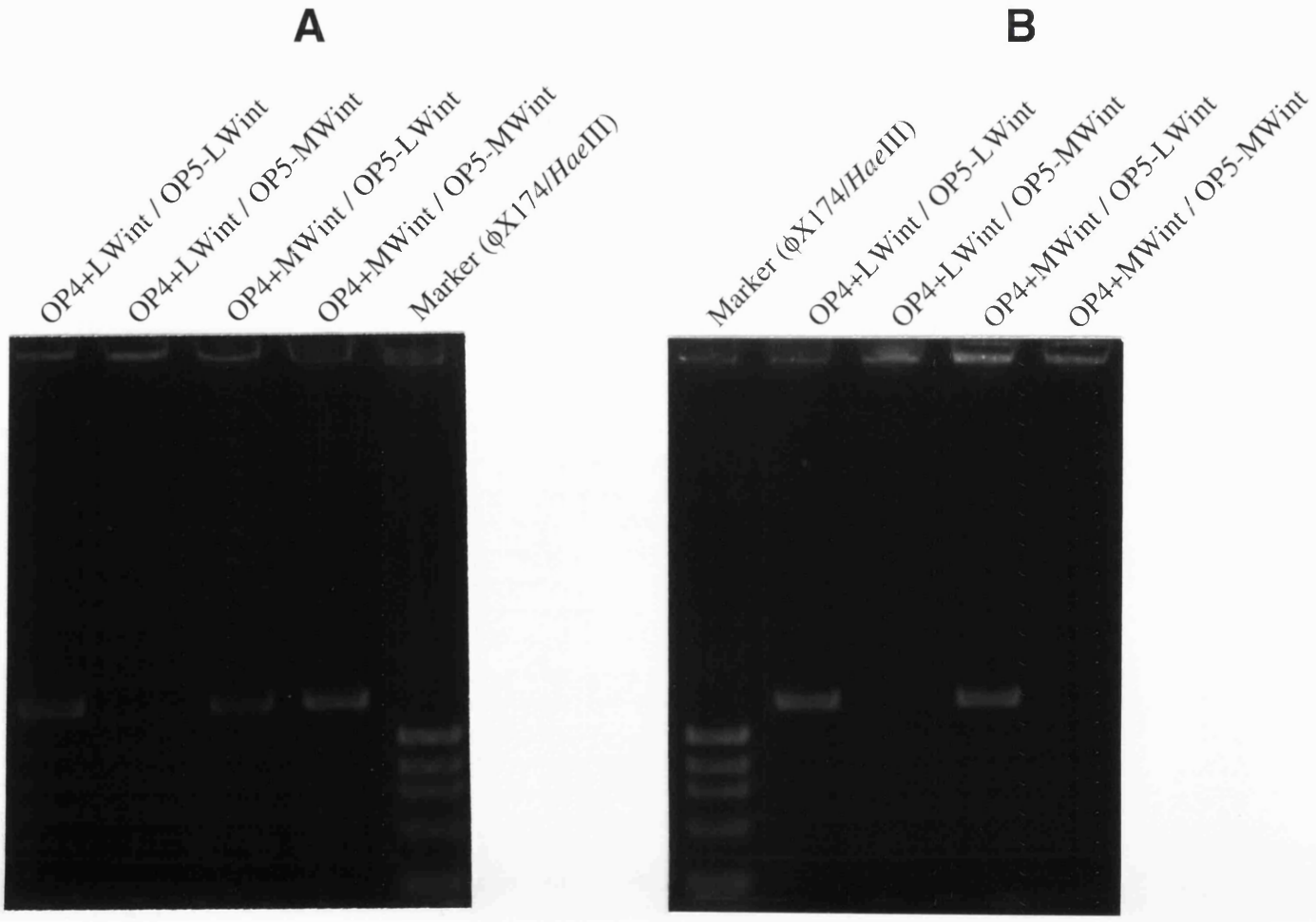


Figure 5.11 Agarose gels used to detect the presence of specific opsin arrangements. (details as per Figure 5.9). **A**. Subject PSJ. **B**. Subject SSJ.

The same experiments reveal that the opsin complement of subject SSJ also contains a hybrid opsin gene of the same type as seen with subject PSJ (Figure 5.11b lane 3). Interestingly, the primer pair Op4+MWint/Op5-MWint failed to amplify a target product, strongly suggesting that no MW opsin 5 sequence is present. This complements the results of exon 5 sequencing where no MW sequences were found. Thus SSJ must possess at least two opsin genes; one of a LW class and the other a 5'MW-3'LW with a cross-over in intron 4. Deuteranomalous individuals with good colour discrimination can also be of this type. Thus, there is a degree of overlap between different genotypes and phenotypes.

5.8 Discussion

The present study failed to isolate the MW opsin gene fragment with a point deletion, previously reported by Ibbotson (1991) to exist in the genome of PSJ. The PCR amplification procedure, using exon 3 primers, alone should have generated the deleted fragment, as it had in the original investigation. This did not happen, even though over 70 clones generated from independent PCR reactions were sequenced. Specific PCR primers were also utilised. These too should have succeeded in priming any opsin gene carrying a deletion. Direct sequencing of the target PCR products may have detected the presence of the deleted segment, but did not. During the course of this work a clone from chimpanzee clearly generated a sequence that showed a single base deletion in exon 5 of a LW opsin gene (data not shown). When the same clone was sequenced along the opposite strand the deletion could no longer be detected. Further clones from this region did not reveal the existence of the point deletion. Clearly, the single occurrence was most likely an artefact of the cloning or sequencing process. This leaves only two possible outcomes: either all these procedures failed to detect the deleted fragment, or that there is no deleted exon 3 present in the genome of PSJ.

Hybrid opsin sequences have been encountered consistently upon examination of human LW and MW genes (Nathans *et al.*, 1986a; Deeb *et al.*, 1992; Winderickx *et al.*, 1992; Neitz and Neitz, 1995).

The tandem nature of the X-linked opsin gene arrays, in conjunction with the high degree of sequence homology between these genes, appears to predispose them to unequal homologous recombination events (Nathans *et al.*, 1986a; Vollrath *et al.*, 1988; Feil *et al.*, 1990). On the basis that the original model proposed by Nathans *et al.* (1986a) is correct, then the 5' end of the LW opsin gene will in each case abut unique flanking DNA, while each MW gene would reside as an island flanked by 24 kb of almost identical intergenic DNA. Crossover events involving the regions between genes (intergenic recombination), in the presence of the asymmetry in the arrangement of the two classes of opsins genes, leads to the duplication and/or deletion of MW but not LW pigment genes. Changes in the number of opsin genes within arrays are a consequence of this mechanism. As Figure 1.13 illustrates, expansion of one array results in the contraction of the other. Intragenic recombination events (crossovers within genes) result in the generation of hybrid pigment genes (see below). An analysis of a substantial number of human males exhibiting variant red/green colour vision suggests that these recombination events are responsible for the vast majority (>97%) of variant genotypes (Nathans *et al.*, 1992); a reflection of the high frequency with which hybrid genes are present in the normal population.

What are the consequences of intragenic recombinations between LW and MW opsin genes? Before understanding the various outcomes it is necessary to quickly refresh knowledge of the terms employed. If the respective terms LW^{+} and MW^{+} are used to refer to the fully normal LW and MW photoreceptor phenotypes, LW^{-} and MW^{-} denote the absence or non functionality of the respective cones, and LW' and MW' designate pigments exhibiting abnormal phenotypes, then it is possible to indicate the colour phenotype of an individual with reference to these terms. Subjects whose colour vision is normal would be designated LW^{+} MW^{+}. Protanopes would be $LW^{-}MW^{+}$, while deuteranopes would be $LW^{+}MW^{-}$. Subjects who require different ratios of

red/green primaries in Rayleigh tests from those selected by the majority of trichromats are referred to as anomalous trichromats, for which there exist two categories; protanomalous individuals are abbreviated as LW'MW^{+}, for they possess LW pigments with abnormal properties along with normal MW pigments, whereas, deuteranomalous individuals, who have abnormal MW pigments, such that they have the spectral properties atributed to LW pigments, and normal LW pigments, are LW^{+}MW'.

Intragenic recombination events involving LW and MW pigment genes that replace the normal LW pigment gene with a 5'LW-3'MW hybrid gene produce both LW^{-}MW^{+} (protanopes) and LW'MW^{+} (protanomalous observers; adopting the Nathans model where a single LW opsin 5' region is present). Another recombination product where the LW gene is unaffected, but in which a normal MW pigment gene is replaced by a 5'MW-3'LW hybrid, depends on the number of MW opsin genes present in the array. If the array contains additional MW genes then deuteranomaly (LW^{+}MW') or normal colour vision (LW^{+}MW^{+}) can result, with the order of the opsins playing a major role. Should there be no other MW pigment genes, the phenotype of the carrier may be LW^{+}MW' (deuteranomaly) or LW^{+}MW^{-} (deuteranopia) depending on the position along the opsin gene at which the crossover takes place (see Figure 1.13).

Theoretical considerations, supported by limited experimental data (Neitz *et al.*, 1989; Nathans *et al.*, 1989c; Deeb *et al.*, 1992; Deeb *et al.*, 1993) suggest that a defined number of different hybrid pigments are present at appreciable frequencies in the human gene pool. The nucleotide differences between the LW and MW genes are not distributed evenly along the length of each of the four exons (2 to 5), but occur localised to certain regions within these exons (illustrated in Figures 1.7, 1.9 and 3.3). This makes it far more likely that a chance recombination event will fall outside of these regions (that is, either in the ends of the exons or more likely in the introns), and thus tend to maintain exons as either LW or MW derived. Given the data presented here and that obtained by others (Winderickx *et al.*, 1992c), that exon 3 is polymorphic, at site 180, in both the LW and MW opsin genes, then the number of common hybrid pigments can be increased.

Analysis of the absorption maxima of the anomalous pigments of LW'MW$^+$ (protanomalous) and LW$^+$MW' (deuteranomalous) subjects, by means of psychophysical tests in the early 1970's, demonstrated that the λ_{max} values of these pigments lay between the absorption maxima of the normal LW and MW pigments (Piantanida and Sperling, 1973b; Piantanida and Sperling, 1973a; Rushton *et al.*, 1973). More recently a series of studies utilising cloned cDNA (Merbs and Nathans, 1992b; Asenjo *et al.*, 1994) has demonstrated that there is indeed a pattern to the hybrid pigments with crossover points between exons, and that the expression of hybrid genes may be the major cause of colour vision deficiencies.

The only evidence that can account for the anomalous trichromacy of subject PSJ is the presence of only a single type of hybrid gene, between exons 4 and 5, which clearly indicated the existence of an exon 4 MW to exon 5 LW hybrid. All other exonic sequences conformed to previously reported sequences from colour normal subjects. The absence of MW exon 5 sequences effectively means that SSJ lacks photopigments that can be classed as MW; all his photopigments will be of the LW class.

Do the different genetic complements of the two subjects result in the same colour phenotype, as suggested by psychophysical testing? The answer to this question depends on the way genes in the X-linked opsin array are expressed within individual cone cells, as well as, the retina as a whole. Consider a situation where a person's opsin complement consists of say ten genes, which represent not only LW and MW classes but also hybrid genes encoding pigments that absorb at intermediate λ_{max} values. MSP data on individual cone cells points to the expression of one opsin gene per cell (Bowmaker, 1980). Other evidence suggests the expression of only two opsin genes from a multi-gene array (Winderickx *et al.*, 1992a). If these arguments represent the *in vivo* situation of opsin gene expression then one can postulate the following for the photopigment complements of SSJ and PSJ. Both subjects express a LW opsin gene and a 5'MW-3'LW hybrid gene. The MW gene of PSJ is possibly not expressed, perhaps because it is either missing important regulatory regions or because the

regulatory mechanism chooses to express the hybrid gene over this one, possibly as a result of its position 3' to the expressed genes.

Both PSJ and SSJ probably possess two photopigments with the same spectral properties, which will then impart on both of them the same colour phenotype, even though the genotypes are not identical. Why then are PSJ and SSJ not ordinary anomalous observers? As mentioned earlier, the fact that their Rayleigh match values are so consistent between tests requires an alternative explanation. One possible explanation could be that the different classes of cone cells possess differing densities of photopigments. Alternatively, the differences in exons 2 to 4 produce an adequate separation for good discrimination.

5.9 General conclusions

The colour anomalies of the four human subjects studied here are each due to genetic alterations in the normal organisation and arrangement of opsin genes. The close homology of the two X-chromosome opsins promotes the formation of hybrid genes through unequal cross-over events. Since the number of genes in the array can vary as well as the hybrid nature of each gene, it would appear that a vast number of possible arrangements is possible. Colour vision defects of a particular category, namely anomalous trichromacy, are the result of such hybrid genes and polymorphisms.

Technological refinements offer new methods of estimating the numbers and ratios of genes that remedy limitations inherent in quantitative Southern analysis. An alternative theory to that of Nathan's (see section 1.11.2) has been proposed by Neitz and Neitz (1995; Neitz *et al.*, 1995) based on both new research data and re-analysis of existing data. Essentially this new model postulates the existence of more genes in the X- linked arrays, in addition to the presence of hybrid genes, in both colour normal and colour defective observers. It also states that each spectral class of pigments can consist of spectrally distinct subtypes. All observers, be they normal or colour defective, draw from the same pool of photopigments.

Considering that homology between the LW and MW opsin genes extends upstream to base - 195 (which is 236 bp from the first amino acid) any cross-over up to this point, between a LW gene and a MW gene, will place the unique 5' flanking region of the LW opsin adjacent to the MW opsin; resulting in the generation of a gene with the spectral properties of a MW opsin and the distal regulatory region of a LW opsin. The reciprocal product, where a MW distal regulatory region is juxtapositioned upstream of a opsin with LW spectral properties, can also result.

Knowledge of the exact gene arrangement of the X-linked opsin genes will have to await a more systematic approach, such as chromosome walking. DNA from both normal observers and those afflicted with colour-vision deficiencies will need to be examined to enable a full and accurate account to be compiled.

CHAPTER 6

Molecular Genetics of the Colour Blindness of John Dalton

6.1 Introduction

John Dalton (1766-1844; Figure 6.1), is famous for his mathematical formulae describing the properties of gases, the measurement of molecular mass (kiloDalton, kD or kDa), and amongst others for his detailed observations and theory of colour-blindness.

At a very early age Dalton yearned to be a physician. In pursuit of this goal he took a keen interest in botany, chemistry, and mathematics, which were then a prerequisite for entry to medical school. In addition he maintained a keen interest in his hobbies of astronomy and optics. During his many observations of natural phenomena Dalton noted that he could not distinguish hues as others appeared to. He is noted as often having to ask others of the colours of objects they saw (Dalton, 1798).

Dalton became fully aware of his colour vision defect when accidentally observing the colour of the cranesbill, *Geranium zonale*, (Figure 6.2) in differing lighting conditions. In a letter addressed to Elihu Robinson, an acquaintance, in the spring of 1793, he wrote: "The flowers of most of the Cranesbills appear to me in the day, almost exactly *sky blue*, whilst others call them *deep pink*; but happening once to look at one in the night by candle light I found it of a colour as different as possible from day light; it seemed then very near yellow, but with a tincture of red" (Lonsdale, 1874). This lead him to pursue the matter further. By observing numerous objects and meticulously recording his observations, and from comparing his opinions of hues with those of others, he came to know nearly twenty persons similarly affected with the same defect. Amongst them was his own brother, Jonathan. Interestingly, Dalton recorded that "I have not heard of one female subject to this peculiarity" (Dalton, 1798).

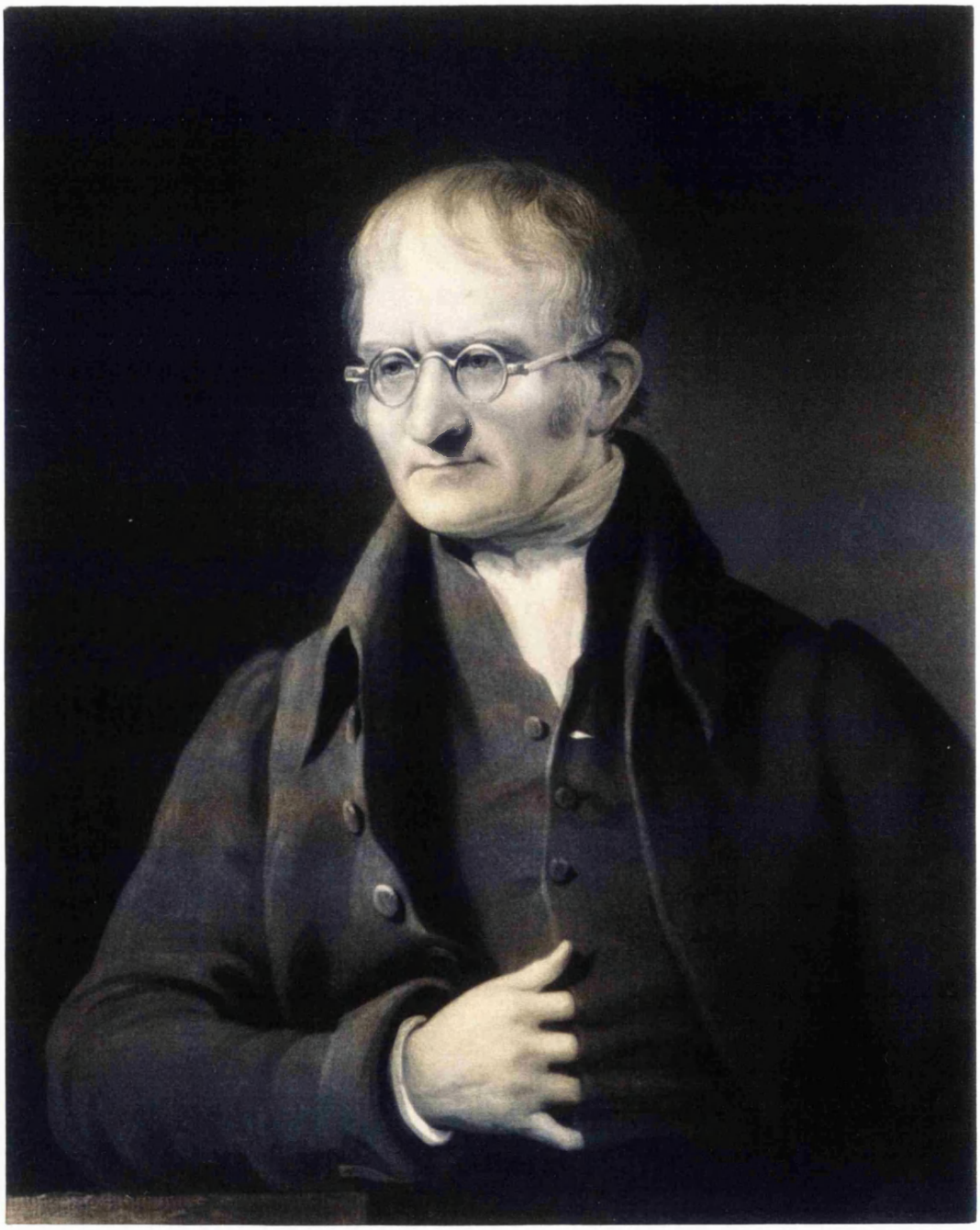


Figure 6.1 John Dalton. An engraving dated 1836, reproduced by permission of the British Library.

Over two centuries have passed since Dalton discussed his colour defect in a paper entitled "Extraordinary Facts relating to the Vision of Colours; with Observations", before the Manchester Literary and Philosophical Society on October 31st, 1794 (Dalton, 1798). At this time the Cranesbill was identified as *Geranium zonale*; and from contemporary descriptions (Martyn, 1777) it is fairly certain that Dalton was referring to the plant now named *Pelargonium zonale*, a native of South Africa cultivated in England as early as 1710 (van der Walt, 1977), (Figure 6.2). Subsequently, a great deal of interest was generated in the subject of defective colour vision. So much so that the term "daltonism" was used to describe the defect and still persists today as the term for colour blindness in languages as diverse as Spanish and Russian. Many in Britain objected to the association of a visual defect with the name of their distinguished countryman (Whewhell, 1841), but Dalton himself was unmoved.

In general, Dalton found that he confounded crimsons with blues, and scarlets with greens. "*Red* and *scarlet* form a genus with me totally different from pink". The vermilion colour of sealing wax was a good match to the outer face of the leaf of the laurel (*Prunus laurocerasus*), while a crimson ribbon matched the colour that others called "mud" (Dalton, 1798; Lonsdale, 1874). In the solar spectrum he saw only two main hues, one of which corresponded to the normal observer's red, orange, yellow and green, while the second corresponded to blue and violet. And in a phrase that was often to be repeated, he described the red end of the solar spectrum as "little more than a shade or defect of light".

In explanation of his defect, Dalton himself proposed that his eye must contain a blue filter, selectively absorbing long wavelengths. Since no coloration was visible by external inspection, he suggested that it was his vitreous that was tinted. Holding to the end of his life that his eye must contain a blue filter, he gave instructions that his eyes should be examined upon his death. He died on 27th July, 1844, and the following day his medical attendant, Joseph Ransome, conducted a post mortem (Henry, 1854). Ransome dissected one eye, and found the aqueous to be "perfectly pellucid", as was the vitreous, while the lens was slightly amber-coloured "as usual in persons of

Figure 6.2 *Pelargonium zonale*. A specimen from stock collected in South Africa that almost certainly corresponds to Dalton's "*Geranium zonale*".

advanced age". The second eye Ransome left very largely intact, removing the posterior pole by a vertical section and examining scarlet and green through it, as though through a lens. Colours remained distinct when observed in this way and thus Ransome's observations showed conclusively that Dalton's own hypothesis, that colour blindness was due to a pre-retinal filter, was mistaken.

Ransome recorded, possibly under pressure to do so by his peers, (Henry, 1854) that the defect was the result of a "deficient development" of one of the convolutions of the frontal lobe, part of the cerebral cortex. Present knowledge of colour blindness points to the absence or the alteration of one of the photosensitive pigments of the retina, and not to any defect within the brain.

From Dalton's own description of his colour blindness, Thomas Young, the architect of the wave theory of light, postulated that colour defects resulted from "the absence or paralysis of those fibres of the retina, which are calculated to perceive red". In modern terms therefore, Dalton's colour blindness was thought to arise from the absence of red receptors. This would have made him a protanope rather than a deuteranope, in which the green receptor is absent.

The remains of one eye has survived, stored dry between watch glasses, first in the possession of Dalton Hall (Brockbank, 1944) and then of the Manchester Literary and Philosophical Society. The Manchester Museum of Science and Industry presently have care of the eye.

DNA begins to deteriorate immediately after cell death (Herrmann and Hummel, 1994), resulting in DNA damage, particularly as a result of background radiation. However, the rate at which DNA damage is acquired is dependent upon both its environment, and the time elapsed since cell death. DNA which has sustained damage for thousands of years can still be extracted and used as template for PCR amplification (Pääbo, 1993). However, in those cases where successful amplification has been reported the length of the products seldom exceeds several hundred bp (Hagelberg *et al.*, 1991; Herrmann and Hummel, 1994).

The DNA in the tissue of Dalton's remaining eye tissue has been subject to these damaging agents since 1844. Considering the favourable conditions for preservation of the tissue, it was not unreasonable to assume that DNA fragments of a longer length should be amplifiable.

It is generally accepted that PCR amplification has revolutionised the field of ancient DNA (aDNA; Pääbo, 1989). Prior to the utilisation of the PCR reaction (Saiki *et al.*, 1985; Mullis *et al.*, 1986), all reports pertaining to the successful extraction of DNA from tissues relied on conventional DNA extraction methodology (Hunan Medical College, 1980; Higuchi *et al.*, 1984; Pääbo, 1985), succeeded by cloning and sequencing. The problem associated with these cloning techniques was that they required large quantities of the precious preserved material, and produced unreliable and non-reproducible sequence data. Only with the advent of the PCR method in the mid-1980's by Kary B. Mullis (1986 and 1987), did the study of aDNA become a reality, and revolutionise the search for aDNA, as well as aRNA, since the amount of source material needed for successful cloning was a fraction of that needed by previous protocols. Additionally, the very nature of the PCR reaction favours the generation of fair lengths of recoverable nucleic acids. Consider the following: for conventional cloning strategies attempts are made to recover any remaining molecules of nucleic acids as they exist. This means that only just that material which has survived in the tissues can be cloned, including all the damage sustained by the DNA. Upon cloning, the damaged molecules of DNA can cause many problems, and thus reduce the efficiency of recovery. With the PCR amplification process, regions of the surviving nucleic acids are targeted by the primer pairs. These priming sites require that at least the 3' end of the primer hybridises to its target region (the degree of homology being dependent upon the kinetics of the reaction), before any amplification can proceed. Once amplification from a primed site starts it may or may not reach its target length, that is across to the complementary priming site. If the target distance is not too large and damage to the DNA molecules is restricted, then successful amplification will normally proceed without problems from the first cycle of the PCR reaction. However,

should there exist substantial damage to the DNA or the distance across which amplification is desired be too large then the chances of success are remote, but not impossible. As Figure 6.3 illustrates, with PCR amplification it is possible to bridge the gap between critical DNA damage, by template swapping during the initial rounds of PCR, and thus generate a template for cloning.

6.2 Methods

6.2.1 Sample collection

The remains of Dalton's eyes (Figure 6.4) are still in the possession of the Manchester Literary and Philosophical Society, Manchester, who kindly gave permission on two occasions for small sections to be removed from the eye tissue for DNA analysis in this study.

Using sterile techniques at all times, samples were carefully cut by scalpel blades from the internal regions of the desiccated tissue, tentatively identified as peripheral retina. These retina samples were placed in sterile 1.5 ml microcentrifuge tubes and stored at room temperature. These same tubes were used to transport the samples to the laboratory where they remained until the analysis was begun.

6.2.2 Initial attempts to isolate DNA

Scrutiny of the literature suggested that it should be possible to isolate DNA from favourably preserved biological material (Pääbo, 1989; Herrmann and Hummel, 1994). The method of choice for extraction from preserved soft tissues utilised proteinase K and SDS, and phenol/chloroform (Pääbo *et al.*, 1988; Herrmann and Hummel, 1994). This method has been successfully used to amplify the tiny quantities of DNA that can be isolated from ancient soft and bony tissues. The first successful application using Proteinase K was carried out by Pääbo *et al.* (1988) who amplified short fragments of mitochondrial DNA from a 7 000-year-old brain in a skeleton buried in a peat bog at a site in Windover, Florida, USA. More recently, nuclear genes have been amplified from the same human source (Lawlor *et al.*, 1991).

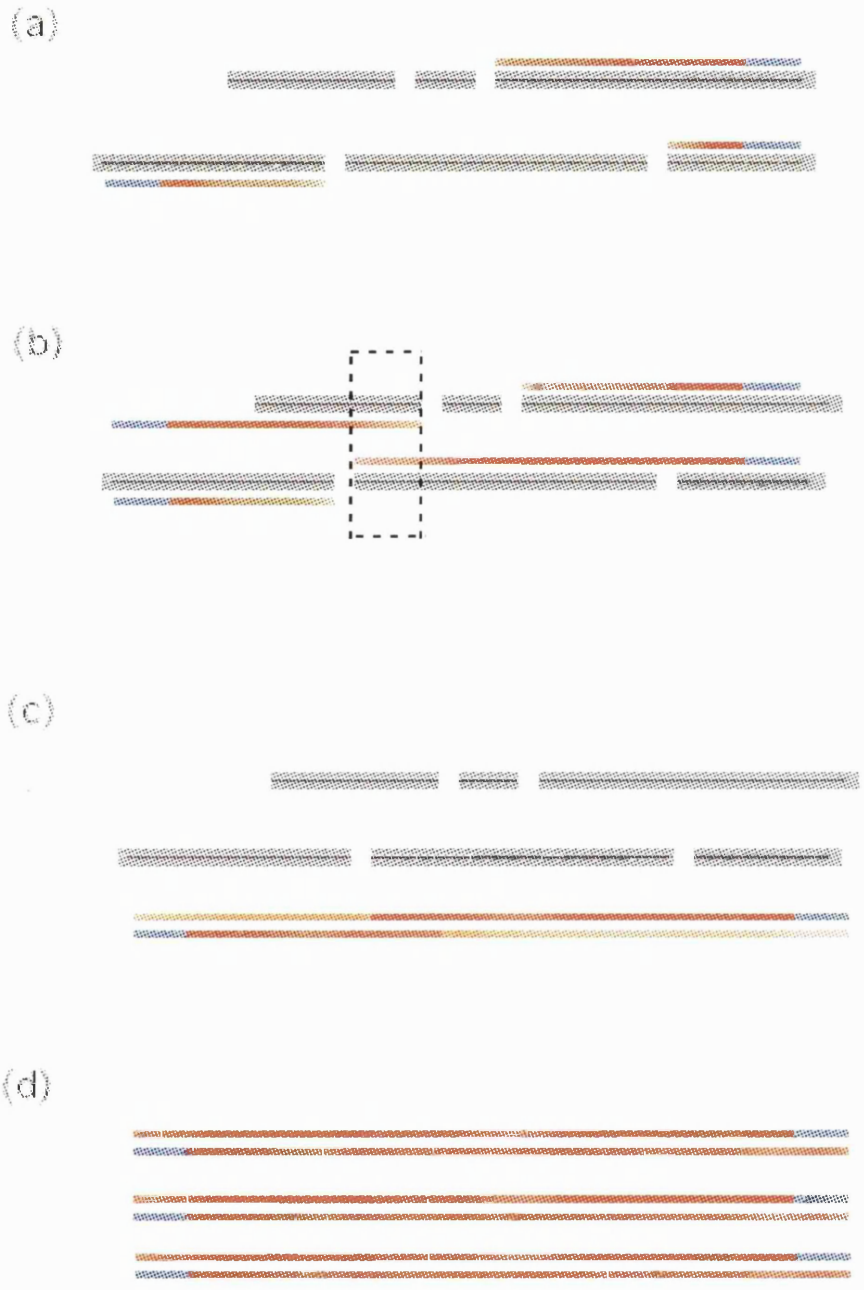


Figure 6.3 A schematic diagram illustrating the ability of the PCR process to amplify across DNA damage or breaks. (a) depicts damaged double stranded DNA molecules (grey lines), where each is broken in different places. The PCR primers (blue) bind to their target templates and during the first round of extension are extended to the break points. (b) during the next round of extension, following the denaturation step (after the strands separated, and rehybridised) the DNA strands are again extended along their new template molecules, until the next break point. (c) eventually a point is reached where sufficient overlap between the newly generated strands (dashed box) continues the reaction to completion. (d) the products are of the target length.

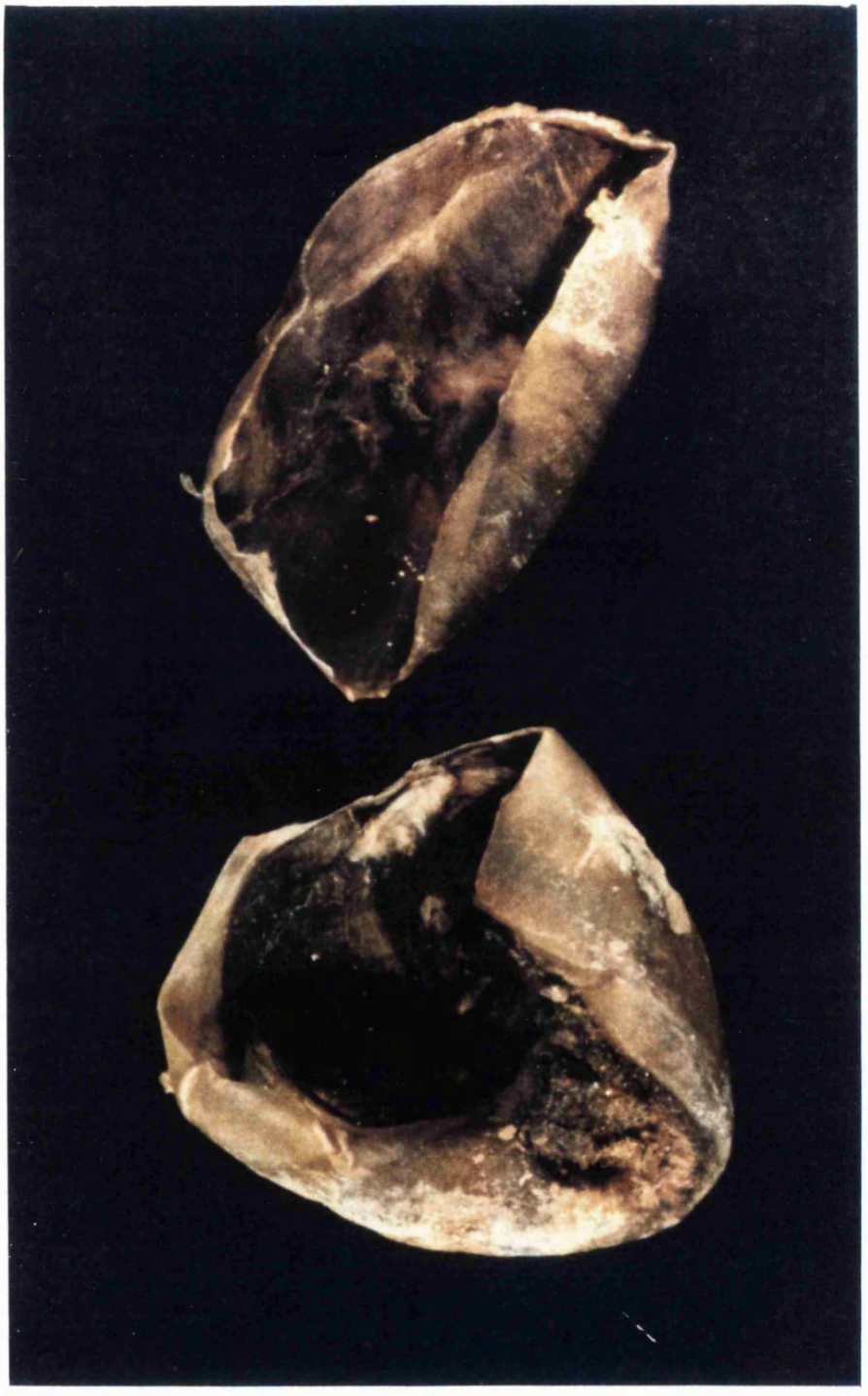


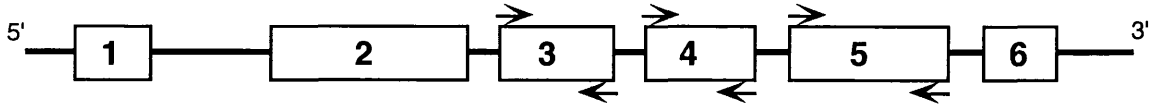
Figure 6.4 The preserved eye fragments of John Dalton, photographed in 1982. Samples were taken from the concave surfaces.

6.2.3 Nucleic acid extraction using proteinase K / SDS

For each of three independent attempts small sections of eye tissue (approximately 3 mm x 2 mm x 1 mm) were cut away, using disposable sterile scalpel blades, from the main samples stored in microcentrifuge tubes. Each piece was placed carefully in a sterile 0.5 ml microcentrifuge tube, using freshly flamed forceps. Proteinase K and SDS solutions were added to the sample as stated in section 2.2.2.

6.2.4 PCR amplification of extract

A 1 µl aliquot of the resuspended precipitate was included in PCR reactions utilising each of three pairs of non-specific exonic opsin primers (Op3+/Op3-; Op4+/Op4-; Op5+/Op5-).



These primer combinations were selected so as to generate products of small size (exon 3 primer pair - 168 bp; exon 4 primer pair - 165 bp; and exon 5 primer pair - 239 bp). It is clear from the literature that when aDNA template is of the order of thousands of years old, it seldom yields amplified products larger then a few hundred base pairs (Hagelberg, 1994). However the tissue involved in these experiments was only 150 years old and had been preserved in a favourable manner.

The annealing step of the PCR reaction was performed at 10°C below the standardised temperature at which these primer pairs were routinely used to yield product (see section 3.2). The reason for reducing the annealing temperature by this degree was to encourage primer hybridisation to any template present and generate products, whether they be from the targeted region or not. The amplification of the desired product would indicate that the DNA template had been successfully recovered from the eye tissue.

Unfortunately, no product was visualised following gel electrophoresis of PCR products on 1.8 % agarose gels (section 2.3.3, data not shown). Thus, initial attempts

to isolate DNA by a standard phenol/chloroform extraction method after proteinase K digestion proved ineffective.

6.2.5 Successful amplification of DNA

During the course of experimentation with the proteinase K/SDS method it became clear that either a modification to the phenol/chloroform procedure needed to be adopted or an alternative protocol had to be employed to overcome the problems associated not only with the excessive amounts of protein in the tissue but also the inhibitory factors which are known to be associated with aDNA (Hagelberg *et al.*, 1991; Herrmann and Hummel, 1994). At this stage a new reagent, GeneReleaser™ (BioVentures, Inc.) was brought to our attention. The manufacturers promoted it's use by a simple two step protocol that can isolate DNA from single cells (bacterial cultures or cultured cells), or tissue samples. GeneReleaser™ serves two purposes, it avoids the need to purify DNA and it sequesters products that might inhibit polymerase activity. This reagent was used with considerable success in isolating amplifiable DNA from Dalton's eye tissue.

6.2.6 DNA extraction using GeneReleaser™

Small pieces of eye tissue were cut away from the main sample, as previously described, and deposited in 20 µl of GeneReleaser™ in a 0.2 µl microcentrifuge tube. This tube was placed in a Perkin Elmer 9600 PCR machine and temperature cycled exactly as specified in the GeneReleaser™ instruction booklet:

65 °C for 30 sec

8 °C for 30 sec

97 °C for 180 sec

8 °C for 60 sec

65 °C for 180 sec

97 °C for 60 sec

65 °C for 60 sec

80 °C for 1 hour

Upon completion of the above cycling regime the reagent containing tube was either held at 80 °C whilst a 50 µl aliquot of PCR reaction mix was added to it, or stored at 4 °C until PCR was possible.

6.2.7 PCR amplification of opsin exon 4

The primer pair Op4+/Op4-, was routinely used to non-specifically amplify LW and MW opsin exon 4 (see section 3.2). This pair was selected for the initial trial because their amplified product was shorter than those amplified with the other primer sets available.

The PCR reaction mix containing opsin primer pair Op4+/Op4- was pre-heated to 94 °C, added directly to a 0.2 ml microcentrifuge tube held at 80°C, containing a fragment of Dalton's eye which had been temperature cycled with GeneReleaser. The PCR reaction mix was made up as in section 2.3.2, except that the magnesium concentration was increased to 9 mM, and 2 µl of Perfect Match (Section 2.3.2) was included.

A Perkin Elmer 9600 thermal cycler was programmed to perform a three step PCR cycling profile: 94 °C for 20 sec, 52 °C for 30 sec, 72 °C for 120 sec, for 35 cycles. Prior to cycling the tube was held at 94 °C for 3 minutes to denature the template, and post-cycling a 10 minute hold at 72 °C was performed to allow "polishing-off" of the PCR products. Upon completion of PCR, a 15 µl aliquot of the reaction was combined with 6 µl of loading buffer, and size fractionated on a 1.2 % low melting point agarose gel, in 0.5 X TAE.

1 µl of the primary PCR reaction product was included in the set up of a second round reaction which contained the following:

200 µM each of dATP, dCTP, dGTP and dTTP

0.5 units of *Taq* polymerase

9 mM $MgCl_2$

30 µM of each primer

2 µl Perfect Match (Stratagene)

1 µl Primary PCR reaction product

q.s. sterile distilled water

Total volume of 50 µl.

The PCR parameters were adjusted to optimal levels by increasing the annealing temperature to 56°C. All other parameters were as for the first round.

Following PCR cycling 15 µl of the second round PCR product were combined with 6 µl of loading buffer and electrophoresed on a 1.2 % low melting point agarose gel, as before (section 2.3.3).

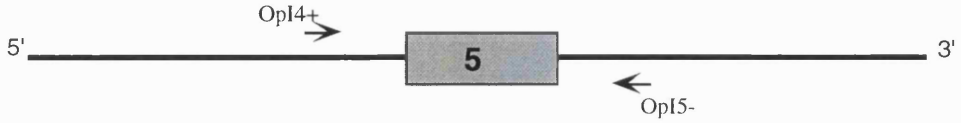
Amplified fragments were recovered from the excised gel band by eluting overnight in 50 µl sterile water (section 2.4.4.1). After TA cloning (Invitrogen) into pCRII vector (section 2.4.5), transformation and selection of positive clones (section 2.4.6), Wizard preps (section 2.4.6.1), the plasmid inserts were cycle-sequenced with *Taq* polymerase, dye-tagged dideoxy nucleotides and Sp6 sequencing primer (section 2.7.4). The products of the reaction were detected in an Applied Biosystems Model 373a DNA Sequencer System (section 2.7.6).

6.2.8 Amplification of opsin exon 3 and exon 5 sequences

Exons 3 and 5 of the X-linked opsin genes were amplified using the primer pairs and PCR parameters employed in section 3.2. Once again a second PCR amplification was necessary to generate fragments (see section 6.2.7), which were size fractionated on agarose gels.

6.2.9 Amplification using intronic opsin 5 primers

Beginning with another small fragment of Dalton's eye tissue, which had been temperature cycled in GeneReleaser™ (section 6.2.6), a primary PCR reaction was carried out utilising the primer pair OpI4+ / OpI5-. As illustrated below these primers were designed to hybridise to intronic regions (intron 4 and intron 5) approximately 100 bp, upstream and downstream, of both the LW and MW opsin exon 5. The length of the amplified fragment was predicted to be 440 bp.

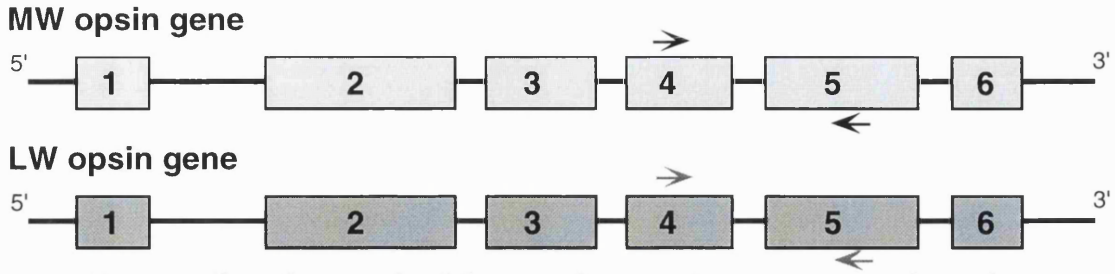


The PCR set up was the same as that given in section 6.2.7, substituting the primer pair with OpI4- / OpI5-. The parameters of the PCR reaction were as follows: Denature at 94°C for 3 minutes. 35 cycles of 94°C for 30 seconds, 68°C for 30 seconds, and 72°C for 180 seconds. A final extension of 72°C for 10 minutes was included.

The set-up and parameters for the secondary PCR reaction were the same as for the primary set-up. Amplification of the correct product was confirmed once an aliquot of the secondary PCR was size fractionated in a 1.2 % low melting point agarose gel (data not shown). This target band was excised, eluted, ligated, cloned, and sequenced as described for the previous fragments (section 6.4.2).

6.2.10 PCR Amplification Utilising Gene Specific Primers

The gene-specific set comprise a primer pair (Op4+LWint/Op5-LWint) that is specific to the LW exons 4 and 5 and the other (Op4+MWint/Op5-MWint) that is specific to the MW exons 4 and 5 (see below and section 5.2.5). These primer pairs produce a fragment of about 1.7 kb. In other combinations (Op4+LWint/Op5-MWint and Op4+MWint/Op5-LWint), they generate fragments from any hybrid gene that contains a MW exon 4 and LW exon 5 or vice versa.



Four small sections, each of 5 mm x 2 mm x 1 mm, were cut from the eye tissue and placed in individual 0.2 ml microcentrifuge tubes. GeneReleaser (20 µl) was added to each tube prior to temperature cycling (as described in section 6.2.6). First round PCR was conducted utilising the following primer combinations:

Tube a	Op4+LWint / Op5-LWint
Tube b	Op4+LWint / Op5-MWint
Tube c	Op4+MWint / Op5-MWint
Tube d	Op4+MWint / Op5-LWint

Each tube contained a 50 µl reaction mix (section 2.3.2)

The following thermal profile was employed

94°C for 180 sec denaturing step

94°C for 20 sec
62°C for 30 sec
72°C for 120 sec
} for 35 cycles

final step of 72°C for 600 sec.

Upon completion, 1 µl of the primary PCR reaction product was included in the secondary PCR reaction. The set up and cycling of the second PCR was the same as that of the first. A 15 µl aliquot of the PCR products from each tube was loaded in adjacent lanes of a 1.2 % low-melting-point agarose gel run in 0.5 X TAE and eletrophoresed.

Appropriate control DNA templates were employed. These were subjected to just one PCR amplification as per the secondary reactions stated above.

6.2.11 Southern hybridisation

The gel in which Dalton's exon 4 to exon 5 PCR products were size fractionated was trimmed and the DNA within it transferred to a nylon membrane by Southern transfer (Southern, 1975), exactly as stated in section 2.4.7.3. The membrane was stored at room temperature between 3MM paper whilst the probe was prepared.

Components from the Pharmacia Oligonucleotide labelling kit were used to generate radioactive single stranded randomly primed polynucleotides. The template was generated by PCR of human genomic (KSD) DNA with the primer pair OpI4+/OpI5-. The quality and quantity of product was determined by electrophoresis on a 3% agarose gel run in 1X TAE. 1 µl of the PCR product, estimated to contain approximately 50 ng of target DNA, was diluted with 33 µl of double distilled water in accordance with the instructions accompanying the labelling kit. The DNA was denatured at 95°C for 3 minutes, then placed immediately on ice. 1 µl of 10 X reagent buffer, supplied in the kit, 5 µl of 32P (dCTP) (Amersham - readyview), and 1 µl of DNA polymerase were added to the template before incubation at 37°C for 45 minutes. Labelled probe was separated from unincorporated components by passing through a Nap 50 column (as per section 2.4.8.4), prior to boiling at 100°C for 5 minutes. After cooling on ice for 15 minutes, 15 ml of pre-hybridisation solution was added to the probe.

The 15 ml of Church's solution containing the radioactive probe was added to a Hybaid bottle containing the filter. Hybridisation was carried out overnight at 65°C in a mini Hybaid oven, with rotation. The hybridisation solution was expelled from the Hybaid bottle. The filter was rinsed with a solution comprising 0.5 X SSC and 0.1 % SDS, heated to 65°C, then washed for 60 minutes at 65°C, with one change of solution (section 2.4.8.5), then autoradiographed. The membrane was sealed in an plastic envelope and placed in an autoradiography cassette. The film was exposed to the membrane for 12 minutes, then developed (section 2.4.8.6).

6.3 Results

6.3.1 DNA isolation utilising proteinase K / SDS

The proteinase K / SDS protocol for the isolation of DNA from non-living material, as used by Pääbo (1988), proved ineffective. All four independent attempts resulted in failure. During the phenol extraction phase a large quantity of protein precipitated at the interface, which was not surprising owing to the dehydrated nature of the tissue and the composition of the optic cup from which the tissue originated. It is possible that some of the nucleic acids present may have co-precipitated with the protein.

No DNA was visualised upon precipitation in absolute ethanol (section 2.2.2.), nor was any detected following spectrophotometric measurements of the extracted products (data not shown). It is possible that the yield of nucleic acids from tissue of this age would be so small that only PCR amplification would be able to detect it. However, even following subsequent PCR amplification of aliquots of the isolated products, no bands were seen following size fractionation on agarose gels (section 2.3.3.

6.3.2 Extraction of DNA using GeneReleaser™

The protocol included with the GeneReleaser™ reagent required that following the initial thermal cycling a subsequent PCR reaction be performed. These first round PCR amplifications did not generate products that could be visualised upon agarose gels. But, following a second round of PCR amplification using several µl of the first round product, on each occasion a faint band of the target size could be seen (data not shown). It was known that performing a second PCR using a tiny amount of the primary PCR product can aid in the generation of target bands. Inhibitory substances present in the initial PCR reaction are diluted to levels where their negative effects are overcome.

Throughout negative PCR controls (no DNA template present) were always run alongside the DNA containing template PCR samples. These would have aided in the detection of any contamination from external sources. After staining the gel in ethidium bromide (section 2.3.3) and examining it under ultra-violet light the negative control lane just shows the presence of non-extended primers.

6.3.3 Sequence of exon 3, 4, and 5 clones

Of the 12 exon 4 clones sequenced, only four included the correct opsin exon 4 product. Comparison of these four sequences to previously published human LW and MW opsin gene sequences (Nathans 1986a) suggested that they were all derived from the LW opsin gene alone (Figure 6.5). The sequence differences between the MW and LW genes at nucleotides 730, 738 to 740 and 747 in exon 4, indicate that this gene codes for an LW visual pigment. No similarity to the published MW sequence from this region was obtained (Figure 6.6).

Due to the small number of clones obtained and sequenced, it was not possible to state with absolute certainty that the MW gene was absent from the eye tissue. To demonstrate that the MW sequence was indeed absent one would either have to sequence more clones from this region to a statistically significant level, or generate sequence data from other regions which also differentiate between the two opsin genes.

Thus, exon 3 and exon 5 were also successfully amplified. Amplification required but a single attempt from preserved eye material with exon 3 primers, while three attempts were needed to clone and sequence exon 5 regions. 15 clones corresponding to opsin 5 five were obtained. With each of these clones all 11 nucleotides that differentiate the LW exon 5 from the MW (see Figure 6.5b) were identical to the LW sequence. Five opsin exon 3 clones were also obtained, all of which showed homology to the second most common variant of the LW sequence (Winderickx *et al.*, 1993; Figure 6.5c), which differs from the most common by possessing a G at nucleotide position 538, and results in the coding of an alanine

A Exon 4

```
MW -     GGC CCA GAC GTG TTC AGC GGC AGC TCG TAC CCC GGG GTG CAG TCT TAC ATG ATT GTC CTC ATG GTC ACC - 719
LW -     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Dalton - ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

MW -     TGC TGC ATC ACC CCA CTC AGC ATC ATC GTG CTC TGC TAC CTC - 761
LW -     ... ... ... .T. ... ... GCT ... ... A.. ... ... ... ...
Dalton - ... ... ... .T. ... ... GCT ... ... A.. ... ... ... ...
```

B Exon 5

```
MW -     AAG GAA GTG ACG CGC ATG GTG GTG GTG ATG GTC CTG GCA TTC TGC TTC TGC TGG GGA CCA TAC GCC TTC - 899
LW -     ... ... ... ... ... ... ... ... ... ... ... A.. T.T ..G .A. ... G.. ... ... ... ..C ... A.. ...
Dalton - ... ... ... ... ... ... ... ... ... ... ... A.. T.T ..G .A. ... G.. ... ... ... ..C ... A.. ...

MW -     TTC GCA TGC TTT GCT GCT GCC AAC CCT GGC TAC CCC TTC CAC CCT TTG ATG GCT GCC CTG CCG GCC TTC - 965
LW -     ... ... ... ... ... ... ... ... ... ..T ... G.. ... ... ... ... ... ... ... ... ... ... .A.
Dalton - ... ... ... ... ... ... ... ... ... ..T ... G.. ... ... ... ... ... ... ... ... ... ... .A.

MW -     TTT GCC AAA AGT - 980
LW -     ... ... ... ...
Dalton - ... ... ... ...
```

C Exon 3

```
LWa -    ATT TCC TGG GAG AGG TGG CTG GTG TGC AAG CCC TTT GGC AAT GTG AGA TTT GAT GCC AAG CTG GCC ATC - 551
LWb -    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Dalton - ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

LWa -    GTG GGC ATT GCC TTC TCC TGG ATC TGG TCT - 581
LWb -    ... ... ... ... ... ... ... ... ... G..
Dalton - ... ... ... ... ... ... ... ... ... G..
```

Figure 6.5 Sequence of amplified fragments from (**A**) exon 4, (**B**) exon 5, and (**C**) exon 3 of Dalton's DNA compared to those of the MW and LW genes originally reported by Nathans *et al.* (1986a). LWa and LWb are, respectively, the first and second most common varients of the LW exon 3 reported by Winderickx *et al.*, (1993).

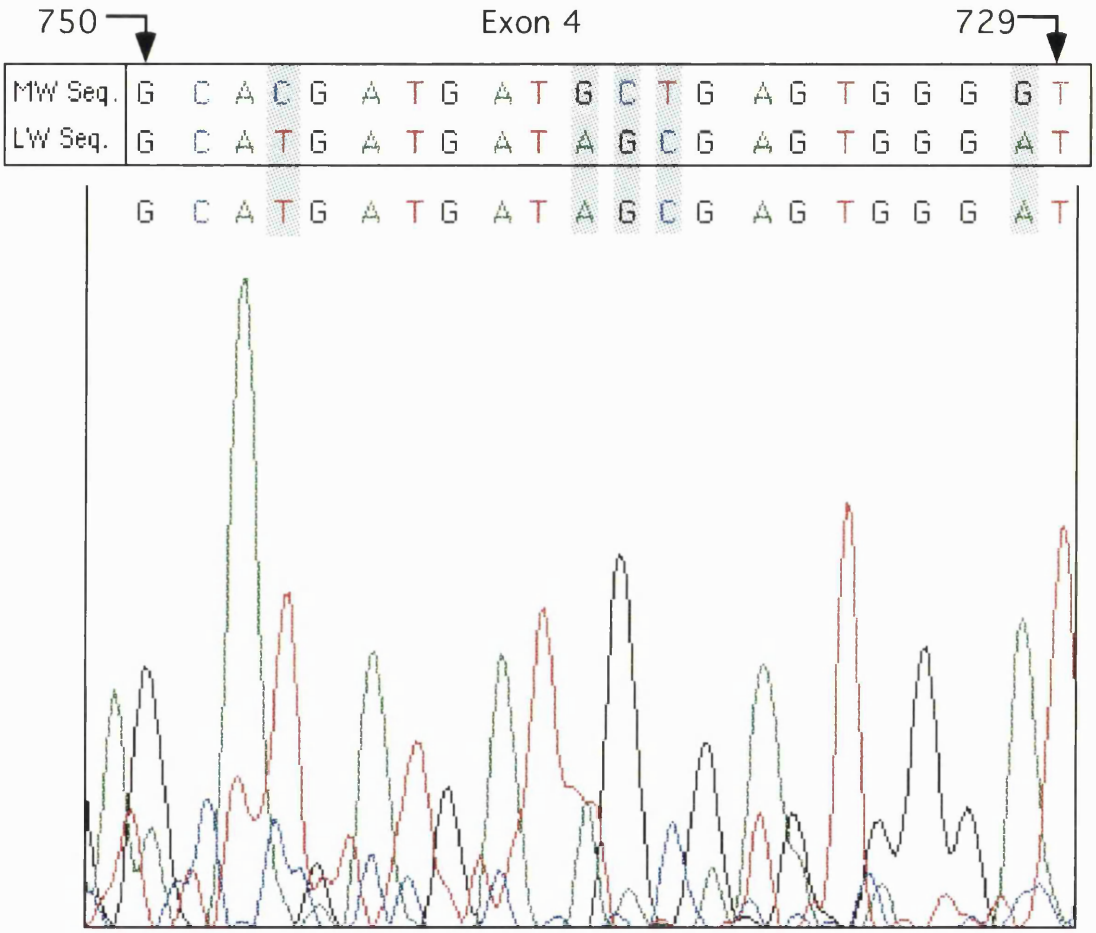


Figure 6.6 An electropherogram showing the sequence data from an informative region of exon 4. A comparison of this sequence with the corresponding regions of the human LW and MW (Nathans *et al.*, 1986a) opsin genes (top) clearly identifies this region as the LW gene. Note that the sequences are of the complementary strand.

instead of serine. Table 11 summarises the data for the number of clones analysed, and confirms that all were from the LW gene(s).

The presence of the LW opsin sequences from exons 3, 4, and 5 strongly indicated that only the LW opsin gene was present in Dalton's DNA. However, this fact alone could not rule out the remote possibility that the MW opsin gene was present but was for some unknown reason not being cloned. So it was decided that further PCR amplifications were required to support the initial findings and to reduce the possibility that contamination was responsible for the generation of these opsin products.

TABLE 11. Number of each class of cone opsin clone sequenced

Amplified region	Number of clones analysed	MW	LW
Exon 3	5	0	5
Exon 4	5	0	5
Exon 5	15	0	15
Intron 4, exon 5, intron 5	2	0	2

6.3.4 Amplification from intronic primers

Prior to experimentation on Dalton's eye material, work had been conducted on a plasmid generated from an LW cDNA opsin clone (pUC1801; a kind gift from J. Nathans). Naturally there was a possibility that airborne contamination of the PCR reactions had occurred with this template. (This possibility was considered to be very slight since PCR amplification of negative water controls were always included. These too would have been subject to contamination by the plasmid). The most unequivocal way to discount this contamination problem was to demonstrate the existence of intronic sequences from Dalton's DNA, which by definition would be absent from any contaminating plasmid. This was done by using the primer pair OpI4+/OpI5-.

Examination of the sequence data from two clones confirmed that: a) the correct intron / exon boundaries were present at both ends, and b) the coding sequence was a LW gene (Figure 6.7a).

6.3.5 Exon 4 to exon 5 PCRs

In an effort to further demonstrate that Dalton's DNA only possessed the gene for the LW opsin protein, gene specific primers pairs, originally used in the analysis of DNA from colour anomalous trichromats (see chapter 5.2.4), were employed. These four primers when employed in particular combinations allow the amplification of regions which extend from the centre of exon 4 across intron 4 to exon 5. Each primer will primer to either the LW or the MW sequence.

Visualisation of the gel after ethidium bromide staining revealed the presence of a faint target band in just one of the lanes, that corresponding to the primer pair Op4+LWint / Op5-LWint (Figure 6.8a). Other bands were also generated at differing positions but none of these were of the desired size.

6.3.6 DNA controls

Appropriate positive controls had to be selected to confirm the specificity of these primer pairs, in addition to the negative DNA control. Experimentation with

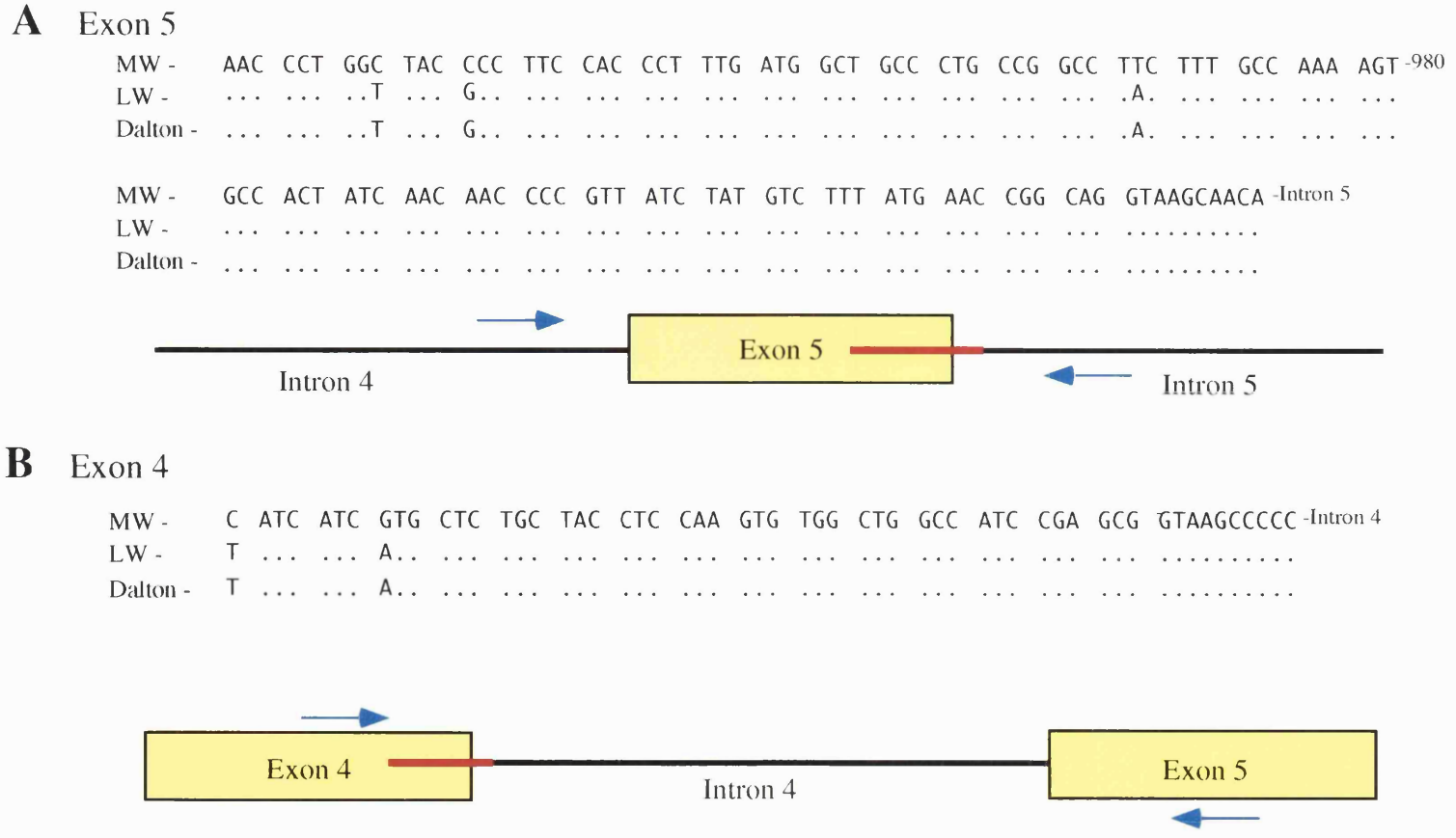


Figure 6.7 Sequence of amplified fragments from (**A**) exon 5, generated with the primer pair I4+/I5- and (**B**) exon 4, using primer pair LW4+/LW5-. Diagrammatic representations are made of the regions sequenced. Blue arrows are primers and regions in red correspond to the sequence displayed. DNA compared to those of the MWand LW genes originally reported by Nathans *et al.* (1986a).

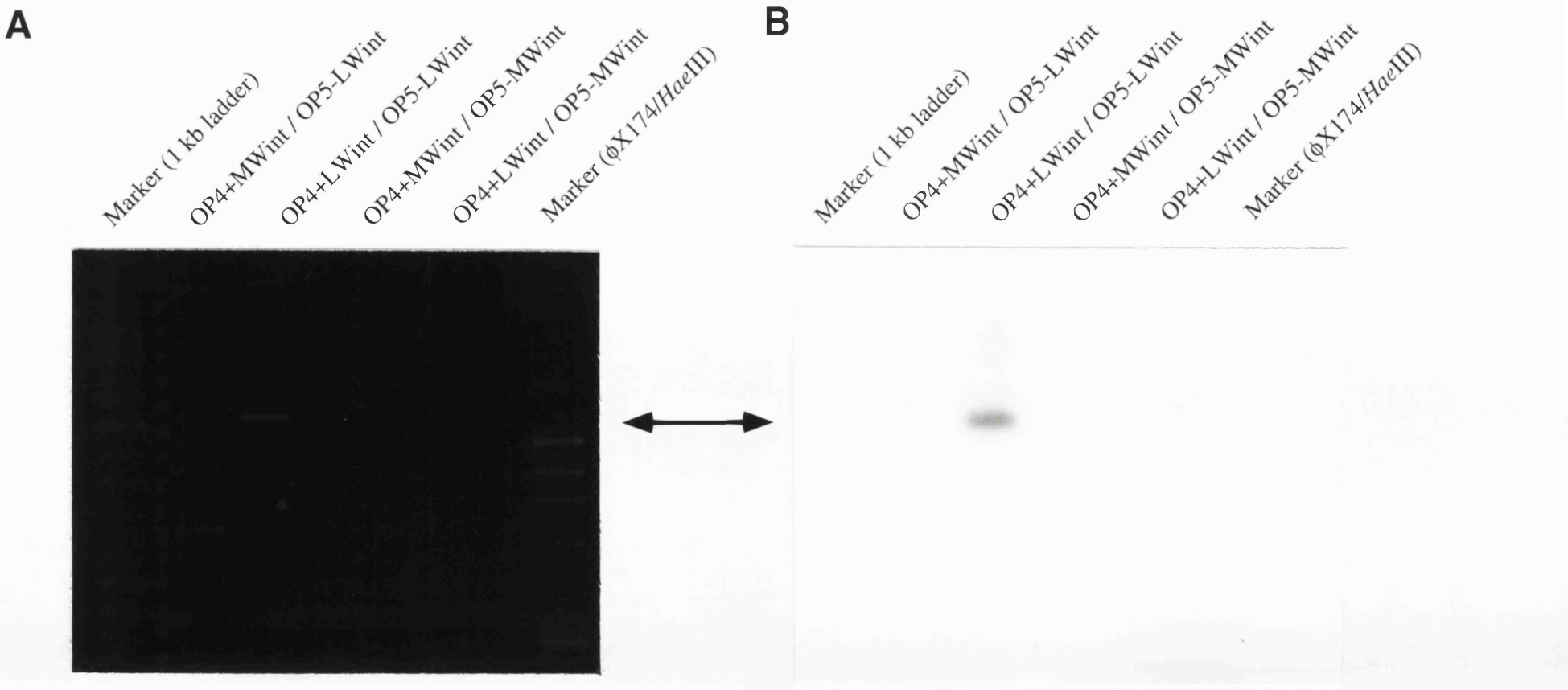


Figure 6.8 **A**. PCR products obtained from Dalton's DNA, using the primer pairs indicated. The expected product size of 1.7 kb is indicated by the arrow. **B**. Southern blot of the gel in panel A.

control DNA's was carried out only when work on Dalton's DNA was completed. Samples of blood from three human males was obtained and processed as outlined in section 2.2.1, to prepare gDNA. The first of these subjects (KSD) was known to be a colour normal observer whose DNA had previously been subject to considerable analysis. The second subject had anomalous trichromacy, while the third was a confirmed colour defective deuteranope.

The gDNA's from each of these subjects were amplified through one round of PCR, with the same four primer sets as those employed for Dalton's samples. Figure 6.9 shows that as expected, the DNA from the normal observer yielded fragments of the correct size only when the Op4+LWint/Op5-LWint and Op4+MWint/Op5-MWint combinations were used. The known anomalous trichromat produced in addition a fragment with the Op4+LWint/Op5-MWint combination, indicating the presence of a hybrid gene; but the Op4+LWint/Op5-LWint primer pair alone amplified a fragment corresponding to the LW opsin from the DNA of a known deuteranope.

The identity of the fragment amplified from Dalton's DNA, by the Op4+LWint/Op5-LWint primer pair, was then confirmed by cloning and sequencing (Figure 6.7). Partial sequences from 10 separate clones show that the fragment was amplified from an opsin gene that coded for LW visual pigment and included intron 4 sequence.

6.3.7 Conformation of the identity of the PCR product generated using primer pair Op4+LWint/Op5-LWint

Was the 1.7 kb band of the correct origin? Were there present undetected amplified products in other lanes, which would contradict the data obtained so far? These questions were resolved by a simple hybridisation experiment (section 6.2.11). Probing with the human opsin exon 5 probe (section 6.2.11) only "lit-up" the target band in the lane containing the product of the primer pair Op4+LWint/Op5-LWint. A photograph of the resultant film is shown in Figure 6.8b.

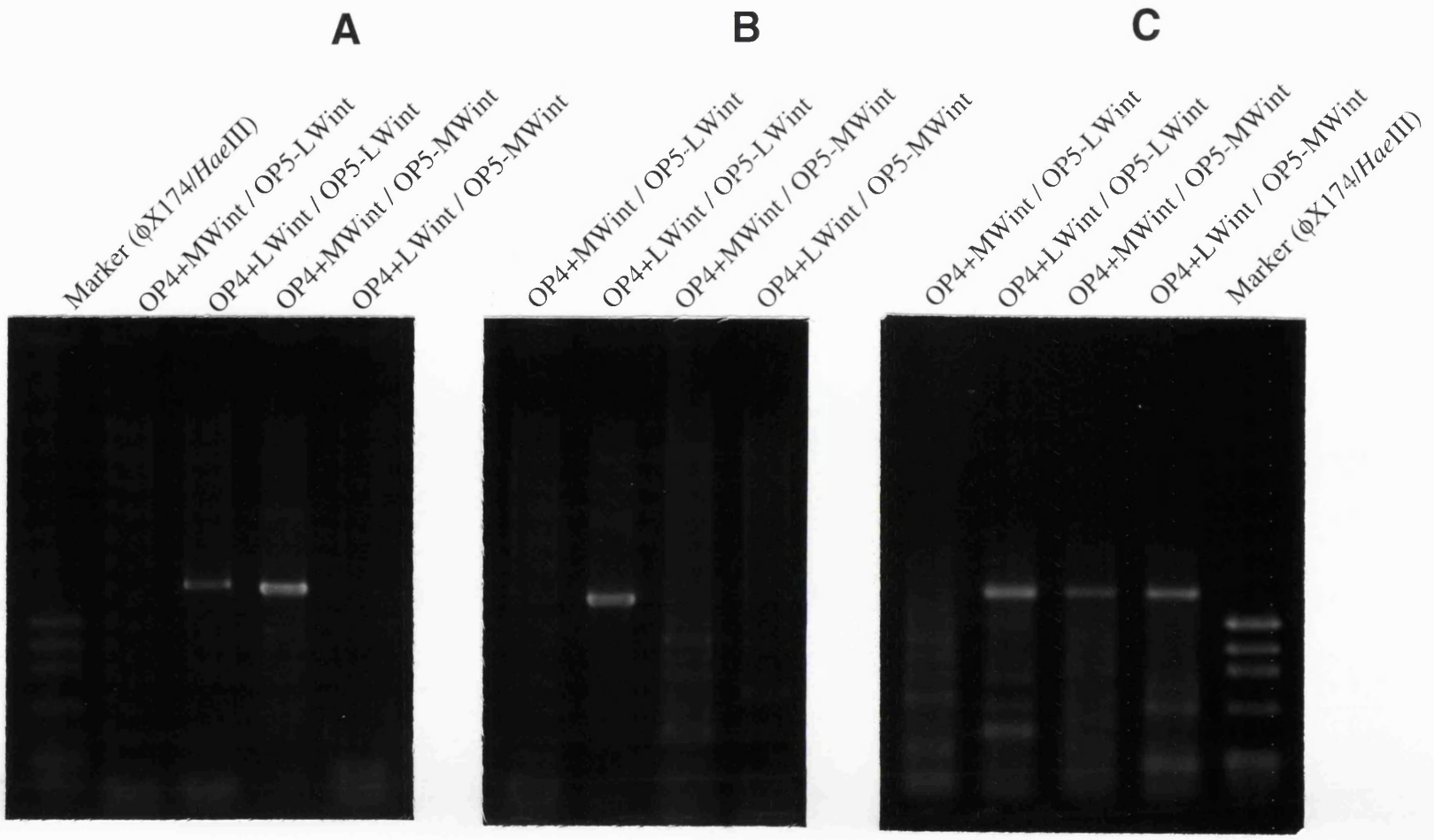


Figure 6.9 Amplification products obtained following the use of the same primers as Figure 6.7. **A**. Normal human male. **B**. Male deuteranope. **C**. Anomalous trichromat.

6.3.8 Contamination

Analyses of old DNA by PCR are bedevilled by contamination. Due in part to the very nature of the polymerase chain reaction and the very small quantities of recoverable nucleic acids from the tissue samples themselves it was of paramount importance to prevent the possibility of exogenous DNA contaminating any stage of the sampling, extraction, or amplification procedures. Contamination with only a few molecules of foreign DNA can lead to a spurious result with all amplified products derived from the contaminating source (Hagelberg *et al.*, 1991; Hagelberg, 1994).

Considerable precautions were adopted to minimise the risk of contamination. Great care was exercised when handling the tissue samples. Disposable gloves and sterile implements were used at all stages. Sampling from the preserved eyes was conducted from what were deduced to be the inner face of the eye cups, where the retinal surface would have resided, and where contamination, from handling, would be minimal. One cannot rule out the possibility of handling contamination during the initial medical examination of the eyes by Dalton's physician, Ransome, nor during subsequent stages of storage, for instance by museum staff. Once in the laboratory the samples were always handled in a room which had never previously been employed as a laboratory. New equipment was sequestered for exclusive use with these experiments, including positive displacement pipettes. Solutions were prepared from newly purchased reagents.

A further argument against contamination can be brought however, namely that only contamination by the DNA from another deuteranope would have given the same result. As far as one was aware, at the time of the initial PCR work, no one in the laboratory had any defect in colour vision and certainly none were dichromats, and no samples of DNA from other dichromats were present in the laboratory until after the analysis of Dalton's DNA was complete. The deuteranope that was used for test purposes was not known until after all the PCR amplifications on Dalton's tissue had been conducted.

6.4 Discussion

The consistent amplification of just the LW opsin gene from the remains of the eye, and the lack of amplification of the MW opsin gene confirms that Dalton was colour blind, but contradicts the previous belief that he was a protanope.

One each occasion a successful PCR amplification was performed, with non-specific primer pairs, sequencing of the target fragments always produced sequence data which matched one of the published variants of the LW opsin gene of human (Nathans *et al.*, 1986a; Nathans *et al.*, 1986b). The results fit the genotype of a deuteranope.

Dalton's observation that his brother also observed colours as he did, and that his sister was quite normal (Wright, 1967) suggests that their mother most probably was a carrier for the defect. One can not rule out the possibility that she was an affected homozygote, in which case she would have been one of the few (0.5%) of Caucasian women known to be colour blind, and certainly Dalton, with his keen eye for scientific detail, would have acknowledged this fact.

6.4.1 Assessment of Dalton's phenotype

From the data it is possible to confirm that Dalton was indeed a dichromat but, contrary to previous interpretations (Young, 1807; Helmholtz, 1896; Abney, 1913; Wright, 1967), he was a deuteranope with a single LW opsin gene that coded for a visual pigment with the shorter spectral sensitivity of the two common LW variants.

Protanopes and deuteranopes can be distinguished by their spectral sensitivity and by their colour matches. The classical sign of protanopia is the foreshortening of the red end of the spectrum, owing to the absence of the LW receptor. Subjects are interrogated and the answers they provide allows phenotypes to be assigned. Fortunately a number of scientists were able to question Dalton directly concerning his view of the spectrum. Amongst them were the physicists Sir John Herschel and Sir David Brewster, who both reported that he did not see the spectrum as foreshortened at long wavelengths. Herschel in a letter dated May 20, 1833, writes directly to Dalton:

"It is clear to me that you and all others so affected perceive as light every ray, which others do" (Henry, 1854). And Brewster in his "Letters on Natural Magic" writes: "In all those cases which have been carefully studied... making personal observations, namely,... Mr. Dalton,... the eye is capable of seeing the whole of the prismatic spectrum, the red space appearing to be yellow", (Brewster, 1842). Similarly, Richard Taylor, the scientific publisher, remarks in a footnote that "Dr Dalton has never stated that the spectrum he saw was shorter than the spectrum seen by others" (Wartmann, 1846).

How does one reconcile Dalton's comment that red appears as "little more than a shade, or defect of light" (Dalton, 1798)? For deuteranopes, the red part of a spectrum will appear dim since the regions that look yellow, orange and red to normal observers are for them all of the same hue, but the red is of reduced luminosity than the yellow and orange regions to which it is juxtaposed. Moreover, the red region does not offer to the deuteranope the Farbenglut, the extra vividness of saturated colours, which derives from the LW/MW opponent signal and which accentuates the brightness of long wavelengths in the case of the normal observer (Kohlrausch, 1923).

In summary, contemporary reports of Dalton's spectral sensitivity favour a diagnosis of deuteranopia, and his colour matches are consistent with deuteranopia. Thus the molecular finding, that Dalton retains the LW and lacks the MW gene, is fully compatible with the historical evidence for his phenotype. The work presented here has been published (Dulai *et al.*, 1995; Hunt *et al.*, 1995).

6.4.2 A modern deuteranopes view of a *Pelargonium zonale* leaf

Dalton's detailed account of his observations of the leaves of the plant *Pelargonium zonale* have been instrumental in interpreting his phenotype. If we assume that this was the very same species which we can cultivate today, then it would be possible to show this plant to an modern deuteranope and analyse his/her response. Following the completion of the PCR experimentation on Dalton's eye tissue, a male deuteranope did happen to attend the laboratory and volunteered to donate some of his

blood (for the trans-exonic PCR experiments; see Figure 6.9), and to participate in non-invasive colour vision studies. He was shown a specimen of *Pelargonium zonale* (see Figure 6.2). In Northern daylight, he described the colour as similar to that of the sky, but slightly desaturated. In candlelight a few minutes later, he described the colour as red! Just as Dalton had done 200 years earlier.

6.4.3 Is recovered DNA of any interest?

aDNA is defined as any DNA from a dead organism or parts thereof, as well as extracorporeally encountered DNA from a living organism. Generally the autolytic processes that commence soon after cell death result in the degradation of nucleic acids into small fragments, too damaged to be of any benefit. However, occasionally the environment surrounding the organic remains is conducive to preserving the tissue, and the nucleic acids therein. The advent of ancient DNA (aDNA) technology enabled the genetic basis of John Dalton's colour blindness to be deduced, and may aid in resolving additional historical riddles.

CHAPTER 7

Conclusions

This research were undertaken to further our knowledge of the molecular biology of the LW and MW opsin genes of higher primates, and cast light on the mechanisms which underlie the regulation of these genes.

The specific objectives of the project have been mostly achieved;

- The OW primate X-chromosome linked opsin genes have been sequenced across the functionally important regions from exon 3 to 5. The same amino acids that determine the majority of the spectral differences between the human MW and LW opsins (sites 180, 277, and 285) are seen in all the species examined (Dulai *et al.*, 1994). It is possible to predict that the colour vision of these non-human primates must resemble the human situation. The different life-styles and feeding habits of the primates have not impacted on their colour perceptions. Multiple opsin sequences have been demonstrated in chimpanzee and talapoin, indicating that multiple opsin genes do exist in the opsin arrays from Cercopithecoid and Hominoid primates. Opsin sequence comparisons employing phylogenetic trees supports the view that the LW and MW genes arose from a duplication event in the OW primates following their separation from the NWM.
- The proximal promoter regions of the X-linked opsin genes of OW and NW primates have been sequenced. Comparisons show that there is very little variation in this region, suggesting that the mechanisms which discriminate between the expression of LW or MW opsin genes must lie elsewhere. These 5' flanking regions do show the presence of *cis*-acting elements expected from genes which are regulated in a temporal and spatial manner. The next step in the analysis of these upstream regulatory regions would be to perform gel retardation (mobility shift) and foot-printing assays. Proteins employed in these assays could be isolated from

several lines of retinoblastoma cells (Y79 & Weri-RBI) which have been characterised and show the expression of the relevant opsin genes. These experiments are already underway and the results are promising in that they show proteins retarding fragments upstream of human SW opsin.

- The presence and expression of hybrid genes has been implicated as the cause of colour vision anomalies for two types of colour phenotype. In order to determine if the type c hybrid gene of subject YT was being expressed he should be examined further by means of pschophysical testing. Conceivably one might do this by selective adaptation, e.g. he might make a deuteranomalous match following a far red bleach. The existance of a hybrid gene of this type may provide the explainination for a number of anomalies.
- Following successful extraction and amplification of John Dalton's DNA, his genotype has been revealed. This has lead to the revision of his phenotype to deuteranopia (Dulai *et al.*, 1995; Hunt *et al.*, 1995; Mollon *et al.*, 1996).

The two greatest hurdles encountered in the examinations of the X-linked opsin genes were both technical. First, the inability to identify the number and order of the opsin genes along an array, due in main, to the lack of an appropriate cloning vector, caused many frustrations. These arrays may reach lengths of 400 kb. It is essential for a complete analysis that their length and order be properly determined, since this will provide essential knowledge on the effect that gene order has on expression, as well as allowing the role played by various hybrid genes to be determined. Second, more reliable methodology needs to be developed to allow flanking DNA regions to be sequenced. The methods employed in this work, although functional, are still in their infancy. The Vectorette protocol has now been substantially modified by the MRC Human Genetics Unit, Edinburgh. The new protocol, known as Splinkerette, overcomes many of the drawbacks of the original Vectorette method.

Great potential exists in the field of aDNA. The work presented here is probably the first recorded instance where a hereditary defect has been identified in an historical figure. As testified by the substantial interest generated in the field of aDNA the ability to extract genetic information from suitably preserved material is not novel (Pääbo, 1993) however, this case is thus unique in that it has extensive documentation to support a previously recorded phenotype. Other historical figures of the past have often left fragments of themselves behind (and living subjects may purposefully archive pieces of themselves), provided the conditions specified below are met, it should become possible to answer unsolved questions about the genetic basis of illnesses and/or mental aberrations, as more is understood of the genetic basis of human variation and disease,

The following criteria must be met before an investigation in molecular biography is conducted: (a) suitably preserved material which yields uncontaminated DNA must exist, (b) the source of the material must be correctly documented, (c) ethical constraints must be adhered too (permission must be granted by responsible authorities to conduct the work), (d)phenotypic evidence must exist that suggests a specific genetic disorder or anomaly in the historical subject, and, perhaps most importantly, (v) the sequence must be known for the gene or genes that determine the condition in question. In the case presented, the conditions were very favourable: the eye tissue was in a dry state that facilitated survival of DNA; there were written references to the removal of the tissue and its subsequent ownership; Dalton explicitly left his eyes for the study of daltonism; he also provided a detailed account of his phenotype; and the psychophysical and molecular bases of dichromacy are rather well understood. Only a few other historical figures will provide so attractive a subject for molecular biography.

Abbreviations

ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
℃	degrees centigrade
cDNA	complementary deoxyribonucleic acid
CIA	chloroform / iso-amylalcohol
dATP	deoxy-adenosine-5'-triphosphate
dCTP	deoxy-cytidine-5'-triphosphate
dd	double distilled
ddATP	dideoxy-adenosine-5'-triphosphate
ddCTP	dideoxy-cytidine-5'-triphosphate
ddGTP	dideoxy-guanosine-5'-triphosphate
ddH2O	double distilled water
ddTTP	dideoxy-thymidine-5'-triphosphate
dGTP	deoxy-guanosine-5'-triphosphate
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxy-thymidine-5'-triphosphate
EDTA	ethylene-diamine-terta acetic acid
g	gram(s)
IPTG	isopropyl-B-D-thiogalactoside
Kb	kilobase(s)
LW	long-wavelength-sensitive
LW/MW	long-wavelength / middle-wavelength sensitive
M	molar
MCC	microcentifuge
MW	middle-wavelength-sensitive
ug	microgram
ml	millilitre
µl	microlitre
MSP	microspectrophotometry
MW	middle-wavelength-sensitive
ng	nanogram

nm	nanometre
OD	optical density
PCR	polymerase chain reaction
PEG	polyetheleneglycol
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDW	sterile distilled water
SSC	sodium chloride / sodium citrate buffer
ssDNA	single-stranded DNA
SW	short-wavelength-sensitive
TAE	Tris / acetate / EDTA buffer
TBE	Tris / borate / EDTA buffer
TEMED	N,N',N',N'-tetramethlethylenediamine
X-Gal	5-bromo-4-chloro-3-indolyl-B-D-galactoside

Appendices

Appendix A - Buffers

Church's solution (Church, 1984)

0.5 M	sodium phosphate buffer (pH7.2)
7 %	SDS
1 mM	EDTA

Denhardts solution (Denhardt, 1966)

For 500 ml of 50X:

5 g	ficoll
5 g	polyvinylpyrrolidone
5 g	bovine serum albumin

made up to 500 ml in SDW.

Loading buffer A

For 39 ml 10X:

12 ml	glycerol
0.06 g	bromophenol blue
0.06 g	xylene cyanol FF
6 ml	50X TAE

made up to 30 ml in SDW.

Loading buffer B

For 30 ml 10X:	7.6 g	ficoll (Mr 400 kD)
	0.75 g	Orange G, 0.25M

NE buffer

150 mM	sodium chloride
25 mM	EDTA

Pre-hybridisation solution A

For 25 ml:	5 ml	30% PEG
	7.5 ml	20X SSC
	1.25 ml	100X Denhardts solution
	1.25 ml	10% SDS
	made up to 25 ml in SDW.	

200 µl of heat denatured (5 minutes in a boiling water bath, 5 minutes on ice) sonicated Herring testis DNA (5 mg/ml) was added immediately before use.

Southern blot solutions

Denaturing Solution:	43.83 g	sodium chloride (1.5M)
	10 g	sodium hydroxide (0.5M)
	made up to 500 ml in water.	

Neutralising solution:	43.83 g	sodium chloride (1.5M)
	30.28 g	Tris pH 7.2 (0.5M)
	186 g	EDTA (0.001M)
	made up to 500 ml in water.	

SSC

For 1 litre:	175.3 g	sodium chloride
	88.2 g	sodium tri-chloride
	made up to 1 litre in SDW.	

TAE buffer

For 500 ml of 50X:	121 g	Tris (0.04M)
	28.6 ml	glacial acetic acid
	9.3 g	EDTA pH 8.0 (0.001M)
	made up to 500 ml in distilled water.	

TBE buffer

For 1 litre of 10X:	108 g	Tris (0.045M)
	55 g	boric acid
	9.5 g	EDTA (0.001M)
	made up to 1 litre in distilled water.	

TE buffer

For 1 litre of 10X:	12.11 g	Tris (0.01M)
	3.7 g	EDTA (0.001M)
	made up to 1 litre in distilled water.	

Appendix B - MEDIA

Antibiotics

Ampicillin

A 50 mg/ml stock solution of Ampicillin was made in distilled water, filter sterilised through a 0.22-micron filter and stored at -20°C.

Kanamycin

A 10 mg/ml stock solution of Kanamycin was made in distilled water, filter sterilised through a 0.22-micron filter and stored at -20°C.

Tetracycline

A 5 mg/ml stock solution of tetracycline was made in 95% ethanol and stored at -20°C.

Laurita Bautaria (LB)

For 1 litre of broth:	10 g	tryptone
	5 g	yeast extract
	10 g	sodium chloride
	1 ml	1M sodium hydroxide

made up to 1 litre with distilled water. Sterilised by autoclaving at 15 psi for 20 minutes.

For 1 litre of agar for plates:	15 g	Agar added.
For 1 litre of top agar:	7.5 g	Agar added.

LBA

For LBA, Ampicillin (final concentration 100 ug/ml) was added to cooled (approximately 50°C), autoclaved LB.

Appendix C: Primate Common Name Index

Taxonomy Of The Primate Order--Common Name Index

This index matches the various common names for different primates to their taxonomic (genus and species) name.

Tree shrews continue to be included in this taxonomy.

Data retrieved from Internet WWW site http://uakari.primate.wisc.edu:70/1/pin/taxon

Angwatibo	Arctocebus calabarensis
Ape, Barbary	Macaca sylvanus
Aye-aye	Daubentonia madagascariensis
Baboon, chacma	Papio ursinus
Baboon, gelada	Theropithecus gelada
Baboon, Guinea	Papio papio
Baboon, hamadryas	Papio hamadryas
Baboon, olive	Papio anubis
Baboon, sacred	Papio hamadryas
Baboon, yellow	Papio cynocephalus
Barrigudo	Lagothrix lagotricha
Beeloh	Hylobates klossi
Bonobo	Pan paniscus
Bushbaby, Allen's	Galago alleni
Bushbaby, eastern needle-nailed	Euoticus inustus
Bushbaby, fat-tailed	Otolemur crassicaudatus
Bushbaby, greater	Otolemur crassicaudatus
Bushbaby, lesser	Galago senegalensis
Bushbaby, thick-tailed	Otolemur crassicaudatus
Bushbaby, western needle-nailed	Euoticus elegantulus
Caparro	Lagothrix lagotrichia
Capuchin, black-capped	Cebus apella
Capuchin, brown	Cebus apella
Capuchin, brown pale-fronted	Cebus albifrons
Capuchin, robust tufted	Cebus apella robustus
Capuchin, tufted	Cebus apella
Capuchin, weeping	Cebus nigrivittatus
Capuchin, white-faced	Cebus capucinus
Capuchin, white-throated	Cebus capucinus
Capuchin, yellow breasted	Cebus apella xanthrosternus
Chimpanzee, common	Pan troglodytes
Chimpanzee, pygmy	Pan paniscus
Cobego	Cynocephalus variegatus
Colobus, Angolan	Colobus angolensis
Colobus, bay	Colobus badius
Colobus, black	Colobus satanas
Colobus, black and white	Colobus abyssinicus
Colobus, east central African red	Colobus rufomitratus
Colobus, king	Colobus polykomos
Colobus, Kirk's	Colobus kirkii
Colobus, olive	Colobus verus
Colobus, red	Colobus badius
Colobus, west central African red	Colobus pennanti
Colobus, western black and white	Colobus polykomos
Colobus, western red	Colobus badius badius
Colobus, Zanzibar red	Colobus kirkii
Cuxius	Chiropotes satanas
Douc	Pygathrix nemaeus
Douroucouli	Aotus trivirgatus
Drill	Mandrillus leucophaeus
Dwarf lemur, fat-tailed	Cheirogaleus medius
Dwarf lemur, fork-marked	Phaner furcifer
Dwarf lemur, greater	Cheirogaleus major
Dwarf lemur, hairy-eared	Allocebus trichotis
Galago, Demidoff's	Galagoides demidovii
Galago, Demidoff's dwarf	Galagoides demidovii
Galago, dwarf	Galagoides demidovii
Galago, Garnett's	Otolemur garnetti
Galago, Thomas'	Galago thomasi
Galago, Zanzibar	Galago zanzibaricus
Gelada	Theropithecus gelada
Gentle lemur, broad-nosed	Hapalemur simus
Gentle lemur, grey	Hapalemur griseus
Gibbon, agile	Hylobates agilis
Gibbon, black	Hylobates concolor
Gibbon, Bornean	Hylobates muelleri
Gibbon, capped	Hylobates pileatus
Gibbon, crested	Hylobates concolor gabriellae
Gibbon, dark-handed	Hylobates agilis
Gibbon, hoolock	Hylobates hoolock
Gibbon, Kloss's	Hylobates klossi
Gibbon, lar	Hylobates lar
Gibbon, Mentawai	Hylobates klossi
Gibbon, pileated	Hylobates pileatus
Gibbon, silvery	Hylobates moloch
Gibbon, white-browed	Hylobates hoolock
Gibbon, white-cheeked	Hylobates concolor leucogenys
Gibbon, white-handed	Hylobates lar
Gorilla, eastern lowland	Gorilla gorilla graueri
Gorilla, lowland	Gorilla gorilla gorilla
Gorilla, mountain	Gorilla gorilla beringei
Grivet	Cercopithecus aethiops sabaeus
Guariba	Alouatta seniculus
Guenon, crowned	Cercopithecus pogonias
Guenon, de Brazza's	Cercopithecus neglectus
Guenon, Dent's	Cercopithecus denti
Guenon, diademed	Cercopithecus mitis
Guenon, dryas	Cercopithecus dryas
Guenon, greater spot-nosed	Cercopithecus nictitans
Guenon, L'Hoest's	Cercopithecus lhoesti
Guenon, lesser spot-nosed	Cercopithecus petaurista
Guenon, lesser white-nosed	Cercopithecus petaurista
Guenon, owl-faced	Cercopithecus hamlyni
Guenon, Preuss's	Cercopithecus preussi
Guenon, red-bellied	Cercopithecus erthytogaster
Guenon, red-eared	Cercopithecus erythrotis
Guenon, russet-eared	Cercopithecus erythrotis
Guenon, Sclater's	Cercopithecus sclateri
Guenon, sun-tailed	Cercopithecus solatus
Guenon, white-collared	Cercopithecus mitis
Guenon, white-nosed	Cercopithecus nictitans
Guenon, white-throated	Cercopithecus mitis
Guereza	Colobus guereza
Howler, black	Alouatta caraya
Howler, black and red	Alouatta belzebul
Howler, brown	Alouatta fusca
Howler, devil-handed	Alouatta belzebul
Howler, Guatemalan	Alouatta villosa
Howler, mantled	Alouatta palliata
Howler, red	Alouatta seniculus
Indri	Indri indri
Langur, banded	Presbytis melalophos
Langur, capped	Presbytis pileata
Langur, douc	Pygathrix nemeaus

Langur, dusky	Presbytis obscurus
Langur, entellus	Presbytis entellus
Langur, Francois'	Presbytis francoisi
Langur, golden	Presbytis geei
Langur, Hanuman	Presbytis entellus
Langur, Hose's	Presbytis hosei
Langur, maroon	Presbytis rubicunda
Langur, Mentawai	Presbytis potenziani
Langur, Nilgiri	Presbytis johnii
Langur, Phayre's	Presbytis phayrei
Langur, purple-faced	Presbytis senex
Langur, purple-faced	Presbytis thomasi
Langur, sacred	Presbytis entellus
Langur, silvered	Presbytis cristata
Langur, spectacled	Presbytis obscura
Langur, Sunda	Presbytis aygula
Langur, white-faced	Presbytis frontata
Langur, white-fronted	Presbytis frontata
Lar	Hylobates lar
Leaf monkey, banded	Presbytis melalophos
Leaf monkey, capped	Presbytis pileata
Leaf monkey, dusky	Presbytis obscura
Leaf monkey, Everett's	Presbytis hosei
Leaf monkey, Francois'	Presbytis francoisi
Leaf monkey, Gee's	Presbytis geei
Leaf monkey, golden	Presbytis geei
Leaf monkey, grey	Presbytis hosei
Leaf monkey, Hose's	Presbytis hosei
Leaf monkey, Javan	Presbytis comata
Leaf monkey, John's	Presbytis johnii
Leaf monkey, maroon	Presbytis rubicunda
Leaf monkey, Mentawai Island	Presbytis potenziani
Leaf monkey, Phayre's	Presbytis phayrei
Leaf monkey, pig-tailed	Simias concolor
Leaf monkey, purple-faced	Presbytis senex
Leaf monkey, red	Presbytis rubicundus
Leaf monkey, silvered	Presbytis cristata
Leaf monkey, Thomas's	Presbytis thomasi
Leaf monkey, white-fronted	Presbytis frontata
Lemur, avahi	Avahi laniger
Lemur, black	Lemur macaco
Lemur, blue-eyed	Lemur fulvus flavifrons
Lemur, broad-nosed gentle	Hapalemur simus
Lemur, brown	Lemur fulvus
Lemur, brown lesser mouse	Microcebus rufus
Lemur, collared	Lemur fulvus collaris
Lemur, Coquerel's mouse	Microcebus coquereli
Lemur, crowned	Lemur coronatus
Lemur, Edward's sportive	Lepilemur edwardsi
Lemur, fat-tailed dwarf	Cheirogaleus medius
Lemur, flying	Cynocephalus variegatus
Lemur, fork-marked dwarf	Phaner furcifer
Lemur, golden bamboo	Hapalemur aureus
Lemur, greater dwarf	Cheirogaleus major
Lemur, grey gentle	Hapalemur griseus
Lemur, grey lesser mouse	Microcebus murinus
Lemur, hairy-eared dwarf	Allocebus trichotis
Lemur, mongoose	Lemur mongoz
Lemur, red-bellied	Lemur rubriventer
Lemur, red-tailed sportive	Lepilemur ruficaudatus
Lemur, ringed-tailed	Lemur catta
Lemur, rufous mouse	Microcebus rufus
Lemur, ruffed	Varecia variegatus
Lemur, Sanford's	Lemur fulvus sanfordi
Lemur, Sclater's black	Lemur macaco flavifrons
Lemur, sportive	Lepilemur mustelinus
Lemur, white-fronted	Lemur fulvus albifrons
Lemur, woolly	Avahi laniger
Lesser mouse lemur, brown	Microcebus rufus
Lesser mouse lemur, grey	Microcebus murinus
Lion tamarin, black	Leontopithecus chrysopygus
Lion tamarin, golden	Leontopithecus rosalia rosalia
Lion tamarin, golden-headed	Leontopithecus chrysomelas
Loris, pygmy slow	Nycticebus pygmaeus
Loris, slender	Loris tardigradus
Loris, slow	Nycticebus coucang
Macaque, Assamese	Macaca assamensis
Macaque, Barbary	Macaca sylvanus
Macaque, bear	Macaca arctoides
Macaque, bonnet	Macaca radiata
Macaque, booted	Macaca ochreata
Macaque, Celebes black	Macaca nigra
Macaque, crab-eating	Macaca fascicularis
Macaque, cynomolgus	Macaca fascicularis
Macaque, Japanese	Macaca fuscata
Macaque, lion-tailed	Macaca silenus
Macaque, long-tailed	Macaca fascicularis
Macaque, Moor	Macaca maurus
Macaque, Pere David's	Macaca thibetana
Macaque, pig-tailed	Macaca nemestrina
Macaque, red-faced stump-tailed	Macaca arctoides
Macaque, rhesus	Macaca mulatta
Macaque, stump-tailed	Macaca arctoides
Macaque, Sulawesi booted	Macaca ochreata
Macaque, Sulawesi crested	Macaca nigra
Macaque, Taiwanese	Macaca cyclopis
Macaque, Tibetan stump-tailed	Macaca thibetana
Macaque, Tonkean	Macaca tonkeana
Macaque, toque	Macaca sinica
Mandrill	Mandrillus sphinx
Mangabey, agile	Cercocebus galeritus
Mangabey, black	Cercocebus aterrimus
Mangabey, black crested	Cercocebus aterrimus
Mangabey, cherry-capped	Cercocebus torquatus
Mangabey, golden-bellied	Cercocebus galeritus
Mangabey, grey-cheeked	Cercocebus albigena
Mangabey, mantled	Cercocebus albigena
Mangabey, smoky	Cercocebus torquatus
Mangabey, sooty	Cercocebus torquatus
Mangabey, Tana River	Cercocebus galeritus
Mangabey, white-cheeked	Cercocebus albigena
Mangabey, white-collared	Cercocebus torquatus
Mangabey, white-crowned	Cercoebus torquatus
Marimonda	Ateles belzebuth
Marmoset, bare-eared	Callithrix argentata
Marmoset, black-pencilled	Callithrix pencillata
Marmoset, black-tailed	Callithrix argentata
Marmoset, buffy-headed	Callithrix flaviceps

Marmoset, buffy-tufted-ear	Callithrix aurita
Marmoset, common	Callithrix jacchus
Marmoset, Goeldi's	Callimico goeldii
Marmoset, pygmy	Cebuella pygmaea
Marmoset, Rio Maues	Calithrix mauesi
Marmoset, Santarem	Callithrix humeralifer
Marmoset, silvery	Callithrix argentata
Marmoset, white-fronted	Callithrix geoffreyi
Marmoset, Wied'sblacktufted-earr	Callithrix kuhli
Mona, crowned	Cercopithecus pogonias
Mona, Wolf's	Cercopithecus wolfi
Monkey, Allen's swamp	Allenopithecus nigroviridis
Monkey, Banded leaf	Presbytis melalophos
Monkey, black spider	Ateles paniscus
Monkey, black-handed spider	Ateles geoffreyi
Monkey, blackish squirrel	Saimiri vanzolinii
Monkey, blue	Cercopithecus mitis
Monkey, brown snub-nosed	Rhinopithecus bieti
Monkey, brown-headed spider	Ateles fusciceps
Monkey, Campbell's	Cercopithecus campbelli
Monkey, capped leaf	Presbytis pileata
Monkey, Chinese snub-nosed	Rhinopithecus roxellanae
Monkey, common spider	Lagothrix lagotricha
Monkey, common squirrel	Saimiri sciureus
Monkey, de Brazza's	Cercopithecus neglectus
Monkey, diana	Cercopithecus diana
Monkey, Dusky leaf	Presbytis obscura
Monkey, Everett's leaf	Presbytis hosei
Monkey, Francois' leaf	Presbytis francoisi
Monkey, Gee's leaf	Presbytis geei
Monkey, Geoffrey's spider	Ateles geoffreyi
Monkey, Goeldi's	Callimico goeldii
Monkey, golden leaf	Presbytis geei
Monkey, golden snub-nosed	Rhinopithecus roxellanae
Monkey, greater spot-nosed	Cercopithecus nictitans
Monkey, greater white-nosed	Cercopithecus nictitans
Monkey, green	Cercopithecus aethiops
Monkey, Hamlyn's owl-faced	Cercopithecus hamlyni
Monkey, Hendee's woolly	Lagothrix flavicauda
Monkey, Hose's leaf	Presbytis hosei
Monkey, Humboldt's woolly	Lagothrix lagotricha
Monkey, hussar	Erythrocebus patas
Monkey, Java	Macaca fascicularis
Monkey, Javan leaf	Presbytis comata
Monkey, John's leaf	Presbytis johnii
Monkey, L'Hoest's	Cercopithecus lhoesti
Monkey, long-haired spider	Ateles belzebuth
Monkey, Lowe's	Cercopithecus campbelli
Monkey, maroon leaf	Presbytis rubicunda
Monkey, Mentawai Island	Presbytis potenziani
Monkey, Mentawai Is. snub-nosed	Simias concolor
Monkey, mona	Cercopithecus mona
Monkey, moustached	Cercopithecus cephus
Monkey, nisnas	Erythrocebus patas
Monkey, patas	Erythrocebus patas
Monkey, Phayre's leaf	Presbytis phayrei
Monkey, pig-tailed leaf	Simias concolor
Monkey, Preuss's	Cercopithecus preussi
Monkey, proboscis	Nasalis larvatus
Monkey, purple-faced leaf	Presbytis senex
Monkey, red	Erythrocebus patas
Monkey, red-backed squirrel	Saimiri oerstedi
Monkey, red-tailed	Cercopithecus ascanius
Monkey, roloway	Cercopithecus diana
Monkey, samango	Cercopithecus mitis albogularis
Monkey, savannah	Cercopithecus aethiops
Monkey, Schmidt's	Cercopithecusascaniusschmidti
Monkey, silvered leaf	Presbytis cristata
Monkey, snow	Macaca fuscata
Monkey, Sykes'	Cercopithecus mitis kolbi
Monkey, tantalus	Cercopithecus aethiopstantalus
Monkey, Thomas's leaf	Presbytis thomasi
Monkey, Tonkin snub-nosed	Rhinopithecus avunculus
Monkey, white-fronted leaf	Presbytis frontata
Monkey, widow	Callicebus torquatus
Monkey, woolly spider	Brachyteles arachnoides
Monkey, yellow-tailed woolly	Lagothrix flavicauda
Monkey, Zaire Diana	Cercopithecus salongo
Mouse lemur, brown lesser	Microcebus rufus
Mouse lemur, Coquerel's	Microcebus coquereli
Mouse lemur, grey lesser	Microcebus murinus
Mouse lemur, rufous	Microcebus rufus
Needle-nailed bushbaby, eastern	Euoticus inustus
Needle-nailed bushbaby, western	Euoticus elegantulus
Night monkey	Aotus trivirgatus
Orabazu	Callicebus moloch
Orangutan	Pongo pygmaeus
Orangutan, Sumatran	Pongo pygameus abelii
Owl monkey	Aotus trivirgatus
Owl monkey, Peruvianred-necked	Aotus nancymai
Potto	Perodicticus potto
Potto, golden	Arctocebus calabarensis
Proboscis monkey	Nasalis larvatus
Redtail	Cercopithecus ascanius
Roloway monkey	Cercopithecus diana
Sahui de collar	Callicebus torquatus
Saki, bearded	Chiropotes satanas
Saki, black bearded	Chiropotes satanas
Saki, monk	Pithecia monachus
Saki, red-nosed	Chiropotes albinasus
Saki, shaggy	Pithecia hirsuta
Saki, white-faced	Pithecia pithecia
Saki, white-nosed bearded	Chiropotes albinasus

Siamang	Symphalangus syndactylus
Sifaka, Coquerel's	Propithecus verreauxi coquereli
Sifaka, diademed	Propithecus diadema
Sifaka, Verreaux's	Propithecus verreauxi
Simakobu	Simias concolor
Snub-nosed monkey, brown	Rhinopithecus bieti
Snub-nosed monkey, Chinese	Rhinopithecus roxellanae
Snub-nosed monkey, golden	Rhinopithecus roxellanae
Snub-nosed monkey,Mentawai Is.	Simias concolor
Snub-nosed monkey, Tonkin	Rhinopithecus avunculus
Spider monkey, black	Ateles paniscus
Spider monkey, black-faced	Ateles paniscus chamek
Spider monkey, black-handed	Ateles geoffreyi
Spider monkey, brown-headed	Ateles fusciceps
Spider monkey, common	Lagothrix lagotricha
Spider monkey, Geoffrey's	Ateles geoffreyi
Spider monkey, long-haired	Ateles belzebuth
Spider monkey, red-faced	Ateles paniscus paniscus
Spider monkey, western black	Ateles paniscus chamek
Spider monkey, woolly	Brachyteles arachnoides
Squirrel monkey, blackish	Saimiri vanzolinii
Squirrel monkey, common	Saimiri sciureus
Squirrel monkey, red-backed	Saimiri oerstedi
Swamp monkey, Allen's	Allenopithecus nigroviridis
Talapoin, northern	Miopithecus sp. (unnamed)
Talapoin, southern	Miopithecus talapoin
Tamarin, bare-faced	Saguinus bicolor
Tamarin, black and red	Saguinus nigricollis
Tamarin, black-mantle	Saguinus nigricollis
Tamarin, cottontop	Saguinus oedipus
Tamarin, emperor	Saguinus imperator
Tamarin, golden lion	Leontopithecus rosalia rosalia
Tamarin, midas	Saguinus midas
Tamarin, mottle-faced	Saguinus inustus
Tamarin, moustached	Saguinus mystax
Tamarin, negro	Saguinus midas
Tamarin, pied	Saguinus bicolor
Tamarin, red-capped	Saguinus mystax pileatus
Tamarin, red-chested	Saguinus labiatus
Tamarin, red-handed	Saguinus midas
Tamarin, Rio Napo	Saguinus graellsi
Tamarin, rufous-naped	Saguinus geoffreyi
Tamarin, saddleback	Saguinus fuscicollis
Tamarin, white-footed	Saguinus leucopus
Tamarin, white-lipped	Saguinus labiatus
Tarsier, Horsfield's	Tarsius bancanus
Tarsier, Phillipine	Tarsius syrichta
Tarsier, spectral	Tarsius spectrum
Tarsier, western	Tarsius bancanus
Titi, dusky	Callicebus moloch
Titi, masked	Callicebus personatus
Titi, northern masked	Callicebuspersonatusmelanochi
Titi, white-handed	Callicebus torquatus
Titi, widow	Callicebus torquatus
Tree shrew, common	Tupaia glis
Tree shrew, feather-tailed	Ptilocercus lowii
Tree shrew, large	Tupaia tana
Tree shrew, lesser	Tupaia minor
Tree shrew, mountain	Tupaia montana
Tree shrew, nicobar	Tupaia nicobarica
Tree shrew, north smooth-tailed	Dendogale murina
Tree shrew, slender	Tupaia garcilis
Tree shrew, small	Tupaia javanica
Tree shrew, striped	Tupaia dorsalis
Tupaia, Everett's	Urogale everetti
Uakari, bald	Cacajao calvus
Uakari, black	Cacajao melanocephalus
Uakari, black-headed	Cacajao melanocephalus
Uakari, golden-backed	Cacajaomelanocephalusouakar
Uakari, red	Cacajao rubicundus
Uakari, white	Cacajao calvus
Vervet	Cercopithecusaethiopspygerth
Woolly monkey, Hendee's	Lagothrix flavicauda
Woolly monkey, Humboldt's	Lagothrix lagotricha
Woolly monkey, yellow-tailed	Lagothrix flavicauda

NO COMMON NAMES HAVE BEEN ASSIGNED TO:

Aotus azarae
Aotus brumbacki
Aotus inflatus
Aotus lemurinus griseimembra
Aotus lemurinus lemurinus
Aotus nigriceps
Aotus vociferans
Callicebus donacophilus
Cercopithecus polykomos dollmani
Cercopithecus polykomos vellerosus
Colobus badius temminckii
Colobus tholloni
Leontopithecus caissara
Lepilemur dorsalis
Lepilemur leucopus
Lepilemur microdon
Lepilemur septentrionalis
Microchoerus erinaceus
Microchoerus erinaceus edwardsi
Necrolemur antiquus
Necrolemur zitteli
Pithecia albicans
Pithecia hirsuta
Propithecus tattersalli
Saguinus inustus
Saguinus leucopus

Appendix D: Taxonomic Classification Of Primates

TAXONOMIC CLASSIFICATION: ORDER PRIMATES
as presented by the American Museum of Natural History
(Data retrieved from the Internet WWW site at **http://uakari.primate.wisc.edu:70/1/pin/taxon**

—

Primates
Cheirogaleidae
Cheirogaleinae
Allocebus
Allocebus trichotis
Cheirogaleus
Cheirogaleus major
Cheirogaleus medius
Microcebus
Microcebus coquereli
Microcebus murinus
Microcebus rufus
Phanerinae
Phaner
Phaner furcifer
Lemuridae
Eulemur
Eulemur coronatus
Eulemur fulvus
Eulemur macaco
Eulemur mongoz
Eulemur rubriventer
Hapalemur
Hapalemur aureus
Hapalemur griseus
Hapalemur simus
Lemur
Lemur catta
Varecia
Varecia variegata
Megaladapidae
Lepilemur
Lepilemur dorsalis
Lepilemur edwardsi
Lepilemur leucopus
Lepilemur microdon
Lepilemur mustelinus
Lepilemur ruficaudatus
Lepilemur septentrionalis
Indridae
Avahi
Avahi laniger
Indri
Indri indri
Propithecus
Propithecus diadema
Propithecus tattersalli
Propithecus verreauxi
Daubentoniidae
Daubentonia
Daubentonia madagascariensis
Loridae
Arctocebus
Arctocebus aureus
Arctocebus calabarensis
Loris
Loris tardigradus
Nycticebus
Nycticebus coucang
Nycticebus pygmaeus
Perodicticus
Perodicticus potto
Galagonidae
Euoticus
Euoticus elegantulus
Euoticus pallidus
Galago
Galago alleni
Galago gallarum
Galago matschiei
Galago moholi
Galago senegalensis
Galagoides
Galagoides demidoff
Galagoides zanzibaricus
Otolemur
Otolemur crassicaudatus
Otolemur garnettii
Tarsiidae
Tarsius
Tarsius bancanus
Tarsius dianae
Tarsius pumilus
Tarsius spectrum
Tarsius syrichta
Callitrichidae
Callimico
Callimico goeldii
Callithrix
Callithrix argentata
Callithrix aurita
Callithrix flaviceps
Callithrix geoffroyi
Callithrix humeralifer
Callithrix jacchus
Callithrix kuhlii
Callithrix penicillata
Callithrix pygmaea
Leontopithecus
Leontopithecus caissara
Leontopithecus chrysomela
Leontopithecus chrysopygus
Leontopithecus rosalia
Saguinus
Saguinus bicolor
Saguinus fuscicollis
Saguinus geoffroyi
Saguinus imperator
Saguinus inustus
Saguinus labiatus
Saguinus leucopus
Saguinus midas
Saguinus mystax
Saguinus nigricollis
Saguinus oedipus
Saguinus tripartitus
Cebidae
Alouattinae
Alouatta
Alouatta belzebul
Alouatta caraya
Alouatta coibensis
Alouatta fusca
Alouatta palliata
Alouatta pigra
Alouatta sara
Alouatta seniculus
Aotinae
Aotus
Aotus azarai
Aotus brumbacki
Aotus hershkovitzi
Aotus infulatus
Aotus lemurinus
Aotus miconax
Aotus nancymaae
Aotus nigriceps
Aotus trivirgatus
Aotus vociferans
Atelinae
Ateles
Ateles belzebuth
Ateles chamek
Ateles fusciceps
Ateles geoffroyi
Ateles marginatus
Ateles paniscus
Brachyteles
Brachyteles arachnoides
Lagothrix
Lagothrix flavicauda
Lagothrix lagotricha
Callicebinae
Callicebus
Callicebus brunneus
Callicebus caligatus
Callicebus cinerascens
Callicebus cupreus
Callicebus donacophilus
Callicebus dubius
Callicebus hoffmannsi
Callicebus modestus
Callicebus moloch
Callicebus oenanthe
Callicebus olallae
Callicebus personatus
Callicebus torquatus
Cebinae
Cebus
Cebus albifrons
Cebus apella
Cebus capucinus
Cebus olivaceus
Saimiri
Saimiri boliviensis
Saimiri oerstedii
Saimiri sciureus
Saimiri ustus
Saimiri vanzolinii
Pitheciinae

Cacajao
Cacajao calvus
Cacajao melanocephalus
Chiropotes
Chiropotes albinasus
Chiropotes satanas
Pithecia
Pithecia aequatorialis
Pithecia albicans
Pithecia irrorata
Pithecia monachus
Pithecia pithecia
Cercopithecidae
Cercopithecinae
Allenopithecus
Allenopithecus nigroviridis
Cercocebus
Cercocebus agilis
Cercocebus galeritus
Cercocebus torquatus
Cercopithecus
Cercopithecus ascanius
Cercopithecus campbelli
Cercopithecus cephus
Cercopithecus diana
Cercopithecus dryas
Cercopithecus erythrogaster
Cercopithecus erythrotis
Cercopithecus hamlyni
Cercopithecus lhoesti
Cercopithecus mitis
Cercopithecus mona
Cercopithecus neglectus
Cercopithecus nictitans
Cercopithecus petaurista
Cercopithecus pogonias
Cercopithecus preussi
Cercopithecus sclateri
Cercopithecus solatus
Cercopithecus wolfi
Chlorocebus
Chlorocebus aethiops
Erythrocebus
Erythrocebus patas
Lophocebus
Lophocebus albigena
Macaca
Macaca arctoides
Macaca assamensis
Macaca cyclopis
Macaca fascicularis
Macaca fuscata
Macaca maura
Macaca mulatta
Macaca nemestrina
Macaca nigra
Macaca ochreata
Macaca radiata
Macaca silenus
Macaca sinica
Macaca sylvanus
Macaca thibetana
Macaca tonkeana
Mandrillus
Mandrillus leucophaeus
Mandrillus sphinx
Miopithecus
Miopithecus talapoin
Papio
Papio hamadryas
Theropithecus
Theropithecus gelada
Colobinae
Colobus
Colobus angolensis
Colobus guereza
Colobus polykomos
Colobus satanas
Nasalis
Nasalis concolor
Nasalis larvatus
Presbytis
Presbytis comata
Presbytis femoralis
Presbytis frontata
Presbytis hosei
Presbytis melalophos
Presbytis potenziani
Presbytis rubicunda
Presbytis thomasi
Procolobus
Procolobus badius
Procolobus pennantii
Procolobus preussi
Procolobus rufomitratus
Procolobus verus
Pygathrix
Pygathrix avunculus
Pygathrix bieti
Pygathrix brelichi
Pygathrix nemaeus
Pygathrix roxellana
Semnopithecus
Semnopithecus entellus
Trachypithecus
Trachypithecus auratus
Trachypithecus cristatus
Trachypithecus francoisi
Trachypithecus geei
Trachypithecus johnii
Trachypithecus obscurus
Trachypithecus phayrei
Trachypithecus pileatus
Trachypithecus vetulus
Hylobatidae
Hylobates
Hylobates agilis
Hylobates concolor
Hylobates gabriellae
Hylobates hoolock
Hylobates klossii
Hylobates lar
Hylobates leucogenys
Hylobates moloch
Hylobates muelleri
Hylobates pileatus
Hylobates syndactylus
Hominidae
Gorilla
Gorilla gorilla
Homo
Homo sapiens
Pan
Pan paniscus
Pan troglodytes
Pongo
Pongo pygmaeus

Appendix E: DECLARATION

I declare that

a) this thesis submitted for the Degree of Doctor of Philosophy is composed by myself, and

b) the work herein is my own, or that the author involved is clearly stated.

Kanwaljit Singh Dulai B.Sc. with Distinction, B.Sc. Hons. (Genetics).

Publications

Hunt, D. M., Williams, A. J., Ibbotson, R. A., **Dulai, K. S.,** Bowmaker, J. K. and Mollon, J. D. (1993b). Sequence evolution of visual pigment genes in primates. *Invest. Ophthalmol. Visual Sci. (Suppl.)* **34**: 749.

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 Research* **34**(19): 2483-2491.

Fitzgibbon, J., Pilz, A., Gayther, S., Appukuttan, B., **Dulai, K. S.,** Delhanty, J. D., Helmkamp, G., Jr., Yarbrough, L. R., & Hunt, D. M. (1994). Localization of the gene encoding human phosphatidylinositol transfer protein (PITPN) to 17p13.3: a gene showing homology to the Drosophila retinal degeneration B gene (rdgB). *Cytogenet Cell Genet.* **67**(3): 205-207.

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.

Dulai, K. S., Mollon, J. D. and Hunt, D. M. (1995). Molecular biography of John Dalton: The genetic basis of his color blindness. *Biosystems Reporter* :

References

Abney, W. (1913). *Researches in colour vision and the trichromatic theory.* London, Longmans, Green & Co.

Ahmad, I., Yu, X. and Barnstable, C. J. (1994). A cis-acting element, T alpha-1, in the upstream region of rod alpha- transducin gene that binds a developmentally regulated retina-specific nuclear factor. *J Neurochem* **62**(1): 396-9.

Ali, M. A. and Klyne, M. A. (1985). *Vision in Vertebrates.* New York, Plenum Press.

Alkorta, I. and Du, P. (1994). Sequence divergence analysis for the prediction of seven-helix membrane protein structures: II. A 3-D model of human rhodopsin. *Protein Eng* **7**(10): 1231-1238.

Applebury, M. L. (1991). Molecular determinants of visual pigment function. *Current Opinion Neurobiology* **1**(2): 263-269.

Applebury, M. L. (1994). Relationships of G-protein-coupled receptors. A survey with the photoreceptor opsin subfamily. *Soc Gen Physiol Ser* **49**: 235-48.

Archer, S. N., Hope, A. and Partridge, J. (1995). The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. *Proceedings of the Royal Society London B* **262**: 289-295.

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

Balding, D. J., Nichols, R. A. and Hunt, D. M. (1992). Detecting gene conversion: primate visual pigment genes. *Proc R Soc Lond [Biol]* **249**(1326): 275-80.

Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *Embo J* **12**(4): 1693-703.

Barlow, D. J. and Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**: 601-619.

Barnes, W. M. (1994). PCR amplification of up to 35 kb DNA with high fidelity and high yield from l templates. *Proceedings of the National Academy of Sciences USA* **91**: 2216-2220.

Barnstable, C. J., Blum, A. S., Devoto, S. H., Hicks, D., Morabito, M. A., Sparrow, J. R. and Treisman, J. E. (1988). Cell differentiation and pattern formation in the developing mammalian retina. *Neurosci Res Suppl* **8**(41): S27-41.

Batzer, M. A. and Deininger, P. L. (1991). A human-specific subfamily of Alu sequences. *Genomics* **9**(3): 481-487.

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J. and et, a. (1994). African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A* **91**(25): 12288-12292.

Bauer, K. (1993). Primate phylogeny studies by comparative determinant analysis. *Exp. Clin. Immunogenet.* **10**: 56-60.

Bennett, J., Sun, D. and Kariko, K. (1995). Sequence analysis of the 5.34-kb 5' flanking region of the human rhodopsin-encoding gene. *Gene* **167**: 317-320.

Beris, P., Kitundu, M. N., Baysal, E., Oner, C., Lanclos, K. D., Dimovski, A. J., Kutlar, F. and Huisman, T. H. (1992). Black beta-thalassemia homozygotes with specific sequence variations in the 5' hypersensitive site-2 of the locus control region have high levels of fetal hemoglobin. *Am J Hematol* **41**(2): 97-101.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**(6): 1513-23.

Blackwell, H. R. and Blackwell, O. M. (1961). Rod and cone receptor mechanisms in typical and atypical congenital achromatopsia. *Vision Research* **1**: 62-107.

Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N. and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. *Science* **250**(4984): 1149-51.

Blackwell, T. K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**(4984): 1104-10.

Bowmaker, J. K. (1984). Microspectrophotometry of vertebrate photoreceptors. *Vision Research* **24**(11): 1641-1650.

Bowmaker, J. K. (1990). Visual pigments: colour vision, opsin structure and evolution. *Seminars in the Neurosciences* **2**: 25-32.

Bowmaker, J. K. (1991). Evolution of visual pigments and photoreceptors. *Vision and Visual Dysfunction, 2, Evolution of the Eye and Visual System*. London, Macmillan. 63-81.

Bowmaker, J. K., Astell, S., Hunt, D. M. and Mollon, J. D. (1991). Photosensitive and photostable pigments in the retinae of Old World monkeys. *J. exp. Biol.* **156**: 1-19.

Bowmaker, J. K., Dartnall, H. J. A. and Mollon, J. D. (1980). Microspectrophotometric demonstration of four classes of photoreceptor in an old world primate, *Macaca fascicularis. J. Physiol.* **298**: 131-143.

Bowmaker, J. K., Govardovskii, V. I., Shukolyukov, S. A., Zueva, L. V., Hunt, D. M., Sideleva, V. G. and Smirnova, O. G. (1994). Visual pigments and the photic environment: the cottoid fish of Lake Baikal. *Vision Research* **34**(5): 591-605.

Bowmaker, J. K., Jacobs, G. H. and Mollon, J. D. (1987). Polymorphism of photopigments in the squirrel monkey: a sixth phenotype. *Proc R Soc Lond B Biol Sci* **231**(1264): 383-390.

Boynton, R. M. (1979). *Human Color Vision*. New York, Holt, Rinehart, Winston.

Brewster, D. (1842). *Letters on Natural Magic, addressed to Sir Walter Scott, Bart.* London, J. Murray.

Briggs, M. R., J. T. Kadonaga, et al. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science **234**(4772): 47-52.

Brini, A. T., Lee, G. M. and Kinet, J. P. (1993). Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors. *J Biol Chem* **268**(2): 1355-61.

Brockbank, E. M. (1944). *John Dalton: Some unpublished letters of personal and scientific interest with additional information about his colour vision & atomic theories*. Manchester, Manchester University Press.

Cachon-Gonzalez, M. B., Delhanty, J. D., Burn, J., Tsioupra, K., Davis, M. B., Attwood, J. and Chapman, P. (1991). Linkage analysis in adenomatous polyposis coli: the use of four closely linked DNA probes in 20 UK families [see comments]. *J Med Genet* **28**(10): 681-5.

Caspersson, T. (1940). Methods for the determination of the absorption spectra of cell structures. *J. R. Microsc.* **60**: 8-25.

Caterina, J. J., Ciavatta, D. J., Donze, D., Behringer, R. R. and Townes, T. M. (1994a). Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Res* **22**(6): 1006-1011.

Caterina, J. J., Donze, D., Sun, C. W., Ciavatta, D. J. and Townes, T. M. (1994b). Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression. *Nucleic Acids Res* **22**(12): 2383-2391.

Caterina, J. J., Ryan, T. M., Pawlik, K. M., Palmiter, R. D., Brinster, R. L., Behringer, R. R. and Townes, T. M. (1991). Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. *Proc Natl Acad Sci U S A* **88**(5): 1626-30.

Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967). Phylogenetic analysis: Models and estimation procedures. *American Journal of Human Genetics* **19**: 233-257.

Chan, T., Lee, M. and Sakmar, T. P. (1992). Introduction of hydroxyl-bearing amino acids causes bathochromic spectral shifts in rhodopsin. Amino acid substitutions

responsible for red-green color pigment spectral tuning. *J Biol Chem* **267**(14): 9478-9480.

Chang, B. S. W., Crandall, K., Carulli, J. and Hartl, D. L. (1995). Opsin phylogeny and evolution: A model for blue shifts in wavelength regulation. *Molecular Phylogenetics and Evolution* **4**(1): 31-43.

Chang, J. C., Liu, D. and Kan, Y. W. (1992). A 36-base-pair core sequence of locus control region enhances retrovirally transferred human beta-globin gene expression. *Proc Natl Acad Sci U S A* **89**(7): 3107-10.

Cheng, S., Chang, S.-Y., Gravitt, P. and Respess, R. (1994). Long PCR. *Nature.* **369:** 684-685.

Chester, N. and Marshak, D.R. (1993). Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction. *Anal Biochem.* **209**(2): 284-90.

Chiu, M. I. and Nathans, J. (1994a). Blue cones and cone bipolar cells share transcriptional specificity as determined by expression of human blue visual pigment-derived transgenes. *J Neurosci* **14**(6): 3426-3436.

Chiu, M. I. and Nathans, J. (1994b). A sequence upstream of the mouse blue visual pigment gene directs blue cone-specific transgene expression in mouse retinas. *Vis Neurosci* **11**(4): 773-80.

Chiu, M. I., Zack, D. J., Wang, Y. and Nathans, J. (1994). Murine and bovine blue cone pigment genes: cloning and characterization of two new members of the S family of visual pigments. *Genomics* **21**(2): 440-443.

Chou, Q., Russell, M., Birch, D.E., Raymond, J. and Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low- copy-number amplifications. *Nucleic Acids Res.* **20**(7): 1717-23.

Church, G.M. and Gilbert, W. (1984). Genomic Sequencing. *Proc Natl Acad Sci USA.* **81:** 1991-1995.

Cobbs, W. H., Barkdoll, A. E. and Pugh, E. (1985). Cyclic GMP increases photocurrent and light sensitivity of retinal cones. *Nature* **317**: 64-66.

Cohen, G. B., Oprian, D. D. and Robinson, P. R. (1992). Mechanism of activation and inactivation of opsin - role of glu(113) and lys(296). *Biochemistry* : 12592-12601.

Collins, S., Bolanowski, M. A., Caron, M. G. and Lefkowitz, R. J. (1989). Genetic regulation of beta-adrenergic receptors. *Annu Rev Physiol* **51**: 203-15.

Courey, A. J. and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55**(5): 887-98.

Dalton, J. (1798). Extraordinary facts relating to the vision of colours. *Memoirs of the Manchester Literary and Philosophical Society* **5**: 28.

Damasio, A. R., Damasio, H., Ferro, J. M. and Castro-Caldas, A. (1974). Recovery from hemianopia in man: evidence for collicular vision? *Lancet* **Jul 13, 2**(872): 110.

Darnell, J., Jr. (1982). Variety in the level of gene control in eukaryotic cells. *Nature* **297**(5865): 365-71.

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

Deeb, S. S. (1993). Genetic determinants of visual functions. *Current Opinion in Neurobiology* **3**: 506-512.

Deeb, S. S., Jorgensen, A. L., Battisti, L., Iwasaki, L. and Motulsky, A. G. (1994). Sequence divergence of the red and green visual pigments in great apes and humans. *Proc Natl Acad Sci USA* **91**(15): 7262-7266.

Deeb, S. S., Lindsey, D. T., Hibiya, Y., Sanocki, E., Winderickx, J., Teller, D. Y. and Motulsky, A. G. (1992). Genotype-phenotype relationships in human

red/green color-vision defects: molecular and psychophysical studies. *Am J Hum Genet* **51**(4): 687-700.

Deeb, S. S., Winderickx, J., Sanocki, E., Teller, D. Y., Lindsey, D. T. and Motulsky, A. G. (1993). Rayleigh match range in protan subjects is highly correlated with amino-acid-sequence of the red-green hybrid opsin. *Invest. Ophthalmol. Visual Sci.* : 910-910.

DesJardin, L. E. and Hauswirth, W. W. (1996). Developmentally important DNA elements within the bovine opsin upstream region. *Invest Ophthalmol Vis Sci* **37**(1): 154-165.

DesJardin, L. E., Lockwood, M. K. and Hauswirth, W. W. (1995). Bovine opsin gene expression exhibits a late fetal to adult regulatory switch. *J Neurosci Res* **40**(6): 728-736.

DesJardin, L. E., Timmers, A. M. and Hauswirth, W. W. (1993). Transcription of photoreceptor genes during fetal retinal development. Evidence for positive and negative regulation. *J Biol Chem* **268**(10): 6953-60.

Devereux, J., Haerberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res.* **12**: 387-395.

Dominguez, O. and Lopez-Larrea, C. (1994). Gene walking by unpredictably primed PCR. *Nucleic Acids Research* **22**(15): 3247-3248.

Dowling, J. E. (1987). *The Retina.* Cambridge, MA, Belknap Press.

Drummond-Borg, M., Deeb, S. S. and Motulsky, A. G. (1989). Molecular patterns of X chromosome-linked color vision genes among 134 men of European ancestry. *Proc Natl Acad Sci U S A* **86**(3): 983-987.

Du, P. and Alkorta, I. (1994). Sequence divergence analysis for the prediction of seven-helix membrane protein structures: I. Comparison with bacteriorhodopsin. *Protein Eng* **7**(10): 1221-1229.

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 Research* **34**(19): 2483-2491.

Dulai, K. S., Mollon, J. D. and Hunt, D. M. (1995). Molecular biography of John Dalton: The genetic basis of his color blindness. *Biosystems Reporter* :

Dynan, W. S., Sazer, S., Tjian, R. and Schimke, R. T. (1986). Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. *Nature* **319**(6050): 246-8.

Dynan, W. S. and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* **32**(3): 669-80.

Dynan, W. S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA- binding proteins. *Nature* **316**(6031): 774-8.

Eckert, K. A. and Kunkel, T. A. (1990). *Nucleic Acids Research* **18**: 3739-3744.

Fager, L. Y. and Fager, R. S. (1979). Halide control of colour of the chicken cone pigment iodopsin. *Experimental Eye Research* **29**: 401-408.

Faisst, S. and Meyer, S. (1992). Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res* **20**(1): 3-26.

Farber, D. B. (1995a). From mice to men - the cyclic-gmp phosphodiesterase gene in vision and disease. *Invest. Ophthalmol. Visual Sci.* **36**: 261.

Farber, D. B. (1995b). From mice to men - the cyclic-gmp phosphodiesterase gene in vision and disease - the proctor lecture. *Invest. Ophthalmol. Visual Sci.* **36**: 263-275.

Feil, R., Aubourg, P., Heilig, R. and Mandel, J. L. (1990). A 195-kb cosmid walk encompassing the human Xq28 color vision pigment genes. *Genomics* **6**: 367-373.

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**(1): 79-80.

Francois, J. (1961). *Heredity in Ophthalmology.* St Louis, Mosby.

Franke, R.R., Sakmar, T.P., Graham, R.M. and Khorana, H.G. (1992). Structure and Function in Rhodopsin : Studies of the interation between the Rhodopsin Cytoplasmic Domain and Transducin. *J Biol Chem.* **267**: 14767-14774.

Fukada, Y. (1993). Roles of lipids linked to a- and g-subunits of photoreceptor G protein. *Frontiers of Photobiology* : 195-200.

Fukada, Y., Takao, T., Ohguro, H., Yoshizawa, T., Akino, T. and Shimonishi, Y. (1990). Farnesylated g-subunit of photoreceptor G protein indispensable for GTP-binding. *Nature* **346**: 658-660.

Gidoni, D., Dynan, W. S. and Tjian, R. (1984). Multiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* **312**(5993): 409-13.

Goodall, J. (1986). The Chimpanzees of Gombe. *Patterns of Behavior.* Cambridge, MA: Harvard University Press.

Goodman, M., Bailey, W. J., Hayasaka, K., Stanhope, M. J., Slightom, J. and Czelusniak, J. (1994). Molecular evidence on primate phylogeny from DNA sequences. *American Journal of Physical Anthropology* **94**: 3-24.

Ghosh, D. (1990). A relational database of transcription factors. *Nucleic Acids Res.* **18**(7)**:** 1749-56.

Grether, W. F. (1939). Colour vision and colour blindness in monkeys. *Mongr. Comp. Psychol.* **15**: 1-38.

Guillemot, F., Lo, L. C., Johnson, J. E., Auerbach, A., Anderson, D. J. and Joyner, A. L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**(3): 463-76.

Hagelberg, E. (1994). Dried samples: hard tissues. *Ancient DNA.* New York, Springer-Verlag.

Hagelberg, E., Gray, I. C. and Jeffreys, A. J. (1991). Identification of the skeletal remains of a murder victim by DNA Analysis. *Nature* **352**: 427-429.

Hargrave, P. A. (1982). Rhodopsin chemistry, structure and topography. *Prog. Retinal Res.* **1**: 1-51.

Hargrave, P. A., McDowell, J. H., Curtis, D. R., Wang, J. K., Juszczak, E., Fong, S. L., Mohanna Roa, J. K. and Argos, P. (1983). The structure of bovine rhodopsin. *Biophys. Struct. Mech.* **9**: 235-244.

Hargrave, P. A., McDowell, J. H., Feldmann, R. J., Atkinson, P. H., Pao, J. K. M. and Argos, P. (1984). Rhodopsin's protein and carbohydrate structure: selected aspects. *Vision Research* **24**: 1487-1499.

Helmholtz, H. V. (1896). *Handbuch der Physiologischen Optik.* Hamburg, Vos.

Henry, W. C. (1854). *Memoirs of the life and scientific researches of John Dalton.* London, Cavendish Society.

Herrmann, B. and Hummel, S. (1994). *Ancient DNA: Recovery and Analysis of Genetic Material from Paleontological, Archaeological, Museum, Medical, and Forensic Specimens.* New York, Springer-Verlag.

Hess, J., Perez-Stable, C., Wu, G. J., Weir, B., Tinoco, I., Jr. and Shen, C. K. (1985). End-to-end transcription of an Alu family repeat. A new type of polymerase-III-dependent terminator and its evolutionary implication. *J Mol Biol* **184**(1): 7-21.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. and Wilson, A. C. (1984). DNA sequences from the Quagga, an extinct member of the horse family. *Nature* **312**: 282-284.

Hodgkinson, C. A., Moore, K. J., Nakayama, A., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. and Arnheiter, H. (1993). Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**(2): 395-404.

Hoey, T., Weinzierl, R. O., Gill, G., Chen, J. L., Dynlacht, B. D. and Tjian, R. (1993). Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. *Cell* **72**(2): 247-60.

Honig, B., Dinur, U., Nakanishi, K., Balogh-Nair, V., Gawinowicz, M. A., Arnaboldi, M. and Motto, M. G. (1979). An external point-charge model for wavelength regulation in visual pigments. *J. Am. Chem. Soc.* **101**: 7084-7086.

Honig, B., Greenburg, A. D., Dinur, U. and Ebrey, T. G. (1976). Visual-pigment spectra: Implications of the protonation of the retinal Schiff base. *Biochemistry* **15**: 4593-4599.

Hope, A. J., Partridge, J. C., Dulai, K. S. and Hunt, D. M. (1996). Molecular adaptations of rod opsins to the deep marine environment of deep-sea fishes. *In preparation.*

Hubbart, J. (1777). An account of persons who could not distinguish colours. *Philosophical Transactions of the Royal Society* **67**: 260-265.

Hunan Medical College (1980). *Study of an ancient cadaver in Mawangtui Tomb No. 1 of the Han Dynasty in Changsha.* Beijing, Ancient Memorial Press.

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.

Hunt, D. M., Fitzgibbon, J., Slobodyanyuk, S. and Bowmaker, J. K. (1994). Spectral tuning and molecular evolution of rhodopsin pigments in cottoid fish from different photic environments in lake baikal. *Invest. Ophthalmol. Visual Sci.* : 1708-1708.

Hunt, D. M., Williams, A. J., Bowmaker, J. K. and Mollon, J. D. (1993a). Structure and evolution of the polymorphic photopigment gene of the marmoset. *Vision Research* **33**(2): 147-54.

Hunt, D. M., Williams, A. J., Ibbotson, R. A., Dulai, K. S., Bowmaker, J. K. and Mollon, J. D. (1993b). Sequence evolution of visual pigment genes in primates. *Invest. Ophthalmol. Visual Sci. (Suppl.)* **34**: 749.

Ibbotson, R. (1991). The molecular biology of colour vision in the Old World monkeys and humans. Ph.D. University of London.

Ibbotson, R. E., Hunt, D. M., Bowmaker, J. K. and Mollon, J. D. (1992). Sequence divergence and copy number of the middle- and long-wave photopigment genes in Old World monkeys. *Proc R Soc Lond B Biol Sci* **247**(1319): 145-154.

Imagawa, M., Chiu, R. and Karin, M. (1987). Transcription factor AP-2 mediates induction by two different signal- transduction pathways: protein kinase C and cAMP. *Cell* **51**(2): 251-60.

Innis, M. A. (1990). *PCR protocols - a guide to methods and applications.* Academic Press.

Ishihara, S. (1968). Ishihara Plates. Tokyo, Kanehara Shuppan Co. Ltd.

Jacobs, G. H. (1977). Visual sensitivity: significant within-species variations in a nonhuman primate. *Science* **197**(4302): 499-500.

Jacobs, G. H. (1984). Within-species variation in visual capacity among squirrel monkeys (*Saimiri sciureus*): colour vision. *Vision Research* **24**:1267-1277.

Jacobs, G. H. (1986). Colour vision variants in non-human primates. *Trends Neurosci.* **9**:320-323.

Jacobs, G. H. (1993). The distribution and nature of colour vision among the mammals. *Biology Review* **68**: 413-471.

Jacobs, G. H., Bowmaker, J. K. and Mollon, J. D. (1981). Behavioural and microspectro-photometric measurements of colour vision in monkeys. *Nature* **292**(5823): 541-3.

Jacobs, G. H. and Deegan II, J. F. (1993). *Invest. Ophthal. Vis. Sci.* **34**: 749.

Jacobs, G. H. and Deegan II, J. F. (1994). *American Journal of Primatology* **33**: 217.

Jacobs, G. H. and Neitz, J. (1987). Inheritance of color vision in a new world monkey (*Saimiri sciureus*). *Proceedings of the National Academy of Sciences USA* **84**: 2545-2549.

Jacobs, G. H., Neitz, J. and Deegan, J. F. (1991). Retinal receptors in rodents maximally sensitive to ultraviolet light. *Nature* **353**: 655-656.

Jacobs, G. H., Neitz, M., Deegan II, J. F. and Neitz, J. (1996). Emergence of routine trichromatic colour vision in New World monkeys. *In Press* .

Jones, K. A. and Tjian, R. (1985). Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription in vitro. *Nature* **317**(6033): 179-82.

Jones, N. C., Rigby, P. W. and Ziff, E. B. (1988). Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. *Genes and Development* **2**: 267-281.

Jorgensen, L., A., D., 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. *Proceedings of the National Academy of Science, USA* **87**: 6512-6516.

Jukes, T. H. and Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*. New York, Academic Press. 21-132.

Karnik, S. S. and Khorana, H. G. (1990). Assembly of functional rhodopsin requires a disulphide bond between cysteine residues 110 and 187. *The Journal of Biological Chemistry* **265**(29): 17520-17524.

Kenter, A. L. and Birshtein, B. K. (1981). *Chi*, a promoter of generalized recombination in lambda phage, is present in immunoglobulin genes. *Nature* **293**(5831): 402-4.

Khorana, H. G. (1992). Rhodopsin, photoreceptor of the rod cell. An emerging pattern for sturcture and function. *J. Biol. Chem* **267**(1): 1-4.

Kielman, M. F., Smits, R. and Bernini, L. F. (1994). Localization and characterization of the mouse alpha-globin locus control region. *Genomics* **21**(2): 431-433.

Kim, J. H., Yu, C. Y., Bailey, A., Hardison, R. and Shen, C. K. (1989). Unique sequence organization and erythroid cell-specific nuclear factor- binding of mammalian theta 1 globin promoters. *Nucleic Acids Res* **17**(14): 5687-700.

Kleinschmidt, J. and Hárosi, F. I. (1992). Anion sensitivity and spectral tuning of cone visual pigments *in situ*. *Proceedings of the National Academy of Sciences USA* **89**: 9181-9185.

Knowles, A. (1980). The chloride effect in chicken red cone receptors. *Vision Research* **20**: 475-483.

Kobilka, B. (1992). Adrenergic receptors as models for G protein-coupled receptors. *Annual Review of Neuroscience* **15**: 87-114.

Kohlrausch, A. (1923). Pflügers Archive ges. *Physiol. Menschen Tiere* **200**: 216-219.

Kominami, R., Muramatsu, M. and Moriwaki, K. (1983). A mouse type 2 Alu sequence (M2) is mobile in the genome. *Nature* **301**(5895): 87-89.

Koop, B. F., Miyamoto, M. M., Embury, J. E., Goodman, M., Czelusniak, J. and Slightom, J. L. (1986). Nucleotide sequence and evolution of the orangutan epsilon globin gene region and surrounding Alu repeats. *J Mol Evol* **24**(1-2): 94-102.

Kosower, E. M. (1988). Assignment of groups responsible for the opsin shift and light absorptions of rhodopsin and red, green, and blue iodopsins (cone pigments). *Proceedings of the National Academy of Sciences USA* **85**: 1076-1080.

Koutalos, Y. and Yau, K.-W. (1993). A rich complexity emerges in phototransduction. *Current Opinion in Neurobiology* **3**: 513-519.

Krowczynska, A. M., Rudders, R. A. and Krontiris, T. G. (1990). The human minisatellite consensus at breakpoints of oncogene translocations. *Nucleic Acids Res* **18**(5): 1121-7.

Kulozik, A. E., Kohne, E. and Kleihauer, E. (1988). The molecular basis of hereditary persistence of fetal hemoglobin (HPFH). Clinical importance of the hemoglobin switching mechanism with special reference to Corfu delta beta zero thalassemia. *Monatsschr Kinderheilkd* **136**(11): 751-757.

Kumar, S., Tamura, K. and Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis. Pennsylvania, The Pennsylvania State University.

Kuwata, O., Imamoto, Y., Okano, T., Kokame, K., Kojima, D., Matsumoto, H., Morodome, A., Fukada, Y., Shichida, Y., Yasuda, K., Shimura, Y. and Yoshizawa, T. (1990). The primary structure of iodopsin, a chicken red-sensitive cone pigment. *FEBS Letters* **272**: 128-132.

Land, M. F. (1984). Mulluscs (Eyes). *Photoreception and Vision in Invertebrates.* New York, Plenum Press. 699-725.

Latchman, D. S. (1995). *Gene Regulation: A Eukaryotic Perspective.* London, Chapman and Hall.

Lawlor, D. A., Dickel, C. D., Hauswirth, W. W. and Parham, P. (1991). Ancient HLA genes from 7,500-year-old archaeological remains. *Nature* **349**: 785-788.

Lem, J., Applebury, M. L., Falk, J. D., Flannery, J. G. and Simon, M. I. (1991). Tissue-specific and developmental regulation of rod opsin chimeric genes in transgenic mice. *Neuron* **6**(2): 201-10.

Lerea, C. L., Bunt-Milam, A. H. and Hurley, J. B. (1989). Alpha transducin is present in blue-, green-, and red-sensitive cone photoreceptors in the human retina. *Neuron* **3**(3): 367-76.

Lerea, C. L., Somers, D. E., Hurley, J. B., Klock, I. B. and Bunt-Milam, H. (1986). Identification of specific transducin a subunits in retinal rod and cone photoreceptors. *Science* **234**: 77-80.

Li, Q. L., Zhou, B., Powers, P., Enver, T. and Stamatoyannopoulos, G. (1990). Beta-globin locus activation regions: conservation of organization, structure, and function. *Proc Natl Acad Sci U S A* **87**(21): 8207-8211.

Li, S. and Rosen, J. (1994). Distal regulatory elements required for rat whey acidic protein gene expression in transgenic mice. *Journal of Biological Chemistry* **269**(29): 14235-14243.

Li, W.H., Wu, C.I. and Luo, C.C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* **2**(2): 150-74.

Lin, S. W., Imamoto, Y., Fukada, Y., Shichida, Y., Yoshizawa, T. and Mathies, R. A. (1994). What makes red visual pigments red? A resonance Raman microprobe study of retinal chromophore structure in iodopsin. *Biochemistry* **33**(8): 2151-60.

Linnaeus, C. (1758). *Systema naturae*. Holmiae, Laurentii Salvii.

Lonsdale, H. (1874). *The Worthies of Cumberland*. London, Routledge.

Lutze, M., Cox, N. J., Smith, V. C. and Pokorny, J. (1990). Genetic studies of variation in Rayleigh and photometric matches in normal trichromats. *Vision Research* **30**(1): 149-162.

Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X chromosome. *American Journal of Human Genetics* **14**: 135-148.

Madisen, L. and Groudine, M. (1994). Identification of a locus control region in the immunoglobulin heavy- chain locus that deregulates c-myc expression in plasmacytoma and Burkitt's lymphoma cells. *Genes Dev* **8**(18): 2212-26.

Maniatis, T., Goodbourn, S. and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* **236**(4806): 1237-45.

Maniatis, T. and Weintraub, H. (1992). Gene expression and differentiation [editorial]. *Curr Opin Genet Dev* **2**(2): 197-8.

Martin, R. D. (1990). *Primate Origins and Evolution: A Phylogenetic Reconstruction.* London, Chapman & Hall.

Martyn, T. (1777). *Thirty-eight plates with explanations: intended to illustrate Linnaeus's System of Vegetables.* London, J. White.

Mathur, E. (1991). Perfect Match. *Strategies* **4**(3): 38.

McKnight, S. L. and Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217**(4557): 316-24.

McKnight, S. L., Kingsbury, R. C., Spence, A. and Smith, M. (1984). The distal transcription signals of the herpesvirus tk gene share a common hexanucleotide control sequence. *Cell* **37**(1): 253-62.

Meireles, C. M., Schneider, M. P., Sampaio, M. I., Schneider, H., Slightom, J. L., Chiu, C. H., Neiswanger, K., Gumucio, D. L., Czelusniak, J. and Goodman, M. (1995). Fate of a redundant gamma-globin gene in the atelid clade of New World monkeys: implications concerning fetal globin gene expression. *Proc Natl Acad Sci U S A* **92**(7): 2607-2611.

Merbs, S. L. and Nathans, J. (1992a). Absorption spectra of human cone pigments. *Nature* **356**(6368): 433-435.

Merbs, S. L. and Nathans, J. (1992b). Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science* **258**(5081): 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**(5): 706-710.

Millchamp, N. J. (1990). Retinal degeneration in the dog and cat. *Vet Clin North Am Small Anim Pract* **20**(3): 799-835.

Miller, I. J. and Bieker, J. J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* **13**(5): 2776-86.

Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W. and Tjian, R. (1991). Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev* **5**(1): 105-19.

Mitchell, P. J., Wang, C. and Tjian, R. (1987). Positive and negative regulation of transcription in vitro: enhancer- binding protein AP-2 is inhibited by SV40 T antigen. *Cell* **50**(6): 847-61.

Mollon, J. D. (1991a). Hue and the heptahelicals. *Nature* **351**: News and views.

Mollon, J. D. (1991b). Evolution of the Eye and Visual System. *Evolution of the Eye and Visual System*. Boca Raton, CRC Press.

Mollon, J. D. (1991c). Uses and evolutionary origins of primate colour vision. *Vision and Visual Dysfunction*. London, MacMillan. 306-319.

Morabito, M., Yu, X. and Barbstable, C. J. (1991). Characterization of developmentally regulated and retina-specific nuclear proteins binding to a site in the upstream region of the rat opsin gene. *The Journal of Biological Chemistry* **266**(15): 9667-9672.

Morris, A., Bowmaker, J. K. and Hunt, D. M. (1993). The molecular basis of a spectral shift in the rhodopsins of two species of squid from different photic environments. *Proc R Soc Lond B Biol Sci* **254**(1341): 233-240.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Horbor Symposium of Quantitative Biology* **51**: 263-273.

Mullis, K. B. and Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. *Meth. Enzymol* **155**: 335-350.

Nakayama, T. A. and Khorana, H. G. (1991). Mapping of the amino acids in membrane-embedded helices that interact with the retinal chromophore in bovine rhodopsin. *J. Biol. Chem.* **29**: 9746-9752.

Napier, J. R. and Napier, P. H. (1985). *The Natural History of the Primates*. London, British Museum (Natural History).

Nathans, J. (1990a). Determinants of visual pigment absorbance: Role of charged amino acids in the putative transmembrane segments. *Biochemistry* **29**(4): 937-942.

Nathans, J. (1990b). Determinants of visual pigment absorbance: Identification of the retinylidene Schiff's base counterion in bovine rhodopsin. *Biochemistry* **29**: 9746-9752.

Nathans, J. (1994). In the eye of the beholder: visual pigments and inherited variation in human vision. *Cell* **78**(3): 357-360.

Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., Klingaman, R. and Fishman, G. (1989). Molecular genetics of human blue cone monochromacy. *Science* **245**(4920): 831-838.

Nathans, J. and Hogness, D. S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* **34**(3): 807-814.

Nathans, J. and Hogness, D. S. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc Natl Acad Sci U S A* **81**(15): 4851-4855.

Nathans, J., Maumenee, I. H., Zrenner, E., Sadowski, B., Sharpe, L. T., Lewis, R. A., Hansen, E., Rosenberg, T., Schwartz, M., Heckenlively, J. R. and et, a. (1993). Genetic heterogeneity among blue-cone monochromats. *Am J Hum Genet* **53**(5): 987-1000.

Nathans, J., Merbs, S. L., Sung, C.-H., Weitz, C. J. and Wang, Y. (1992). Molecular genetics of human visual pigments. *Annual Review of Genetics* **26**: 403-424.

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**(4747): 203-210.

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**(4747): 193-202.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**(3): 443-53.

Neitz, J. (1986). Variations in color matching among humans with normal color vision. University of California, Santa Barbara.

Neitz, J. and Jacobs, J. H. (1986). Polymorphism of the long-wavelength cone in normal human colour vision. *Nature* **323**: 623-625.

Neitz, J. and Neitz, M. (1994). Color Vision Defects. *Molecular Genetics of Inherited Eye Disorders*. Chur, Harwood Academic Publishers.

Neitz, J., Neitz, M. and Jacobs, G. H. (1993). More than three different cone pigments among people with normal color vision. *Vision Research* **33**(1): 117-122.

Neitz, J., Neitz, M. and Jacobs, G. H. (1989). Analysis of fusion gene and encoded photopigment of colour-blind humans. *Nature* **342**(6250): 679-682.

Neitz, M. and Neitz, J. (1995). Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* **267**: 1013-1016.

Neitz, M., Neitz, J. and Jacobs, G. H. (1991). Spectral tuning of pigments underlying red-green color vision. *Science* **252**(5008): 971-974.

Neitz, M., Neitz, J. and Jacobs, G. H. (1995). Genetic basis of photopigment variations in human dichromats. *Vision Research* **35**(15): 2095-2103.

Nowak, R. M. (1991). *Walker's Mammals of the World*. Baltimore, The John Hopkins University Press.

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**(48): 11367-11372.

Oseroff, A. R. and Callender, R. H. (1974). Resonance Raman spectroscopy of rhodopsin in retinal disk membranes. *Biochemistry* **30**: 11367-11372.

Ovchinnikov, I., Abdulaev, N. G., Feigina, M., Artamonov, I. D. and Bogachuk, A. S. (1983). [Visual rhodopsin. III. Complete amino acid sequence and topography in a membrane]. *Bioorg Khim* **9**(10): 1331-1340.

Ovchinnikov, Y. (1982). Rhodopsin and bacteriorhodopsin: structure-function relationships. *FEBS letters* **148**(2): 179-191.

Pääbo, S. (1985). Molecular Cloning of Ancient Egyptian Mummy DNA. *Nature* **314**: 644-645.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A* **86**(6): 1939-43.

Pääbo, S. (1993). Ancient DNA. *Sci Am* **269**(5): 86-92.

Pääbo, S., Gifford, J. A. and Wilson, A. C. (1988). Mitochondrial DNA sequences from a 7000-year-old brain. *Nucleic Acids Research* **16**: 9775-9787.

Partridge, J. C. (1986). Microspectrophotometry of vertebrate photoreceptors. University of Bristol.

Partridge, J. C. (1989). The visual ecology of avian cone oil droplets. *J. Comp. Physiol. A* **165**: 415-426.

Partridge, J. C. and DeGrip, W. J. (1991). A new template for thodopsin (vitamin A1 based) visual pigment. *Vision Research* **31**: 619-630.

Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence comparisons. *Proc Natl Acad Sci USA*. **85**: 2444-2448.

Peng, Y.-W., Robishaw, J. D., Levine, M. A. and Yau, K.-W. (1992). Retinal reds and cones have distinct G protein b and g subunits. *Proceedings of the National Academy of Sciences USA* **89**: 10882-10886.

Perichon, B., Ragusa, A., Lapoumeroulie, C., Romand, A., Moi, P., Ikuta, T., Labie, D., Elion, J. and Krishnamoorthy, R. (1993). Inter-ethnic polymorphism of the beta-globin gene locus control region (LCR) in sickle-cell anemia patients. *Hum Genet* **91**(5): 464-468.

Piantanida, T. P. and Gille, J. (1992). Methodology-specific Rayleigh-match distributions. *Vision Research* **32**(12): 2375-2377.

Piantanida, T. P. and Sperling, H. G. (1973a). Isolation of a third chromatic mechanism in the deuteranomalous observer. *Vision Research* **13**:2049-2058.

Piantanida, T. P. and Sperling, H. G. (1973b). Isolation of a third chromatic mechanism in the protanomalous observer. *Vision Research* **13**:2033-2047.

Porter, C. A., Sampaio, I., Schneider, H., Schneider, M. P., Czelusniak, J. and Goodman, M. (1995). Evidence on primate phylogeny from epsilon-globin gene sequences and flanking regions. *J Mol Evol* **40**(1): 30-55.

Post, R. A. (1962). Population differences in red and green colour vision deficiencies: a review and a query on selection relaxation. *Eugenics Quart.* **9**: 131-146.

Poulat, F. and Berta, P. (1991). Determination of sex and Y chromosome. *Ann Endocrinol* **52**(6): 410-414.

Prestridge, D.S. (1991). SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. *Comput Appl Biosci.* **7**(2): 203-206.

Pugh, B. F. and Tjian, R. (1990). Mechanism of transcriptional activation by Sp1: Evidence for coactivators. *Cell.* **61**:1187-1197

Pugh, B. F. and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev* **5**(11): 1935-45.

Rayer, B., Naynert, M. and Stieve, H. (1990). Phototransduction: different mechanisms in vertebrates and invertebrates. *J Photochem Photobiol B* **7**(2-4): 107-48.

Rayleigh, L. (1881). Experiments on Colour. *Nature* **Nov, 17**: 64-66.

Reitner, A., Sharpe, L. T. and Zrenner, E. (1991). Is colour vision possible with only rods and blue-sensitive cones? *Nature* **352**(6338): 798-800.

Reyniers, E., Vanthienen, M. N., Meire, F., Boulle, K., Devries, K., Kestelijn, P. and willems, P. J. (1995). Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* : 323-328.

Rogers, J. (1994). Levels of the genealogical hierarchy and the problem of hominoid phylogeny. *American Journal of Physical Anthropology* **94**: 81-88.

Rohlich, P., van Veen, T. and Szel, A. (1994). Two different visual pigments in one retinal cone cell. *Neuron* **13**(5): 1159-66.

Rushton, W. (1963). A cone pigment in the protanope. *J. Physiol* **168**: 345-359.

Rushton, W. A. H., Powell, D. S. and White, K. D. (1973). Pigments in anomalous trichromats. *Vision Research* **13**: 2017-2031.

Ruvolo, M. (1994). Molecular evolutionary processes and conflicting gene trees: the hominoid case. *American Journal of Physical Anthropology* **94**: 89-113.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985). Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.

Saitou, N. and Nei, M. (1987). The Neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**(4): 406-425.

Sakmar, T. P., Franke, R. R. and Khorana, H. G. (1989). Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc Natl Acad Sci U S A* **86**(21): 8309-8313.

Sambrook, T., Fritsh, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual.* **1-3**, CSH Lab Press USA.

Sanger, F., Nicklen, S.T. and Coulsen, A.R. (1977). DNA sequencing with chain termination inhibitors. *Proc Natl Acad Sci USA.* **74**: 5463.

Schertler, G. X., Villa, C. and Henderson, R. (1993). Projection structure of rhodopsin. *Nature* **362**: 770-772.

Shen, M. R., Batzer, M. A. and Deininger, P. L. (1991). Evolution of the master Alu gene(s). *J Mol Evol* **33**(4): 311-320.

Shyue, S. K., Hewett-Emmett, D., Sperling, H. G., Hunt, D. M., Bowmaker, J. K., Mollon, J. D. and Li, W. H. (1995). Adaptive evolution of color vision genes in higher primates. *Science* **269**(5228): 1265-1267.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments seperated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-518.

Stemmer, W. and Morris, S. (1992). Enzymatic inverse PCR: a restriction site independent, single-fragment method for high-efficiency, site-directed mutagenesis. *BioTechniques* **13**(2): 217-220.

Stiles, W. S. and Burch, J. M. (1959). N.P.L. colour-matching investigation: final report (1958). *Optica Acta* **6**: 1-26.

Stryer, L. (1988). *Biochemistry.* New York, Freeman.

Szél, A., Röhlich, P., Caffé, A. R., Juliusson, B., Aguirre, G. and Veen, T. V. (1992). Unique topographic separation of two spectral classes of cones in the mouse retina. *The Journal of Comparative Neurology* **325**: 327-342.

Szél, A., Veen, T. V. and Rohlich, P. (1994). Retinal cone differentiation. *Nature* **370**: 336.

Takao, M., Yasui, A. and Tokunaga, F. (1988). Isolation and sequence determination of the chicken rhodopsin gene. *Vision Res* **28**(4): 471-80.

Tjian, R. and Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**(1): 5-8.

Tokunaga, F., Iwasa, T., Miyagishi, M. and Kayada, S. (1990). Cloning of cDNA and amino acid sequence of one of chicken cone visual pigments. *Biochem Biophys Res Commun* **173**(3): 1212-1217.

Tovée, M. J. (1994). The molecular genetics and evolution of primate colour vision. *Trends Neurosci* **17**(1): 30-37.

Tovée, M. J., Bowmaker, J. K. and Mollon, J. D. (1992). The relationship between cone pigments and behavioural sensitivity in a New World monkey (*Callithrix jacchus jacchus*). *Vision Research* **32**(5): 867-878.

Travis, D. S., Bowmaker, J. K. and Mollon, J. D. (1988). Polymorphism of visual pigments in a callitrichid monkey. *Vision Research* **28**: 481-490.

Trumpp-Kallmeyer, S., Hoflack, J., Bruinvels, A. and Hibert, M. (1992). Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors. *J Med Chem* **35**(19): 3448-62.

Tuan, D., Soloman, W., Li, Q. and London, I. M. (1985). The b-globin gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA* **82**: 6384-6388.

Turner, D. L. and Cepko, C. L. (1987). A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**(6126): 131-6.

Turner, D. L., Snyder, E. Y. and Cepko, C. L. (1990). Lineage-independent determination of cell type in the embryonic mouse retina. *Neuron* **4**(6): 833-45.

van der Walt, J. J. A. (1977). *Pelargoniums of southern Africa*. Cape Town, Purnell.

Vogt, K. and Kirschfeld, K. (1984). Chemical identity of the chromophores of fly visual pigment. *Naturewissenchaften* **71**: 211-213.

Vollrath, D., Nathans, J. and Davis, R. W. (1988). Tandem array of human visual pigment genes at Xq28. *Science* **240**(4859): 1669-1672.

von-Schantz, M., Szel, A., van-Veen, T. and Farber, D. B. (1994). Expression of phototransduction cascade genes in the ground squirrel retina. *Invest Ophthalmol Vis Sci* **35**(5): 2558-66.

Wahls, W. P., Wallace, L. J. and Moore, P. D. (1990). Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell* **60**(1): 95-103.

Wang, J. K., McDowell, J. H. and Hargrave, P. A. (1980). Site of attachment of 11-cis retinal in bovine rhodopsin. *Biochemistry* **19**: 5111-5117.

Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., Gearhart, J. and Nathans, J. (1992). A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**(3): 429-40.

Wang, Z., Asenjo, A. B. and Oprian, D. D. (1993). Identification of the Cl(-)-binding site in the human red and green color vision pigments. *Biochemistry* **32**(9): 2125-30.

Wartmann, E. (1846). *Taylor's Scientific Memoirs* **4**: 156-187.

Watanabe, T. and Raff, M. C. (1990). Rod photoreceptor development in vitro: intrinsic properties of proliferating neuroepithelial cells change as development proceeds in the rat retina. *Neuron* **4**(3): 461-7.

Weitz, C. J., Miyake, Y., Shinzato, K., Montag, E., Zrenner, E., Went, L. N. and Nathans, J. (1992). Human tritanopia associated with two amino acid substitutions in the blue-sensitive opsin. *Am J Hum Genet* **50**(3): 498-507.

West, R. W. and Dowling, J. E. (1975). Anatomical evidence for cone and rod-like receptors in the gray squirrel. ground squirrel, and prairie dog retinas. *J. Comp. Neurol.* **159**(4): 439-460.

Whewhell, W. (1841). *Eleventh Meeting of the British Association for the Advancement of Science*. Athenaeum,

Wikler, K. C. and Rakic, P. (1990). Distribution of photoreceptor subtypes in the retina of diurnal and nocturnal primates. *The Journal of Neuroscience* **10**(10): 3390-3401.

Willard, C., Nguyen, H. T. and Schmid, C. W. (1987). Existence of at least three distinct Alu subfamilies. *J Mol Evol* **26**(3): 180-186.

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**(6): 2039-2045.

Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A. G. and Deeb, S. S. (1993). Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. *Hum Mol Genet* **2**(9): 1413-21.

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 U S A* **89**(20): 9710-4.

Winderickx, J., Lindsey, D. T., Sanocki, E., Teller, D. Y., Motulsky, A. G. and Deeb, S. S. (1992c). Polymorphism in red photopigment underlies variation in colour matching. *Nature* **356**(6368): 431-433.

Wong, D.M., Weinstock, P.H. and Wetmur, J.G. (1991). Branch capture reactions: displacers derived from asymmetric PCR. *Nucleic Acids Res.* **19**(9): 2251-9

Wright, W. D. (1967). *The Rays are not Coloured.* London, Hilger.

Yau, K.-W. (1994). Phototransduction mechanism in retinal rods and cones. *Investigative Ophthalmology and Visual Science* **35**(1): 9-32.

Yokoyama, S. (1995). Amino acid replacements and wavelength absorption of visual pigments in vertebrates. *Mol. Biol. Evol.* **12**(1): 53-61.

Yokoyama, R. and Yokoyama, S. (1990). Convergent evolution of the red- and green-like visual pigment genes in fish, Astyanax fasciatus, and human. *Proc Natl Acad Sci U S A* **87**(23): 9315-9318.

Yoshizawa, T. and Kuwata, O. (1991). Iodopsin, a red-sensitive cone visual pigment in the chicken retina. *Photochem Photobiol* **54**(6): 1061-70.

Young, R. W. (1971). The renewal of rod and cone outer segments in the rhesus monkey. *J. Cell. Biol.* **49**: 303-318.

Young, T. (1807). *A course of lectures on natural philosophy and mechanical arts.* London, J. Johnson.

Yu, J., Bock, J. H., Slightom, J. L. and Villeponteau, B. (1994). A 5' beta-globin matrix-attachment region and the polyoma enhancer together confer position-independent transcription. *Gene* **139**(2): 139-145.

Zhang, Y. N., El-Sayed, M. A., Bonet, M. L., Lana, J., Chang, M., Ni, B. and Needleman, R. (1993). Effects of genetic replacement of charged and H-bonding residues in the retinal pocket on Ca2+ binding to deionized bacteriorhodopsin. *Proceedings of the National Academy of Sciences USA* **90**: 1445-1449.

Zhukovsky, E. A. and Oprian, D. D. (1989). Effect of carboxylic acid side chains on the absorption maximum of visual pigments. *Science* **246**(4932): 928-930.

Zhukovsky, E. A., Robinson, P. R. and Oprian, D. D. (1992). Changing the location of the Schiff base counterion in rhodopsin. *Biochemistry* **31**(42): 10400-5.

Zietkiewicz, E., Richer, C., Makalowski, W., Jurka, J. and Labuda, D. (1994). A young Alu subfamily amplified independently in human and African great apes lineages. *Nucleic Acids Res* **22**(25): 5608-5612.

Zvyaga, T. A., Fahmy, K. and Sakmar, T. P. (1994). Characterization of rhodopsin-transducin interaction: a mutant rhodopsin photoproduct with a protonated Schiff base activates transducin. *Biochemistry* **33**(32): 9753-61.

Additional References

Bunt-Milam, A. H. and Saari, J. C. (1983). Immunocytochemical localization of two retinoid-binding proteins in vertebrate retina. *J Cell Biol* **3**(97): 703-712.

Chen J., Tucker C. L., Woodford B., Szel A., Lem J., Gianella-Borradori A., Simon M. I. and Bogenmann E. (1994). The human blue opsin promoter directs transgene expression in short-wave cones and bipolar cells in the mouse retina. *Proc Natl Acad Sci U S A* **91**: 2611-5.

DesJardin L. E. and Hauswirth W. W. (1996). Developmentally important DNA elements within the bovine opsin upstream region. *Invest Ophthalmol Vis Sci* **37**: 154-165.

Feinberg A. P. and Vogelstein B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132**: 6-13.

Feinberg A. P. and Vogelstein B. (1984). "A technique for radiolabeling DNA restrictionendonuclease fragments to high specific activity". Addendum. *Anal. Biochem.* **137**: 266-267.

Higgins, D. G., Bleasby and Fuchs. (1991). Clustral V. *Comput. Applic. BioSci.* **8**: 189-191

Hurvich, L. M. (1981). *Color Vision.* Sunderland: Sinaver Ass. Inc.

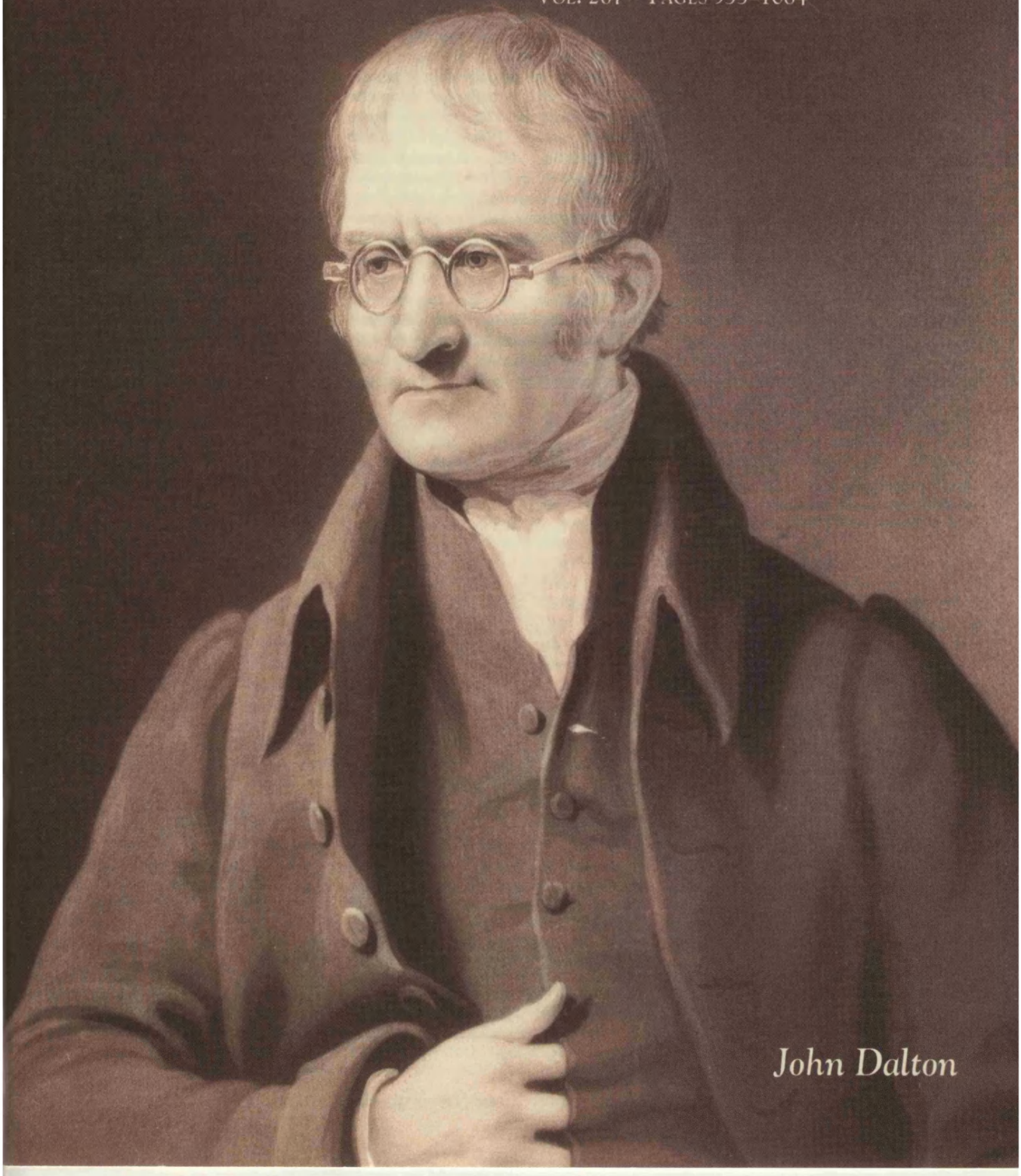
Mollon, J. D. (1993). Mixing genes and mixing colors. *Current Biology* **3**(2): 82-85.

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

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

SCIENCE

17 FEBRUARY 1995 $7.00
VOL. 267 • PAGES 933–1064



John Dalton

The Chemistry of John Dalton's Color Blindness

David M. Hunt,* Kanwaljit S. Dulai, James K. Bowmaker, John D. Mollon

John Dalton described his own color blindness in 1794. In common with his brother, he confused scarlet with green and pink with blue. Dalton supposed that his vitreous humor was tinted blue, selectively absorbing longer wavelengths. He instructed that his eyes should be examined after his death, but the examination revealed that the humors were perfectly clear. In experiments presented here, DNA extracted from his preserved eye tissue showed that Dalton was a deuteranope, lacking the middlewave photopigment of the retina. This diagnosis is shown to be compatible with the historical record of his phenotype, although it contradicts Thomas Young's belief that Dalton was a protanope.

Two centuries ago the chemist John Dalton analyzed his own color blindness in his first lecture to the Manchester Literary and Philosophical Society (*1*), and "daltonism" has since become the name for the condition in many languages (*2*). Dalton judged red sealing wax to be a good match for the outer face of a laurel leaf, and a crimson ribbon matched the color that others called "mud" (*1*, 3). In the solar spectrum he saw only two main hues, one of which corresponded to the normal observer's red, orange, yellow, and green, whereas the second corresponded to blue and violet. He was particularly surprised to observe that the pink flowers of a cranesbill (*Geranium zonale*) (*4*), which appeared "sky-blue" to him by daylight, looked "very near yellow, but with a tincture of red" (*3*) by candlelight (Fig. 1). Of his immediate acquaintances, only his brother shared his astonishment at this failure of color constancy (*3*).

In an explanation of his color deficiency, Dalton proposed that the vitreous humor of his eye was tinted blue, selectively absorbing longer wavelengths. He instructed that after his death his eyes should be dissected to confirm his hypothesis. He died at age 78 on 27 July 1844, and on the following day an autopsy was done by his medical attendant, Joseph Ransome (*5*, *6*). Ransome collected the humors of one eye into watch glasses and found them to be "perfectly pellucid," the lens itself exhibiting the yellowness expected in someone of Dalton's age (*6*). He shrewdly left the second eye almost intact, slicing off the posterior pole and noting that scarlet and green objects were not distorted in color when seen through the eye. Thus, Ransome found no support for Dalton's hypothesis that color blindness was due to a preretinal filter. Ransome did not discard the eyes but stored them only in air, and fragments of them have survived to this day (Fig. 2). Originally in the possession of Dalton Hall, the eyes passed into the keeping of the Manchester Literary and Philosophical Society (*7*) who gave us permission to take small samples for a reexamination of Dalton's color blindness by DNA analysis.

At the time of Dalton's death, the most popular alternative to his own hypothesis was that daltonism arose from a cerebral defect. In his autopsy report, Ransome recorded (*6*) a "deficient development" of the phrenological organ of color (which corresponded to one of the convolutions of the frontal lobe). Today we know that hereditary color blindness seldom arises from either a preretinal filter or a central defect, but rather is due to the absence or the alteration of one of the photosensitive pigments of the retina. Interestingly, an explanation of this kind had already been proposed by 1794, for although we owe to Dalton the first systematic analysis of the confusions that occur in color blindness, the actual existence of the condition was well enough known for King George III to have raised it in conversation with Fanny Burney at court in 1785 (*8*). A little earlier a German general science magazine reported the theory advanced by a shadowy figure called Giros von Gentilly (*9*) who held that there were three types of "molecule" or "membrane" in the retina, corresponding to three kinds of light. Color blindness arose when one or two of the three types of molecule was either paralyzed or constitutively overactive (*10*).

Thomas Young, author of the wave theory of light and the decipherer of the Rosetta stone, may have heard of von Gentilly's ideas while a medical student in Goettingen. Young combined a three-receptor theory of vision with a realization that the physical correlate of hue is a continuous variable. To explain Dalton's color blindness, Young postulated "the absence or paralysis of those fibers of the retina, which are calculated to perceive red" (*11*). In supposing that it was the red-sensitive fibers that were affected, Young was critically influenced by one phrase in Dalton's description of the solar spectrum: "that part of the image which others call red, appears to me little more than a shade, or defect of light" (*1*).

After much debate, the three-receptor theory of von Gentilly and of Young is today generally accepted. We now know that daylight vision depends on three types of photopigments, segregated into three classes of cone cell, with peak wavelength sensitivities (λ_{max}) near 420 nm (shortwave), 530 nm (middlewave, MW), and 560 nm (longwave, LW) (*12*, *13*). Color vision depends on a neural comparison of the quantum catches of the different cone classes. The photopigments are embedded in the multiply enfolded membranes of the cone outer segments, and each photopigment consists of a protein moiety bound to retinal, a derivative of vitamin A1. The proteins are members of the superfamily of

Fig. 1. *Pelargonium zonale*. A specimen from stock collected in South Africa that almost certainly corresponds to Dalton's "*Geranium zonale*" (*4*).

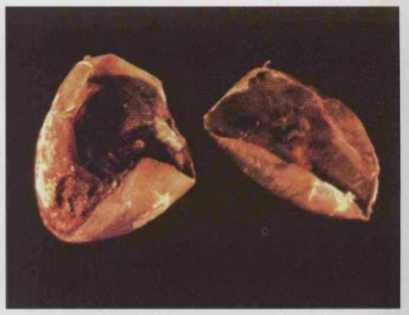


Fig. 2. The preserved eye fragments of John Dalton, photographed in 1982.

D. M. Hunt and K. S. Dulai are in the Department of Molecular Genetics, Institute of Ophthalmology, University of London, London, EC1V 9EL, UK. J. K. Bowmaker is in the Department of Visual Science, Institute of Ophthalmology, University of London, London, EC1V 9EL, UK. J. D. Mollon is in the Department of Experimental Psychology, University of Cambridge, Cambridge, CB2 3EB, UK.

*To whom correspondence should be addressed.

G protein–coupled receptor molecules, or heptahelicals (G proteins are guanosine triphosphate–binding proteins); they exhibit seven transmembrane helices, which are linked by intra- and extracellular loops and form a palisade around the chromophore retinal (*14*).

Modern spectrophotometric measurements have confirmed that the dichromatic form of color blindness commonly arises from the lack of either the LW or the MW photopigment (Fig. 3) [the former condition is termed protanopia and the latter deuteranopia (*12, 13*)]. Molecular biology has shown that most of the inherited forms of color deficiency are associated with changes in the genes on the X chromosome that code for the protein moieties of the LW and MW visual pigments (*15, 16*). The MW and LW genes are both composed of six coding regions or exons, and such changes are thought to arise from illegitimate pairing between these highly homologous genes, followed by crossing-over (*16*).

Although there are 15 amino acid differences between MW and LW opsins (*15*), the greater part of the spectral shift in sensitivity between MW and LW visual pigments is the result of substitutions at sites 180, 277, and 285 (*17–20*), with five other sites having smaller effects (*21*). Site 180, however, is polymorphic (*22, 23*) in both the MW and LW opsin genes: The substitution of serine for alanine in an otherwise MW sequence increases the λ_{max} by about 2 nm, and the substitution of alanine for serine in an otherwise LW sequence decreases the λ_{max} by 4 to 7 nm (*20, 21*). We have used sites 277 and 285 (both coded by exon 5), together with gene-specific differences in exon 4, to determine whether Dalton possessed a MW or a LW opsin gene or both.

We took several small samples of tissue from Dalton's preserved eye from regions tentatively identified as peripheral retina, and these samples were used for the amplification of opsin genes by the polymerase chain reaction (PCR) (*24*). Two different primer sets were used (*25*): One set is gene-specific (*26*) and amplifies either from MW or from LW exons, whereas the second set amplifies from both genes (*18, 27, 28*). The gene-specific set comprises a primer pair (R4+, R5–) that is specific to the LW exons 4 and 5 and another (G4+, G5–) that is specific to the MW exons 4 and 5 (Fig. 4). These primer pairs produce a fragment of about 1.7 kb. In other combinations (R4+, G5– and G4+, R5–), they generate fragments from any hybrid gene that contains a MW exon 4 and LW exon 5 or vice versa. Only primer pair R4+, R5– generated a fragment of the correct size from Dalton's DNA (Fig. 5A), indicating that a LW but no MW opsin gene is present. Southern (DNA) blot analysis and probing of the PCR products with an exon 5 probe confirmed that only this set of primers amplified an opsin gene fragment (Fig. 5B). The specificity of these primers was confirmed with a number of test DNA samples (Fig. 5C). The DNA from a normal observer yielded fragments of the correct size only when the R4+, R5– and G4+, G5– combinations were used. DNA from a known anomalous trichromat produced, in addition to these fragments, a fragment with the R4+, G5– combination, indicating the presence of a hybrid gene. And lastly, only the R4+, R5– primer pair amplified a fragment from the DNA of a known deuteranope. The identity of the fragment amplified from Dalton's DNA by the R4+, R5– primer pair was then confirmed by cloning and sequencing (*29*). Partial sequences from 10 separate clones (Fig. 6) show that the fragment was amplified from an opsin gene that coded for a LW visual pigment.

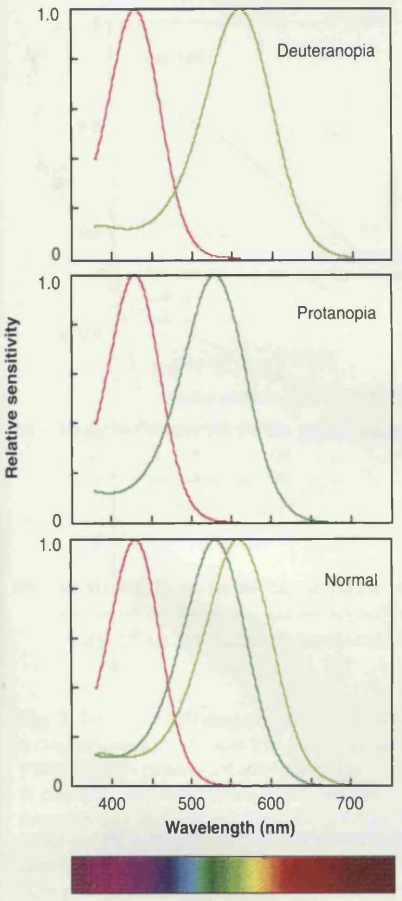


Fig. 3. The spectral sensitivities of the three types of cone photoreceptors in the normal eye (bottom panel), protanopia (middle panel), and deuteranopia (top panel).

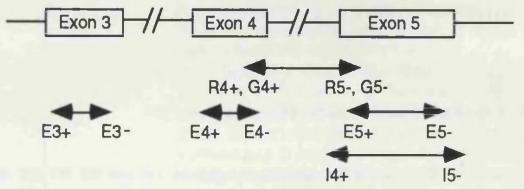


Fig. 4. Primer pairs (*25*) used for the amplification of opsin gene fragments. Only three of the six exons that make up the LW and MW opsin genes are shown. Primer pairs R4+, R5– and G4+, G5– are specific for the LW and MW opsin genes, respectively (*26*), and amplify from base 721 in exon 4, through intron 4, to base 868 in exon 5. Primer pairs E3+, E3–; E4+, E4–; E5+, E5– (*19, 28*); and I4+, I5– (*28*) are not gene-specific and amplify both the LW and MW genes. The pair E3+, E3– amplifies from base 480 to base 596 in exon 3, E4+, E4– amplifies from base 650 to base 761 in exon 4, E5+, E5– amplifies from base 830 to base 983 in exon 5, and I4+, I5– amplifies from intron 4, across exon 5, and into intron 5.

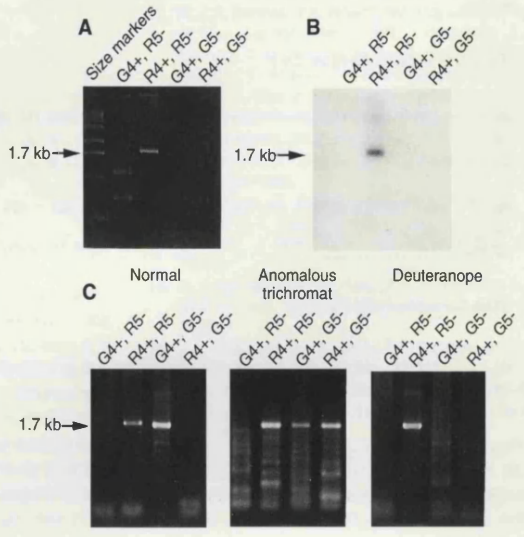


Fig. 5. Amplification products obtained from different combinations of the gene-specific exon 4 (R4+ or G4+) and exon 5 (R5+ or G5–) primers. (**A**) Amplification products from Dalton's DNA. The size marker bands are 0.506, 1.018, 1.636, 2.036, and 3.054 kb. (**B**) Southern blot of gel in (A). The membrane was hybridized overnight at 65°C with a ^{32}P-labeled exon 5 probe amplified from normal genomic DNA with the nonspecific I4+, I5– primers (*25*). The membrane was washed with 0.5× standard saline citrate and 0.1% SDS for 30 min at 65°C. (**C**) Amplification products from a normal male observer, a known anomalous male trichromat, and a known male deuteranope.

We sought further evidence for the presence of only a LW opsin sequence in Dalton by using nonspecific primer pairs (*18*, *27*, *28*) that do not differentiate between MW and LW genes. These primers also provide information on codon 180 in exon 3, codons 230 and 233 in exon 4, and codons 277 and 285 in exon 5. Four pairs were used, three that amplify within exons 3, 4, and 5, respectively, and a fourth pair that amplifies from intron 4, across exon 5, and into intron 5 (Fig. 4). For each primer pair, the resulting amplified fragment was cloned and sequenced. In all cases, only a single sequence was amplified (Fig. 6): For exons 4 and 5, all of the clones were identical to the sequence of the LW opsin gene reported by Nathans and his colleagues (*15*), whereas all of the exon 3 clones coded for Ala^{180}, the sequence being identical to the second most common variant of this highly polymorphic region of the LW opsin gene (*23*). A consistent result was therefore obtained from 27 different clones that contain fragments generated with a range of nonspecific primers that amplify from both intron and exon regions of the MW and LW opsin genes: Only a single LW opsin sequence is present in Dalton's DNA.

Analyses of old DNA by PCR are bedeviled by contamination (*30*). For the following reasons, however, we do not believe that our result is artifactual. We deliberately took the samples from the interior of the eye, which is unlikely to have been handled. Samples were removed with a sterile scalpel blade and placed immediately into sterile 1.5-ml eppendorf tubes. At all times, sterile surgical gloves were worn. New pipettes and reagents were used in all subsequent procedures, and all experiments were set up in a room that had not been previously used for work on visual pigments. Each PCR used a different sample of eye tissue yet gave the same overall result. And there is a further unusual and salient argument against contamination in the present case: Our result depends on the absence, rather than the presence, of a gene—a gene whose LW fellow is present. Only contamination by the DNA from another deuteranope would have given the same result. However, as far as we are aware, no workers in our laboratory have any defect in color vision and certainly none are dichromats, and no samples of DNA from other dichromats were present in our laboratory until after the analysis of Dalton's DNA was complete. The deuteranope that we used for test purposes was not known to us until after all of the PCR amplifications on Dalton's tissue had been carried out.

Therefore, from our data we can confirm that Dalton was indeed a dichromat, but contrary to previous interpretations, he was a deuteranope with a single LW opsin gene that coded for a visual pigment with the shorter spectral sensitivity of the two common LW variants. Long tradition holds that Dalton was a protanope (*11*, *31*). Can our result be reconciled with the historical evidence for his phenotype?

The classical sign of protanopia is the foreshortening of the red end of the spectrum because of the absence of the LW receptor. However, when the physicists Sir John Herschel and Sir David Brewster each questioned Dalton directly, they both reported that he did not see the spectrum as foreshortened at long wavelengths (*8*, *32*). Similarly, Richard Taylor, the scientific

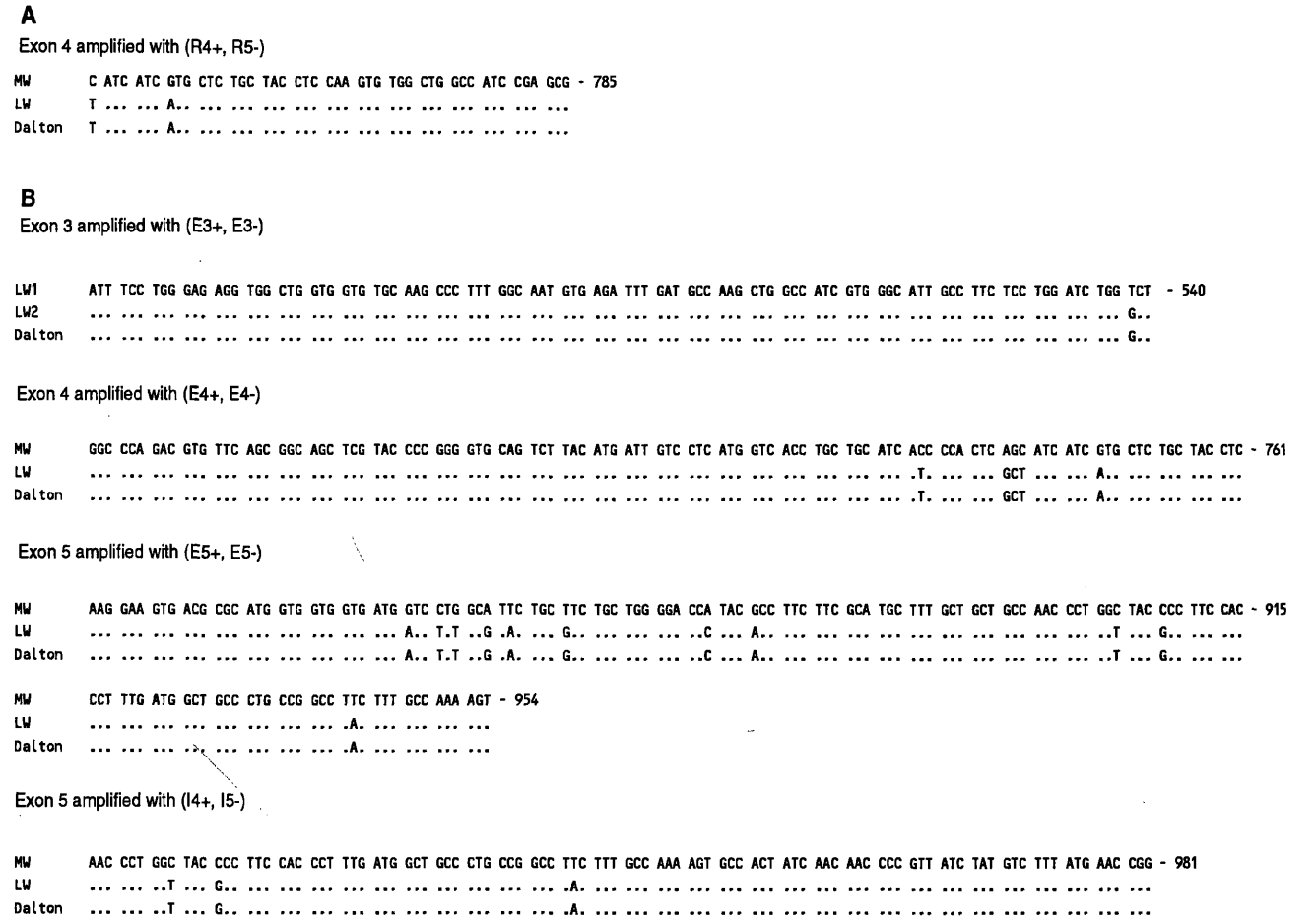


Fig. 6. Sequence of amplified fragments from (**A**) gene-specific primers and (**B**) nonspecific primers compared with those of the MW and LW genes originally reported by Nathans *et al.* (*15*). Five clones of exon 3 generated with the E3+, E3– primers, five clones of exon 4 generated with the E4+, E4– primers, 15 clones of exon 5 generated with the E4+, E4– primers, and two clones of exon 5 generated with the I4+, I5– primers (*25*) were sequenced. LW1 and LW2 are, respectively, the first and second most common variants of LW exon 3 (*23*).

publisher, remarked in a footnote that "Dr. Dalton has never stated that the spectrum he saw was shorter than the spectrum seen by others" (*33*). But what of Dalton's comment that red appears as "little more than a shade, or defect of light" (*1*)? For a deuteranope, the red part of a spectrum will look dim because the regions that look yellow, orange, and red to normal observers are for him all of the same hue, but the red is of lower luminosity than the yellow and orange regions to which it is juxtaposed. Moreover, the red region does not offer to the deuteranope the Farbenglut, the extra vividness of saturated colors, which derives from the LW-MW opponent signal and which accentuates the brightness of long wavelengths in the case of the normal observer (*34*).

Furthermore, the color confusions reported by Dalton are compatible with deuteranopia. Using a Photo Research PR650 spectroradiometer and a Macbeth daylight lamp, we measured samples of 18th-century sealing wax and of the leaves of *Prunus laurocerasus*, which Dalton saw as being of similar hue. The measured chromaticities of these samples are plotted in the Commission Internationale de l'Éclairage (CIE) diagram of Fig. 7: they fall on a confusion line that is at least as compatible with deuteranopia as with protanopia. Dalton reported as blue a number of wild flowers that appear pink to the normal eye, and we measured the chromaticities of several of these flora when illuminated by a daylight lamp. All of them, including the pink of *Pelargonium zonale* and even the crimson pink of the male red campion (*Lychnis dioica*), lie on the blue side of a line passing through the deutan confusion point and the chromaticity of our daylight source, a line that represents the set of chromaticities that appear neutral to a deuteranope. The measurements shown in Fig. 7 do not unequivocally require a diagnosis of deuteranopia, but they do deal with the notion that has corrupted many earlier discussions of Dalton's condition—the false notion that it was because he lacked red cones that he did not see the redness in pink. The fact that pink looks so reddish to the normal eye only serves to remind us of the danger of attaching color names to receptors when we discuss color vision.

But what of the failure of color constancy that prompted Dalton's self-examination? We showed a specimen of *P. zonale* (Fig. 1) to the deuteranope whose DNA was analyzed as a test for the gene-specific primer pairs (Fig. 5C). In northern daylight he described the color as similar to that of the sky, but slightly desaturated. In candlelight a few minutes later, he described the color as red.

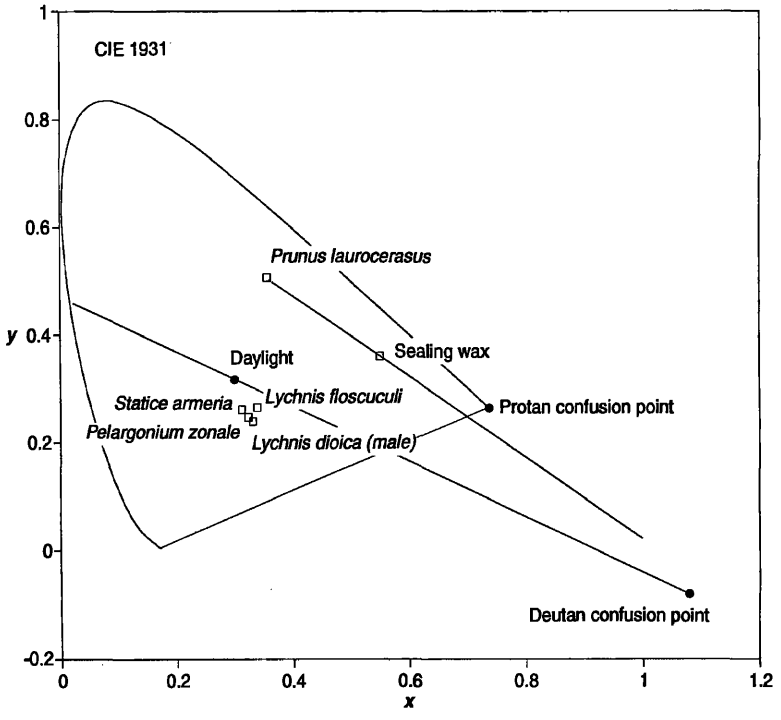


Fig. 7. The CIE (1931) diagram (*35*) showing the chromaticities of objects referred to by Dalton. In such a diagram, subsets of colors that are confusable by a protanope lie on lines that converge at $x = 0.747$, $y = 0.253$ (the protan confusion point). The corresponding deutan confusion point is less well specified, in part because of properties of the diagram and in part because of the genetic heterogeneity of the longwave pigment; we plot the value given by Wyszecki and Stiles (*35*). Sealing wax and the face of a laurel leaf (*P. laurocerasus*) give a confusion line that is at least as compatible with deuteranopia as it is with protanopia. Also shown are several flowers that appeared blue to Dalton: thrift (*Statice armeria*, now *Armeria maritima*), ragged robin (*Lychnis floscuculi*), the red campion (*Lychnis dioica*), and the pink geranium (*Geranium zonale*, now renamed *Pelargonium zonale*). Although these flora all appear pink to the normal eye in daylight, they plot to the left of the deuteranopic confusion line that passes through the chromaticity of our standard daylight source (Macbeth easel lamp), so they would be expected to look bluish not only to a protanope but also to a deuteranope.

REFERENCES AND NOTES

1. The lecture by Dalton was delivered on 31 October 1794. It was published in *Mem. Manch. Lit. Philos. Soc.* **5**, 28 (1798) and reprinted in the *Edinb. J. Sci.* **5**, 88 (1831).
2. French, Spanish, and Russian are among the languages that have the term daltonism. The term *daltonien* was used in print as early as 1827 by P. Prevost [*Bibl. Univers. Sci. B.-Lett. Arts Sci. Arts* **35**, 320 (1827)].
3. Letter from Dalton to Elihu Robinson, dated "2^{d} mo. 20th, 1794," reproduced separately and in H. Lonsdale, *The Worthies of Cumberland*, vol. 5, *John Dalton* (Routledge, London, 1874).
4. The "horsehoe cranesbill" or *Geranium zonale* is illustrated by Thomas Martyn [*Thirty-Eight Plates Intended to Illustrate Linnaeus's System* (J. White, London, 1799)], and we can be fairly certain that the species to which Dalton refers is the one now known as *Pelargonium zonale*, a native plant of South Africa imported to Britain during the 18th century.
5. G. Wilson, *Br. Q. Rev.* **1**, 157 (1845).
6. W. C. Henry, *Memoirs of the Life and Scientific Researches of John Dalton* (Cavendish Society, London, 1854).
7. E. M. Brockbank, *John Dalton* (Falkner, Manchester, 1929).
8. C. Barrett, *Diary and Letters of Madame D'Arblay* (Macmillan, London, 1904). George III told the future Madame D'Arblay, "There are people who have no eye for difference of colour. The Duke of Marlborough actually cannot tell scarlet from green! . . . I do not find though that this defect runs in his family, for Lady Di Beauclerk draws very finely."
9. J. H. Voigt, *Mag. Neuste Phys. Naturgesch. (Gotha)* **1**, 57 (1781). The identity of Giros von Gentilly has been discussed in an entertaining essay by G. Walls, *J. Hist. Med.* **11**, 66 (1956). A solution to the mystery is offered by J. D. Mollon, in *The Dictionary of National Biography: Missing Persons*, C. S. Nicholls, Ed. (Oxford Univ. Press, Oxford, 1993).
10. ". . . die Umwirksamkeit der Molecülen . . . deren außerordentliche Beweglichkeit." This may be the earliest suggestion that a hereditary disorder was due to a molecular defect.
11. T. Young, *A Course of Lectures on Natural Philosophy and the Mechanical Arts* (J. Johnson, London, 1807).
12. W. A. H. Rushton, *Nature* **206**, 1087 (1965).
13. H. J. A. Dartnall, J. K. Bowmaker, J. D. Mollon, *Proc. R. Soc. London, Ser. B* **220**, 115 (1983); J. L. Schnapf, T. W. Kraft, D. A. Baylor, *Nature* **325**, 439 (1987); J. D. Mollon, J. K. Bowmaker, H. J. A. Dartnall, A. C. Bird, in *Colour Vision Deficiencies VII*, G. Verriest, Ed. (Junk, The Hague, 1984), pp. 303–310.
14. P. A. Hargrave, *Prog. Retinal Res.* **1**, 1 (1982).
15. J. Nathans, D. Thomas, D. S. Hogness, *Science* **232**, 193 (1986).
16. J. Nathans, T. P. Piantanida, R. L. Eddy, T. B. Shows, D. S. Hogness, *ibid.*, p. 203.
17. M. Neitz, J. Neitz, G. H. Jacobs, *ibid.* **252**, 971 (1991).
18. R. E. Ibbotson, D. M. Hunt, J. K. Bowmaker, J. D. Mollon, *Proc. R. Soc. London Ser. B* **247**, 145 (1992).
19. A. J. Williams, D. M. Hunt, J. K. Bowmaker, J. D. Mollon, *EMBO J.* **11**, 2039 (1992).
20. S. L. Merbs and J. Nathans, *Science* **258**, 464 (1992); *Nature* **356**, 433 (1992); *Photochem. Photobiol.* **58**, 706 (1993).
21. A. B. Asenjo, J. Rim, D. D. Oprian, *Neuron* **12**, 1131 (1994).
22. J. Neitz and G. H. Jacobs, *Vision Res.* **30**, 621 (1990); J. Winderickx *et al.*, *Nature* **356**, 431 (1992).

23. J. Winderickx *et al.*, *Hum. Mol. Genet.* **2**,' 1413 (1993).
24. Genomic DNA can be isolated from many tissues. However, because only eye tissue has survived, this was used for the amplification of opsin genes by PCR. Small pieces of the dry eye tissue of not more than 1 mm^3 were placed in 0.2-ml eppendorf tubes, and 20 μl of GeneReleaser (BioVentures) was added. GeneReleaser serves two purposes: It avoids the need to purify DNA, and it sequesters products that might inhibit polymerase activity. The tube was heated as follows: 65°C for 30 s, 8°C for 30 s, 97°C for 180 s, 8°C for 60 s, 65°C for 180 s, 97°C for 60 s, 65°C for 60 s, and 80°C for 1 hour. The PCR mix was then either added immediately or after storage at 4°C overnight. The PCR mix contained 200 μM each of 2′-deoxyadenosine 5′-triphosphate, deoxycytidine 5′-triphosphate, deoxyguanosine 5′-triphosphate, and deoxythymidine 5′-triphosphate, 0.5 units of *Taq* polymerase, 9 mM $MgCl_2$, 30 μM of each primer, and 2 μl of Perfect Match (Stratagene), all in a final volume of 50 μl. The tube was heated at 94°C for 180 s before 35 cycles of 94°C for 20 s, 62°C for 30 s, and 72°C for 120 s, and a final step of 72°C for 600 s. No products of this first round of PCR could be visualized on an agarose gel. A second round of PCR was then carried out by the addition of 1 μl of the first-round mix to 49 μl of PCR mix prepared as before but preheated to 94°C. The PCR parameters were as for the first round except that the annealing temperature was adjusted as follows: R4+, R5– and G4+, G4– at 62°C; E3+, E3– and E4+, E4– at 56°C; E5+, E5– at 58°C; and I4+, I5– at 68°C. Blank tubes lacking DNA were carried through both first- and second-round PCRs. No amplified fragments were detected in these tubes. The test DNA samples from a normal male observer, an anomalous male trichromat, and a male deuteranope were isolated by conventional methods from blood samples, and only a single PCR was carried out with the parameters of the second-round PCR.
25. The sequence of the two sets of primer oligonucleotides used for the amplification of the opsin gene fragments are as follows. (i) Gene-specific set: R4+, GCTGCATCATCCCACTCGC; G4+, GCTGCATCACCCCACTCAG; R5–, GACGCAGTACGCAAAGATC; and G5–, GAAGCAGAATGCCAGGACC. (ii) Nonspecific set: E3+, TCACAGGTCTCTGGTCTCTGG; E3–, CTCCAACCAAAGATGGGCGG; E4+, CACGGCCTGAAGACTTCATGC; E4–, CGCTCGGATGGCCAGCCACAC; E5+, GAATTCCACCCAGAAGGCAGAG; E5–, GTCGACGGGGTTGTAGATAGTGGC; I4+, ACGTGGAATTCCCTCTCCTCCTCCCCACAAC; and I5–, ACGTGAAGCTTCAGGTGGGGCCATCACTGCA.
26. S. S. Deeb *et al.*, *Am. J. Hum. Genet.* **51**, 687 (1992).
27. K. S. Dulai, J. K. Bowmaker, J. D. Mollon, D. M. Hunt, *Vision Res.* **34**, 2483 (1994).
28. J. Nathans *et al.*, *Am. J. Hum. Genet.* **53**, 987 (1993).
29. Amplified fragments for sequencing were obtained from excised gel bands of the correct size by eluting the fragments from the bands overnight in 50 μl of sterile water and then heating the fragments to 68°C for 5 min. After TA cloning (Invitrogen) into pCRII, the fragments were cycle-sequenced with *Taq* polymerase, dye-tagged dideoxy nucleotides, and either a T7 or Sp6 sequencing primer. The products of the reaction were visualized in an Applied Biosystems Model 373 DNA Sequencer System.
30. E. Hagelberg, I. C. Gray, A. J. Jeffreys, *Nature* **352**, 427 (1991).
31. W. D. Wright, *The Rays Are not Coloured* (Hilger, London, 1967).
32. D. Brewster, *Letters on Natural Magic, Addressed to Sir Walter Scott, Bart* (J. Murray, London, 1842).
33. E. Wartmann, *Taylor's Sci. Mem.* **4**, 156 (1846).
34. A. Kohlrausch, *Pfluegers Arch. Gesamte Physiol. Menschen Tiere* **200**, 216 (1923).
35. G. Wyszecki and W. S. Stiles, *Color Science* (Wiley, New York, 1982).
36. Supported by the Wellcome Trust and the Medical Research Council. We thank the Manchester Literary and Philosophical Society for permission to take samples from Dalton's eye; the Museum of Science and Industry in Manchester for giving us access to their material; H. Key for the provision of a specimen of *P. zonale*; D. Hull, C. James, H. Key, E. Spong, and E. N. Willmer for botanical advice; I. Cannell for photography; and the Director of the Botanic Garden and the Keeper of Manuscripts, Cambridge University, for access to colorimetric samples.

COVER

John Dalton (1766–1844) is celebrated for developing the atomic theory of chemistry, but his first scientific paper to the Manchester Literary and Philosophical Society in 1794 was an account of how his color perceptions differed from those of other people. After 150 years, amplification of DNA from his preserved eye tissue has revealed the molecular basis for his color blindness. See page 984. [Engraving: Dated 1836, reproduced by permission of the British Library]