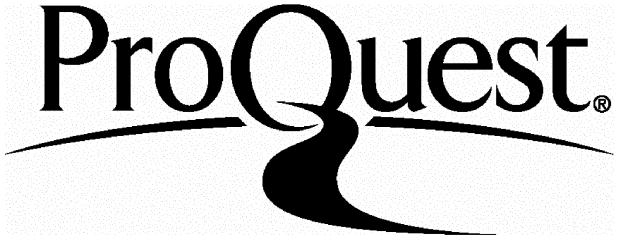
REGULATION, SPECTRAL TUNING AND EVOLUTION OF VISUAL PIGMENT GENES IN FISH

Jill Alison Cowing

Department of Molecular Genetics, Institute of Ophthalmology,
University College London,
University of London

A thesis submitted to the University of London for the degree of Doctor of Philosophy (PhD)

2001



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The opsin genes of mammals have been the subject of a number of studies. As a useful comparison and to increase our understanding of gene regulation, spectral tuning and evolution, I have studied the visual pigment genes of goldfish (*Carassius auratus*) and a number of species of cottoid fish from Lake Baikal in Eastern Siberia.

An unusual feature of the goldfish is that, due to tetraploidy, many of its genes are present as duplicate copies and both may be expressed. This is seen with the green cone opsin genes. The level of expression and localisation to particular cone cells of the two green genes was assessed. It was found that green 2 was expressed at a higher level than green 1, although it is thought that both opsins are expressed within the same subset of cone cells. Regulation of the rod and two green cone opsin genes in goldfish has also been studied by the sequencing of the promoter regions and comparing them to each other and to those of mammals, chicken and *Xenopus*. A number of regions of homology were identified, many of which were known *cis*-acting elements. Additionally, an eight base pair motif, which has not been identified before, was found to be conserved between species.

The wavelength of maximum absorption (λ_{max}) of a visual pigment is determined by the structure of the opsin; changes in the amino acid composition at critical sites can significantly affect the λ_{max}. The species flock of cottoid fish living in Lake Baikal are known to have visual pigments whose λ_{max} is short-wave shifted with increasing depth of habitat. In order to identify amino acid substitutions responsible for these spectral shifts, the blue opsin gene of a representative subset of these fish has been sequenced. Three potential spectral tuning sites were identified at positions 118, 215 and 269. Mutations at these sites were then generated by site-directed mutagenisis and the mutant opsins expressed in 293T cells. After isolation and regeneration with 11-cis retinal, the resulting pigments were analysed spectrophotometrically.

Evolution of the blue visual pigment genes in the cottoid fish has also been investigated by phylogenetic analysis of the blue opsin sequences.

Declaration

I declare that this thesis submitted for the degree of Doctor of Philosophy is my own composition and the data presented herein is my own original work, unless otherwise stated.

Jill Alison Cowing

This thesis is dedicated to

Graham

for his love and inspiration.

I thank you with all my heart.

Acknowledgements

The list of people to thank for their help and inspiration during the years it has taken me to complete my PhD is endless; I never would have managed without all of you. In particular, my heartfelt thanks go to Professor David Hunt who has acted as my principal supervisor and has given me the time to write my thesis. In the twelve years that we have worked together I have been through some very difficult times and his friendship, endless encouragement and belief in me have brought me this far. I must also express my thanks to Professor Jim Bowmaker who has acted as my second supervisor, for providing the MSP data and for answering all those little questions.

Within the Department of Molecular Genetics I owe my gratitude to so many people, however there are several to whom I am particularly indebted. Sue Wilkie, for her help and expertise in getting to grips with spectral tuning and for her advice on just about everything. Phillippa Cottrill, for proof reading the gene regulation chapter and for useful discussions about transcription factors. Suba Poopalasundarum, for her unending patience in teaching me the delights of cell culture and protein purification; thank you for not killing yourself! Miranda VonDornum, for proof reading the evolution chapter and assisting with the construction of phylogenetic trees. Alex Morris, for those wonderful massages. Neil Ebenezer and Inez Stinton, for helping out so much with the general lab duties and making my life a whole lot easier. Madam Z (Vissani) and her small furry creatures, for the many laughs and shoulder to lean on in times of trouble and desperation. Melanie Freedman, who showed me that there was light at the end of the tunnel and for escorting me to the binders. Finally, Shomi Bhattacharya, who somehow talked me into doing a PhD in the first place and for his constant support and encouragement ever since.

In addition to current members of the lab, my love and thanks must also extend to a very special person, Zoë David-Gray, my fab ex-bench mate who has always been there for me; I'd never have made it without you sweetie. Also to Kamal and Rani Duali for their help and friendship over the years and harassing e-mails asking if I've

finished yet! I must also not forget Andy Williams and Rachel Ibbotson who introduced me to the world of molecular biology all those years ago. We had some wonderful times working together and they have remained dear friends ever since.

My appreciation must also be expressed to Anton the fish man for supplying samples of *Cottus gobio* and to Phyllis Robinson for the pMT4 expression vector.

My Mum and Dad too, deserve a special mention for their love and support and for always giving me good advice; I wish I had listened when you said, "don't do any more studying!" To my big sister Karen, I send a thank you hug for letting me camp out at your flat when I have been working late; I do not thank you though for giving me my first hangover!

Last but not least I would just like to say that I have the most wonderful husband in the World. Graham, you are a saint for putting up with such a nag bag of a wife whilst I have been writing my thesis. You are a constant source of amazement to me with your unending enthusiasm for life and I cannot thank you enough for your patience, understanding and love.

Table of contents

List of figures

List of colour plates

List of tables

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

The ability to detect light is common to many life forms from the simplest single cell organism to higher animals and even plants. Vision however, is a multicellular capability which has evolved over a period of millions of years into a very specialised and complex process. Higher animals which rely on the ability to detect predators, prey and mates for their very survival as a species, tend to have the most sophisticated visual sensory systems. Fish, in particular, exhibit enormous diversity in morphology, biology and the habitats that they occupy and hence have a broad range of visual capabilities making them an ideal group in which to study vision.

1.1 Teleost fishes

Of approximately 48,170 living vertebrate species, just over half are fish and these are of immense value to humans not only for food but also for recreational purposes. Our knowledge of their evolutionary relationships is based largely on morphological comparisons and is incomplete although in recent years the use of molecular genetic analysis, especially the polymerase chain reaction (PCR) and sequencing has increased our understanding. Their complex visual systems have evolved to cope with almost every conceivable type of aquatic habitat from dark, deep oceans to clear surface waters and shallow, murky river estuaries. The brilliance of their decoration would suggest that some species, at least, are able to perceive a wide spectrum of colours whether for mating purposes, warding off prey or camouflage.

The rich fossil record of teleost fishes points to their emergence during the middle to late Triassic, around 200 – 220 million years ago. Of all vertebrates, they are the most species rich and diverse group, with 38 orders, 462 families, 4,064 genera and around 23,637 extant species (Nelson, 1994). Teleosts are characterised by their caudal skeleton and have elongated ural neural arches which stiffen the upper tail lobe

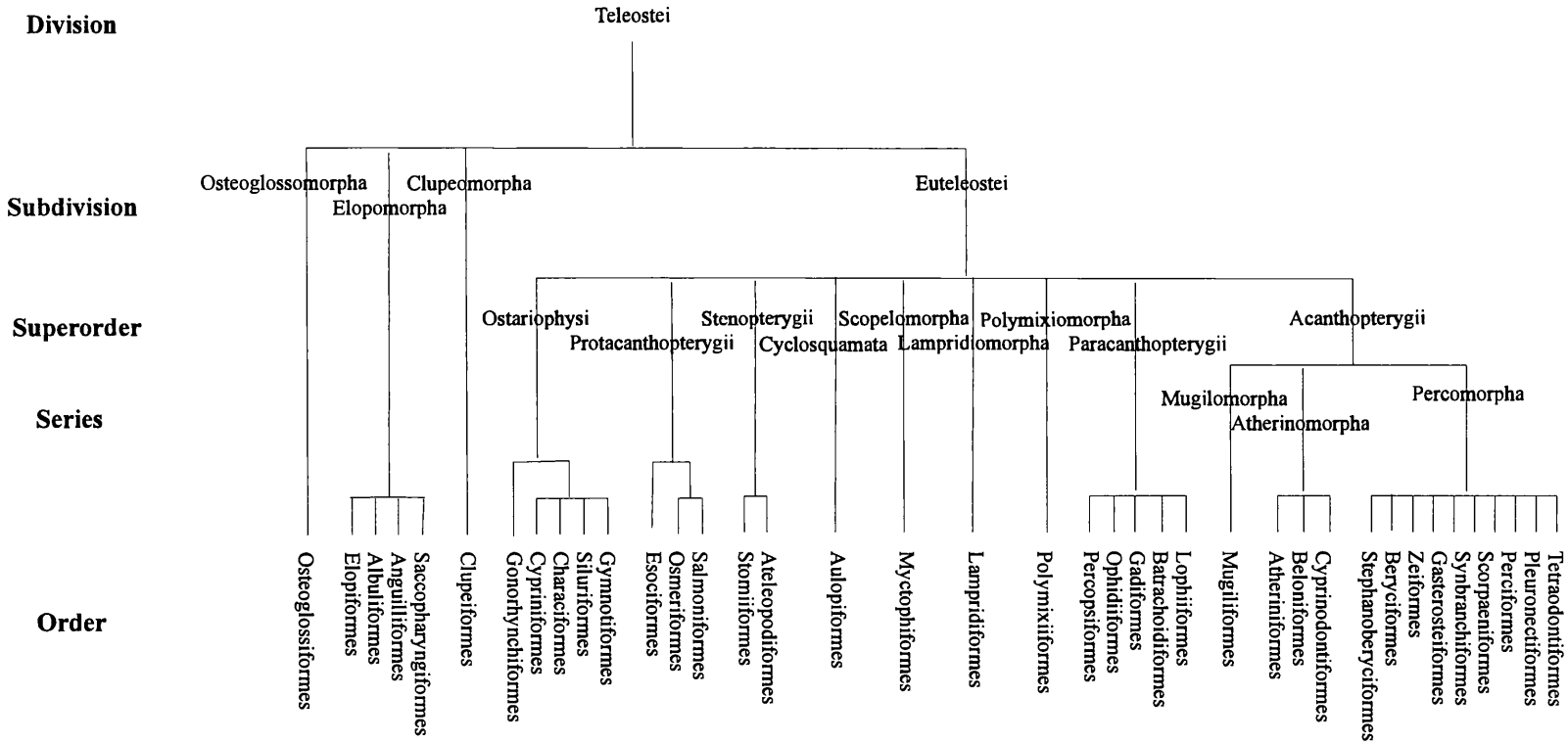


Figure 1.1: Hierarchy of higher categories of teleost fishes (from Nelson 1994)

forming an internally asymmetrical caudal fin. However, unlike other actinopterygians, they also have a caudal fin which is externally symmetrical. Based on their morphology, Patterson (1968) concluded that teleost fishes were of monophyletic origin. Figure 1.1 Shows the hierarchy of teleost fishes based on classifications by Nelson (1973a, 1973b) and Patterson and Rosen (1977).

This thesis focuses on the study of vision at the molecular level in certain teleost fishes. Regulation of the green cone and rod opsin genes has been examined in the goldfish, *Carassius auratus*, a member of the Cypriniformes order. Also studied was the spectral tuning and evolution of blue opsin genes in the cottoid fishes of Lake Baikal; these fish belong to the order, Scorpaeniformes. Other fishes used as comparisons include; the zebrafish *Danio rerio* which is also a Cypriniform, the blind cave fish *Astyanax fasciatus*, a member of the Characiformes order and the killifish, *Oryzias latipes*, a Beloniform.

1.2 Light

Light exists as both an electromagnetic wave, varying in frequency and wavelength (λ), and also as discrete energy packets called photons. For the purpose of colour perception, light is measured in nanometers (nm). The full spectrum ranges form the ultraviolet (around 300nm), through violet, indigo, blue, green, yellow and orange, to red and infrared at about 1100nm. The actual spectrum will vary depending on factors such as cloud cover, altitude and solar elevation, and wavelengths of radiation outside these limits are absorbed chiefly by the ozone layer and water vapour (Endler, 1993). Most living species however, are only able to visually perceive a very restricted spectrum and certain wavelengths, including potentially damaging ultraviolet light, are often prevented from even reaching the retina by the lens and cornea, which can absorb strongly in these regions.

In aquatic environments, the spectrum of penetrating light is further restricted because natural waters are rarely pure and water acts as a monochromator to absorb both long and short-wave light. Contaminants within the water body affect the wavelength of

penetrating light; for example dissolved organic material and phytoplankton may colour the water orange-brown or yellow-green respectively (Levine, 1982) and suspended particles will scatter short wavelengths. With increasing depth, downwelling light rapidly becomes attenuated until a state of darkness is reached.

When light rays travelling through a transparent medium, such as air, pass through a second transparent medium of different density, such as water, they bend at the surface of the two media. This is known as refraction and there are four such points of refraction within the eye: between the air or water and cornea, the posterior surface of the cornea and the aqueous humour, the aqueous humour and the lens, and the posterior surface of the lens and the vitreous humour. The degree of refraction is very precise to ensure correct focusing of an image onto the retina.

1.3 Eye structure and function

The basic function of the eye is to capture light and focus it on to sensory receptor cells in the retina. The front of the eye is formed from transparent, closely interwoven fibres, which are arranged in a regular fashion to form the cornea. The cornea not only protects the eye but can also play a role in focusing of an image. In mammals approximately 70% of the eye's focusing is at the air/cornea interface although it is not adjustable. However, underwater there is no air/cornea interface and eyes here derive no optical benefit because the refractive index of the cornea is almost identical to that of water.

Surrounding the cornea is the sclera, the white of the eye, which gives shape to the eye and protects its inner parts. Behind the cornea is the pupil, whose diameter affects the amount of light entering the eye. This is controlled by the iris which is composed of two bands of muscles, the dilator and sphincter. During bright light conditions the sphincter contracts thus reducing the size of the pupil, whereas in dim light, the dilator contracts and the iris opens, enlarging the pupil. Light then passes through the pupil to the anterior chamber and lens. In many fish however, the lens actually protrudes through the pupil and the iris is unable to restrict images to the central area. With an immobile pupil, light control is by means of retinomotor mechanisms involving

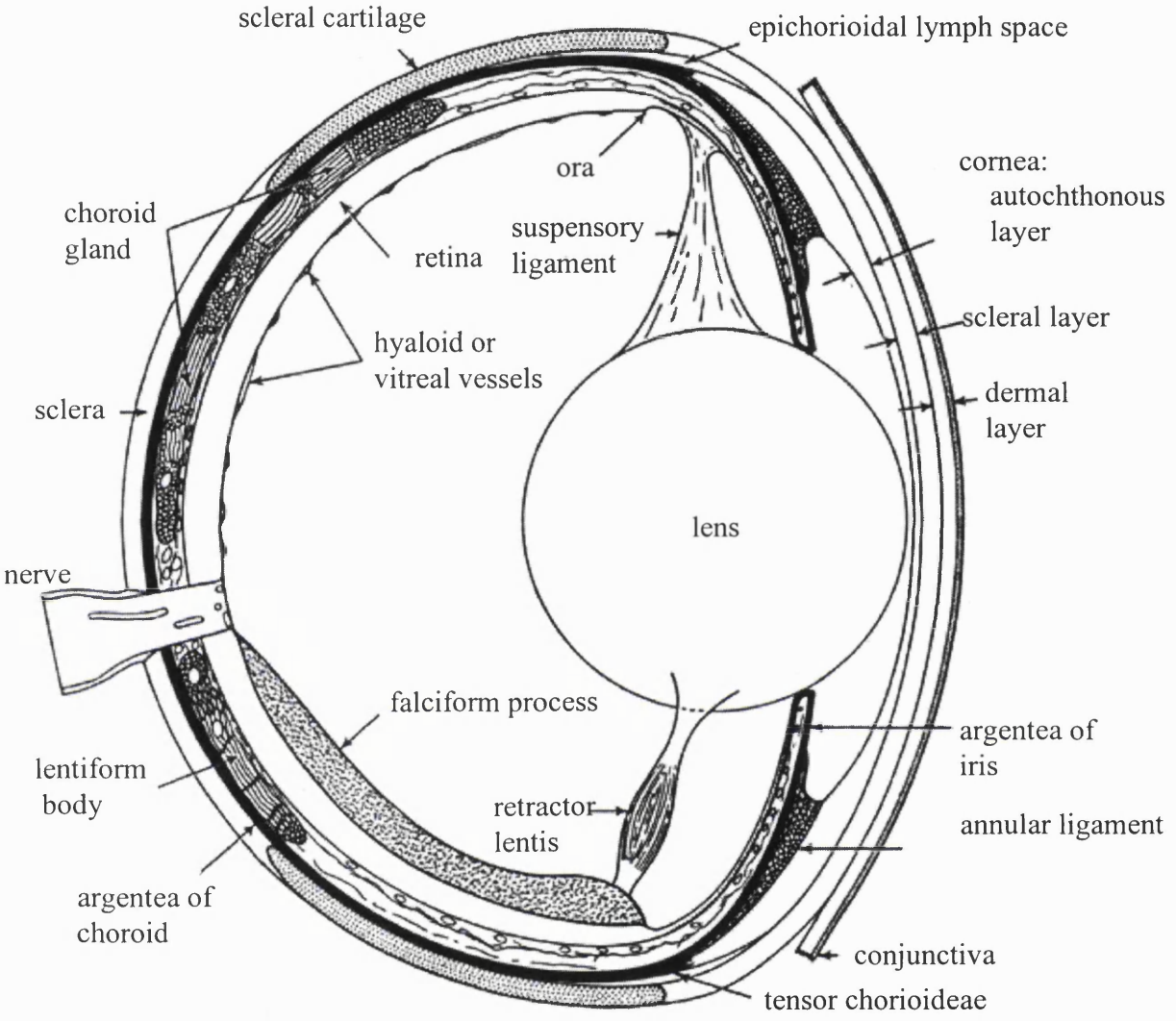


Figure 1.2: Diagramatic vertical section of a typical teleost eye. (Redrawn from Walls 1942).

cellular changes at the level of the retinal pigment epithelium and/or the photoreceptors themselves (Douglas, 1982). The retinomotor movements serve to optimally position the rods for scotopic vision and cones for photopic vision, and also to protect the rods from excessive bleaching during the day.

Within the anterior chamber a watery liquid, aqueous humour, transports oxygen and nutrients to structures of the eye and removes any waste products. This is constantly being produced by ciliary bodies and therefore needs to drain away. If this process is blocked in some way, pressure builds up within the eye leading to permanent visual damage. In humans, this is the condition called glaucoma.

The main purpose of the lens is to focus light on to the retina. Because the refractive power of the cornea is neutralised underwater, the lens in a fish eye must be much more powerful than its counterpart in terrestrial vertebrates. This has been achieved through the evolution of a spherical lens with very high refractive power. As light passes through the aqueous humour and hits the lens its path is altered, it is refracted, and the image is also inverted. Unlike the fixed focusing of the cornea, focusing by the lens can be controlled by its shape or position. Not all fish are able to accommodate but when it does occur it is as a result of movement of the lens within the globe and not by a change in shape. In most fish, movement is parallel to the plane of the pupil and, as such, different regions of the retina are focused at different distances. In the relaxed state the lens lies nearer to the nasal pole of the eye and the temporal part of the retina is focused for near vision and the nasal part for far vision (Fernald & Wright, 1985). In mammals when the lens is flat, the ciliary muscles are relaxed and distant objects are focused onto the retina. However, upon contraction of these muscles the lens becomes more rounded and is able to focus nearer objects. The lens consists of three separate parts; an elastic covering, an epithelial layer and the lens itself. The lens is constructed of proteins called crystallins arranged in layers like those of an onion (Snell & Lemp, 1989). Fibre cells are constantly produced by the epithelial layer and over a period of time may lead to a thickening. Old cells in the centre of the lens become hardened and reduce its ability to focus light. Also, the crystallins suffer from denaturisation with time and lead to a yellowing of the lens with age and an alteration in colour perception.

Beyond the lens is the main body of the eye. This is formed from a clear gelatinous substance, the vitreous humour. After passing through this, light reaches the retina and activates the photoreceptors; the resulting nerve impulses are then conducted to the visual area of the cerebral cortex. Photoreceptors are highly specialized for stimulation by light rays; rods are sensitive to dim light and allow discrimination between different shades of dark and light and allow perception of shapes and movement, whereas cones are specialized for colour vision and visual acuity in bright light. The ratio of rods to cones varies between different species, as does their spatial arrangement within the retina. The ends of the photoreceptors are embedded in the pigment epithelial layer behind the retina. This layer fulfils all the metabolic requirements of the photoreceptors including visual pigment regeneration. Behind this is the choroid layer which, like the pigment epithelial layer, is rich in the black pigment melanin and thus prevents reflection of stray light within the globe of the eyeball. Melanin granules, melanosomes, undergo extensive positional changes in response to ambient light conditions. In intense daylight melanosomes in the apical processes of the pigment epithelium serve to optically isolate cone outer segments and shield rod outer segments. In the dark-adapted state, these melanosomes are moved to the cell bodies of the pigment epithelial cells and the process is reversed (Douglas, 1982; Douglas & Wagner, 1982). Nocturnal or semi-nocturnal animals have few melanosomes. Instead, these are replaced by a reflective coating at the back of the eye called a tapetum, which increases visual sensitivity in low light environments. In fish, there is a correlation between spectral reflectance of the tapetum and the underwater environment. For example, freshwater fish may have longwave reflectors whereas those of deep-sea sharks are shortwave (Partridge, 1999).

1.4 The retina

The neural structure of the retina consists of three layers. Light must pass through two of these, the ganglion cell layer and bipolar cell layer before striking the photoreceptor cells composed of rods and cones. Once receptor potentials are generated by the rods and cones they are transmitted through the cell bodies and excite the bipolar neurons by the release of neurotransmitters. This excitation leads to

depolarisation of the ganglion cells and initiation of nerve impulses which are transmitted directly, via the optic nerve, to the brain.

The relative thickness of the photoreceptor layer as compared to the rest of the neural retina varies between species and habitats. Diurnal species with good vision generally show a ratio of 1:1. In comparison, fishes in habitats with low light levels such as deep sea species, show a reduction in thickness of the inner retina to as low as 15-25% of that of the photoreceptor layer (Munk, 1966). In such fish, the retina may be arranged in more than one layer to form a multi-bank arrangement which may be useful in dim light to increase the probability of photon capture (Locket, 1999).

The morphology of the neural retina shows considerable species differences but in general may be described as follows: there are several forms of ganglion cell some of which are selective for wavelength in high spatial frequencies but have a slow sustained response and others which, although not wavelength sensitive, are sensitive to low spatial frequencies and have a faster conduction velocity. Ganglion cell axons come together, passing through the bipolar and receptor cell layers, to exit the eye at a point called the optic disc. This region produces a blind spot as no photoreceptors can be located here but we are not consciously aware of this due to a process called infilling. The visual system fills in the hole based on the surrounding visual image to give a complete picture.

The receptive field of a ganglion cell is dependent upon the number of photoreceptors connected to it. The excess of photoreceptors to ganglion cells means that frequently more than one photoreceptor is connected to each ganglion cell. This ratio is not constant and varies between species and position within the retina. Although this greatly lowers visual acuity, it permits summation effects to occur so that low levels of light can stimulate a ganglion cell that would not respond had it been connected directly to one photoreceptor. Photoreceptors are also organised in a centre-surround arrangement. Light falling on the centre of a receptive field may result in stimulation of the ganglion cell but when falling on the surrounding ring of photoreceptors may actually inhibit the ganglion cell. This is known as an on centre, off surround cell. Similarly, there are off centre, on surround cells. If the whole receptive field is illuminated there will be little, if any, change in the cells firing rate because both the

on and off regions are stimulated. If however, the edge of an object crosses the field then the presence of a light dark boundary will be signalled and the ganglion cell stimulated. The orientation and shape of the object can then be determined by the comparison of a number of ganglion cell responses in the first cortical visual area within the brain.

There are also two other specialised cell types within the retina, the horizontal and amacrine cells. These form lateral pathways that modify and control the message that is transmitted along the direct pathway. Horizontal cells act to transmit inhibitory signals to bipolar neurons in areas lateral to excited photoreceptors in order to enhance contrast between strongly and weakly stimulated areas of the retina. Horizontal cells also assist in the differentiation of colours. Amacrine cells, excited by bipolar neurons, synapse with ganglion cells and transmit information which signals a change in levels of retinal illumination.

Studies of the neural retina of fish have shown that in species with low light vision, such as the catfish, there are very few amacrine or bipolar cells and it is the horizontal cells which dominate. In comparison, fish with well developed colour vision living at higher levels of illumination show a dominance of amacrine and bipolar cells (Ali & Anctil, 1976).

1.5 Photoreceptors

Photoreceptors absorb light and undergo structural changes that lead to the development of a receptor potential. Retinal photoreceptors, the rods and cones, are so called because of their morphology (figure 1.3). These highly specialised cells each have an outer segment, which is a sensory organelle, and an inner segment, containing the nucleus and biosynthetic machinery. Typically, rods have cylindrical inner and outer segments whereas cones have a shorter, conically formed outer segment. The outer segment is filled with a stack of several hundred flattened membrane discs or lamellae in which the photopigments are embedded. In the rods, these lamellae are free floating discs, whereas in cones they consist of one continuous folded membrane which maintains a connection with the plasma membrane. The

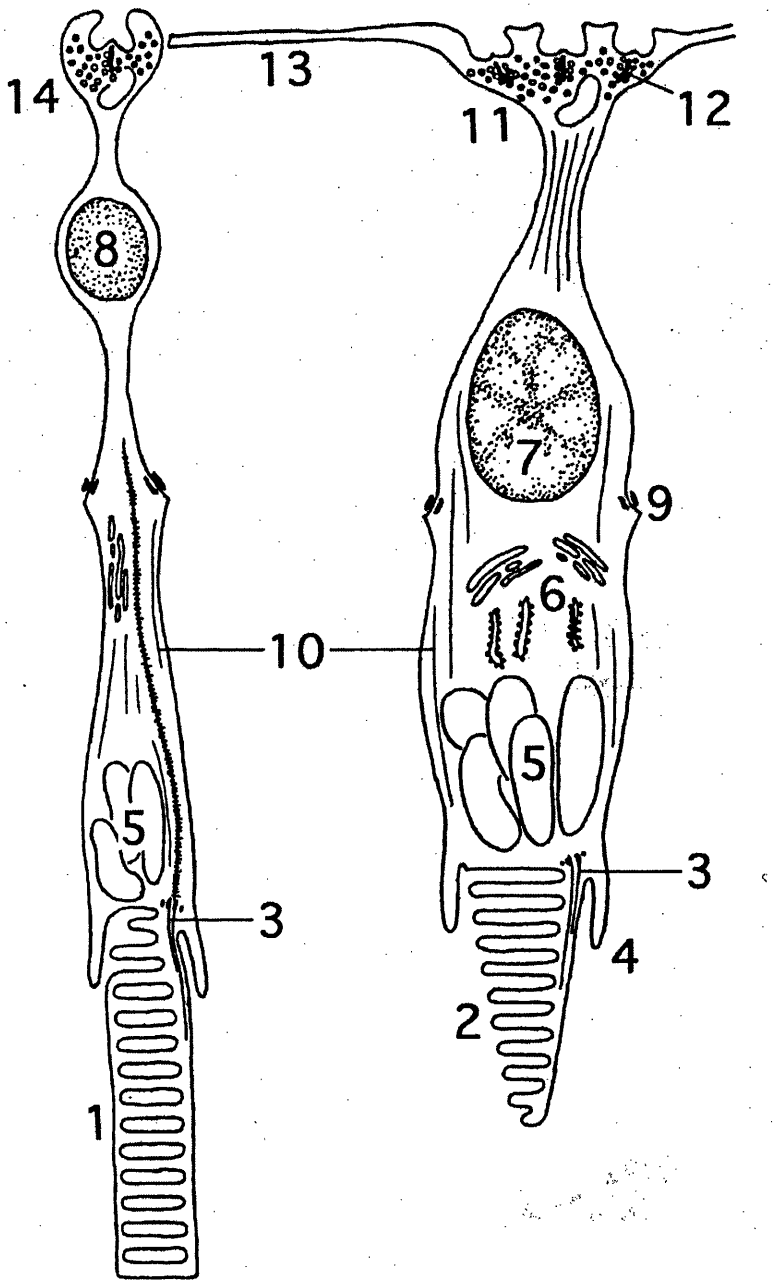


Figure 1.3: Diagrams of a typical rod and cone cell. 1: rod outer segment discs. 2: cone outer segment discs. 3: cilium. 4: cyclal processes. 5: golgi apparatus. 6: myoid. 7: cone nucleus. 8: rod nucleus. 9: outer limiting membrane. 10: microtubules and filaments. 11: cone pedicle. 12: synaptic ribbons. 13: basal filaments. 14: rod spherule (from Locket 1999).

lamellae are continually produced and in the rhesus monkey approximately 90 discs are added per day (Young, 1971). As new discs are produced at the base of the rod outer segment, the older ones move up and are eventually phagocytosed at the distal end by the pigment epithelial cells after approximately ten days. Rhodopsins or porphyropsins, which are the photopigment molecules of vertebrates, are embedded within the lamellar membrane. In each human rod about 100 million tightly packed rhodopsin molecules are present and form approximately 80% of the membrane proteins (Nathans, 1987). The inner segments of rods are small and conical, compared to cones whose inner segments are often bulky containing large mitochondria. In the retina, the development of rods lags behind that of cones and long after cones have matured, the rods continue to be produced (Raymond, 1985). An unusual feature of the fish visual system is that its structure may change into adulthood and the retina is able to regenerate after injury (Pankhurst & Lythgoe, 1983; Cameron & Easter, 1995). In the European eel (*Anguilla anguilla*), for example, the number of rods in the retina increases during maturation and migration to deep sea, whilst the number of cones decreases (Pankhurst & Lythgoe, 1983).

Rod cells are highly sensitive, active in dim light and provide low acuity, monochrome vision. Cones however, although less sensitive, can provide high acuity colour vision with higher levels of background illumination. The switch from rod to cone vision in fish is partly controlled by photomechanical movements of the photoreceptors and pigment epithelium as described previously and also by other factors under the control of the circadian mechanism. Most vertebrates have both rods and cones but the complement may vary greatly depending upon the ecological niche inhabited by the animal. For example, fish living at great depth where there is very little penetrating light or nocturnal animals such as the gecko (*Gekko gekko*) may have pure rod retinas (Bowmaker *et al.*, 1994; Partridge *et al.*, 1988; Kojima *et al.*, 1992). These rods are also elongated to maximise the chances of intercepting a photon of light. In comparison, the goldfish which lives where the light is bright and spectrally broad, has four distinct cone pigments in addition to the rod (Johnson *et al.*, 1993; Bowmaker *et al.*, 1991b) and the American chameleon (*Anolis carolinensis*), which is diurnal, has a retina formed exclusively of cones (Crescitelli, 1972).

Cone cells are classified according to their maximal sensitivity to wavelengths of light (λ_{max}). The short-wave sensitive cones (SWS) respond maximally to light in the blue and ultra violet region of the spectrum whereas the middle and long-wave sensitive cones (MWS and LWS) are generally more responsive to green and red light respectively. In teleost fish where more than a single cone type is present, the longest ones are mostly red-sensitive, those of intermediate length are green sensitive, short single units are blue sensitive and miniature types are ultra-violet sensitive (Stell & Harosi, 1976; Levine & MacNichol, 1982; Lythgoe, 1979). However, in birds and lower vertebrates there are often double cone cells consisting of closely connected, contiguous inner segments. These two members may be twins, which are similar in size, or doubles, which can vary significantly having a larger principal member and a smaller accessory member. The two members may also be of the same cone class or consist of one each of the MWS and LWS cones as in the goldfish (*Carassius auratus*) (Stell & Harosi, 1975).

An additional feature of cone cells in many species is the inclusion of oil droplets which act as ocular filters to enhance colour vision and increase resolution, and also to protect the retina from photo-oxidative damage (Douglas & Marshall, 1999). Although found in many birds (Partridge, 1989; Wright, 2000), reptiles (Lipetz, 1984; Ohtsuka, 1985), some amphibia (Hailman, 1976), monotremes (Ahnelt *et al.*, 1995) and primitive fish (Robinson, 1994), oil droplets are absent in teleost fish and placental mammals.

The distribution of photoreceptors and neurones is not constant across the retina. In the human retina, approximately 6 million cones are concentrated in a small depression, the fovea, which is surrounded by 120 million rod cells. Foveae in deep-sea fish however, have also been discovered which consist entirely of rods (Munk, 1966) or a mixture of rods and cones. Foveae are also common in many teleost and may be fish formed not only from single rods but also from multiple banks of rod cells (Collin, 1999). Other areas such as the area centralis in goldfish (Mednick *et al.*, 1988) have an increased density of photoreceptors are also present but these are not characterised by a depression.

In fish with superior visual capabilities, cone cells are often arranged in a set pattern or mosaic. In many species, single cones are arranged in parallel rows the same distance apart from each other, with double cones occupying a central position. Square or triangular patterns are more common in predators, whereas row mosaics are more often found in schooling species. In the cottoid fish of Lake Baikal, a well-ordered mosaic is present in surface dwelling species. However, in comparison, those fish which live in deeper waters, have a less well organised pattern (Bowmaker *et al.*, 1994). Also, the pattern may change during the development of the eye, for example from a row to a square mosaic in trout (Lyall, 1957), or as a result of retinomotor movements where a square cone mosaic in teleost fish was observed to transform into a row mosaic at night (Kunz & Ennis, 1983). Mosaics have also been identified in mammals (Williams, 1986) and birds (Wilkie *et al.*, 1998) but although there appears to be some degree of regularity, they are distinct from the simple geometric patterns observed in fishes. The human mosaic appears to be more random and varies between normal individuals; the LWS and MWS cones may appear in clusters with the SWS cones sparsely and irregularly distributed (Roorda, 1999). The structural patterns observed are determined not so much by taxonomic relationships but more by the habitat and behavioural characteristics of a given species.

1.6 Visual pigments

All visual pigments are G-protein coupled receptors and consist of an apoprotein, opsin, joined covalently via a protonated Schiff's base linkage at a lysine residue (Lys 296) to retinal, an aldehyde of vitamin A. The opsin consists of seven hydrophobic transmembrane α-helices linked by hydrophilic, straight chain, extra-membrane loops which fold to form a central binding pocket for the chromophore retinal (figure 1.4). The model proposed by Baldwin (1993, 1997) suggests that of the 26 amino acids per helix, only the central 18 are actually embedded within the membrane. Comparisons of these regions between different visual pigments have shown that they are highly conserved not only within a species but also across the animal kingdom.

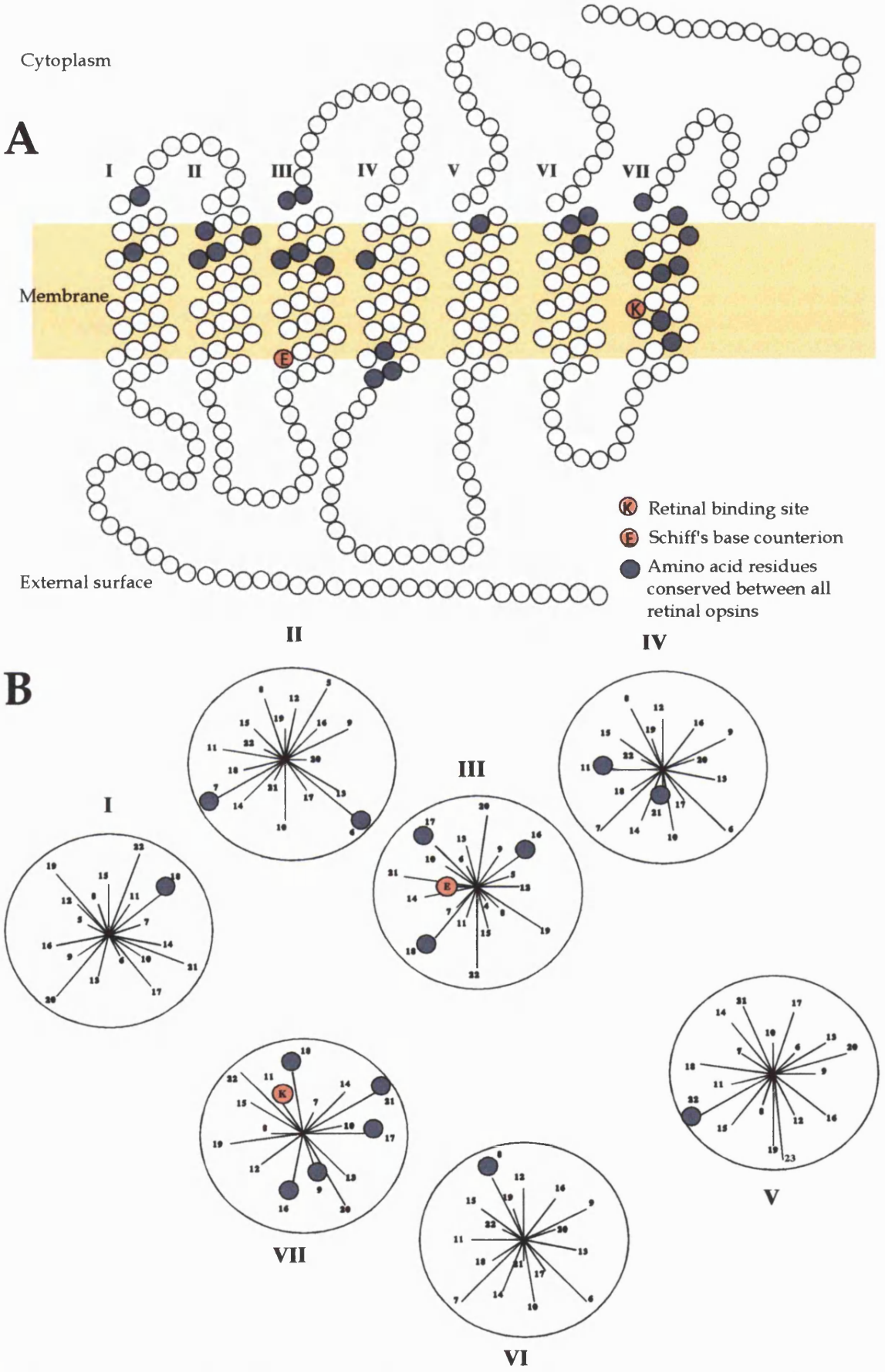


Figure 1.4: Models of opsin. A) 2D structure showing seven transmembrane α-helicies. B) 3D structure derived from Baldwin 1998. Each circle represents an α-helix, each radiating line an amino acid and the length of the line, depth within the helix.

Retinal exists in two forms in vertebrates, as 11-*cis* retinal (A_1) or 11-*cis* 3-dehydroretinal (A_2). In birds and mammals the chromophore is 11-*cis* retinal, forming the visual pigment rhodopsin. In fish, amphibians and reptiles however, a second chromophore 11-*cis* 3-dehydroretinal may also be present forming the porphyropsins. The replacement of retinal with 3-dehydroretinal results in a shift of the λ_{max} to longer wavelengths (Knowles, 1977), the extent of which is governed by the opsin component of the visual pigment; the longer the wavelength of the opsin, the greater the difference between the rhodopsin and its partner porphyropsin. In the rudd, *Scardinius erythrophthalamus*, there is a seasonal shift in the balance of opsins and porphyropsins and hence λ_{max} of the visual pigments. The rods and LWS cones have a λ_{max} of 507nm and 562nm respectively when the chromophore is retinal and a λ_{max} of 535nm and 625nm respectively when the chromophore is 3-dehydroretinal (Whitmore & Bowmaker, 1989). The λ_{max} of the rods in the European eel (*Anguilla anguilla*) also short wave shifts as the eel matures sexually and migrates from freshwater to the deep sea. This is a result of opsin replacement and a shift in the ratio of A_1 to A_2 chromophores (Wood & Partridge, 1993; Archer *et al.*, 1995). The λ_{max} of a visual pigment is therefore governed by the amino acid sequence of opsin together with the chromophore present.

Microspectrophotometry (MSP) is used to measure the characteristic response curve of a visual pigment to the wavelength of incident light (Partridge, 1986). The absorption spectrum of these pigments is broad, varies between classes and is dependent on its concentration within the outer segment; the higher the concentration the wider and flatter the absorption spectra. As the concentration increases a theoretical limit is reached whereby there is 100% absorption of all wavelengths regardless of the λ_{max} (Dartnall, 1957). Visual pigments which absorb at longer wavelengths have a very broad bandwidth as compared to the SWS pigments which have narrower spectra. The principle of univariance states that the apex of the curve shows the wavelength to which a pigment is most sensitive (λ_{max}) and the area under the curve reflects the probability of photon capture for pigments at that wavelength (Rushton, 1963). A typical rhodopsin with a λ_{max} of 500nm has a bandwidth at 50% absorbance of 102nm.

1.7 Structure of the opsin molecule

As previously described, the opsin molecule is folded to form an alternating hydrophilic/hydrophobic structure comprising seven α-helical transmembrane regions linked by extracellular loops. Studies on the crystal structure of frog, bovine and bacterio-rhodopsin show that the helices are tilted to differing degrees and interact to form the binding pocket (Schertler & Hargrave, 1995; Schertler *et al.*, 1993; Unger *et al.*, 1997; Kimura *et al.*, 1997; Palczewski *et al.*, 2000) (figure 1.5). The extracellular end of helix I is highly exposed to the lipid and, as such, is subject to variation within opsin genes. It is tilted to an angle of 28° and is linked to helix II via a short cytoplasmic loop. Helix II is tilted to 27° with the extracellular side interacting with helices I, III and VII; on the intracellular side, it is less exposed and interacts only with helix III. Helix III is tilted to 30° and is buried deep inside the molecule. At its central position on the intracellular side, it has extensive contact with helices II, IV, V and VI and on the intradiscal side, with helices II, IV and VII. This results in the retinal-binding pocket being held closed at the cytoplasmic side and helices IV, VI and VII being kept apart in the part of the structure which is closer to the intracellular surface. The fourth helix is the shortest and least tilted and contacts helices II and III on the intracellular side and III and V on the extracellular side. The structure of helix V was, until recently, unclear due to poor resolution of the electron micrographs in this region. However, Palczewski *et al.* (2000) have now shown that it is almost straight, tilting at an angle of 23° and is exposed to lipid along its entire length. Helix VI is in contact with helices III, V and VII on the intracellular side and helices V and VII on the intradiscal side. It is bent towards helix V on the intradiscal side thus preventing the interior from becoming exposed to the lipid bi-layer. Helix VII is close to helix III in the centre of the molecule just above the region where retinal is probably attached.

Although strong homology exists between all visual opsins, there are specific sites which are conserved to a higher degree due to their functional importance. These

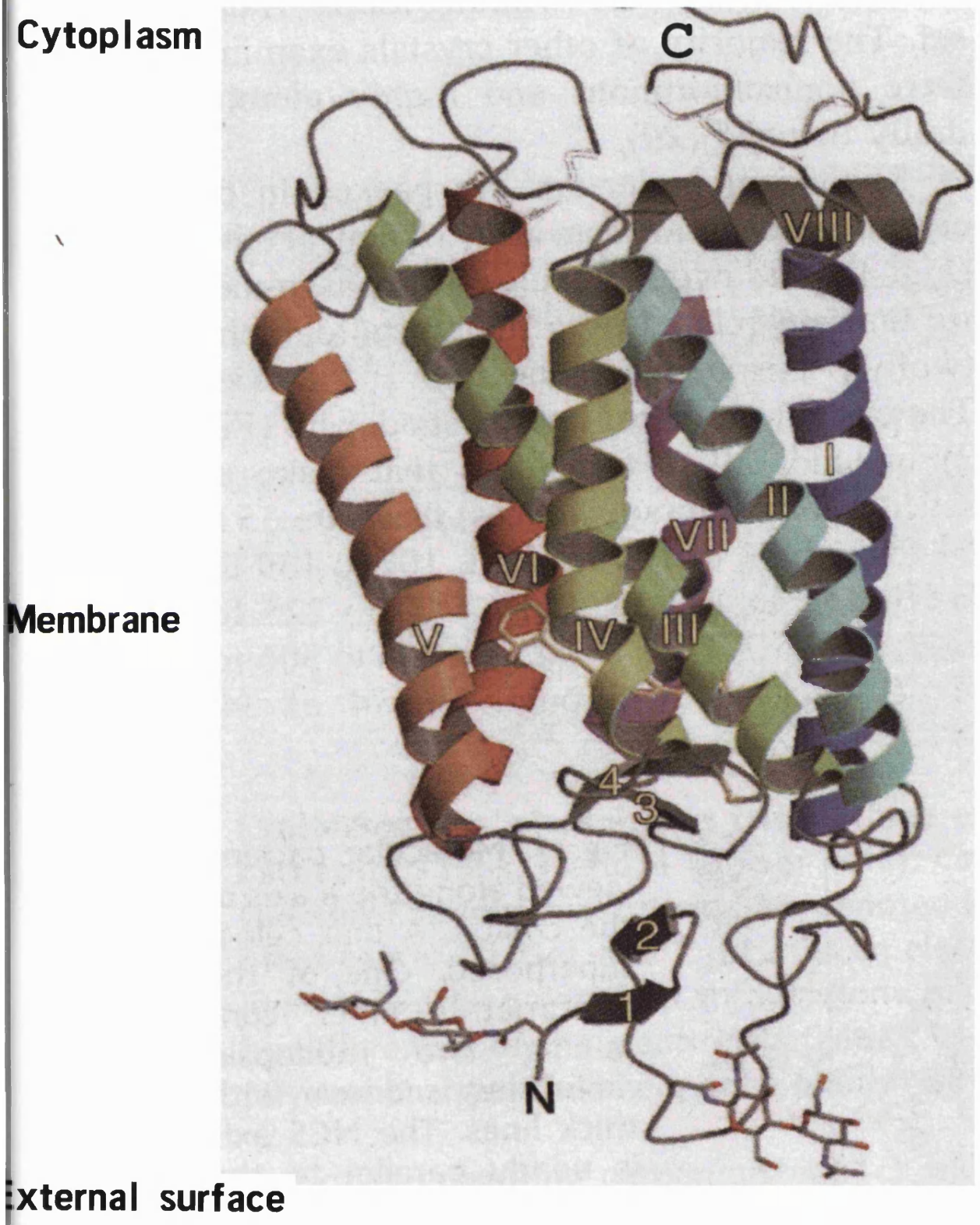


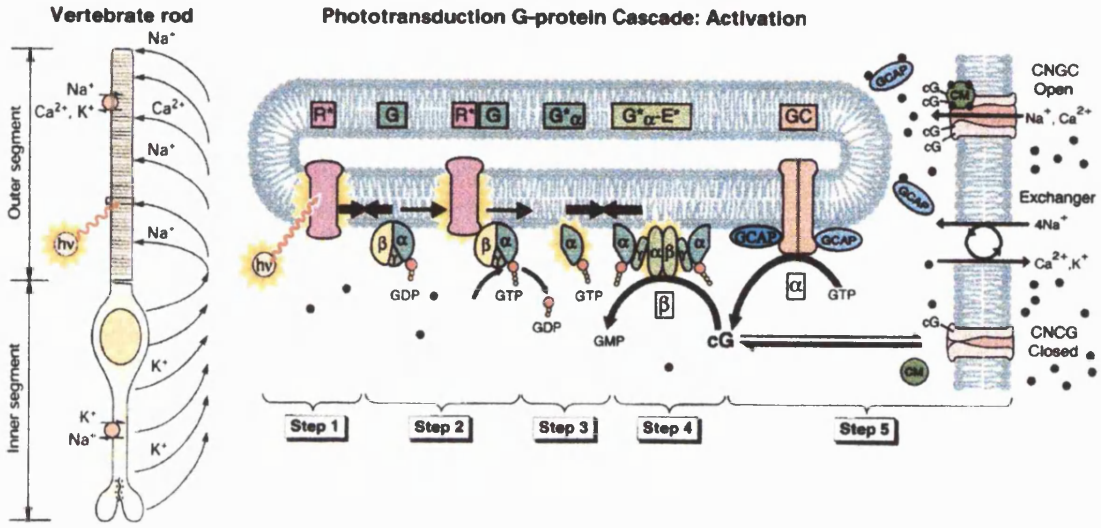
Figure 1.5: Ribbon diagram of rhodopsin showing the seven transmembrane helices (from Palczewski *et al.* 2000) labelled I-VII. The numbers 1-4 indicate β sheets.

include: a lysine residue at position 296, which forms a Schiff's base linkage to the retinal chromophore (Bownds, 1967); glutamic acid at 113, which acts as the counterion to the protonated Schiff's base (Sakmar *et al.*, 1989); a disulphide bridge formed by cysteine residues at 110 and 187 which connects helix III to the second extracellular loop (Karnik *et al.*, 1988; Karnik & Khorana, 1990); glutamic acid at 135 and arginine at 136, which interact with transducin (Franke *et al.*, 1990).

1.8 Phototransduction

The photoreceptor response to light is rapid, adaptable to different light intensities and is controlled generally by the intracellular calcium levels through the phototransduction cascade. The reaction is similar for both rods and cones although many proteins along the phototransduction pathway have a different isoform for each cell type (Yau, 1994). It is perhaps one of the best-understood biochemical pathways and has been extensively reviewed e.g. (Yau, 1994; Pugh, 1999).

In darkness, vertebrate photoreceptors have an inflowing dark current through open cyclic GMP (cGMP) gated cation channels. On receiving photons of light, the isomerization of the photoreceptor chromophore leads to the closure of these channels and the hyperpolarization of the cell and propagation of the neural signal. The initial step in phototransduction is the conversion of the chromophore 11-*cis* retinal to all-*trans* retinal upon activation by a photon of light. During this transition, rhodopsin passes through two intermediate forms called metarhodopsin I and II. It is the photoactivated metarhodopsin II form of rhodopsin which then binds to molecules of G proteins, the transducins, catalysing the exchange of GTP for GDP. One fundamental difference between photoreceptor cells is the more rapid response and lower photosensitivity of cones compared to rods (Lamb & Pugh, 1990). The rate of production and decay of metarhodopsin II is critical to the rate of response and level of sensitivity of the photoreceptor. It has been shown in chicken that the metarhodopsin II intermediate of the green cone opsin has a rate of formation that is



Step 1 Capture of a photon causes rhodopsin to transform into its enzymatically active form R*

Step 2 R* repeatedly contacts molecules of the G protein, catalyzing the exchange of GTP for GDP, producing the active form of G*α

Step 3 G*α subunits bind to the inhibitory γ subunits of phosphodiesterase (E), thereby activating the corresponding α and β catalytic subunits, forming E*'s

Step 4 E*s catalyse the hydrolysis of GMP (cG)

Step 5 The consequent reduction in the cytoplasmic concentration of cGMP leads to the closure of cyclic nucleotide gated channels (CNGCs) and blockage of the inward flux of Na+ and Ca2+

Figure 1.6: Steps in the activation of the vertebrate rod phototransduction cascade (from Pugh 1999).

four times faster and a rate of decay which is fifty times faster than that of rod opsin (Shichida *et al.*, 1994).

Transducin is composed of three subunits, α, β and γ. Upon activation, the α subunit bound to GTP is released and binds to the inhibitory γ subunit of phosphodiesterase. This action activates the corresponding α and β subunits, which catalyse the hydrolysis of cyclic GMP (cGMP). As the cytoplasmic concentration of cGMP is reduced, the cyclic nucleotide gated channels close thus blocking the inward flow of Na^+ and Ca^{2+}. Ca^{2+} is still being pumped out resulting in its dissociation from guanylate cyclase activation proteins (GCAPs), calmodulin and recoverin and a decline in its concentration within the cytoplasm, all of which initiate a feedback response. The loss of Ca^{2+} allows GCAPs to bind to a cytoplasmic domain of guanylate cyclase, thereby activating the enzyme and resulting in an increase in cGMP. The loss of Ca^{2+} from calmodulin causes it to dissociate from the cyclic nucleotide gated channels which in turn increases the channel's affinity for cGMP. The increased binding of cGMP results in the reopening of the CNGCs allowing Ca^2+ and Na+ to re-enter. The disassociation of Ca^2+ from recoverin leads to its disassociation from the membrane and rhodopsin kinase, increasing rhodopsin kinase activity. The photoactivated rhodopsin is subsequently deactivated upon phosphorylation by rhodopsin kinase and binding by arrestin. This pathway is summarised in figure 1.6.

1.9 Colour vision

Colour vision requires a comparison of the outputs of at least two photoreceptor cells, which differ with respect to their absorption spectra, usually a 'red' or 'green' with a 'blue' pigment. In the mammalian world, only a limited number of species have full trichromatic colour vision, i.e. those that have three separate, functioning visual pigments, excluding the rod pigment. These include humans (Nathans *et al.*, 1986b), Old World primates (Bowmaker *et al.*, 1991a; Ibbotson *et al.*, 1992) and one New World primate, *Allouatta* (Jacobs *et al.*, 1996). In addition, female New World primates may also exhibit trichromacy if they posses two different alleles of the

MWS/LWS pigment on their X-chromosomes (Jacobs *et al.*, 1996). In comparison, many fishes, birds and reptiles have superior colour vision. Tetrachromacy has been reported in many species including the goldfish (Bowmaker *et al.*, 1991b) and budgerigar (Wilkie *et al.*, 1998). In addition to LWS, MWS and SWS pigments these species also have a UV sensitive opsin.

The spectral absorbance of a visual pigment is governed not only by the chromophore present (vitamin A1 or A2) but also by the interaction between the chromophore and the opsin. Substitution of only a few critical amino acids within an opsin molecule can have a significant affect upon its absorption spectra (Neitz *et al.*, 1991). This process of spectral tuning is discussed in detail in chapter 4.

1.10 Classification and evolution of opsins

Halobacteria have existed for 1.5 billion years and contain pigments with an opsin-like protein, bacteriorhodopsin and halorhodopsin. Although neither of these have direct sequence homology to rhodopsin, there are topographical similarities including: the site and nature of chromophore attachment, the number of transmembrane helices and positions of the amino and carboxyl termini (Henderson & Schertler, 1990). This would suggest that opsin-like structures were present during the very early stages of evolution and that at one time they were closely related (Yokoyama & Yokoyama, 1989). Visual pigments have been studied extensively using several techniques including molecular analysis, and opsin genes have now been cloned and sequenced from a wide range of species exhibiting a relatively high degree of homology. Rod opsins, in particular, show greater that 70% identity, whereas the cones range from 96% between human red and green down to around 40% for blue and red/green (Nathans *et al.*, 1986b). Evolutionary analysis suggests that two cone pigments were present at the beginning of vertebrate evolution, one sensitive to shorter wavelengths and the other to longer wavelengths (Goldsmith, 1990). It has also been proposed recently that vertebrates acquired colour vision before scotopic vision (Okano *et al.*, 1992) and therefore, it seems likely that animals generally acquired the ability to

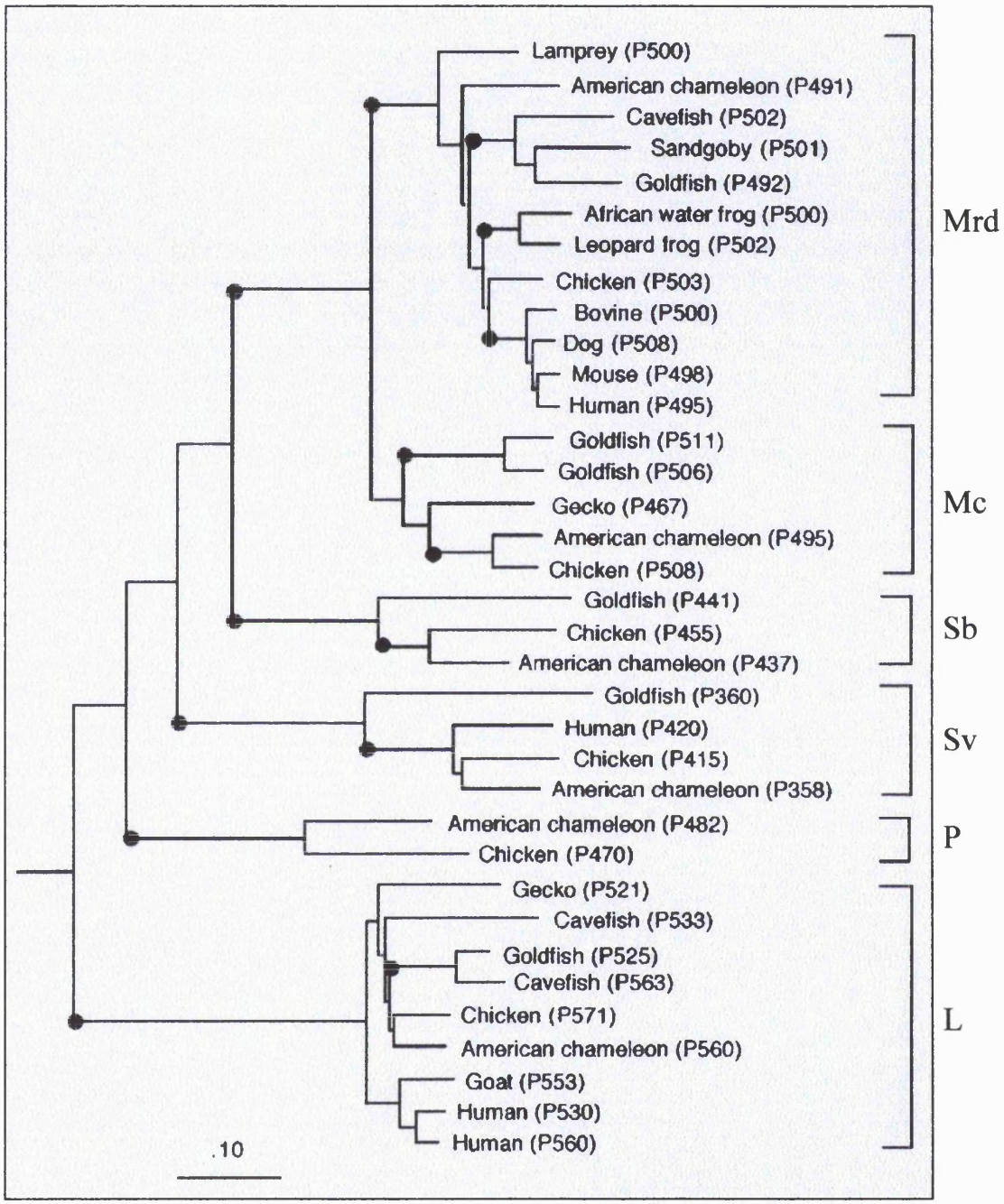


Figure 1.7: Phylogenetic tree of vertebrate opsins constructed using the neighbour-joining method. Numbers in brackets denote λ_{max} values. (From Yokoyama 1997). Mrd: middle wavelength sensitive rod pigments. Mc: middle wavelength sensitive cone pigments. Sb: short wavelength sensitive blue cone pigments. Sv: short wavelength sensitive violet pigments. P: pineal opsin pigments. L: long wavelength sensitive cone pigments. P value in brackets indicates λ_{max}

detect various wavelengths of light before evolving the ability to function at various intensities of light.

Original studies on human opsins were carried out by Nathans *et al.* in 1986 (Nathans *et al.*, 1986b). Since then, studies on the molecular basis of colour vision in vertebrates have developed rapidly. Published sequence data have been used to construct phylogenetic trees which show four distinct cone opsin families in addition to the rods (Okano *et al.*, 1992; Johnson *et al.*, 1993; Chang *et al.*, 1995; Yokoyama, 1997; Tokunaga *et al.*, 1999). Figure 1.7 shows a typical phylogenetic tree of vertebrate opsins. The longer wavelength group (L) forms a branch on its own and includes the goldfish red, human red and green and the chicken iodopsin. Subsequent duplications within the other branch gave rise to the short-wave violet (Sv), short-wave blue (Sb), middle wave sensitive rod-like cone opsins (Mc) and the rod opsins (Mrd). The terminology used for the classification of opsin groups, was proposed by Bowmaker and Hunt in 1999. Evolution of visual pigments, especially within the Sb group, is discussed further in chapter 5.

The Mc cluster is particularly interesting because members of this group share characteristics of both rod and cone opsins. Goldfish green 1 and 2 have been directly linked to cone cells by spectroscopic analysis and *in situ* hybridisation (Johnson *et al.*, 1993) and, like cone pigments, they are rapidly inactivated by hydroxylamine. However, their similarities to rod opsins include a higher amino acid sequence identity to rod opsins than other cone opsins, two potential asparagine glycosylation sites at positions 2 and 15 (Fukada *et al.*, 1979) and a conserved histidine residue at 211 (Weitz & Nathans, 1992).

The intron/exon structure of visual opsins is conserved within all groups except for the L-group of pigments and the rod opsins of bony fish. The primate L-group consists of six exons separated by five introns and has been localised to the X-chromosome (Nathans *et al.*, 1986a; Fitzgibbon *et al.*, 1994). The intron/exon structures of other members of this group remain unpublished. In comparison, the rod opsins of the cottoid fish are intronless (Fitzgibbon *et al.*, 1995). All other retinal opsins are formed of five exons interrupted by four introns. Sex chromosomes in teleost fish have yet to be identified, as karyotyping of the zebrafish, *Danio rerio*, and

pufferfish, *Tetraodon fluviatilis*, failed to find any difference in the chromosome complement between males and females (Amores & Postlethwait, 1999; Mandrioli *et al.*, 2000). Therefore, whether any opsins are sex linked in fishes remains unknown until more genetic information is available.

CHAPTER 2

MATERIALS AND METHODS

2.1 Extraction of genomic DNA from liver samples

Liver samples were either removed and immediately processed, placed in liquid nitrogen before being stored at -80°C, or chopped up and stored at 4°C in absolute ethanol. The liver was then weighed and ground up with liquid nitrogen using a pestle and mortar. The following protocol applies to each gram of tissue. The tissue was resuspended in 5ml of ice-cold NE buffer and spun at 4000 X g for 7 minutes. The supernatant was removed and the pellet resuspended in a further 5ml NE before spining again. This was repeated until the supernatant was clear, approximately 3 times. The washed pellet was then resuspended by vortexing in 5ml NE containing proteinase K to a final concentration of 400μg/ml. 500μl of 10% SDS was then added and the sample mixed gently before overnight incubation at 50°C in a waterbath. The following day an equivalent volume of buffered phenol was added and the sample mixed for 20 minutes on a shaking platform. To separate the aqueous phase, the sample was spun at 200 X g for 10 minutes and the upper layer was removed to a fresh tube. This was repeated once more and then an equal volume of chloroform/isoamyl alcohol was added. The sample was spun again and the upper aqueous phase removed to a fresh tube. To precipitate the DNA, 5M NaCl was added to a final concentration of 0.4M together with 2 volumes of absolute ethanol, the tube was mixed by inversion for 2 minutes and the DNA recovered by spooling or, if this was not possible, by centrifugation at 3500 X g for 10 minutes. The DNA was resuspended in 1ml sterile distilled water and stored at 4°C overnight. The next day the DNA was again precipitated by adding 500μl of 8M ammonium acetate and 6 volumes of absolute ethanol. The spooled DNA was washed in 70% ethanol and then allowed to air dry for several hours before being resuspended in 1ml of sterile distilled water.

2.2 Determination of DNA concentration and purity

A 1:100 dilution of the DNA was prepared using distilled water and the absorbance measured at 260 and 280nm by ultraviolet spectrophotometry. The amount of ultraviolet radiation absorbed by the DNA solution is directly proportional to the amount of DNA present. At 260nm, an absorbance of 1.0 is equivalent to 50μg of double stranded DNA per ml. The following formula was used to determine concentrations:

$$\mu g \text{ DNA/ml} = \frac{A_{260} \text{ X } 50 \text{ X} 100}{1}$$

The ratio of the absorbances at 260 and 280nm is used to determine the DNA purity. A pure sample will have a ratio 1.8, anything less that this is an indication of contamination with either protein or phenol.

2.3 The polymerase chain reaction (PCR)

The standard PCR was used to selectively amplify chosen regions of DNA by simultaneous primer extension of complementary strands of DNA. This method was originally devised by Mullis at the Cetus Corporation (Mullis *et al.*, 1986; Mullis & Faloona, 1987). Generally, the reactions were performed as follows unless otherwise stated with a typical 25μl reaction mixture containing; 1μl DNA, 30 pmol of each primer, 4μl dNTP mix, 2.5μl of a 10X NH_4 based reaction buffer, 1.5mM $MgCl_2$, 1 unit of *Taq* polymerase and sterile distilled water to a final volume of 25μl. The buffers and *Taq* polymerase used were supplied by either Promega or Bioline.

The thermal cycling conditions varied according to the melting temperature of the primer (Tm) but routinely the following parameters were used: denaturation at 94°C for 5 minutes; 35 cycles of 94°C for 20 seconds, annealing of primer at Tm°C for 20 seconds and elongation at 72°C for 30 seconds; finally there was a further elongation at 72°C for 7 minutes.

2.4 Primer design

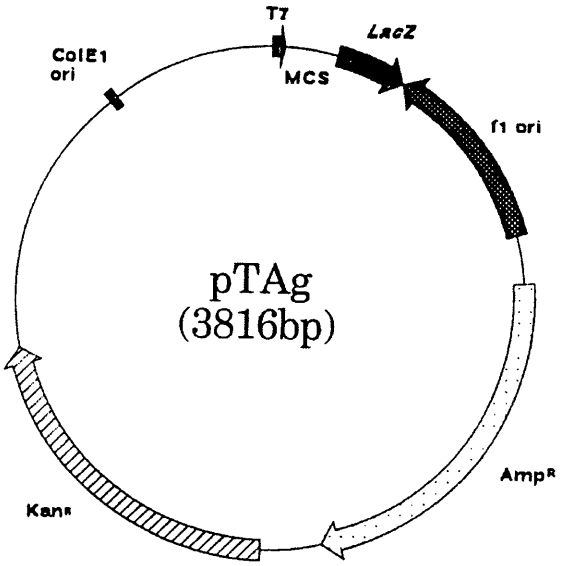
Oligonucleotide primers were designed taking the following parameters into consideration: the G+C content should be around 50% of the total bases; there should be minimal secondary structure; the primer pairs should have low complementarity to each other, especially in the 3' region and the primer length should be around 19-20 bases. The primers were all designed 'by eye' although there are now programs available for this purpose.

2.5 Agarose gel electrophoresis

PCR products were routinely visualised and sized by agarose gel electrophoresis. Gels were prepared with low melting temperature agarose at a concentration of 1.5% and run in 1 X TAE buffer with ϕX174/*Hae*III molecular size standards (Promega). After electrophoresis the gel was stained in a solution of ethidium bromide (0.5μg/ml) for approximately 10 minutes before visualisation and photography on an ultraviolet transilluminator.

2.6 Cloning of PCR products

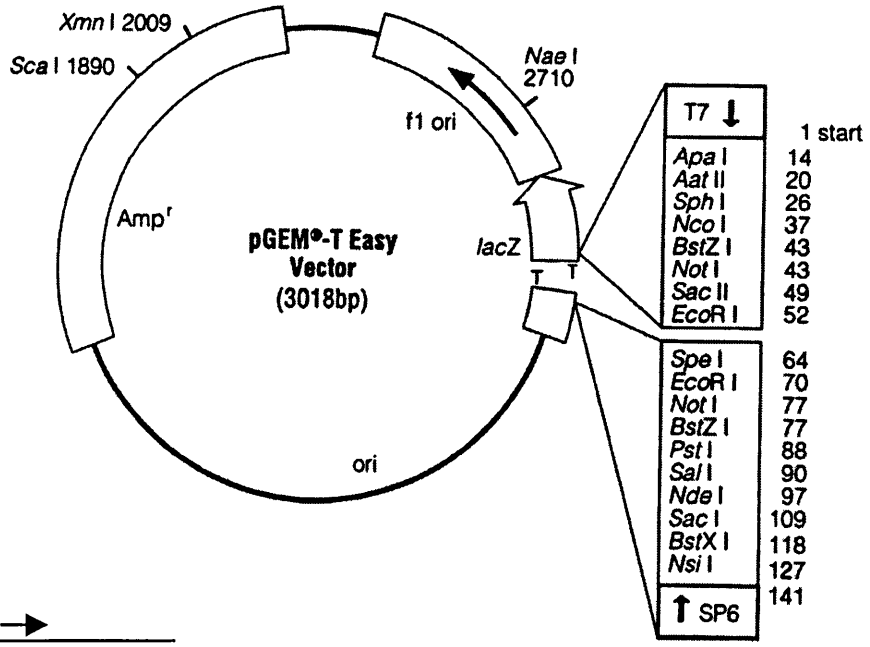
The pTAg (R & D Systems) and pGEM®-T Easy Vector System (Promega) were used for the cloning of PCR products. These plasmids possesses 3'-T overhangs at the insertion site and greatly enhance the efficiency of ligation into a plasmid (figure 2.1). PCR reactions which showed a single clean band after electrophoresis, were ligated directly into the pGEM vector. If however, multiple bands were present, the desired band was excised form the agarose using a scalpel blade and placed in a 1.5ml tube. The gel band was washed briefly in 500μl distilled water and then suspended in a further 50μl and left over-night at 4°C to elute the DNA. The ligation reactions were set up using 1μl 10X ligase buffer, 50ng pGEM®-T Easy Vector, 3 Weiss units T4 DNA ligase, 1μl PCR product or 5μl of an overnight eluate, and distilled water to a final volume of 10μl. The reactions were mixed and incubated overnight at 4°C.



pTAg 5' primer → T7 promoter →

GCTATG ACCATG ATTACG CCAAGC TCTAAT ACGACT CACTAT AGGGAA AGCTCG GTACCA
CGCATG CTGCAG ACGCGT TACGTA TCGGAT CCAGAA TTCGTG ATT//A TCTGTC GACAAG
CTTCTC GAGCCT AGGCTA GCTCTA GACCAC ACGTGT GGGGGC CCGAGC TCGCGG CCGCAC
AATTCA CTGGCC GTCGTT TTACAA

← pTAg 3' primer



pTAg 3' →

TTGTAA AACGAC GGCCAG TGAATT GTAATA CGACTC ACTATA GGGCGA ATTGGG CCCGAC
GTCGCA TGCTCC CGGCCG CCATGG CGGCCG CGGGAA TTCGAT //ATCA CTAGTG AATTCG
CGGCCG CCTGCA GGTCGA CCATAT GGGAGA GCTCCC AACGCG TTGGAT GCATAG CTTGAG
TATTCT ATAGTG TCACCT AAATAG CTTGGC GTAATC ATGGTC ATAGCT

← pTAg 5'

Figure 2.1: Vectors used for cloning of PCR products

For the transformation reaction, 1µl of the ligation mix was added to 25µl of JM109 competent cells, which had been thawed on ice. This was incubated on ice for 20 minutes before a heat shock at 42°C for 50 seconds. The tube was then returned to ice for 2 minutes after which 500µl of SOC medium was added. The transformation mix was incubated at 37°C with shaking for 90 minutes, plated onto LB plates containing ampicillin, IPTG and X-Gal and incubated at 37°C overnight. The following day positive, white colonies were screened for the presence of the desired insert by colony PCR.

2.7 Colony PCR

A reaction mix to screen 40 colonies was prepared containing the following: 10X NH_4 reaction buffer, 40µl; 64µl dNTP mix, 8units *Taq* polymerase; 20pmol of each primer, pTag 3' and pTag 5' (table 2.1); 600nmol $MgCl_2$ and distilled water to 400µl. This was distributed into 10µl aliquots, each of which was inoculated with a desired colony. The reaction conditions used were: one cycle of 94°C/90 seconds, 50°C/30 seconds and 72°C/2 minutes; followed by 30 cycles of 94°C/20 seconds, 50°C/20 seconds and 72°C/2 minutes. The PCR products were loaded on to a 1.5% low melting temperature agarose gel, electrophoresed and visualised. Clones containing an insert of the correct size were selected for mini plasmid preparations.

Primer	**Sequence**
pTAg 5'	5'-GCTATGACCATGATTACGCCAA-3'
pTAg 3'	5'-TGTAAAACGACGGCCAGTGAA-3'

Table 2.1: Primers used for colony PCR and sequencing of inserts in pTAg and pGEM®-T Easy vectors.

2.8 Plasmid mini and maxi preparations

For the mini-preps, a 5ml overnight broth culture of LB plus ampicillin (100μg/ml) was pelleted by centrifugation at 4,000rpm for 10 minutes in the swing-out rotor of a benchtop preparative centrifuge. Small-scale plasmid isolations were then prepared using the Wizard (Promega) or Qiaquick (Qiagen) mini prep kits following the manufacturer's standard protocols. Plasmid DNAs were eluted into 50μl of sterile distilled water or 10mM Tris, pH 8.5 and stored at -20°C.

For larger scale plasmid preps, the Qiagen maxi kit was used. A 5ml starter culture was grown for 8 hours and used to inoculate a flask containing 100ml LB plus ampicillin. This was grown overnight with shaking at 37°C and was then centrifuged as above to pellet the bacteria. Again, the manufacturer's instructions were followed and the resulting pellet of plasmid DNA was resuspended in 200μl of 10mM Tris, pH8.5. The DNA was quantified by spectrophotometry at 260nm and stored at -20°C.

2.9 DNA sequencing

A majority of the sequencing in this project was performed using the ABI PRISM® BigDye cycle sequencing ready reaction kit and an ABI 373a sequencer. However, early sequencing was carried out either manually or on a Pharmacia ALF or ABI 373a using the older PRISM® kits as described below.

2.9.1 Manual sequencing using the Sanger dideoxy chain termination method

The method used was modified from that described by Sanger *et al* (1997) and involved the use of [α-^{35}S]dATP and a genetically engineered T7 polymerase (Pharmacia Biotech) which lacks 3' to 5' exonuclease activity. 1.0-2.0μg of plasmid DNA, in a total volume of 32μl, was denatured with 8μl 2M NaOH at room temperature for 10 minutes. To precipitate the DNA, 7μl 3M Na acetate (pH4.8) and

4µl distilled water were added followed by 120µl 100% absolute ethanol. The solution was mixed, placed at -20°C for 30 minutes and spun at 13,000rpm in a microcentrifuge to form a pellet. The DNA pellet was rinsed in 70% ethanol, air dried and resuspended in 10µl distilled water. The T7 sequencing kit (Amersham Pharmacia Biotech) was used for the sequencing reactions following manufacturer's instructions. 3µl of the final mix was loaded onto a 6% polyacrylamide gel.

2.9.2 Polyacrylamide gel electrophoresis for sequencing

Sequencing gels were run using the BioRad Sequi-Gen GT system. The plate attached to the buffer chamber was siliconised with Sigmacote (Sigma), to allow for easy removal of the gel. For the 38 X 50cm plates a gel was prepared containing the following: 132ml 8.3M urea, 20ml 10X TBE and 48ml 25% Sequagel concentrate (19:1 acrylamide: bisacrylamide, National Diagnostics). To 50ml of this mix, 300µl each of 25% ammonium persulphate and TEMED were added, quickly mixed and used to pour the plug. For the gel, 600µl of 25% ammonium persulphate and 66µl TEMED were added to the remaining 150ml. The gel was run in 1 X TBE buffer at approximately 100W for 2 - 3 hours. When the gel temperature was stable at around 55°C, samples were heat denatured at 90°C for 5 minutes and loaded. After electrophoresis, the gel was fixed in a solution of 10% methanol/ 10% acetic acid for 5 minutes. A sheet of Whatman 3MM was placed over the gel and gently pressed down to allow attachment of the gel to the paper. The gel was then 'peeled' off from the plate, covered with cling film and heat dried under vacuum for 3 - 4 hours before overnight exposure to Fuji RX film. The following day the x-ray film was developed using an automatic film processor and the sequence read.

2.9.3 Automated sequencing using the ABI 373a sequencer

Automated sequencing involves the use of flourescent dyes which are detected when activated by a laser thus alleviating the need to use radioisotopes. A number of commercially available kits were available for use with the sequencer but since a

recent upgrade, only the BigDye Terminator cycle sequencing ready reaction kit can be used. The original kits were the ABI PRISM™ dye terminator cycle sequencing kit which used AmpliTaq® and the ABI PRISM™ dye terminator cycle sequencing ready reaction kit which used AmpliTaq® DNA polymerase, FS. The BigDye also uses AmpliTaq® DNA polymerase, FS to give a much more even peak intensity pattern and also to virtually eliminate the the 5'→3' nuclease activity of AmpliTaq DNA polymerase in the older kits. The BigDye terminators contain highly sensitive dyes which are 2-3 times brighter than the rhodamine in the PRISM™ kits and this, coupled with the reduction in background noise, gives a greatly improved level of sensitivity.

For the Dye Terminator kits the reactions were set used up as follows: reaction mix 9.5µl with AmpliTaq® DNA polymerase or 8.0µl with AmpliTaq® DNA polymerase FS, 3.2µmol sequencing primer, 1µg template DNA and distilled water to a final volume of 20µl. The reactions were cycled on a Perkin Elmer 9600 PCR machine using the following parameters for a total of 25 cycles: 96°C/ 10 seconds, 50°C/ 5 seconds and 60°C/ 4 minutes. Unincorporated dye was removed prior to loading on the gel by precipitation with 3µl of 2M Na acetate and 50µl of absolute ethanol on ice for 10 minutes. The sample was centrfuged at 13,000RPM for 15 minutes and the pellet rinsed in 70% ethanol before being vacuum dried and resuspended in 4µl loading dye. For the kit which used AmpliTaq® as opposed to AmpliTaq® FS, an additional phenol/chloroform extraction was performed prior to precipitation.

The BigDye™ reaction was set up using 4µl of sequencing mix, 500ng DNA (approximately 2.5µl of a Qiagen plasmid mini prep), 1.6pmol of primer and distilled water to a final volume of 10µl. The mix was cycled as follows: 96°C/ 10 seconds, 50°C/ 5 seconds and 60°C/ 4 minutes, for a total of 25 cycles. The samples were ethanol precipitated by adding 25µl of absolute ethanol and incubated on ice for 10 minutes. They were spun at 13,000RPM for 15 minutes, washed in 150µl 70% ethanol and re-spun for 10 minutes. The ethanol was aspirated and the pellet vacuum dried for 6 minutes. 3µl of loading dye was added and the samples denatured at 93°C for 3 minutes immediately prior to loading.

Sequencing gels for the ABI were prepared using 40ml of Sequagel 6, 10ml of Sequagel complete buffer reagent (National diagnostics) and 40mg of ammonium persulphate. The running buffer was pre-mixed TBE (National Diagnostics) diluted to a final concentration of 1 X and the gel was run for 12 hours. Collected data were analysed using MacVector™ versions 3.5, 4.0, 4.1 and 6.0 (Oxford Molecular Group, 1996) and GeneWorks™ version 4.45 (Oxford Molecular Group, 1997), and sequence alignments performed using AssemblyLIGN™ versions 1.0.6 and 1.0.7. (Oxford Molecular Group, 1996).

2.9.4 Automated sequencing using the Pharmacia ALF

Reactions for ALF sequencing utilise standard dideoxy sequencing methods. The primer is extended by T7 DNA polymerase in four separate dideoxy reactions (A,C,G and T) creating four separate populations of flourescently labelled, chain terminated fragments. The AutoRead™ sequencing kit (Pharmacia) was used together with either a Fluore-dATP labelling mix or primers labelled with 5'-Fluorescein. Manufacturers instructions were followed exactly. A 6% gel mix, prepared using acrylamide:bis (19:1, BioRad), 7M urea and 0.6X TBE, was deionised by shaking with ion exchange resin (BioRad) and filtered. 50ml of this was mixed with 50mg of APS and 23µl TEMED and used to pour the gel. The gel was pre-run in 1X TBE until it stabilised at 45°C before the denatured samples were loaded.

Sequence analysis was performed using MacVector™, GeneWorks™ and AssemblyLIGN™ as detailed above.

2.10 Sequence alignment and analysis

Sequence alignments were performed with ClustalV (Higgins *et al.*, 1992) using fixed gap and floating penalties of 10 and unweighted for transitions. Geneworks™ was used to search for conserved motifs and MacVector™ for the production of dot matrix plots. Common transcription factor binding sites were identified with the aid of the

Transcription Element Search System (TESS) database available via the Internet (http://www.cbil.upenn.edu/tess/index.html) (Schug & Overton, 1997).

2.11 Isolation of polyA+ mRNA

The QuickPrep™ micro mRNA purification kit (Pharmacia) was used for the direct isolation of polyadenylated RNA from retinal tissue. Retinae from six fish were homogenised in 400µl extraction buffer. 800µl of elution buffer was added and the mix transferred to a microcentrifuge tube. Samples were spun at 13,000RPM for one minute in an MSE Microcentaur. In a separate tube, 1ml of oligo(dT)-cellulose (25mg/ml suspended in a storage buffer containing 0.15% Kathon CG®) was spun as previously and the supernatant discarded. The supernatant from the retina preparation was used to resuspend the oligo(dT)-cellulose pellet and the suspension mixed gently for three minutes by inverting the tube. Samples were centrifuged for 30 seconds at 13,000RPM and the supernatant discarded. The pellet was resuspended and washed in 1ml of high salt buffer and spun for 30 seconds at 13,000RPM. The supernatant was again discarded and the washes repeated for a total of five times. Pellets were then resuspended in 1ml of low salt buffer, centrifuged for 30 seconds at 13,000RPM and the supernatant discarded. The low salt wash was repeated for a total to two washes. Pellets were resuspended in 300µl low salt buffer and the slurry transferred to a micro spin column fitted into a microcentrifuge tube. The column was spun for 5 seconds at 13,000RPM and the pellet washed further with 3 X 0.5ml of low salt buffer, centrifuging in between each wash. The eluate was discarded and the column given a final spin to remove any traces of buffer. mRNA was eluted from the column using 0.2ml of elution buffer prewarmed to 65°C

2.12 Production and digestion of goldfish green cone cDNAs

The Quick Prep Micro mRNA Purification kit (Pharmacia) was used for the isolation of polyA$^+$ RNA from fresh retinal tissue of six goldfish as above. cDNA was

synthesised using the 3' RACE system (Gibco BRL). An 819bp fragment of both green cone cDNAs was amplified with a forward and reverse primer pair:

Primer	**Sequence**
Forward	5'-GTCAACCTGGCTGTGGCTGGT-3'
Reverse	5'-TGCTGGAGATACAGAGGACA-3'

Table 2.2: Primers used to amplify goldfish green cone opsins

Half of the forward primer was end labelled with [γ32P]dATP using T4 polynucleotide kinase in a 10X reaction buffer. 200ng of each primer was used in a PCR reaction containing 100 ng cDNA, 0.2 mM of dATP, dCTP, dGTP and dTTP, 1.5 mM $MgCl_2$, 0.5u of *Taq* polymerase and NH_4 reaction buffer in a total volume of 25 µl to amplify an 819bp fragment. The PCR parameters were, 94°C for 1 min, 35 cycles at 94°C for 20 sec, 58°C for 20 sec, and 72°C for 20 sec, followed by a final extension of 10 min at 72°C. The resulting product, a mixture of amplified fragments from both green cone opsin cDNAs, was purified using a QIAquick PCR purification kit (Qiagen) and aliquots digested to completion with 10u of *Sac*I or *Eco*RI in 1 X One-Phor-All buffer (Pharmacia). Half of each digest was electrophoresed as previously and the gel stained with ethidium bromide; an additional 5u of restriction enzyme was added to the remaining aliquot and the sample re-incubated and electrophoresed to ensure that the samples had been digested to completion. The relative quantity of DNA in each of the 819bp bands was determined by scintillation counting. The bands were excised and melted at 60°C in 500µl H_2O. 5ml of Ecoscint A (National Diagnostics) was added and the samples counted in a Packard Liquid Scintillation Analyser. The results were analysed for any difference between the means of paired samples using the t-test (Welkowitz *et al.*, 2000), and the ratio of green 1 to green 2 expression was determined.

2.13 Determination of the transcription start site using 5' RACE (rapid amplification of cDNA ends)

RACE allows the amplification of unknown sequences at the 5' and 3' ends of the mRNA to obtain the full-length cDNA. Using this method, the 5' end, and hence the transcription start site, of the goldfish green 1 and 2 and the rod opsin genes was determined.

Goldfish retinal poly A^+ RNA was isolated as above and used to produce the first strand cDNA. A reaction mix was prepared containing the following: 4µl synthesis buffer; 2µl each of 10mM dATP, dCTP, dGTP and dTTP in Tris-HCl pH 7.5 @20°C; 1µ of 12.5µM cDNA synthesis primer (green 1, green 2 and rod, table 2.3); AMV reverse transcriptase 20 units (in 200mM potassium phosphate, 2mM dithiotreitol, 0.2% (v/v) Triton® X-100, 50% glycerol (v/v), pH 7.2), poly A^+ RNA 1µl, H_2O to 20µl. The samples were incubated at 55°C for 60 minutes followed by 65°C for 10 minutes and then purified using a High Pure PCR product purification kit (Gibco BRL). A poly A^+ tail was added to the 5' end of the first strand cDNA using terminal transferase as follows: purified cDNA 19µl, 10X reaction buffer 2.5µl, 2mM dATP 2.5µ. The mix was incubated at 94°C for 3 minutes then chilled on ice. 10 units of terminal transferase were added and the mix incubated at 37°C for 20 minutes followed by 70°C for 10 minutes. The dA-tailed cDNA was amplified using PCR with nested gene specific primers. 5µl of the tailed cDNA was mixed with 1µl of the oligo d(T)-anchor primer (37.5µM), 1µl of the nested specific primer (green 1 N1, green 2 N1 and rod N1, 12.5µM), 1µl deoxynucleotide mix (as above), 5µl 10X reaction buffer and H_2O to a final volume of 50µl. The PCR conditions used were: 94°C for 5 minutes followed by 35 cycles of 94°C for 30 seconds, 54°C (for rod) or 62°C (for green 1&2) for 20 seconds, 72°C for 10 seconds and a final elongation of 72°C for 7 minutes. A 1µl aliquot of the primary PCR was used with a further set of nested primers (green 1 N2, green 2 N2 and rod N2 together with an anchor primer, 12.5µM, table 2.3) in a second round of PCR to obtain a specific product. The PCR conditions used were: 94°C for 5 minutes; 35 cycles of 94°C for 30 seconds, 59°C (for

Primer	Sequence
Green 1	5'-CTCACTAGCCCGGTCCTGTT-3'
Green 2	5'-CTCACTAGACCTGTCCTGTT-3'
Rod	5'-TGCCAGTGGCATTGGACATA-3'
Oligo dT	5'-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV-3'
Green 1 N1	5'-GGACATGGGGACGTAGAAGT-3'
Green 2 N1	5'-GGACAAGGGGACGTAGAAGT-3'
Rod N1	5'-ACGTAGAACATATCTCCCCTC-3'
Green 1 N2	5'-TTCCCTCAGTGCCCTTCAT-3'
Green 2 N2	5'-TTCCCTCAGTGCCATTCAT-3'
Rod N2	5'-CTCCCTCTGTACCGTTCAT-3'
Anchor	5'-GACCACGCGTATCGATGTCGAC-3'

Table 2.3: Primers used in 5' RACE for amplification of goldfish rod and green cDNAs.

rod) or 63°C (for green 1&2) for 20 seconds, 72°C for 5 seconds and a final elongation of 72°C for 7 minutes. The resulting product was cloned into the pGEM®-T Easy Vector System (Promega) using T4 DNA ligase and a 10X ligation buffer (300mM Tris-HCl pH 7.8, 100mM $MgCl_2$, 100mM DTT, 10mM ATP). After transformation into JM109 competent cells and selection on LB agar plates containing ampicillin (100ug/ml), X-Gal (80ug/ml) and IPTG (0.5mM), white positive colonies were picked and grown overnight in LB broth plus ampicillin. The resultant culture was used for the preparation of plasmid DNA using the Qiaprep spin miniprep kit (Qiagen). Ten clones of each sample were sequenced and compared. The transcription start site was positively identified when the sequence in at least three clones terminated at the same nucleotide position.

2.14 Amplification and sequencing of the 5' flanking regions of the rod and green cone opsin genes

The method of walking PCR (Dominguez & Lopez-Larrea, 1994) was used to amplify the upstream regions of the rod and green cone opsin genes from genomic DNA. The gene-specific and universal primers used are listed in table 2.4. The outer reverse gene-specific primers, complementary to regions of exon 1 of the rod and green cone opsin genes respectively, were used with each of three universal primers, UNI33, UNI28 and WP33. Primary annealing and elongation steps were carried out at very low stringency to allow annealing of the universal primer to an upstream region. The annealing temperature (T_a) was then increased to match the T_a of the outer gene-specific primer. All PCRs were performed in triplicate. PCRs were carried out as follows: 100 ng template DNA was added to a mixture of 0.5 mM universal primer, 0.2 mM of dATP, dCTP, dGTP and dTTP, 1.5 or 3.0 mM $MgCl_2$ and 10 X NH_4 reaction buffer in a total volume of 23 µl. After 1 min at 94°C, 0.5u *Taq* polymerase was added and the temperature was slowly raised in steps of 2 min at 15°C, 10 min at 25°C, 1 min at 72°C and 1 min at 90°C. 0.5 mM of the outer gene-specific primer was then added (to give a final volume of 25 µl) and denatured at 94°C for 1 sec, followed

A

PRIMER	SEQUENCE
UNI33	5'-TTTTTTTTTTTTTTGTTTGTTGTGGGGGGGTT-3'
WP33	5'-TTTTTTTTTTTTTTCTTTCTTCTCCCCCCCTT-3'
UNI28	5'-TTTTTTTTTTGTTTGTTGTGGGGGTGT-3'
UNI17	5'-TTTTTGTTTGTTGTGGG-3'

B

PRIMER	SEQUENCE	T_a°C
1st walk rod out	5'-CAGGCAGGCGTATGCCCATGGC-3'	74
1st walk rod in	5'-GTAGTACTGGGGATAGTCGTAC-3'	66
2nd walk rod out	5'-CGGCAGGACTGTGGAGAAACGAG-3'	74
2nd walk rod in	5'-ATGTTGTGCTCAGCCCCTTTCTG-3'	70
3rd walk rod out	5'-GCTATTGTGATGACCTGTCC-3'	60
3rd walk rod in	5'-CTGGGATCTTGGAAGATATA-3'	56

C

PRIMER	SEQUENCE	T_a°C
1st walk green 1&2 out	5'-ACCAGCCACAGCCAGGTTGAC-3'	68
1st walk green 1 in	5'-TTTGTGTTGAGCTGTAACC-3'	54
1st walk green 2 in	5'-TTTGTGTTGAGCTGTACAAA-3'	54
2nd walk green 1 out	5'-TACCCGGTAGAGCTTTGGATC-3'	64
2nd walk green 2 out	5'-CACCCAGTAAAGCTTGGGATC-3'	64
2nd walk green 1 in	5'-TAGGCCACGATTTTGATTG-3'	54
2nd walk green 2 in	5'-TTGGTTGTCAGTACAAAGTG-3'	56

Table 2.4: Primers used for walking PCR. **A**: universal walking primers. **B**: goldfish rod opsin walking primers. **C**: goldfish green opsin 1 & 2 walking primers. Out is the outer primer, in is the inner nested primer. T_a is the annealing temperature (calculated where A/T = 2°C, C/G = 4°C).

by 30 cycles of 94°C for 10 sec, 30 sec at the T_a of the outer gene-specific primer, and 72°C for 30 sec, with a final extension of 1 min at 72°C.

A smear of amplified product could be visualised after electrophoresis on a 0.5% low melting point agarose gel using a 20 mM Tris-acetate, 0.5 mM EDTA buffer pH 8.0 and ethidium bromide staining. An aliquot of the primary PCR was then diluted 1 in 1000 and used in a secondary PCR with nested primers as follows: 1µl diluted primary PCR, 0.5 mM nested universal primer (UNI17), 0.5mM inner reverse gene-specific primer (table 2.4), 0.2 mM of dATP, dCTP, dGTP and dTTP, 1.5 or 3.0 mM $MgCl_2$ and 10 X NH_4 reaction buffer in a total volume of 24 µl. After 10 sec at 94°C, 0.25u *Taq* polymerase was added. The reaction was then cycled 30 times at 94°C for 5 sec, 1 sec at the T_a of the inner gene-specific primer, 72°C for 30 sec, followed by a final extension of 1 min at 72°C. A clear band could be visualised after separating the secondary PCR product on a 0.5% low melting point agarose gel as before. The band was cut from the gel and the DNA eluted in 50 µl of water overnight. This was then sequenced either manually, directly on an Applied Biosystems 373A sequencer using dye terminator cycle sequencing with a Ready Reaction Kit (Perkin Elmer) or, after cloning into pTag (R&D Systems Ltd), on a Pharmacia ALF sequencer using Autoread and Autofluor kits. Cloned fragments were sequenced with fluorescently labelled pTag 3' and 5' primers (table 2.1) plus internal sequencing primers complementary to regions of the upstream sequences. To extend the walk upstream, new gene-specific primers were designed close to the 5' end of the sequence and the process repeated.

2.15 Sequencing the blue opsin gene from the Cottoid *fish of Lake Baikal*

Fish were collected either by trawling Lake Baikal at night at various depths down to 1200m or by local fishermen who netted inshore species in the early hours of the morning. The eyes were enucleated and extracts used for microspectrophotometry (Bowmaker *et al.*, 1994). Genomic DNA was isolated from fresh liver tissue by standard phenol-chloroform methods and transported back to the Institute of Ophthalmology for use in the PCR and for sequence analysis.

Samples of *Cottus gobio*, a cottoid species (commonly known as a bullhead) native to Britain, were kindly donated by the Institute for Freshwater Ecology in Dorset. Fish were dark adapted overnight and the eyes enucleated under dim red light. Retinae were removed and used in the preparation of poly A+ RNA and synthesis of the first strand cDNA by 3' RACE.

Numerous combinations of the 3' anchor primer and primers designed to the goldfish blue opsin sequence were used together with various *Taq* polymerases and reaction buffers to amplify a fragment of the bullhead cDNA. Eventually, a primer designed from exon 1 of the goldfish blue gene (GFB1+) paired with the 3' anchor primer gave a smear when run on an agarose gel. A set of nested primers, GFB2+ and GFB4-, were then used to obtain a discreet band which was eluted from the agarose and cloned into the pGEM-T easy cloning vector. Primers used for 3' RACE are shown in table 2.5. The cloned product was sequenced using the ABI PRISM™ dye terminator cycle sequencing ready reaction kit. From the bullhead blue opsin sequence obtained, exon specific primers were designed to amplify the complementary sequences from the Lake Baikal Cottoid genomic DNA. The 5' sequence was produced using the walking PCR method as described previously. All further primers used for amplification and sequencing are detailed in tables 2.5 and 2.6.

Translation of nucleotide sequences and alignments were performed using Geneworks™ and MacVector™.

2.16 In vitro mutagenesis

In order to generate oligonucleotide-directed mutants, the Promega Altered Sites® II and the Stratagene QuikChange™ site-directed mutagenesis systems were used. Section 2.17 details the Altered Sites® II method and section 2.18, the QuikChange™ method. A summary of the steps involved in site-directed mutagenisis is shown in figure 2.2. Preliminary steps to prepare the DNA for mutagenesis are described below.

Primer	Sequence
Oligo d(T)-anchor	5'-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTT-3'
Anchor primer	5'-GACCACGCGTATCGATGTCGAC-3'
GFB 1+	5'-ACAACCTCTCAGCCTACAGC-3'
GFB2+	5'-TGGCTGGTCATTTGCAAACC-3'
GFB4-	5'-CGTCACCTCCCTCTCTGCCT-3'
BHB1+	5'-CCTAATTTACATCCAGGTC-3'
BHB1a+	5'-TCTCATAGAATGTGTCTC-3'
BHB2+	5'-GTCATTTGCAAACCACTTGG-3'
BHB2-	5'-CATAGCATGGTCGGGCTTG-3'
BHB3+	5'-GTACATCCCAGAAGGCCTC-3'
BHB3-	5'-TGCAGGCCTTCTGGGATG-3'
BHB4+	5'-CGTCATCTATGTTCTCCTC-3'
BHB4-	5'-GGGTGGAGGCAGATTCTGC-3'
BHB4a-	5'-GGGTGGAGGCAGATTCTGC-3'
BHB4b-	5'-CACAGGAGCACTGCAGGCC-3'
BHB5-	5'-TCATGCATGACCGGAACTG-3'
BHB5a-	5'-GTGGCAGAGAAGTCAATGG-3'

Table 2.5: Primers used to amplify the blue cone opsin from *Cottus gobio* cDNA, *Cottus kesslerei* gDNA, *Cottocomephorus inermis* gDNA, *Limnocottus eurystomas* gDNA and *Batrachocottus nicolskii* gDNA. Oligo d(T) and anchor primer were used for 3' RACE. GFB primers were designed from goldfish blue cone opsin sequence and BHB from *Cottus gobio* blue cone opsin.

Primer	Sequence	
1st walk CK out	5'-CACAGGAGCACTGCAGGCC-3'	64°C
1st walk CK in	5'-GGGCTTGAAAACAAAGGTA-3'	54°C

Primer	Sequence
CK internal+	5'-TGGGCTGTTGTCATCTGTA-3'
CK internal-	5'-ATCTTGCATGCTAGTGGTC-3'
BHB internal+	5'-CAATACCGTCCGTCTTTTCC-3'

Table 2.6: Primers used for walking PCR and sequencing of the blue cone opsin gene from cottoid fishes. Walk = primers used for walking PCR. Internal = primers used for sequencing cloned PCR products.

The *Cottus gobio* blue opsin cDNA was used as the wild type template for creation of the mutants. A PCR reaction was performed using the cDNA with an annealing temperature of 60°C and primers which add restriction sites for Eco RI and Sal I as shown below:

Primer	Sequence
BULL MUT +	5'-GCGCGAATTCCACC**ATGAAGCACGGTCGTGTCACGGAG**-3'
BULL MUT -	5'-CGGCGTCGACGC**AGCAGGCCCAACTTTGGAGAC**-3'

Table 2.7: Primers used to add *Eco RI* and *Sal I* restriction sites to *Cottus gobio* blue opsin cDNA sequence. Sequence underlined shows the restriction sites, *Eco RI* in the + primer, *Sal I* in the – primer. Coding sequence of the blue opsin gene in *Cottus gobio* is shown in bold.

For expression of the opsin protein, the *Cottus gobio* blue cDNA was cloned into an expression vector pMT4. This vector is a derivative of the mammalian expression vector pMT2 and additionally carries sequence of the bovine rhodopsin 1D4 epitope, including the stop codon, downstream and in frame of the Sal I site (Franke *et al.*, 1988). Both the pMT4 vector and the *Cottus gobio* blue cDNA in pGEM®-T Easy, were digested with *EcoRI* and *Sal I* in a one-phor-all buffer (Pharmacia) according to manufacturers instructions. The digests were electrophoresed on a 1.5% low melting temperature agarose gel run in 1 X TAE buffer and the products visualised after staining with ethidium bromide. The desired bands were excised from the gel and the DNA purified using a QIAquick gel extraction kit (Qiagen). DNA was eluted into 30µl of elution buffer and stored at -20°C. The *Cottus gobio* blue cDNA was then ligated into the pMT4 vector as follows: 1µl of 10X ligation buffer, 1.5units of T4 DNA ligase, 1µl pMT4 and 7.5µl *Cottus gobio* blue cDNA. The mix was incubated

at 4°C overnight and used in a transformation reaction as described above but with DH5α competent cells and plating on to tetracycline LB plates. White colonies were screened for the presence of the correct insert using colony PCR and BHB2+ and BHB5- primers (table 2.4). Three positive colonies were selected and plasmid DNA isolated from each using the QIAprep mini prep kit (Qiagen). Each clone was sequenced in both forward and reverse orientations using the primers listed below in table 2.8, with the ABI Big Dye™ in order to ensure that no mutations had been introduced during amplification. One clone was chosen for subsequent mutagenesis.

Primer	Sequence
1+	5'-ATGAAGCACGGTCGTGTCAC-3'
1-	5'-AGTTGAGGTGGGATCGGAGC-3'
2+	5'-GTCCTCCTCTGTTTGGATG-3'
2-	5'-CATAGCATGGTCGGGCTTG-3'
3+	5'-CCCCTCTTAATCCTTGTCT-3'
3-	5'-TGTAGCAACACTTGGCGG-3'
4+	5'-GGTCATCTATGTTCTCCTC-3'
4-	5'-ATTCTGCTTGGGCCTTTGC-3'
5-	5'-AGGCCAACCCATTCCCAT-3'

Table 2.8: Primers used for sequencing cottoid blue opsins

2.17.1 in vitro mutagenisis - Promega Altered Sites® II

The pALTER®-1 vector has genes for both ampicillin and tetracycline resistance and the protocol alternately repairs one and inactivates the other, thus allowing subsequent rounds of mutagenesis on the same plasmid without subcloning.

2.17.1 Cloning of Cottus gobio blue cDNA into the mutation vector pALTER®-1

In order to mutate the bullhead cDNA, it was first cloned into the pALTER®-1 vector. The full-length cDNA was approximately 1200 base pairs in length. To increase the success rate of mutation and to make the task of sequencing easier, the bullhead DNA was digested with PstI to generate two fragments of similar size.

Digests were prepared using a mixture of SalI, EcoRI and PstI to release the bullhead DNA from pMT4 and to cut it into two fragments. Aliquots of the pALTER®-1 vector were also digested with SalI/PstI and EcoRI/PstI to generate the appropriate sticky ends for cloning. Each digest was run on a 1.5% low melting temperature agarose gel and the desired bands excised. DNA was purified from the agarose using the QIAquick gel extraction kit and eluted into 50µl of elution buffer. Ligation reactions were set up using 3µl of the vector, 1µl of the bullhead DNA fragment, 1.5 units T4 DNA ligase and 1µl of 10X ligation buffer in a total volume of 10µl. The samples were incubated overnight at 4°C and then used to transform JM109 competent cells (as described previously). Initially the pALTER®-1 vector is tetracycline resistant so transformants were plated on to LB agar plates containing tetracycline, IPTG and X-Gal. After incubation overnight at 37°C, positive white colonies were screened for the presence of an insert of the correct size by colony PCR using M13 forward and reverse primers. Positive clones containing each of the *Cottus gobio* DNA fragments were grown overnight in LB broth plus tetracycline and the plasmid DNA extracted using the QIA prep mini prep kit.

2.17.2 Design of mutagenic oligonucleotides

General considerations for the design of mutagenic primers are as follows: for single base mutations, the oligonucleotide should be 17-20 bases in length with the mismatch located in the centre; where two or more mismatches are to be created the primer needs to be longer, at least 25 bases. The oligonucleotide must also anneal to the same strand as the antibiotic repair oligonucleotide *i.e.* the antisense strand.

Primers were designed to create four separate mutants (table 2.9). Two were in helix three to change threonine at position 118 into glycine or alanine; one in helix five at 215 to change proline into glycine, and lastly, at position 269 in helix six, to change threonine to alanine. The primers used are shown below with the mutations to be introduced in bold type; BULL 5 GLY and BULL 6 ALA were both PAGE purified.

Primer	Sequence
BULL 3 GLY	5'-GAAGGTTTTTTAGCA**GG**ACTTGGCGGTATG-3'
BULL 3 ALA	5'-GTTTTTTAGCA**G**CACTTGGCG-3'
BULL 5 GLY	5'-CCTGCCACTTTACTGTT**GG**CCTCTTAATCCTTG-3'
BULL 6 ALA	5'-GCTGGATGCCTTAC**G**CCTGCTTTGCGCT-3'

Table 2.9: Mutagenic oligonucleotide primers used with Altered Sites® II. Nucleotides in bold indicate mutations to be introduced.

2.17.3 Phosphorylation of oligonucleotides

Prior to use the primers were 5'-phosphorylated with T4 polynucleotide kinase. 200pmol of each primer was incubated at 37°C for 30 minutes with 10 units of T4 polynucleotide kinase (Advanced Biotechnologies), 0.5µl 100mM ATP and 10µl 5X

polynucleotide kinase buffer in a total volume of 50µl. The reaction was terminated by a further incubation at 70°C for 10 minutes.

2.17.4 Denaturation of double stranded DNA

The double stranded plasmid DNA was denatured before use in the mutagenesis reaction. 2µg of plasmid DNA (15µl) was mixed with 2µl of 2M NaOH/2mM EDTA in a total volume of 20µl and incubated at room temperature for 5 minutes. To precipitate the DNA, 1.3µl of 3M sodium acetate pH4.8 and 75µl 100% ethanol were added and the mix incubated at -70°C for 30 minutes before centrifugation in a microcentrifuge at 13,000rpm for 15 minutes. The pellet was stored overnight under ethanol at -70°C. The following day the sample was centrifuged as before, washed with 200µl 70% ethanol and the supernatant removed using a stretched Pasteur pipette. The pellet was dissolved in 100µl 10mM Tris.HCl pH 8.5 and a 10µl aliquot analysed on a 0.7% agarose gel together with a sample of the undenatured DNA to confirm that it had been denatured. The sample was kept on ice until the annealing reaction.

2.17.5 Annealing reaction and mutant strand synthesis

For annealing of the oligonucleotide and synthesis of the mutant DNA strand the protocol available with the Altered Sites II kit was followed. In the mutagenesis reaction, the denatured DNA is mixed with the mutagenic oligonucleotide, a repair oligonucleotide (which repairs the ampicillin resistance), a knockout oligonucleotide (which knocks out the tetracycline resistance) and an annealing buffer. The reaction mix is heated to 75°C and cooled slowly to room temperature to allow annealing of the oligonucleotides. The annealing reactions are then placed on ice and a synthesis buffer, T4 DNA polymerase and T4 DNA ligase are added. The mix is then incubated at 37°C for 90 minutes to allow synthesis and ligation of the mutant strand.

2.17.6 Transformation of ES1301 mutS *competent cells*

Manufacturers instructions were followed for the transformation although only 30µl of the ES1301 *mutS* cells were used per reaction. After heat shock and incubation on ice, 300µl of SOC broth was added and the reaction incubated at 37°C for 30 minutes. Overnight cultures were prepared by adding 300µl of each transformation to 5ml LB broth plus ampicillin and incubating at 37°C.

The following day, plasmid DNA was purified using the QIAprep mini prep kit and eluted into 50µl of elution buffer.

2.17.7 Transformation into JM109 competent cells

Because ES1301 *mutS* cells are *recA+*, inserts containing highly repetitive sequences may be unstable. As a consequence of this plasmids were transformed into JM109 cells for long-term maintenance after mutation. JM109 cells were thawed on ice and divided into 50µl aliquots. 0.2 and 1.0µl of each plasmid DNA preparation was added to the cells and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 50 seconds and returned to ice for 2 minutes. 500µl of SOC medium was added and the samples incubated at 37°C with shaking for 60 minutes. The 500µl was then spread onto LB plates containing ampicillin which were incubated overnight at 37°C.

2.17.8 Screening for positive mutants

Plasmid DNA was purified from five colonies of each plate and sequencing reactions prepared using the BigDye Terminator cycle sequencing ready reaction kit with SP6 and M13 forward primers. The mutants were identified by comparison of the sequence to that of the wild type bullhead blue opsin cDNA using AssemblyLIGN™.

2.18 In vitro mutagenesis - Stratagene QuikChange™

Creating mutations a sites 215 and 269 was not successful using the Altered Sites II kit, instead the QuikChange kit was used. This involves using complementary primers containing the desired mutation and a high fidelity polymerase, PfuTurbo.

2.18.1 Design of mutagenic oligonucleotides

General consideration for the design of mutagenic oligonucleotides for use with the QuikChange kit are as follows: both primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid; primers should be between 25 and 45 bases in length with a GC content of 40%, terminating in one or more G or C bases and have with a Tm of at least 78°C; the desired mutation should be in the middle of the primer and they must be PAGE purified. The primers are shown below with the mutations to be introduced, in bold type.

BULL 5 GLY-	5'-CAAGGATTAAGAGG**CC**AACAGTAAAGTGGCAGG-3'
BULL 5 GLY+	5'-CCTGCCACTTTACTGTT**GG**CCTCTTAATCCTTG-3'
BULL 6 ALA-	5'-GCGCAAAGCAGG**C**GTAAGGCATCCAGC-3'
BULL 6 ALA+	5'-GCTGGATGCCTTAC**G**CCTGCTTTGCGC-3'

Table 2.10: Mutagenic oligonucleotide primers used with QuikChange™. Nucleotides in bold indicate mutations to be introduced.

2.18.2 Annealing and mutant strand synthesis

To generate the mutant DNA strand, the protocol available with the QuikChange kit was followed. 50μg of DNA in the pALTER-1 plasmid (prepared as in section 2.17.1) was used in a reaction with the primer pairs, BULL 5 GLY-/BULL 5 GLY+ and BULL 6 ALA-/BULL 6 ALA+. The reaction mix was cycled as follows: denaturation at 95°C for 30 seconds followed by 16 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 68°C for 13 minutes.

2.18.3 Digestion of PCR products with Dpn I

The PCR product was digested with 10u of Dpn I at 37°C for 1 hour. This enzyme is specific for methylated and hemimethylated DNA and so digests the non-mutated parental DNA.

2.18.4 Transformation into JM109 competent cells

The mutated DNA is in the pALTER-1 vector, which is tetracycline resistant. Therefore, the Epicurian Coli XL1-Blue supplied with the kit could not be used because they too are tetracycline resistant. Instead, JM109 competent cells were used and the standard transformation procedure followed as described above (section 2.17.7). Cells were spread onto LB plates containing tetracycline, IPTG and X-Gal, and positive colonies were screened by colony PCR, using M13 forward and reverse primers, to ensure that they contained a DNA insert of the correct size. The mutated DNA was released from the pALTER-1 vector and ligated into pMT4.

2.19 Cloning of mutant DNA into the pMT4 expression vector

Aliquots of the pMT4 vector were digested with SalI/EcoRI to generate the appropriate sticky ends for cloning. The digests were run on a 1.5% low melting temperature agarose gel, the desired bands excised and the DNA purified using the

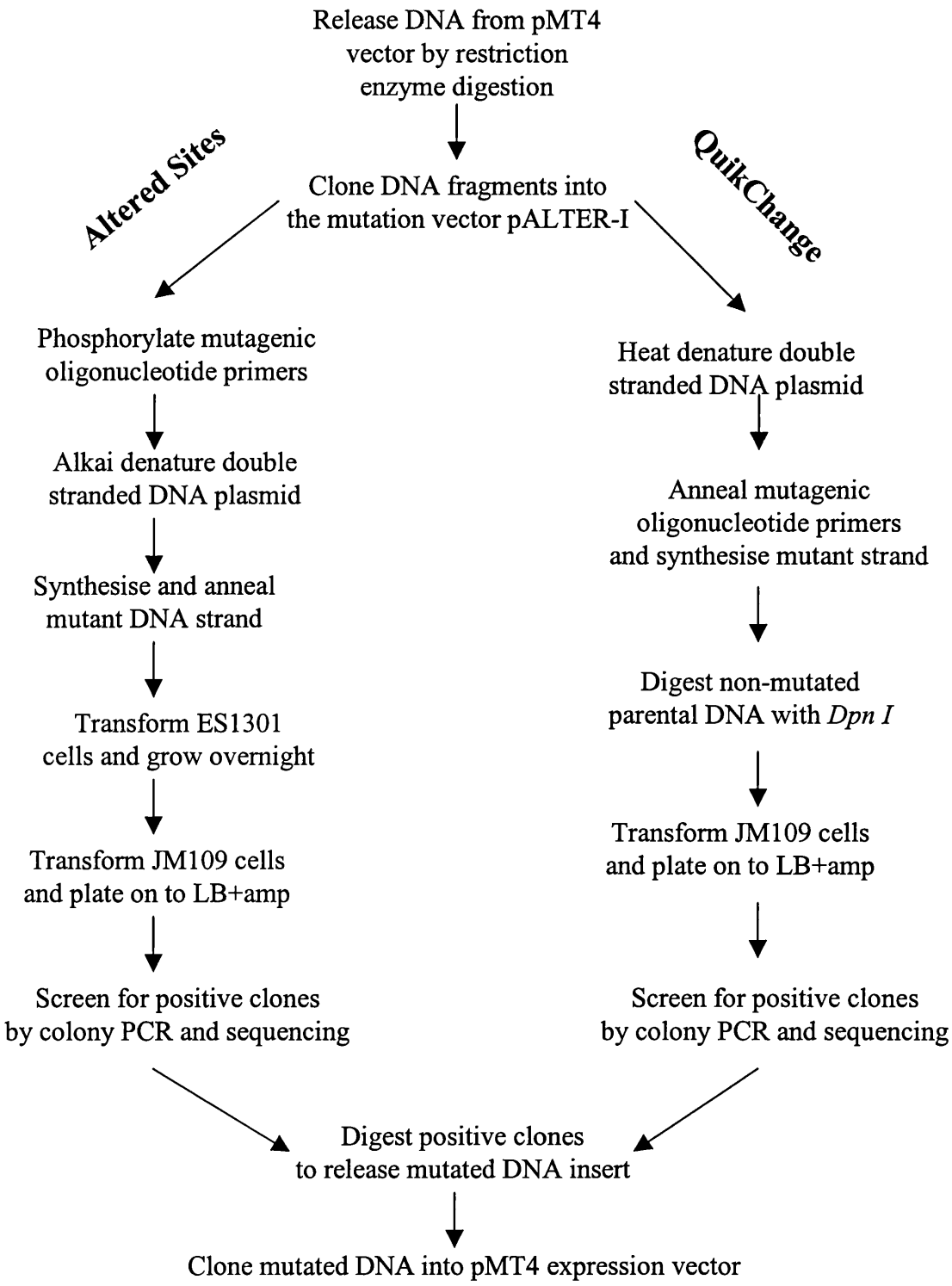


Figure 2.2: Flow diagram of steps involved in site directed mutagenisis using Altered Sites and QuikChange kits

QIAquick gel extraction kit. The mutated *Cottus gobio* DNA was released from the p-ALTER I vector by digestion with SalI/PstI and EcoRI/PstI and the DNA purified as above. These mutated fragments were ligated into pMT4 together with the unmutated gene fragment to create clones containing the full-length bullhead blue opsin cDNA sequence, each with a separate mutation.

2.20 Maxi plasmid preparations

The mutated bullhead DNA ligated into pMT4 was used to transform DH5α cells. For the bulk preparation of plasmid DNA, Qiagen maxi kits were used. Manufacturers protocols were followed using a 5ml starter culture inoculated into 100ml LB plus ampicillin. The resulting pellets were resuspended in 200μl of elution buffer and the DNA concentration estimated by determining the absorbance at 260nm.

2.21 Growth and maintenance of 293T cell line

293T cells were obtained from the European Collection of Cell Cultures, Salisbury, Wiltshire. All manipulations were carried out in a Microflow class II biological safety cabinet. Cells were routinely cultured in Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX II™ (Gibco BRL) supplemented with 10% foetal calf serum (Sigma) and 1% penicillin/streptomycin (Gibco BRL). Cultures were maintained at 37°C in a humidified incubator with 5% CO_2. The 293T cells were grown to 100% confluence before sub dividing. To sub divide the cells, the media was aspirated from each flask and 4ml PBS added. This was removed and 1ml of trypsin/EDTA (Gibco BRL) added, followed by incubation for 10 minutes at 37°C. Cells were resuspended by vigorous pipeting and mixed with 9ml of DMEM (including FCS and antibiotics). 1ml of the cell suspension was then added to 14ml fresh medium in a new 250ml flask. Media was replenished every 2 days until cell cultures were ready for splitting again. One flask was needed per transfection to produce 10 X 60mm plates.

2.22Transfection and harvesting of 293T cells

The day prior to transfection, each flask of cells was split and plated on to 10 x 60mm dishes as follows. Media was removed from the dishes and the cells washed with 4ml phosphate buffered saline (PBS). This was aspirated and 1ml of trypsin/EDTA (Gibco BRL) added. The cells were incubated at 37°C for five minutes until they began to detach from the flask, and were then transferred to a 15ml tube. DMEM (including FCS and antibiotics) was added to bring the volume to 10ml. 1ml of this suspension was added to each of 10 X 60mm dishes containing 8ml DMEM (including FCS and antibiotics). Cells were incubated for approximately 24 hours until they showed 50% confluence.

5μg of plasmid DNA was used per plate for the transfection. For 10 plates, 50μg of DNA was pooled with 2ml of DMEM (without additives). 400μl of Lipofectamine (Gibco BRL) was added and the mix vortexed briefly before incubation at room temperature for 45 minutes. Media was removed from the cells and each dish washed gently with 5ml DMEM (without additives). To the transfection mix, 48ml of DMEM (without additives) was added and 5ml of this transferred to each washed plate. Cells were incubated at 37°C for 24 hours. The media was removed and replaced by 8ml DMEM (including FCS and antibiotics) followed by incubation for a further 24 hours.

During incubation, cells should reach approximately 80% confluence for harvesting. Each plate was washed twice with 5ml PBS to remove traces of the media. 1ml of PBS was then added per plate and the cells suspended using a cell scraper. The 10 X 1ml aliquots were transferred to a 15ml tube, which was spun at 1000g for 5 minutes. The supernatant was removed and cells were resuspended in 15ml PBS. The cells were spun again and washed once more with 15ml PBS. After the final spin, the supernatant was removed and the pellet stored at -70°C.

2.23 Western blotting of expressed opsin proteins

Western blotting is the transfer of proteins separated by SDS-PAGE onto nitro-cellulose membranes. In order to confirm expression of the wild type and mutant opsin proteins, membranes were isolated from small-scale cell preparations and western blots produced to visualise expression levels. Cell pellets were produced as above for the wild type, each of the mutants and from untransfected cells, but from the equivalent of only two plates each. Pellets were resuspended in 2ml cold homogenisation buffer and incubated on ice for 30 minutes before being lysed by passing through an 0.5mm gauge needle five times. The lysate was spun at 2000rpm at 4°C for 5 minutes in a benchtop microcentrifuge and the supernatant transferred to a clean tube. NaCl was added to achieve a final concentration of 0.25M and the samples mixed by inversion. Each sample was then centrifuged at 13,000rpm for 1 hour at 4°C to pellet the membrane. The supernatant was discarded and the pellet resuspended in 100µl 1X PBS. 7.5µl of sample buffer was added to each sample and heated to 100°C for 10 minutes. 20µl aliquots were loaded onto a 15% polyacrylamide gel together with 20µl of a rainbow size marker and 5µl of a rod outer segment preparation. The gel was run at 150V for approximately 2 hours in a running buffer following the method of Laemmli (1970).

The proteins were transferred to a Hybond-C membrane using a trans-blot cell (Bio-Rad) at 20V for 1 hour. The membrane was then soaked overnight in a 50ml solution of blocking agent. The agent was discarded and the blot washed 3 times in 1 X TBS before adding the anti-1D4 antibody. After incubation at room temperature for 1 hour with gentle shaking, the antibody was removed and the blot washed again 5 times with 1 X TBS. A further 50ml of blocking agent was added and the blot incubated for 30 minutes at room temperature. The blot was again rinsed in 1 X TBS for a total of three washes and 5ml of a 1 X TBS solution added, containing 10µl rabbit anti-mouse alkaline phosphatase antibody. After incubation with gentle shaking for 1 hour at room temperature the blot was washed 5 times with 1X TBS. 20ml of detection buffer was added to the blot and incubated for 5 minutes to bring the pH to 9.5. The buffer was removed and 5ml of detection buffer added containing 100µl of

NBT/BCIP. The blot was left for approximately 1 minute until the bands began to appear and was then immersed in a bath of water/EDTA to quench the reaction.

*2.24 Reconstitution of bullhead blue opsin with 11-*cis *retinal*

The cell pellet from above (section 2.23) was thawed on ice and gently resuspended in 10ml 1X PBS. The reconstitution method was based upon that of Oprian *et al.* (1987). All following procedures were performed in a dark room at 4°C with a red safe light unless otherwise indicated.

*2.24.1 Preparation of 11-*cis *retinal*

One crystal of 11-*cis* retinal was dissolved in 400μl anhydrous ethanol. 1μl of this was diluted to a final volume of 500μl and the absorbance read at 380nm. The concentration was determined by using Baer's law

$$C = \frac{OD_{380}}{\epsilon \, l} \quad \text{nmoles/}\mu\text{l}$$

Where OD = absorbance at 380nm, ϵ = molar extinction coefficient of 11-*cis* retinal (25,000) and l = pathlength.

2.24.2 Solubilisation of the cell pellet

11-*cis* retinal was added to the resuspended pellet to obtain a final concentration of 40μM. The tube was wrapped in foil and placed on a rotary mixer at 4°C for 1 hour. Following incubation, the cells were centrifuged at 3500rpm for 5 minutes at 4°C in a bench top centrifuge. The supernatant was removed, 5ml PBS containing 0.2mg/ml PMSF added and the pellet resuspended completely. 5ml of 2% DM/PBS was then

added to solubilise the cell membranes and the tube placed on a mixer again at 4°C for 1 hour.

2.24.3 Binding of the visual pigment to the 1D4 column

The pigment was purified by immunoaffinity chromatography using an anti-rhodopsin monoclonal antibody (kindly supplied by Dr Phyllis Robinson, University of Maryland, Baltimore) coupled to CNBr-activated Sepharose 4B (Pharmacia). The 1D4 antibody solution was prepared by dissolving 6mg in 10ml 0.1M $NaHCO_3$ pH 8.3/0.5M NaCl and the absorbance was measured at 280nm. To prepare the sepharose, 2g were swollen in 12ml 1mM HCl for 15 minutes. The sepharose was transferred to a glass filter aspirator and washed with 400ml 1M HCl, followed by a further quick wash with 100ml ice-cold 0.1M $NaHCO_3$/0.5M NaCl. The gel was immediately scraped into a tube containing the 1D4 antibody solution and left overnight at 4°C on a rotating platform. The following day the sepharose was centrifuged at 4,000rpm/10 minutes. Supernatant was removed and its absorbance at 280nm determined. The gel was washed twice with 40ml 0.2M glycine pH 8.0 and spun as before. Following incubation with 40ml 0.2M glycine pH 8.0 for two hours at room temperature on a rotating platform, the sepharose was again transferred to a glass filter. The sepharose was washed twice with 40ml 0.1M $NaHCO_3$/0.5M NaCl pH 8.3, twice with 40ml 0.1M Na acetate/0.5M NaCl pH 4.0, twice with 40ml 0.1M $NaHCO_3$/0.5M NaCl pH 8.0 and twice with 40ml PBS/0.02% sodium azide. The gel was scraped into a tube and resuspended with 6ml PBS/0.02% sodium azide for storage at 4°C.

500µl of the 1D4 coupled sepharose column was transferred to a 15ml centrifuge tube and washed three times with 1ml of 1% DM/PBS to remove the sodium azide, spinning at 3,000rpm for 1 minute between each wash. A small amount of supernatant was left on the column after the final spin to prevent it from drying out.

After incubation, the cells were centrifuged at 3,500rpm for 3 minutes. The supernatant was removed and transferred to the tube containing the column. The

sample was placed on the mixer for 2 hours at 4°C to allow binding of protein to the column.

2.24.4 Elution of the visual pigment

Opsin bound to the 1D4 coupled sepharose column was eluted using peptide I. This peptide consists of 18 amino acids corresponding to the carboxy terminus of bovine rhodopsin and inhibits the binding of rhodopsin to the antibody column. The peptide sequence, Asp-Glu-Ala-Ser-Thr-Thr-Val-Ser-Lys-Thr-Glu-Thr-Ser-Gln-Val-Ala-Pro-Ala, was synthesised by Severn Biotech.

Spin columns were prepared by inserting a small amount of glass wool into a 1ml syringe. The mix containing the bound protein was centrifuged at 3,500 rpm for 5 minutes and the supernatant removed. The remaining slurry was transferred to the spin column which was in turn placed inside a 15ml centrifuge tube and spun for approximately 15 seconds at 1,200 rpm. The column was then washed eight times with 1ml of 0.1% DM/PBS, spinning for approximately 15 seconds between each wash. Most of the liquid should pass through the column but it should never be allowed to dry; for the final spin the meniscus should be just above the top of the column resin. The bottom of the column was sealed with Nescofilm and 200µl of the 1D4 epitope peptide I solution added. Samples were incubated on ice for 30 minutes, the seal was then removed and the column spun at 3,500 rpm for 30 seconds. A further 200µl of the peptide solution was added and the tube re-spun. The eluate was removed and used for spectrophotometric analysis.

*2.24.5 Spectrophotometric analysis of opsin protein reconstituted with 11-*cis *retinal*

Samples were analysed using a Unicam UV-visible spectrophotometer fitted with a cooled sample cell holder and Vision 32-bit, version 1.20 software (1999). The scan range was from 190nm – 700nm, using a scan speed of 240nm/min and a data interval setting of 1nm. The lamp change was set at 315nm. After the initial dark spectrum

was obtained, each sample was photobleached by exposure to light for five minutes and scanned again from 190nm – 700nm. Difference spectra were then generated and the λ_{max} of each sample determined by fitting a standard rhodopsin A_1 template to the experimental data (Stavenga *et al.*, 1993). Data from Vision 32 was imported into Axum 6.0 software (MathSoft, 1999) for further analysis and subsequent production of graphs.

2.25 DNA and protein molecular weight markers

ϕX174/Hae III (kb)	**λ Hind III (kb)**	**Rainbow marker (kDa)**
1.358	23.13	Blue 250
1.078	9.416	Red 160
0.872	6.682	Green 105
0.602	4.361	Yellow 75
0.310	2.322	Purple 50
0.281	2.027	Blue 35
0.271	0.564	Orange 30
0.234	0.125	Green 25
0.194		Blue 15
0.118		Pink 10
0.072		

Table 2.11: DNA and protein molecular weight markers used for PAGE and agarose gel electrophoresis.

2.26 Reagents and media

2.26.1 General reagents

- NE: 100mM NaCl, 25mM EDTA, pH 8.0
- 10X NH4 PCR buffer: 160mM $(NH_4)_2SO_4$, 670mM Tris-HCl (pH 8.8 at 15°C), 0.1% Tween-20
- 10X KCl PCR buffer: 100mM Tris-HCl (pH8.8 at 25°C), 500mM KCl, 15mM $MgCl_2$, 1% Triton X-100
- PCR dNTP mix: 1.25mM dATP, 1.25mM dCTP, 1.25mM dGTP, 1.25mM dTTP
- 10X ligation buffer: 300mM Tris-HCl (pH 7.8), 100mM $MgCl_2$, 100mM DTT, 10mM ATP
- 10 X PNK buffer: 500mM Tris-HCl pH7.5, 100mM $MgCl_2$, 50mM DTT and 1mM spermidine.

2.26.2 Electrophoresis and sequencing reagents

- 10X TAE: 0.4M Tris-acetate, 10mM sodium EDTA, pH 8.0
- 10X TBE: 90mM Tris borate, 10mM sodium EDTA, pH 8.3
- 6X agarose gel loading buffer: 0.25%(w/v) bromophenol blue, 0.25%(w/v) xylene cyanol FF, 15% Ficoll in water
- BigDye™ dilution buffer: A 5X solution of Tris HCl pH9.0 and $MgCl_2$ (Applied Biosystems)
- BigDye™ sequencing mix: BigDye™ ready reaction mix diluted 50:50 with BigDye™ dilution buffer

- ABI sequencing gel loading buffer: deionised formamide, 25mM EDTA (pH 8.0) containing 50mg/ml Blue dextran in a ratio of 5:1 formamide to EDTA/Blue dextran.
- Pharmacia ALF sequencing gel loading dye: deionised formamide, dextran blue 2000 5mg/ml
- Stop solution for manual sequencing: 0.3% bromophenol blue, 0.3% xylene cyanol FF, 10mM EDTA pH 7.5, 97.5% deionised formamide
- SequaGel-6® (National Diagnostics): 19:1 acrylamide:methylene bisacrylamide
- SequaGel® (National Diagnostics) complete buffer reagent: 0.89M Tris-borate, 20mM EDTA, pH 8.3, urea

2.26.3 Growth media

- LB (Luria Bertoni) broth: 1%(w/v) tryptone, 0.5%(w/v) yeast extract, 1%(w/v) NaCl, 1mM NaOH.
- LB agar: LB broth, 1.5%(w/v) bacteriological agar. LB agar was supplemented with 0.5mM IPTG, X-Gal 80μg/ml and ampicillin 100μg/ml or tetracycline 10mg/ml (in 80% ethanol) for plating transformants.
- SOC: 2%(w/v) tryptone, 0.5%(w/v) yeast extract, 10mM NaCl, 2.5mM Kcl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose.

2.26.4 RNA extraction solutions (Pharmacia)

- DEPC water: 0.1% diethyl pyrocarbonate, allowed to stand overnight then autoclaved
- Extraction buffer: a buffered aqueous solution containing guanidinium thiocyanate and N-lauryl sarcosine

- Elution buffer: 10mM Tris-HCl pH 7.4, 1mM EDTA
- High salt buffer: 10mM Tris Hcl pH 7.4, 1mM EDTA, 0.5M NaCl
- Low salt buffer: 10mM Tris Hcl pH 7.4, 1mM EDTA, 0.1M NaCl

2.26.5 5' and 3' RACE solutions

- Synthesis buffer: 250mM Tris-HCl, 40mM$MgCl_2$, 150mM KCl, 5mM dithiotreitol, pH8.5 @ 20°C
- 10 X reaction buffer: 100mM Tris-HCl, 15mM $MgCl_2$, 500mM KCl pH 8.3 @ 20°C

2.26.6 Tissue culture media and solutions

- DMEM: Dulbecco's Modified Eagle Medium with GlutaMAX I™ , 4,500mg/l D-glucose without sodium pyruvate. Gibco BRL catalogue No.61965
- DMEM including FCS and antibiotics: 500ml DMEM, 50ml heat inactivated foetal calf serum (First Link (UK) Ltd, cat no. 02-05-850), 5ml penicillin/streptomycin solution 1000iu/ml –10000ug/ml (Gibco BRL, cat no. 15140-114)
- Trypsin-EDTA: 1X liquid of 0.5g Trypsin (1:250), 2.0g EDTA per litre of Modified Puck's Saline A. Gibco BRL catalogue No. 45300
- PBS: Phosphate buffered saline tablets (Dulbecco A). Oxoid catalogue No. BR014G

2.26.7 SDS-PAGE and western blotting reagents

- Homogenisation buffer: 10mM MOPS pH 7.3, 5mM β-mercaptoethanol, 20μg/ml leupeptin, 1mM PMSF

- Sample buffer: 125mM Tris HCl pH 6.8, 4% (v/v) SDS, 20% (w/v) glycerol, 10% (v/v) 2-mercaptoethanol, 0.2% (w/v) bromophenol blue in H_2O.
- Polyacrylamide gel: 15% acrylamide, 0.267% (v/v) bisacrylamide, 375mM Tris HCl pH 8.8, 0.1% (w/v) SDS, 0.075% (w/v) APS, 0.05% (v/v) TEMED.
- Gel running buffer: 0.1M glycine, 50mM Tris, 0.1% (w/v) SDS in H_2O, pH 8.3.
- Blocking agent: 50mM Tris, 150mM NaCl, 5% (w/v) skimmed milk, 0.05% (v/v) Tween-20, pH 7.4.
- Transfer buffer: 25mM Tris, 190mM glycine, 20% (v/v) methanol.
- 1 X TBS: 10mM Tris, 0.9% (w/v) NaCl, 0.05% (v/v) Tween-20, pH 7.4.
- Detection buffer: 100mM Tris pH 9.5, 100mM NaCl, 50mM $MgCl_2$.
- Quenching buffer: 10mM Tris pH 7.5, 1mM EDTA pH 8.0, 150mM NaCl.

*2 26.8 Reagents for reconstitution of opsins with 11-*cis *retinal*

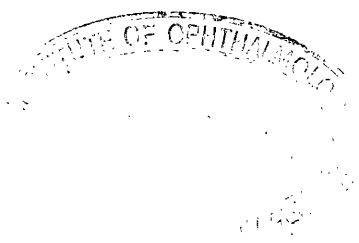
- PBS: Phosphate buffered saline pH7.0. 8g NaCl, 240mg KH_2PO_4 dissolved in 800µl distilled H_2O. Adjust pH to 7.0 with HCl and make volume up to 1l. Autoclave at 15lb/ 20 minutes

- DM/PBS: Dodecyl maltoside, prepared at concentrations of 2, 1 and 0.1% in 1X PBS

- PMSF: Phenylmethylsulphonyl fluoride, prepared at 50mg/ml in propan-2-ol. NB. this solution must be freshly prepared as it will only keep for approximately 30 minutes.

- Anhydrous ethanol: Available from Sigma (cat no. 45,983-6)

- Peptide I: Asp-Glu-Ala-Ser-Thr-Thr-Val-Ser-Lys-Thr-Glu-Thr-Ser-Gln-Val-Ala-Pro-Ala



CHAPTER 3

CHARACTERISATION OF THE 5' PROMOTER REGION OF THE GOLDFISH GREEN AND ROD OPSIN GENES

3.1 Introduction

3.1.1 Development of rod and cone cells

A basic step in understanding cell development is to analyse the molecular basis for gene regulation. The cells which make up the retina are born and differentiate in a defined temporal and spatial sequence (Carter-Dawson & LaVail, 1979). It is thought that these cells arise from common progenitors (Wetts & Fraser, 1988) although the full process following the development from committed progenitor to mature retinal cell is poorly understood. The process involves a cascade of events in which different genes are turned on and off in a precisely regulated manner under the control of a number of transcription factors; the production of only one type of visual pigment in each photoreceptor class is due to selective transcription of one type of visual pigment gene. Although morphological maturation of cone cells preceeds that of rods (Raymond, 1985), studies of opsin expression in the goldfish show that rod opsin is expressed first, within three days of cell birth. In comparison, expression of cone opsins is a stepwise process which requires at least seven days and follows a phenotypic-specific sequence; red, then green, then blue and finally ultraviolet (Stenkamp *et al.*, 1997). Both rods and cones are expressed in the same general retinal location suggesting that cell-cell interactions may coordinate selection of the specific photoreceptor phenotypes (Stenkamp *et al.*, 1997).

Of the opsins, rod opsin regulation has been the most extensively studied, with many of the *cis*-acting DNA elements characterised and some *trans*-acting factors isolated. Many of these are discussed in the following pages and include transcription factors which are either ubiquitous, retinal specific and are involved in various stages of cell

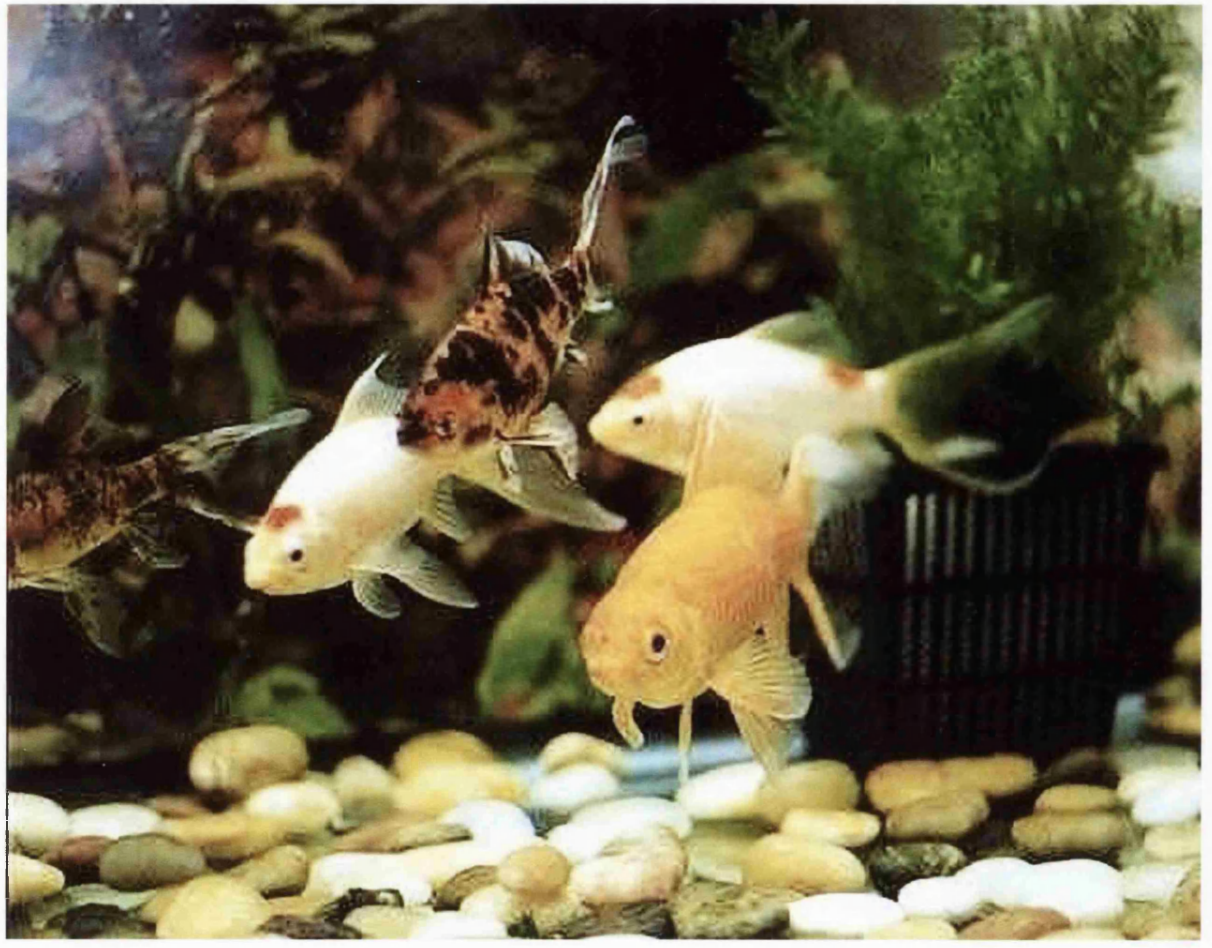


Plate 3.1: The goldfish (*Carassius auratus*). Photograph courtesy of Tara D. Fields

development (DesJardin & Hauswirth, 1996; Boatright *et al.*, 1997a; Yu *et al.*, 1996; Chen & Zack, 1996).

This study focuses on the goldfish (*Carassius auratus*) (plate 3.1). This fish is unusual in that it has two distinct green opsin genes both of which are expressed and represent a group of opsins with both rod and cone-like properties. The aims of this study were to analyse the 5' promoter regions of the goldfish green and rod opsin genes and to compare the sequences in order to identify any novel conserved *cis*-acting DNA elements which may be important in regulating expression of these rod-like green opsins. Known transcription factor binding sites have also been identified and the expression levels of the two green opsins assessed.

3.1.2 Visual pigments in the goldfish retina

The goldfish lives in freshwater and feeds close to the surface where light is bright and spectrally broad and as such, it might be expected that they have a well-developed visual system sensitive to a wide spectrum of light. MSP studies have identified a rod pigment and four types of cone pigment sensitive in the UV, blue, green and red regions of the spectrum (Tsin *et al.*, 1981; Bowmaker *et al.*, 1991b). The cones cells may be divided into four morphological classes arranged into a mosaic within the retina: double cones with a larger principal member having a λ_{max} of 625nm and a shorter accessory member with a λ_{max} of 537nm; long single cones with a λ_{max} of 537 or 625; short single cones with a λ_{max} of 452 and miniature short single cones which have UV sensitivity and a λ_{max} of 355-360nm (Tsin *et al.*, 1981; Stell & Harosi, 1975; Marc & Sperling, 1976; Bowmaker *et al.*, 1991b; Hashimoto *et al.*, 1988). The structure of goldfish rod and cone cells is shown in figure 3.1. Unpublished results by Bowmaker (personal communication) have also shown that the shorter accessory member of 19 double cones analysed, had an average λ_{max} of 536nm. Analysis of 7 long single cones also gave a λ_{max} of 536nm and it was therefore concluded that these cone cells contained the green opsin pigment. Like many freshwater fishes, the goldfish generally utilizes 11-*cis* 3-dehydroretinal as the chromophore although a study by Tsin *et al.* (1981) suggests that increased exposure to light (in this case,

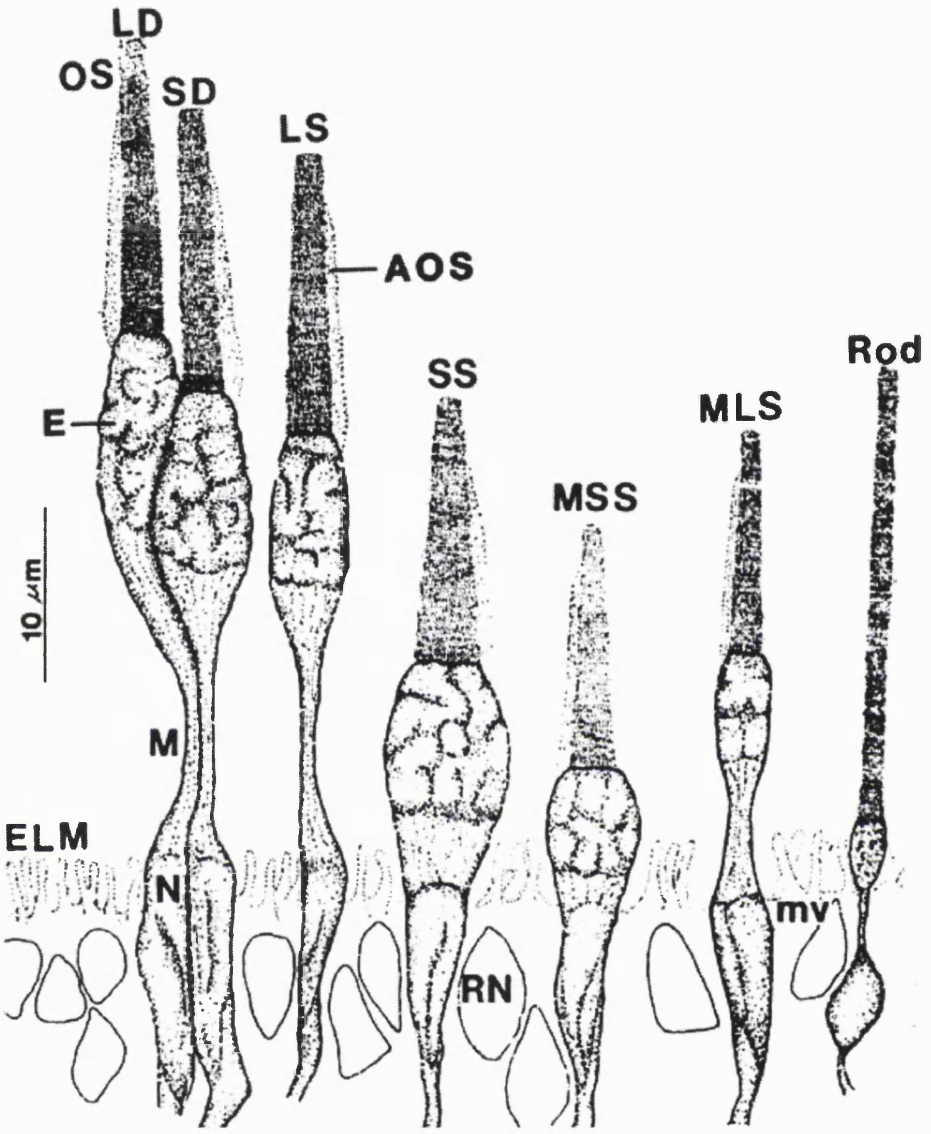


Figure 3.1: Goldfish cone and rod cells, Key: OS, outer segment; AOS, accessory output segment; E, cone ellipsoid; M, cone myeloid; N, cone nucleus; RN, rod nucleus; mv, muller cell microvilli; ELM, external limiting membrane; LD, long double cone; SD, short double cone; LS, long single cone; SS, short single cone; MSS, miniature short single cone; MLS, miniature long single cone. (From Marc& Sperling 1976)

supplied by a tungsten bulb) favours the production of rhodopsin resulting in a short wave shift in the λ_{max} of the pigments (Tsin *et al.*, 1981).

Johnson *et al.* isolated opsins from a goldfish retinal cDNA library and expressed the clones in 293S cells, reconstituting the pigments with 11-*cis* retinal. The λ_{max} values obtained were: 502nm for the rod pigment (492nm when inactivated with hydroxylamine); 511 and 504nm for the two green pigments, 524nm for the red pigment and 441nm for the blue. Expression studies by Raymond *et al.* (1993) have also shown that the 'green' opsins identified are expressed in the green cones (Raymond *et al.*, 1993) and Northern blot analysis of the two green cone opsin cDNAs by Johnson *et al.* (Johnson *et al.*, 1993) concluded that green 1 was transcribed at a higher level that green 2. A particularly interesting feature of the green cone opsins in the goldfish is the presence of two distinct genes (Johnson *et al.*, 1993) which code for proteins that differ in amino acid sequence by only 8.6% but show a 5 nm shift in λ_{max} between the two pigments. These green opsins also share characteristics of both rod and cone opsins; in addition to having a higher identity to rod opsins than to cones, they also have the potential asparagine glycosylation sites at positions 2 and 15 (Fukada *et al.*, 1979). Another fundamental difference between photoreceptor types is the more rapid response, lower photosensitivity and rapid rate of metarhodopsin II decay of cones compared to rods (Lamb & Pugh, 1990; Shichida *et al.*, 1994). Goldfish green cone opsins have a glutamic acid residue at site 122, as do rod opsins (Johnson *et al.*, 1993). It has been shown that the replacement of Glu-122, present in the chicken rod opsin, with Gln-122, present in the chicken green cone opsin, effectively converts the rate of regeneration and decay of metarhodopsin II from rod-like to cone-like (Imai *et al.*, 1997). There is also a conserved histidine residue at 211 (Weitz & Nathans, 1992) which is common to all rod and rod-like green cone opsins (Wright, 2000). However, like many other cone opsins, goldfish green cone opsins are rapidly inactivated by hydroxylamine.

Phylogenetic analysis suggests that gene duplication and divergence within the middle wave sensitive branch of the tree occurred early in vertebrate evolution and gave rise to the green and rod opsin clades (Mrd) (Johnson *et al.*, 1993; Register *et al.*, 1994) (figure 3.2). Chicken green shows an identity of 79% and 77% to goldfish green I and

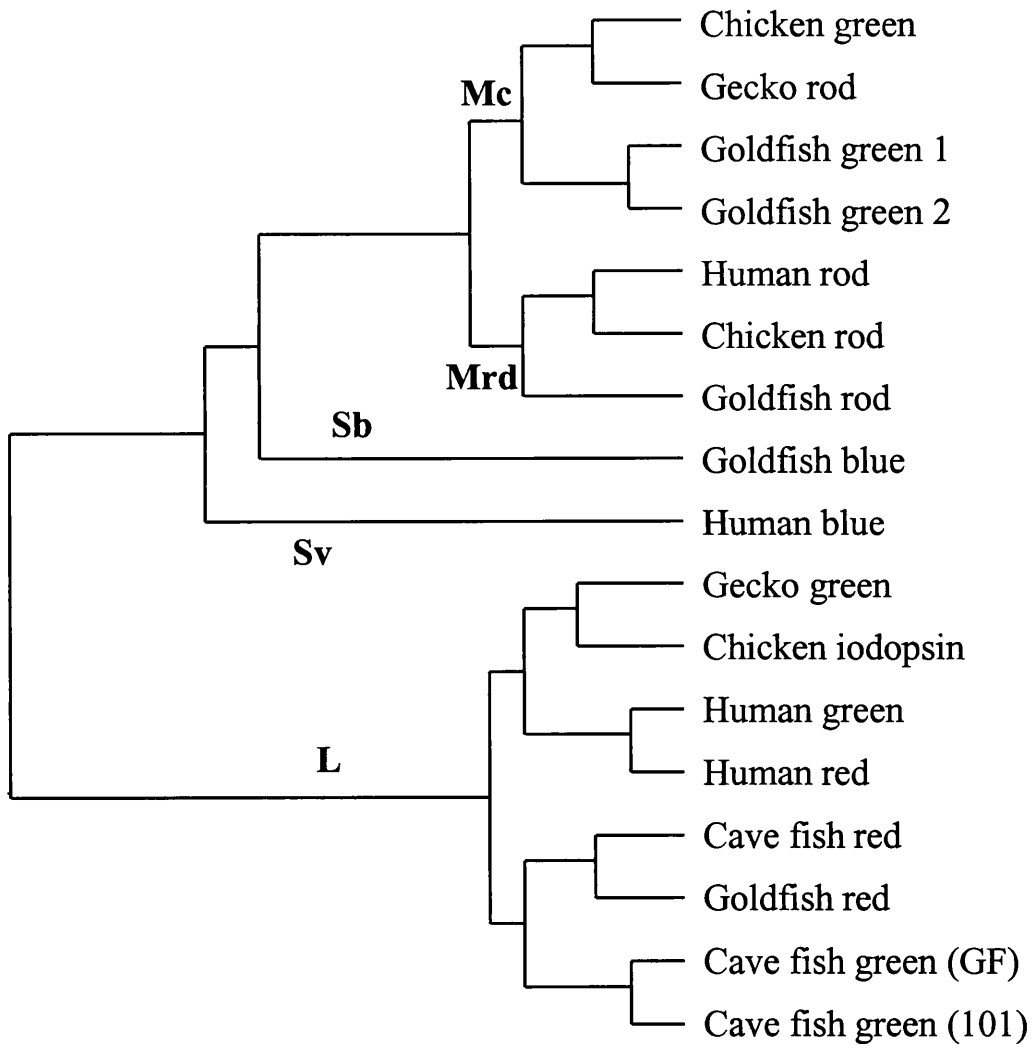


Figure 3.2: Dendogram illustrating the evolutionary relatedness of visual pigments in the goldfish and other vertebrates. reproduced from Johnson *et al.* 1993

green II respectively and 72% and 71% respectively to human rod (Wang *et al.*, 1992a; Okano *et al.*, 1992). In comparison, the cave fish green opsin which belongs to the L branch of the phylogenetic tree, along with mammalian green opsins, showed less than 40% identity to the goldfish green opsin. From this we can conclude that although these goldfish green pigments are expressed in green cone cells (Johnson *et al.*, 1993), in evolutionary terms they are separate and distinct from the L-group of green opsins found in mammals and other fishes.

3.1.3 Tetraploidy in goldfish

The two rod-like green cone opsin genes may have arisen from tetraploidization which has occurred on numerous occasions throughout vertebrate evolution as a result of chromosome doubling (autotetraploidization) or species hybridization (allotetraploidization). This phenomenon may be responsible, in part, for the presence of multigene families in higher metazoans. The common carp (*Caprinus carpio*) is closely related to and readily interbreeds with the goldfish and it is thought that they diverged from a common tetraploid ancestor. After tetraploidization, duplicate genes may continue to be expressed for millions of years although they will gradually accumulate mutations leading to a change in amino acid sequence and may eventually cease to, or change, function (Force *et al.*, 1999). Larhammar and Risinger have analysed carp sequences for genes encoding five products including somatotrophin and gonadotrophin (Larhammar & Risinger, 1994). The divergence rates at silent sites were found to be 19Mya and 12Mya respectively. Taking the average rate, tetraploidization in the carp is estimated to have occurred approximately 16Mya ago resulting in a chromosome complement which is approximately twice that of other cyprinid fishes (Larhammar & Risinger, 1994). More recently a second expressed rod opsin has also been identified in the carp by Su *et al.* (2000) (Gen Bank accession No.AJ012014 & AJ012013) and there remains the possibility that the other opsins may too be expressed in duplicate.

3.1.4 Gene expression and regulation

Given the relatively high sequence similarity in the coding regions of the rod and green cone opsin genes and the fact that the two green opsins arose as a result of

tetraploidization, some level of sequence conservation in the 5' flanking regions of rod and green cone opsin might be expected suggesting that similar regulatory proteins are responsible for regulation during expression. Much of the recent work on opsin regulation has been confined to rod opsins (Morabito *et al.*, 1991; Batni *et al.*, 1996; Sheshberadaran & Takahashi, 1994); studying the goldfish, with its cone rich retina, will help to clarify the mechanisms involved in cell type specific regulation of the entire opsin gene family.

Messenger RNA (mRNA) is synthesised on a DNA template in the nucleus of eukaryotes by the process of transcription, the major point at which gene expression is regulated, and is exported to the cytoplasm where it is translated into protein. Initiation of transcription is governed by many DNA sequence elements including: a core promoter, which contains the binding site for RNA polymerase II; and upstream promoter elements and enhancers which regulate the rate at which RNA polymerase II initiates new rounds of transcription from the core promoter. Transcriptional activation of eukaryotic genes involves the regulated assembly of multiprotein complexes on enhancers and promoters. Even a relatively small number of different transcription factors can be used to regulate complex patterns of gene expression at a high level of specificity. Typically, a transcription factor will contain a specific DNA-binding domain, a multimerization domain which allows the formation of homo or heteromultimers and a transcriptional activation domain. Some transcription factors commonly interact with a variety of promoters, whereas others are specific to a particular gene or class of genes.

3.1.5 The promoter

The promoter region of a gene is immediately upstream of the transcription initiation site and contains characteristic short conserved sequences which are recognised by certain transcription factors and RNA polymerase II. Promoters play an essential role in accurately positioning the start site of transcription. In most eukaryotic genes they are responsible for basic transcriptional processes and, in a more limited number of genes, may result in a tissue specific or signal specific pattern of expression.

The TATA box is generally located about 30 base pairs upstream from the transcription start site and is surrounded by GC rich sequences. Although present in most eukaryotic genes, there is no TATA sequence in the *Xenopus* rod opsin promoter; instead this is replaced by a TTTAAAA motif surrounded by G-rich sequences and has a similar function (Batni *et al.*, 1996). In another study (Shaaban & Deeb, 1998), mutation of the TATA box lead to a significant reduction in levels of promoter activity. The group used Y-79 and WERI cell lines in transfection studies to assess activity of the promoter in driving luciferase reporter gene activity. To their surprise, when the TATA box was deleted, the promoter was still active but at a reduced rate suggesting that it was not essential for expression of human red and green opsins.

For the process of transcription to be initiated, a transcription factor (TFIID) first binds to the TATA element to form a complex, which in turn, acts as a binding site for another transcription factor TFIIB. RNA polymerase II and TFIIF can then be recruited into the complex and finally TFIIE and TFIIH to complete the system. When this complete initiation complex has been assembled, transcription occurs upon activation by ATP (figure 3.3). Once activated, the complex either undergoes a significant conformational change or one or more components of the complex are released (Buratowski *et al.*, 1989).

3.1.6 Upstream promoter elements

Ubiquitous upstream promoter elements include the SP1 and CAAT boxes. The SP1 box is a GC rich region, with the consensus sequence GGGCGG, which enhances low transcriptional activity and can exist in either orientation. The sequence was originally identified in the promoter of the eukaryotic virus SV40 (McKnight & Tjian, 1986) where six copies of the motif were found to bind the Sp1 transcription factor. The transcription factor itself contains three zinc finger motifs and two glutamine rich regions which interact with transcription activation factors within the TFIID complex (Courey & Tjian, 1988). A zinc finger motif consists of a 30 amino acid unit of which 12 residues form a loop projecting out from the surface of the protein. Four conserved

residues, two cysteines and two histidines, anchor the loop at its base and bind the zinc atom which is essential for activity of the Sp1 protein (figure 3.4). Potential SP1 binding sites have been identified in the *Xenopus* rhodopsin promoter (Batni *et al.*, 1996). However, DNase I footprinting of the promoters of human red and green opsin genes (Shaaban & Deeb, 1998), and rat (Morabito *et al.*, 1991) and bovine rod opsin genes (DesJardin & Hauswirth, 1996), failed to highlight any areas where this transcription factor was actually binding. The Sp1 protein has also been shown to interact with different forms of CACC boxes where it is involved in mammalian globin gene regulation (Yu *et al.*, 1991).

The CCAAT box acts as an upstream promoter element and was initially shown to bind two transcription factors, the CCAAT box transcription factor (CTF/NF1) and the CCAAT enhancer binding protein (C/EBP) (Rosenfeld & Kelly, 1986). It has since been shown to bind a number of other constitutively expressed and tissue specific proteins including Ret-1 (DesJardin & Hauswirth, 1996) and a 40 kilodalton protein, possibly CRX (Morabito *et al.*, 1991). It can also function in either orientation and may be located at various distances from the transcription startpoint. In the human (Nathans & Hogness, 1984), bovine (DesJardin & Hauswirth, 1996), mouse (Baehr *et al.*, 1988), chicken (Sheshberadaran & Takahashi, 1994) and *Xenopus* (Batni *et al.*, 1996) rod opsin promoters, this consensus sequence is indeed present. However, in the rat, the site is modified to CTAAT but has still been shown to be developmentally important and binds the retinal specific transcription factor Ret-1 (Morabito *et al.*, 1991). Mutation studies by Morabito *et al.* in which synthetic oligonucleotides were produced with altered binding sites showed a significant decrease in binding of retinal extracts such that in some cases the formation of a retina-specific complex was barely detectable. These experiments therefore confirm the importance of this site in determining the efficiency of opsin promoters.

Promoter elements are often interdigitated with regulatory elements and a number of those involved in opsin expression in mammals have now been identified by either DNase I footprinting or electrophoretic mobility shift assays (EMSA) and their associated *trans* acting factors isolated. These include: Ret 1/PCE 1 (Morabito *et al.*, 1991; Kikuchi *et al.*, 1993) thought to bind CRX, the cone rod homeobox (Chen *et al.*,

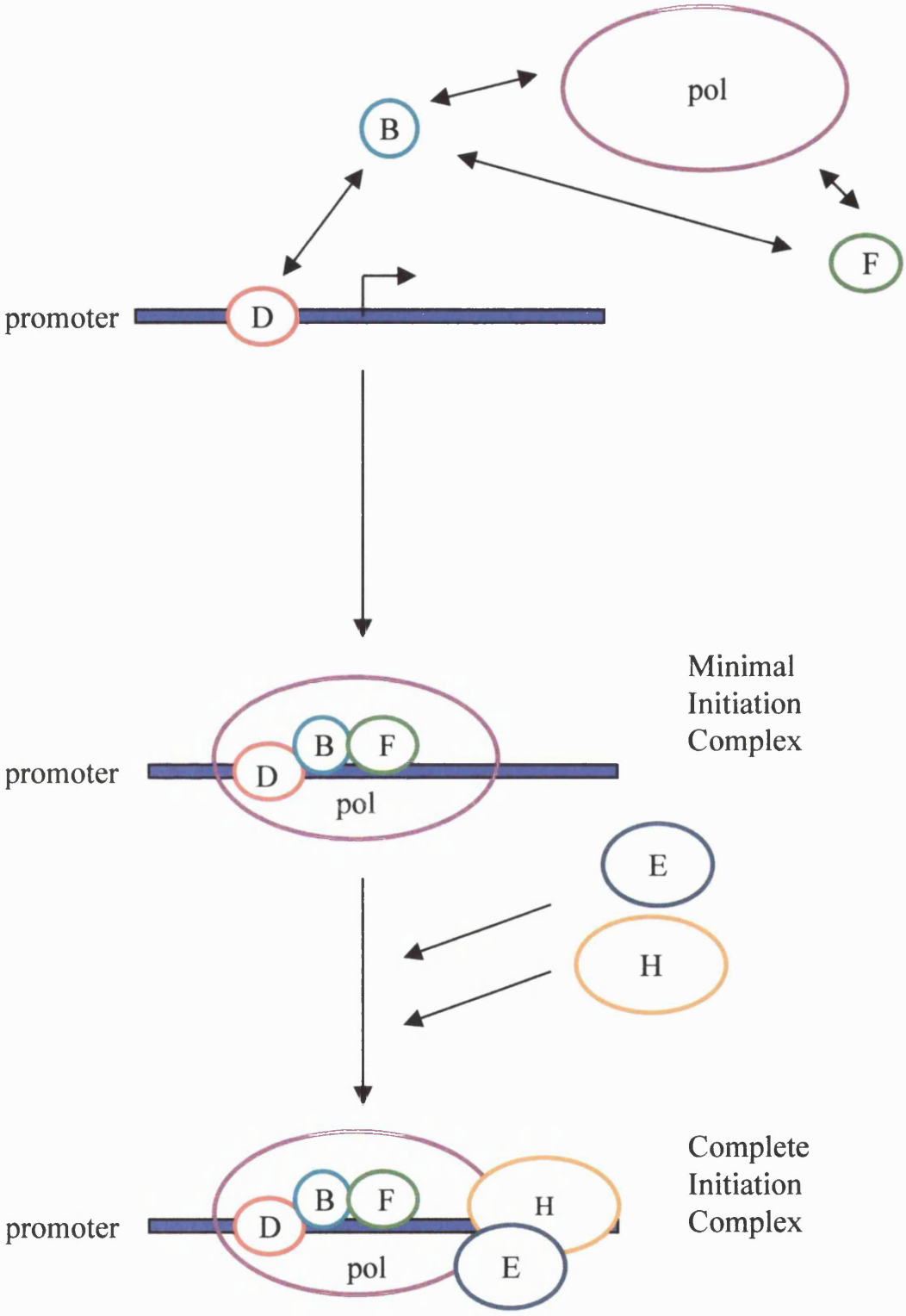


Figure 3.3: Schematic model of the transcription initiation complex assembly. B, D, E, F & H represent transcription factor IIB (TFIIB), TFIID, TFIIE, TFIIF & TFIIH. Pol is RNA polymerase II. Double-headed arrows indicate protein-protein interactions. The bent arrow indicates the initiation site.

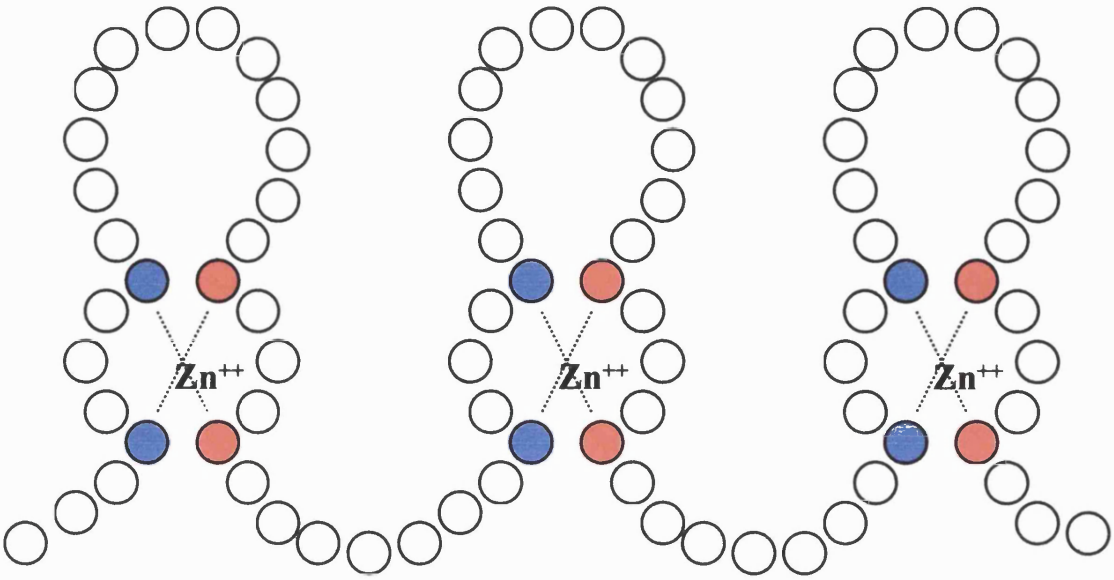


Figure 3.4: The transcription factor Sp1 has a series of three zinc fingers, each with a characteristic pattern of cysteine (blue circles) and histidine (red circles) residues that constitute the zinc binding site (redrawn from Lewin 1990).

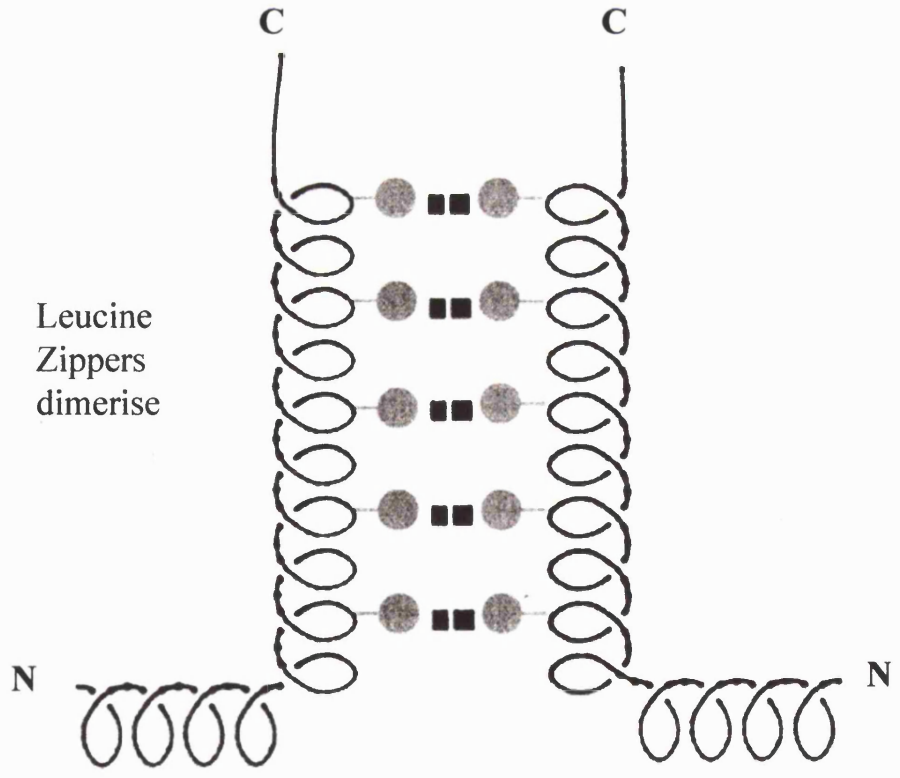


Figure 3.5: Two leucine zippers in parallel orientation forming a dimeric structure (redrawn from Lewin 1990). The grey circles represent the leucine molecules.

1997; Furukawa *et al.*, 1997); Ret 2 and Ret 3 (Yu, 1993); Ret 4, which binds retina specific and more ubiquitously expressed factors including CRX (Chen & Zack, 1996; Chen *et al.*, 1997); Eopsin-1 which binds Mash-1 (Ahmad, 1995); eight C-rich regulatory sequences (CRS) (DesJardin & Hauswirth, 1996); BAT-1 (bovine A/T rich sequence), shown to be important during development (DesJardin & Hauswirth, 1996); and NRE which binds the retinal protein NRL (Rehemtulla *et al.*, 1996).

Ret 1, first identified in the rat opsin promoter (Morabito *et al.*, 1991), was shown to bind developmentally-regulated retinal nuclear proteins, and mutation studies confirmed that it was necessary to drive gene expression in rod photoreceptors (Yu *et al.*, 1996). The PCE 1 site defined in the arrestin promoter, is very similar to and may be identical to Ret 1 hence the term Ret 1/PCE 1 is frequently used (Kikuchi *et al.*, 1993). Within the interphotoreceptor retinoid binding protein (IRBP) 5' promoter region, the Ret 1/PCE-I site is also required for promoter activity. Nested deletion analysis of the promoter showed that only 70bp 5' to the transcription start site was needed to maintain high promoter activity although it was no longer tissue specific. However, if part of this region containing the Ret 1 site was deleted, activity was lost (Boatright *et al.*, 1997a). The importance of Ret 1 in development is currently uncertain. In a study on the rat opsin upstream region, a 20-fold excess in the amount of protein binding to this element was observed in the adult as compared to newborns (Morabito *et al.*, 1991). In comparison, a similar study of the bovine promoter region showed no substantial changes during development (DesJardin & Hauswirth, 1996).

The reverse sequence of Ret 1, GATTAA, binds an OTX-like photoreceptor specific trans-acting factor, the cone rod homeobox (CRX). The gene was first isolated from the mouse retina and plays a crucial role in differentiation of photoreceptor cells, being most highly expressed when rod photoreceptor differentiation is maximal between postnatal day 4 and 9 (Furukawa *et al.*, 1997). In *Drosophila*, the Ret 1 and OTX consensus sequences are functionally equivalent suggesting that CRX might interact with the Ret 1 site in invertebrate opsins too. This theory is supported by the fact that Ret 1 binds a putative 40 kDa factor (Yu *et al.*, 1996) which may be CRX (Furukawa *et al.*, 1997). Human diseases associated with CRX include; blue cone monochromacy, where the OTX binding site is deleted (Nathans *et al.*, 1989;

Furukawa *et al.*, 1997), cone-rod dystrophy 2, which results in photoreceptor degeneration as a result of mutations in CRX itself (Freund *et al.*, 1997) and Leber's congenital amaurosis (Freund *et al.*, 1998).

The Ret 1 site also shows 70% homology to the transducin α-1 (Tα-1) element found in the upstream region of the mouse rod α-T gene (Ahmad *et al.*, 1994). Binding to the Tα-1 site is also developmentally regulated, peaking during the period of photoreceptor maturation between postnatal days 6 and 9. Gel retardation assays using Ret 1 as a competitor however, showed that the Tα-1 site was distinct from that of Ret 1 and that it bound a retinal-specific nuclear factor of approximately 90 kDa (Ahmad *et al.*, 1994).

Of the other Ret elements, Ret 4 has been identified by deletion analysis and shown to bind both retinal specific and ubiquitously expressed protein factors (Chen & Zack, 1996). It has a positive affect on regulation and mutation of this binding site greatly reduces levels of rhodopsin expression. In addition to the Ret 4 site, deletion analysis revealed other regions which bind factors critical for transcription; these included NRE (Chen & Zack, 1996).

Nrl is a member of a subfamily of basic motif-leucine zipper DNA-binding proteins and binds to NRE, an extended AP-1-like site (Rehemtulla *et al.*, 1996). The protein can form homodimers and also heterodimers with members of the c-Fos and c-jun family. In the basic leucine zipper structure, the hydrophobic groups (including leucine) face one side whilst charged groups face the other. The leucines protrude from the α-helix and interdigitate with the leucines of another protein in the reverse orientation to form the dimer (figure 3.5). In transfection studies using primary chick retinal cell cultures, the presence of NRE showed a 3-5 fold increase in the activity of the rhodopsin promoter as compared to constructs with a mutated or deleted site (Kumar *et al.*, 1996), suggesting it plays a role in controlling retinal function. This important site has been identified in several rhodopsin promoter regions including those of the mouse, rat, Chinese hamster, bovine, human, and *Xenopus* (Zack *et al.*, 1991; Morabito *et al.*, 1991; Gale *et al.*, 1992; Nathans & Hogness, 1983; (Nathans & Hogness, 1984; Batni *et al.*, 1996).

A further group of developmentally regulated elements, termed CRS, have been identified within the promoter region of the bovine opsin gene. Eight protected regions, in addition to a Ret 1 and BAT-1 site, were identified using DNase I footprinting techniques (DesJardin & Hauswirth, 1996). These protected elements could be divided into three categories: BAT-1, Ret-1, CRS-2 and CRS-5, which did not change during development; CRS-1, CRS-3, CRS-4 and CRS-7, which showed a reduction in protection during development after 5.5 months gestation; and CRS-6 and CRS-8, which showed a marked reduction in protection in adult extracts as compared to the 5.5 and 7.5 month foetal extracts. With the exception of CRS-8, all other CRS sites were protected both with brain and retinal nuclear extracts suggesting that they are not retinal specific. CRS-8 however, was not protected with the brain nuclear extract and may therefore be a novel retinal specific element.

3.1.7 Enhancer elements

A second type of element that can potentiate gene transcription is the enhancer. This is located in the 5' flanking region of the gene and typically contains a collection of sites which bind activating or suppressing factors to stimulate or depress transcription. Although enhancers lack promoter activity themselves, they may be able to greatly increase levels of gene expression and can be essential for expression. The positive regulatory activity of an enhancer is generally position and orientation independent, as demonstrated by the rhodopsin enhancer region (RER) (Zack *et al.*, 1991).

Transgenic studies have shown that the rhodopsin enhancer region is responsible for high levels of rod opsin expression (Nie *et al.*, 1996). Sequence comparisons between bovine, mouse, rat and human rod opsin sequences, identified a 100 base pair region which showed significant homology (75-85%) between the species (figure 3.6); chicken sequence however, only showed 40-50% homology to this region (Sheshberadaran & Takahashi, 1994). Despite the high sequence homology, the transcription start site of the RER varies between species from around 1.5kb upstream in the mouse to over 1.9kb upstream in bovine (Zack *et al.*, 1991; Nie *et al.*, 1996) and, in addition to its transposition further upstream, the RER of bovine rhodopsin has

cis-Element	Consensus sequence	*trans*-Factor	reference
TATA box	TATA A/T A A/T	TATA binding protein	(Breathnach & Chambon, 1981)
SP 1	GGGCGG	SP 1	(Courey & Tjian, 1988)
CCAAT box	C C/T AAT	CTF/NF1	(Rosenfeld & Kelly, 1986),(McKnight & Tjian, 1986)
Ret 1/PCE 1	CAATTA	*Crx*	(Kikuchi *et al.*, 1993), (Yu & Barnstable, 1994), (Yu *et al.*, 1996), (Chen *et al.*, 1997), (Furukawa *et al.*, 1997)
Ret 2	TCCAGTCACAGCCTGAGGCCACCAGAGTG	unknown	(Yu, 1993)
Ret 3	CCTTAAACTGCTGGAGACCAACTTCCAGCC	unknown	(Yu, 1993)
Ret 4	GGAGCTTAGGGAGGG	*Crx*	(Chen & Zack, 1996), (Chen *et al.*, 1997)
NRE	TGCTGATTCAGCC	Nrl	(Rehemtulla *et al.*, 1996)
AP1	TGATTCA	AP1	(Rehemtulla *et al.*, 1996)
Proximal *glass*-like	ACCCCTGAAAATGGC	unknown	(Sheshberadaran & Takahashi, 1994)
Distal *glass*-like	CTTTTGCAGGGTTTC	unknown	(Sheshberadaran & Takahashi, 1994)
CRS-1	CAGCACGCCCCGCCTTCTCCCCG	unknown	(DesJardin & Hauswirth, 1996)
CRS-2	CTTGGCAACCTGGCAAAGTGCCCCCCTCCCT	unknown	(DesJardin & Hauswirth, 1996)
CRS-3	CTGGCACTAAGCCCCACCCCA	unknown	(DesJardin & Hauswirth, 1996)
CRS-4	TGCCCGGGACCACACCTCCTTGCT	unknown	(DesJardin & Hauswirth, 1996)
CRS-5	GTTCCCAGGGCCCACCTCCCCTCC	unknown	(DesJardin & Hauswirth, 1996)
CRS-6	TCCCCGCCCCCCCAG	unknown	(DesJardin & Hauswirth, 1996)
CRS-7	CCCCACCCACCCGCCA	unknown	(DesJardin & Hauswirth, 1996)
CRS-8	GTCTGGCCACCAGGGG	unknown	(DesJardin & Hauswirth, 1996)
BAT-1	GATTAA	*Crx*	(DesJardin & Hauswirth, 1996)

Table 3.1: Summary of transcription factor binding sites common to many vertebrate opsin promoters

Figure 3.6: Sequence alignments of bovine, human, mouse and rat rhodopsin enhancer regions with the human red/green LCR. Bases which are identical in all four rhodopsin genes are shown in blue. The base position relative to the transcription start site is shown on the right. The ret-3 binding site is also shown by the bold line.

also been inverted (Zack *et al.*, 1991). Further studies of opsin promoter regions using DNase I footprinting, have revealed that parts of the RER appear to be tissue specific; others are restricted to neuronal tissue or are more ubiquitously expressed.

The human red/green locus control region (LCR) (Wang *et al.*, 1992b) sequence shows homology with the RER and is centred approximately 3.5kb upstream of the transcription start site. The LCR contains a 37bp core sequence, which is highly conserved between human, mouse and bovine DNA, embedded in a 200bp region of 70% identity. Studies by Wang *et al.* (1999) have shown that identical transcription regulatory proteins bind to the LCR of both red and green opsins in mammals. In further transgenic studies, this entire region was proven to be essential for expression, and deletion or mutation of the LCR leads to a condition known as blue cone monochromacy in humans where there is a combined loss of red and green cone function (Wang *et al.*, 1992b). It has been proposed that the LCRs and other upstream regulatory elements, including Sp 1, interact with promoters by a looping mechanism which brings them together via protein-protein interactions (Cvekl & Paces, 1992), (Mastrangelo *et al.*, 1991). Transgene constructs generated with 5.4kb or 470bp (Chiu & Nathans, 1994a) and 3.1kb or 1.1kb (Chen *et al.*, 1994) of the human blue pigment promoter were able to direct expression to the blue cones and even a subset of cone bipolar cells. The efficiency of transgene expression did vary however, suggesting that some additional sequences involved in blue pigment gene expression may be missing from the constructs (Chiu & Nathans, 1994a). A construct containing 6.4kb upstream of the mouse blue visual pigment was later shown to direct blue cone cell specific expression and contained all of the necessary sequence elements for efficient transcription (Chiu & Nathans, 1994b).

Also within the RER, lie the Ret-3 site (Yu, 1993) and a sequence that is homologous to the chicken *glass*-like binding domain (Sheshberadaran & Takahashi, 1994). The *glass* gene, first identified in *Drosophila*, encodes a site-specific DNA-binding zinc finger protein. *Glass* is thought to play an important role in development by inducing photoreceptor cell-type specific gene expression in pre-photoreceptor cells from the first stages of their differentiation from a common progenitor cell (Moses *et al.*, 1989). DNase I footprinting identified the 27bp *glass* binding site and has shown that

it confers *glass*-dependent expression on a reporter gene in developing photoreceptor cells (Moses & Rubin, 1991). Studies of the *glass*-like binding sequence in the chicken have also shown that it exhibits tissue-specific binding properties similar to those in *Drosophila,* but it remains to be seen if the proteins which bind to this site in the chicken can actually activate gene transcription in a photoreceptor cell-specific mode. Partial homologies to the glass-binding site were also found in the bovine, mouse, human (Sheshberadaran & Takahashi, 1994) and *Xenopus* (Batni *et al.*, 1996) opsin promoter regions but to date, their role in transcription is undetermined.

3.1.8 Sequence conservation within opsin promoters

Biologically important sequences tend to be conserved between species during evolution. Comparisons between approximately 2kb of the 5' flanking sequences of human, murine and bovine rod opsin genes by Zack *et al.* (1991) found that there were two main regions of conservation. Firstly, the 200bp immediately 5' of the transcription start site and secondly the RER at just over 1kb further upstream, a region of approximately 100bp, showing 75-85% homology. Although the evolutionary distance between the species is similar, the homology observed between the human and bovine sequences is significantly greater than that between the murine sequence and either of the other two. Wang *et al.* (1992b) also compared human red with the murine and bovine sequences and, in addition to the regions mentioned above, found a distal region of homology. This sequence contains the LCR and begins at approximately 3.2kb 5' to the transcription start site of the human red opsin gene, extending to 4.5kb with several interruptions; in the mouse the sequence is inverted. Other comparisons between mammalian promoters have shown similar results (Chiu & Nathans, 1994a; Kumar *et al.*, 1996). However, when chicken and *Xenopus* sequences are included, homology in the 200bp proximal region falls to around 50%, although in localised regions such as the TATA box, Ret 1/PCE and AP-1 sites, homology is nearer 90% (Chen & Zack, 1996; Batni *et al.*, 1996).

In an attempt to determine whether promoter sequences are conserved within all vertebrates, this study examines the 5' regulatory regions of the goldfish rod and

green opsin genes. Fishes are distantly related to mammals and contain the full compliment of opsins including the unique rod-like green opsins, the promoter regions of which have not been studied before. A high degree of homology between promoters would suggest that similar transcription factors are responsible for activation of the goldfish rod and green opsins; lack of homology would suggest that very different factors promote transcription of fish opsins.

3.2 Results

3.2.1 Determination of the transcription start site using 5' RACE

The transcription start sites of the two goldfish green genes were determined using 5' RACE and sequencing. In the primary round of PCR, the dA-tailed cDNA was amplified with each of the gene specific primers, green 1 N1, green 2 N1 and rod N1, and the oligo d(T)-anchor primer, to produce a smear when viewed on an agarose gel. An aliquot of this PCR product was then used in a reaction with each of the nested gene specific primers, green 1 N2, green 2 N2 and rod N2, and the oligo d(T)-anchor primer, to obtain a more specific product (figure 3.7). The bands indicated were excised and the DNA eluted and cloned into the pGEM®-T Easy vector. The plasmids of multiple colonies were sequenced. The transcription start site was confirmed when at least three separate clones gave identical results. The sequencing results with the position of the transcription start site are shown in figure 3.8. The transcription start site for green 1 is 85 bp upstream of the translation start site and for green 2, 67 bp. However, when these sequences are brought into alignment by inserting gaps, the transcription start sites are in positions with similar sequences.

The sequence of the full-length transcript of the carp rod opsin, has been deposited by Su *et al.* on Gen Bank under accession number AJ012014/AJ012013.

3.2.2 Amplification of the 5' flanking region of rod and green cone opsins

Genomic DNA was isolated from goldfish liver and the gene walking method (Dominguez & Lopez-Larrea, 1994) used to generate fragments from the 5' flanking region of the rod and two green cone opsin genes.

1.55 kb of the 5' flanking region of the rod opsin gene was amplified in three gene walks of approximately 250bp, 500bp and 800bp respectively, starting at 67 bps

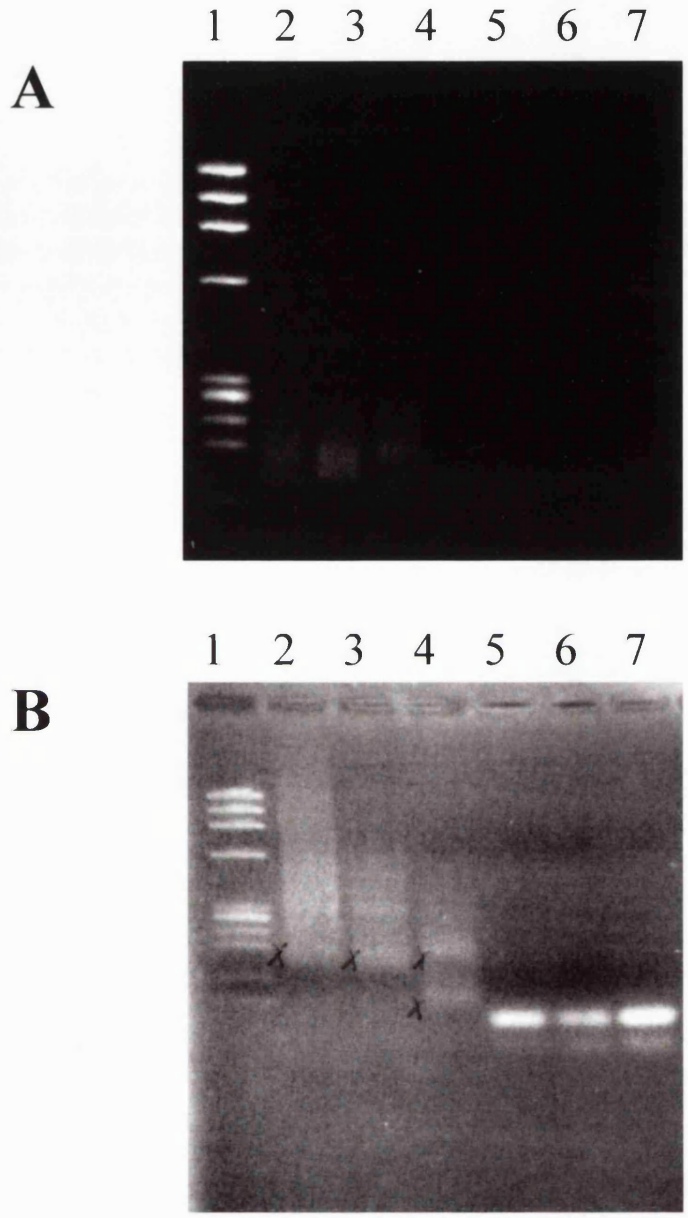


Figure 3.7: 5' RACE of goldfish green 1, green 2 and rod opsin genes to determine the transcription start site. A) Primary PCRs. B) Secondary PCRs. Lane 1: ϕX174/Hae III DNA ladder. Lane 2: green 1. Lane 3: green 2. Lane 4: rod. Lanes 5, 6 & 7: control reactions for green 1, green 2 and rod respectively. Bands marked with an X, were excised.

```
-62   GCCGCGAATT CACTAGTGAT TGACCACGCG TATCGATGTC GACTTTTTTT  Green 1
-69   GCCGCGAATT CACTAGTGAT TGACCACGCG TATCGATGTC GACTTTTTTT  Green 2

-12   TTTTTTTTTC TG-------A AAGAGCAGGT GAAGGGCTGC CACTCAGCTG  Green 1
-19   TTTTTTTTTC TATTTTTTTA AAGAGCAG-- ---------- -ACG-AGA-G  Green 2

+32   TACCACAGCA GCAGATTGCA TCCTTTATTT CCTCTGGATC ACTAGTGGGC  Green 1
+17   TACCACAACA GCAGATTACA CCCTTTATAT ---CTGGATC GCTAGTGTGC  Green 2

+82   AAAGATGAA CGGCACTGAGG GAAATCGAAT TCCCGCGGCC GCCATGGCGG  Green 1
+64   AAAGATGAA TGGCACTGAGG GAAATCGAAT TCCCGCGGCC GCCATGGCGG  Green 2
```

Figure 3.8: Alignments of goldfish green 1 & green 2 opsin 5'RACE sequence cloned into the pGEM®-T Easy Vector. Red sequence shows the pGEM®-T Easy cloning vector; blue, the poly A tail and black, the 5' upstream sequence starting at the transcription start site. Numbers on the left represent the nucleotide position relative to the transcription start site (+1). Spaces have been introduced to bring the sequences into alignment.

downstream from the initiation codon. Primer pairs used for walks 1, 2 and 3 are shown in table 3.2.

	Walk 1	**Walk 2**	**Walk 3**
Primary PCR	1st walk rod out/ UNI 33 or UNI28	2nd walk rod out/ UNI33 or UNI 28	3rd walk rod out/ UNI33 or UNI 28
Secondary PCR	1st walk rod in/ UNI 17	2nd walk rod in/ UNI 17	3rd walk rod in/ UNI 17

Table 3.2: Primer pairs used for walking PCR to amplify the goldfish rod opsin 5' promoter region.

In each case the primary PCR showed a smear when the samples were analysed on an agarose gel. Using nested inner primers for the secondary round of PCR, specific bands were seen for reactions which had included the UNI33 primer in the primary round. An example of the results obtained from the primary and secondary walking PCRs is shown in figure 3.9.

Only two gene walks of approximately 750bp and 1300bp were required to generate over 2 kb of 5' flanking sequence for both green opsin genes. A primer which would amplify both green opsins, 1st walk green 1 & 2 out, was used initially with each of the walking primers, UNI28, UNI33 and WP33. An annealing temperature of 60°C was used for the primary round of PCR and a smear containing multiple bands was observed upon analysis on an agarose gel for all samples except those containing the WP33 primer. The secondary round used nested primers which targeted each gene separately; reverse primers corresponding to bases 179 to 198 of green 1 and 178 to 198 of green 2 were used together with UNI17 as shown in figure 3.10. The gene specific primers for the secondary round of PCR differed at the final three bases of the

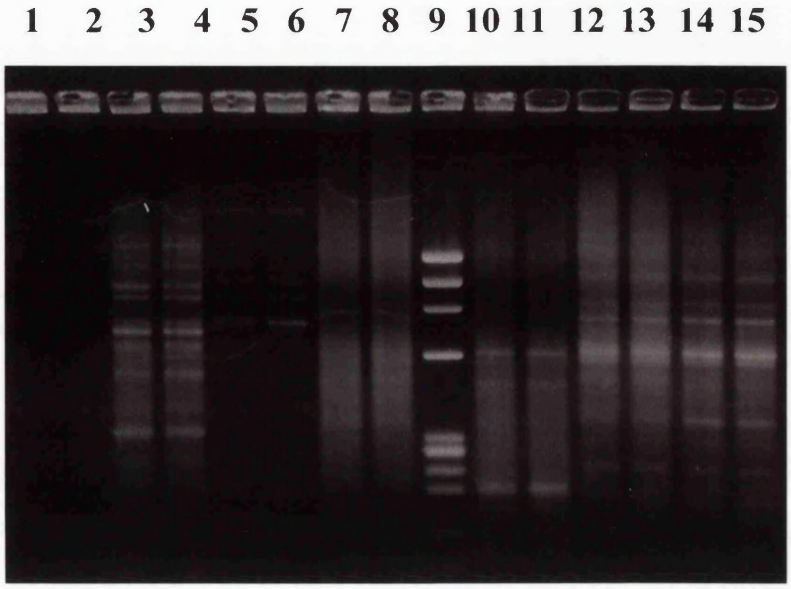


Figure 3.9.1: Primary walking PCR of goldfish rod opsin. Lanes 1 & 2: No DNA controls. Lanes 3 & 4: Primer WP33, 3mM MgCl2. Lanes 5 & 6: Primer WP33, 1.5mM MgCl2. Lanes 7& 8: Primer UNI28, 3mM MgCl2. Lane 9: ϕX174/Hae III DNA ladder. Lanes 10 & 11: Primer UNI28, 1.5mM MgCl2. Lanes 12 & 13: Primer UNI33, 3.0mM MgCl2. Lanes 14 & 15: Primer UNI33, 1.5mM MgCl2.

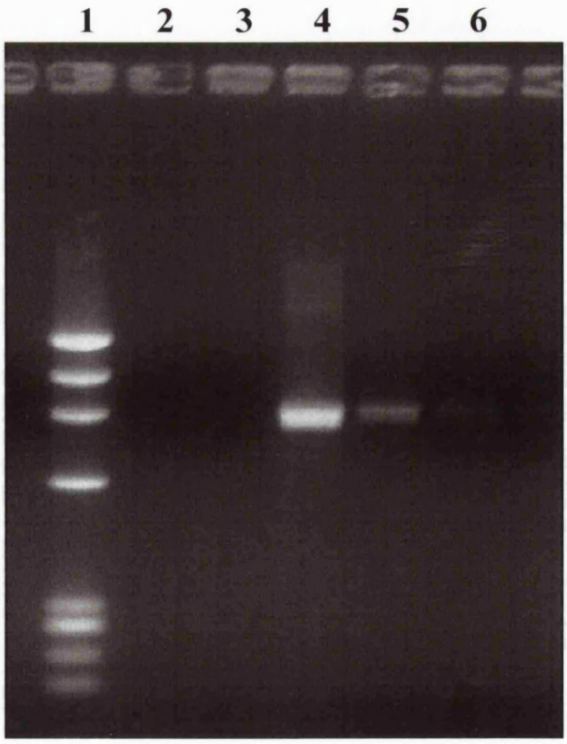


Figure 3.9.2: Secondary walking PCR of goldfish rod opsin. Lane 1: ϕX174/Hae III DNA ladder. Lanes 2 & 3: no DNA control. Lanes 4,5 & 6: secondary walking PCRs of samples in lanes 12, 13 & 14 above respectively

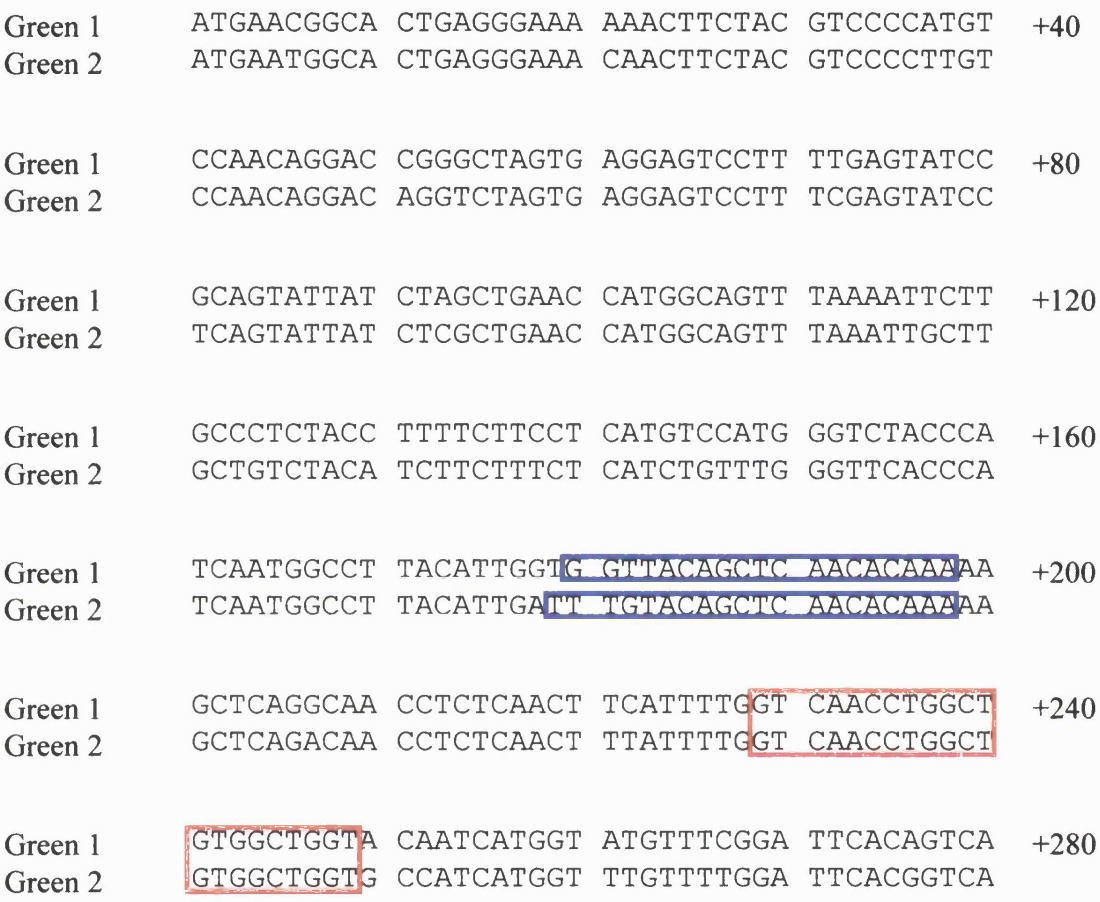


Figure 3.10: Alignment of goldfish green 1 and green 2 opsin genes to show the position of primers used in the first round walking PCR. Sequence boxed in red is the primer, 1st walk green 1 & 2 out, sequences boxed in blue are the nested primers used in the secondary nested PCR, 1st walk green 1 in and 1st walk green 2 in. The numbers to the right indicate the position relative to the translation start site.

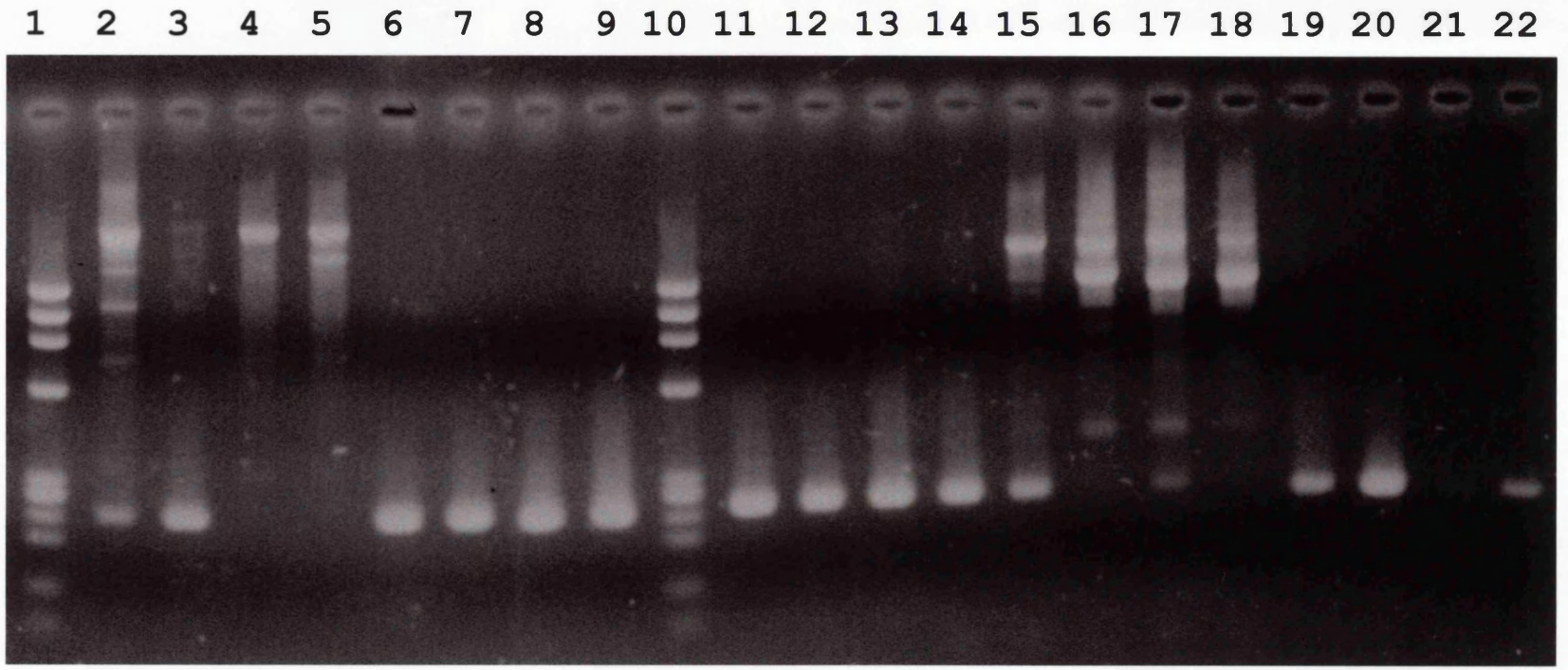


Figure 3.11: Goldfish green opsin 1 & 2 secondary walking PCRs. Lanes 1 &10: ϕX174/Hae III DNA ladder. Lanes 2 – 5: secondary walking PCRs of goldfish green 1 opsin, using UNI28 in primary round. Lanes 6-9: secondary walking PCRs of goldfish green 1 opsin, using UNI 33 in primary round. Lanes 11-14: secondary walking PCRs of goldfish green 2 opsin, using UNI 28 in primary round. Lanes 15-18: secondary walking PCRs of goldfish green 2 opsin, using UNI 33 in primary round. Lanes 19 & 20: No DNA control green 1. Lanes 21 & 22: No DNA control green 2. Lanes 2, 3, 6, 7, 11, 12, 15, 16, 19 and 21: 1.5mM Mg Cl2. Lanes 4, 5, 8, 9, 13, 14, 17, 18, 20 and 22: 3.0mM Mg Cl2

-1454 GCGCTATTTT GGCCTTGTGC TACAGGTCCG AGGCTATTAG CCCCGCCCAG CCAGTTTAAT

-1394 GCACTGTCTC ACACTGGCCC GACAGTGGAA AAGCAGCTAT TAAAGCCTAC TTCACATTGT

-1334 GTGGATATTT TAGGTTCAAA TATTTATCAA TAGATGAACA GAATGCTAAC TGAGACTATA

-1274 **CTAGCCA**GGT ACAGCCAGTA TTTCCCCTGC AGCCTGTGCA GTGTTGATTG TAGAAAGACT

-1214 AGCAAGCAAC TAATGTGCTT AAGGCTCCTC ATTGTAGGGT AACTCTCCTG TAATCTCAGG

-1154 TAAATCTCCT GTAAATC**CTC TTCA**TAAAAT CCACCTGTGT GAAATATGCT GAGCCTTGAG

-1094 CATGTCTA**CC CAGGCC**TTTC ACAGGGACAA GGGTAGCATT AGTCCACCTG ACACTACACA

-1034 ATCTAGACTA TGCAAGAGTG GAAATAAGTT TATGACAAAT ATTATTGTGC TTAGTCCACT

-974 GGTTTTATTA TCCTGAATGA CAAATTACTG ATGAATGAAT TTCTTATTCT TGAATTAATG

-914 TGCTTGATTA TATGTAATTA TTTCTATGAA CATTAGTTAA CTTGGAGAAT TCTGGTTCCT

-854 TCAGGTGTTT GAGAATGACA CCAACATCTG AAAGGTTGTG AGTGCATTTC GCAGGAAACA

-794 TATACTATAT ATATATATAC TATGTATACT TTGAAAGAAT CAAGACCGAT ACATGCCAGA

-734 GCACTTTTTA GGGATTTTTT AGTGTGTGTG AAATGTGAAC TACTTTTCTA TGGTGTTCCT

-674 CACGTATATC TTCCAAGATC CCAGCTTGAC TTGTTGGCAT TAATGCAGGA CAGGTCATCA

-614 CAATAGCATG AAACAGTTTG TGGTTTGATC ACATGAAAGT GTCTTAATTT **TGAATCAG**TT

-554 AAAATAAAAT AAAAACTTTT TTTGTCAAAT GTTTTGCTTA AAAACAGGAA TGTGCTATCT

-494 TGGTTTGAAA TTTTGTTGTC TCACTTGTGC ACTTTGAAAT GTGTCCAAAT **CCAAT**TCGGG

-434 GCATCTTCAT TGAGATGCTA TTTGATGGAA GTCAAGTCAC CTTTATTACA TATA**CAGTTG**

-374 **T**GTCAAAGCA GCTTTACAGT ATTAAATAGG AAAATTGTGT GTAATGATGC AAATGGGACA

-314 ACGGTAAACA CTCAATTTTC AGTTAAATCC AGTTCCATCG TTGAATCCAC TGATGTTCAT

-254 CATACAAGTT GTTGCCCTTA TGTTCTCAAA TAAAGTCCTG TTAGTGTTTT TTTTTTTTCT

-194 TTTTCTATGA CTACTTACCT GCCTTGTGCA GGTACAACCT **TAATAAC**TTG AGAAGTTGCC

-134 AGATATGCAG T**GCAAT**GATG GCTGGGATTA TTTTATTAGC TGTGGCCTAA ATAGACTCAT

-74 GCTGACAGCC TGGAAACATG AGATAATCCC AAGCGAGTCT C**TATAA**AGTG GGGTTCACAC

-14 TTGCCCCATC AAGTCGTAGC ACGGTCCTGC CTCGTTTCTC CACAGTCCTG CCGAGCCATC

+47 CAAACACTAC CACAGAAAGG GGCTGAGCAC AACATCCAAC CGCAACA

Figure 3.12: 5' flanking region of the goldfish rod opsin gene. Sequences in bold show conserved motifs. Numbers on the left represent the nucleotide position relative to the transcription start site (+1). Key: blue, TATA; green, Ret1/PCE1; brown, cMyc; purple, AP-1.

3' end only. Samples were annealed at 54°C and specific bands were observed only for those which had included the UNI33 primer in the primary round (figure 3.11). These bands were excised and eluted from the agarose. Primer pairs used for both walks are shown in table 3.3 below.

	Walk 1	**Walk2**
Green 1 primary PCR	1st walk green 1 & 2 out/ UNI 28 or UNI 33	2nd walk green1 out/ UNI 28 or UNI 33
Green 1 secondary PCR	1st walk green 1 in/ UNI 17	2nd walk green 1 in/ UNI 17
Green 2 primary PCR	1st walk green 1 & 2 out/ UNI 28 or UNI 33	2nd walk green 2 out/ UNI 28 or UNI 33
Green 2 secondary PCR	1st walk green 2 in/ UNI 17	2nd walk green 2 in/ UNI 17

Table 3.3: Primer pairs used for walking PCR to amplify the goldfish green opsin 1 and 2 5' promoter region.

Bands were excised from the gels and the DNA eluted overnight in distilled water. DNA was cloned into the pTag vector and white colonies screened for the presence of an insert of the correct size, using colony PCR with pTag 5' and 3' primers. Those which were positive were sequenced on the Pharmacia ALF sequencer using Autoread and Autofluor kits. From this sequence, gene specific primers were designed and used for the second walk and again a smear of PCR product was observed for the primary round and a discreet band for the secondary round. The WP33 primer was unsuccessful in initial PCRs and so was not used in any further reactions. Aligned

sequences for both the rod, and green 1 and green 2 promoters are shown in figures 3.12 and 3.13

3.2.3 Sequence homology

The sequences of the rod and green opsins were compared for regions of similarity using dot matrix plots produced in Mac Vector. The plots showed no regions of significant similarity between the rod and green opsins but did highlight a region of homology between the two green opsins as shown in figure 3.14. Using Clustal V the sequences were aligned (figure 3.13). A high level of homology is seen between these two sequences for the first 450 bps upstream, interrupted only by two indels and a region of lower identity. Immediately after this, homology falls to around 40% for the next 200 bps, before averaging around 50% for the remainder of the sequenced flanking region.

3.2.4 Conserved motifs

Functionally important sequence motifs are likely to be conserved during evolution and therefore comparisons were made between promoter regions. The goldfish rod and green promoter sequences were compared with rods of other species, including chicken, bovine and *Xenopus*; few regions of similarity were found and these were restricted to a number of short DNA motifs which are detailed in table 3.4. In particular, one heptanucleotide motif was found to be common to all promoter sequences. The sequence, 5'-TGAAGAG-3', was found in both forward and reverse orientation and, in the cases of the two goldfish green opsin promoters, in multiple copies. This sequence forms part of the human retina-specific leucine zipper (NRL) transcription factor binding site (Swaroop *et al.*, 1992) and has been previously identified in the *Xenopus* rod opsin promoter by Batni *et al* (1996).

A second sequence, 5'-TAATAAC-3', was found in all promoters except that for *Xenopus* rod. Footprinting of the bovine rod opsin (DesJardin & Hauswirth, 1996)

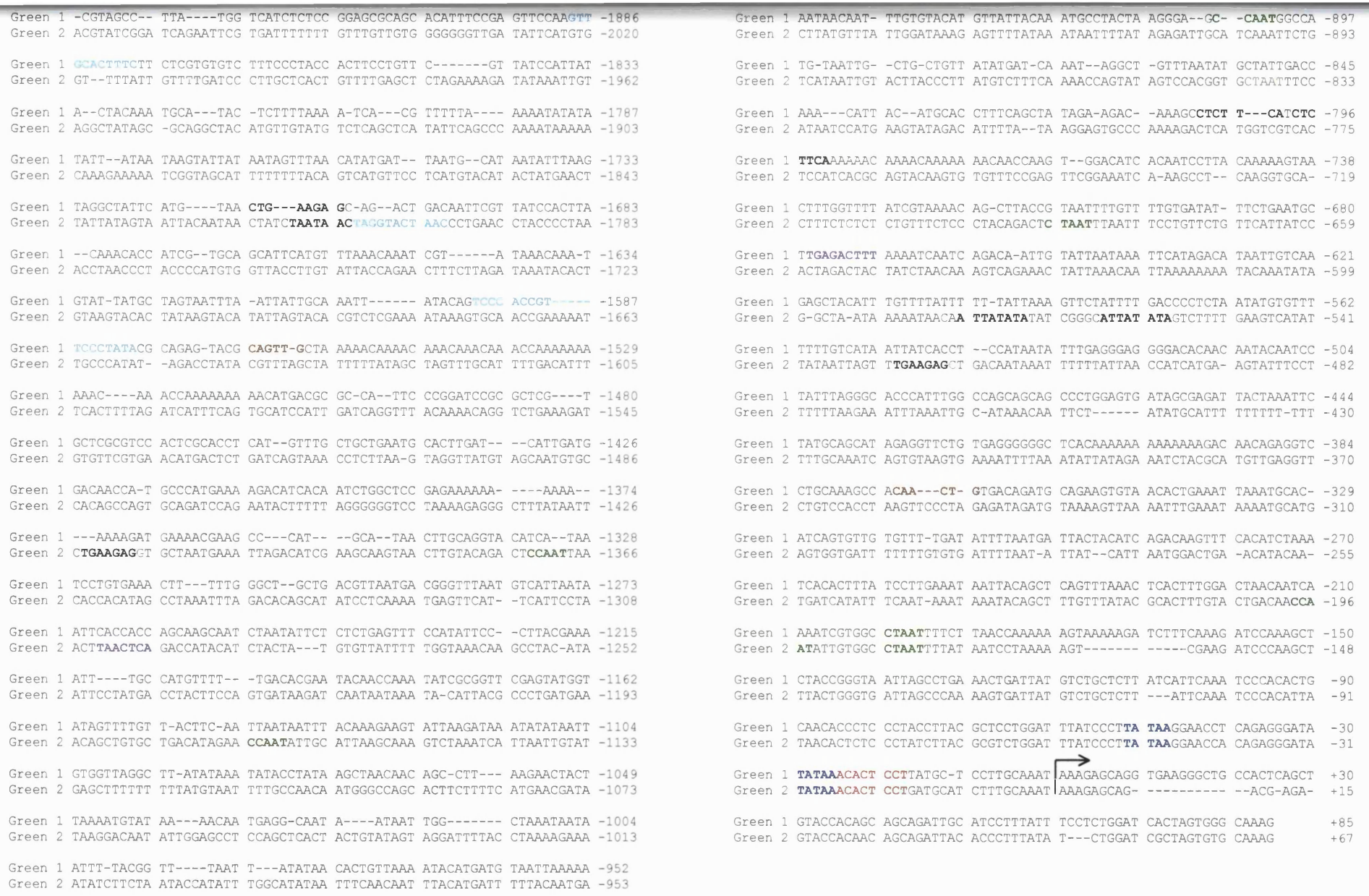
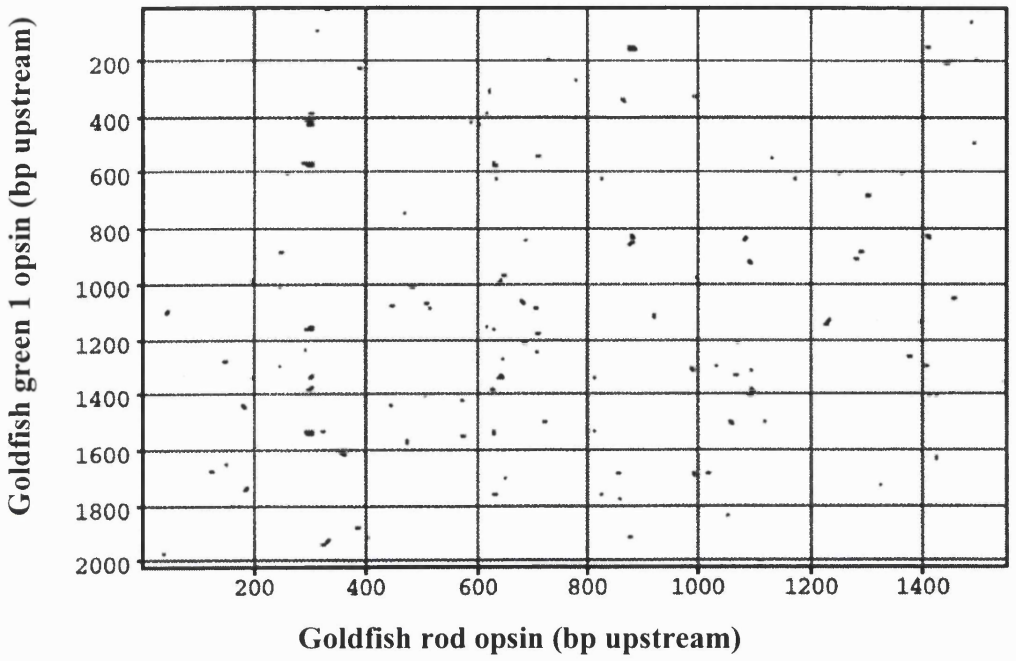


Figure 3.13: Aligned 5' flanking regions of the two green cone opsin genes of the goldfish. Numbers on the right represent nucleotide position relative to the transcription start site (+1). Key: red, Sp-1; blue, TATA; purple, AP-1; green, Ret1/PCE1; brown, cMyc; turquoise, CRS-like. The black arrow indicates the transcription start site.



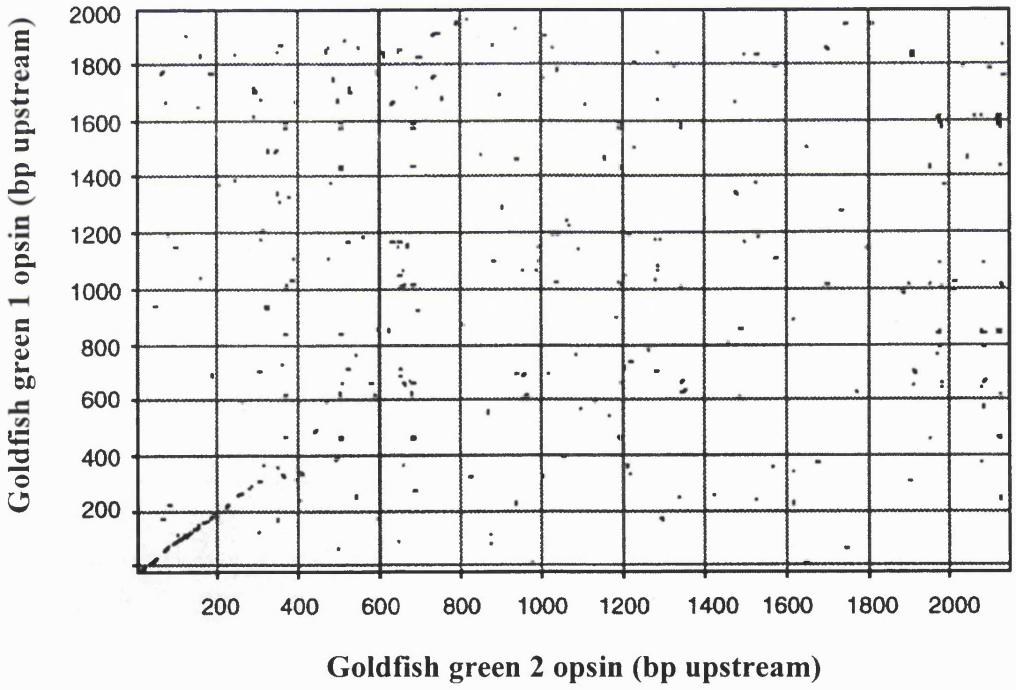


Figure 3.14: DNA identity matricies of goldfish rod, green 1 and green 2 opsins. Window size = 10, Min % score = 90

Sequences present in all opsins examined

Motif sequence	Species and opsin	Location
5'-TGAAGAG-3'	goldfish green 1	-806*, -798*, -1716
	goldfish green 2	-529, -1424
	goldfish rod	-1137*
	chicken rod	-1627
	bovine rod	-2016*
	xenopus rod	-409*
5'-TAATAAC-3'	goldfish green 1	-932*
	goldfish green 2	-1817
	goldfish rod	-154
	chicken rod	-104
	bovine rod	-92

Sequences present in rod opsins only

Motif sequence	Species and opsin	Location
5'-CCCAGGCC-3'	goldfish rod	-1086
	chicken rod	-598*
	bovine rod	-23*
5'-CTAGCCA-3'	goldfish rod	-1274
	chicken rod	-1214*

Table 3.4: Conserved DNA motifs in the 5' flanking regions of the rod and cone opsin genes. Location numbers represent position relative to the transcription start site. Sequences present in reverse orientation are indicated by an asterisk.

highlighted a proximal AT rich sequence (BAT-1) which was protected from DNase I digestion. The BAT-1 region is composed of 20 nucleotides and within this lies this second conserved motif.

The third motif composed of eight nucleotides, 5'-CCCAGGCC-3', was found in the rod promoters of goldfish, chicken and bovine. This core sequence however, was not in any protected regions identified by DNase I footprinting of either the chicken (Sheshberadaran & Takahashi, 1994) or bovine (DesJardin & Hauswirth, 1996) 5' flanking sequences. Finally, a fourth motif was found in goldfish and chicken rod promoters but again, this had not been previously identified as a protected region by DNase I footprinting.

3.2.5 Transcription factor binding sites identified in the rod promoter

Using the TESS database, common transcription factor binding sites were identified in the rod promoter. A TATA box is present at -33 upstream of the translation initiation start codon. A Ret1/PCE1 site (CC/TAAT), shown to be present as a conserved element in the upstream regions of several photoreceptor-expressed genes (Morabito *et al.*, 1991; Sheshberadaran & Takahashi, 1994; Yu *et al.*, 1996; DesJardin & Hauswirth, 1996; Batni *et al.*, 1996; Boatright *et al.*, 1997b; Shaaban & Deeb, 1998; Su *et al.*, 2000), is present at -444, placing it 411 bps upstream of the TATA box. There is also a modified Ret1/PCE1 site, GCAAT, at -123. An NRE-like site is present at -75 and a further potential NRL binding site at -1137. NRL has also been shown to bind to an extended AP-1-like site (Rehemtulla *et al.*, 1996) and in the goldfish rod, a site is found at -564. Additionally, there is a cMyc site at -370. There is also a region which shows limited homology to the LCR (Wang *et al.*, 1992b) although the core sequence CG(A/T/C)TGG found in the RER is not present (figure 3.15).

3.2.6 Transcription factor binding sites identified in the green opsin promoters

The alignment of the 5' regions of the two green cone opsin genes immediately upstream of the coding sequence requires two insertion/deletion (indel) events of 13 and 12 bps respectively. Both genes have two TATA boxes and the effect of the first

GF rod	G GTAAATCTCCTGTAAA TCCTCTTCATAAA ATCCACCT	-1119
Bovine	CACCTTGACCT C TTTAGCGATGGACCTAATCA CAGGGCCAAGCA	-1887
Human	CACCTTAACCT C ATTAGCGTTGGGCATAATCACCA GCCAAGCG	-1746
Mouse	CACCTTGACCT C ATTAACGCTGGTCTTAATCACCAA GCCAAGCT	-1418
Rat	CACCTTGACCT C ATTAACGCTGGTCTTAATCACCAA GCCAAGCT	-1375
LCR	GA CTTGATCTTCTGTTAG CCCTAATCATCAA TT AGCA	-3430

Figure 3.15: Alignment of the goldfish rod opsin LCR-like region with mammalian rod RER's and human red/green LCR. Identical bases are shown in blue.

indel is to align the box at -29 in green 1 with that at -30 in green 2, and the box at -51 in green 1 with that at -52 in green 2. A number of sites for the ubiquitous transcription factor Sp1 are present in both upstream regions, although only one at -25 in green 1 and -26 in green 2 is at an equivalent position in both genes. AP-1 sites are present in green 1 at -678 and in green 2 at -1304. Green 1 has two cMyc sites at -373 and -1567.

The Ret1/PCE1 element is present in the upstream regions of both green cone opsin genes. In green 1, two motifs are present at -199 and at -906, whereas in green 2, six are present at -185, -198, -689, -841, -1172 and -1373. However, the more distal site for green 1 and the five most distal sites for green 2 are considerably more distant than the functional motifs in other photoreceptor genes; it would seem unlikely therefore that these sites are of functional significance. The remaining site for green 1 is 170 bps and the two remaining sites for green 2 are 155 and 168 bps upstream of the TATA box. None of these sites are associated with GC boxes, although there is a limited GC-rich region immediately upstream of the element in green 1 and between the two adjacent elements in green 2.

3.2.7 Additional transcription factor binding sites

In addition to the Ret1/PCE I sites (Morabito *et al.*, 1991; Kikuchi *et al.*, 1993), a number of other binding sites for retina-specific nuclear proteins have been identified in the upstream regions of opsin genes. In a detailed developmental study of the binding of nuclear proteins to 2.1 kb of the 5' flanking region of the bovine gene, 10 CRS sequences were identified (DesJardin & Hauswirth, 1996) that are protected from DNase I digestion during one or more stages of development. A search of the upstream regions of the goldfish rod and green cone opsin genes revealed some limited homology to these elements. As shown in Table 3.5, the goldfish rod upstream region would appear to contain elements with homology to CRS-3, CRS-6 and CRS-8. Putative CRS-3 elements are also present in the 5' flanking region of both green 1 and green 2 and CRS-7 in green 1. However a search of the upstream regions of goldfish rod and green opsins failed to reveal homology to other conserved protein-

binding elements such as; CRS-1, CRS-2, CRS-4 and CRS-5, which have been reported in mammals (DesJardin & Hauswirth, 1996), proximal and distal glass-like sequences in chicken (Sheshberadaran & Takahashi, 1994), and CREB and glass in *Xenopus* (Batni *et al.*, 1996).

CRS-3	CTGGCACTAAG	CRS-7	CCCCACC-CACCCGCCA
rod	AT-GCACTGTC	green 1	TCCCACCGCTCCCTATA
green 1	GTTGCACTTTC		
green 2	TAGGTACTAAC	CRS-8	GTCTGGCC--ACCAGGGG
		rod	CACTGGCCCGAC-AGTGG
CRS-6	TCCCCGCCCCCCCAG		
rod	GCCCCGCCCAGCCAG		

Table 3.5: Goldfish sequences with homology to CRS of bovine rod opsin. Identical nucleotides are shown in blue.

3.2.8 Transcriptional activity of the two green cone opsin genes

Based on Northern blot analysis of the two green cone opsin cDNAs, Johnson *et al.* (1993) concluded that green 1 was transcribed at a higher level than green 2. Since cross-hybridization between gene probes and the two green cone cDNAs prevented an estimate of differential binding, this experiment relied on a small size difference between the two transcripts (1.5 kb for green 1, 1.4 kb for green 2) that is not easily resolved on Northern blots. The expression of these two genes has therefore been reassessed using a PCR method that overcomes these difficulties. cDNA was generated from retinal polyA$^+$ mRNA, isolated from six individual fish by the 3' RACE method, and used in a PCR with forward (radiolabelled) and reverse primers

RACE method, and used in a PCR with forward (radiolabelled) and reverse primers designed to regions that are identical in both genes. The resulting product of 819bp, was digested with restriction enzymes that differentiate the two transcripts. Figure 3.16 shows an agarose gel of the restriction digests. For the *Sac I* digests, the undigested 819bp bands represent relative amounts of green 2 cDNA, the smaller bands being the digestion products of green 1. Similarly for the *Eco RI* digests, the 819bp band represents green1 cDNA and the smaller bands, the digestion products of green 2. The 819bp bands were excised and scintillation counted, the results of which are shown in table 3.6. After scintillation counting these bands, the difference between the means was calculated using the *t* test for paired samples (Welkowitz *et al.*, 2000). A value of $P<0.001$ was obtained indicating that there is a significant difference between the values for green 1 and green 2 and that this difference was highly unlikely to have occurred by chance. A ratio of approximately 1:1.9 for the levels of green 1 : green 2 cDNAs was obtained. This result differs therefore from the previous estimate (Johnson *et al.*, 1993) but, for the reasons discussed above, is likely to be a more accurate estimate of the relative expression of these two genes.

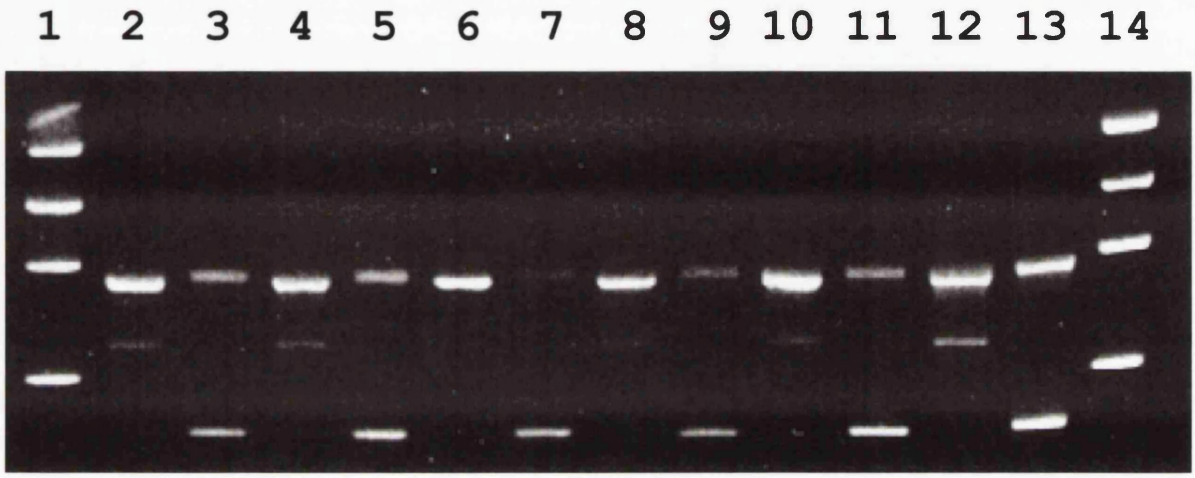


Figure 3.16: *Sac I & Eco RI* digests of γ32p dATP labelled goldfish green opsin PCRs. Lanes 1 & 14, ϕX174/Hae III DNA ladder. Lanes 2 & 3: goldfish 1. Lanes 4 & 5: goldfish 2. Lanes 6 & 7: goldfish 3. Lanes 8 & 9: goldfish 4. Lanes 10 & 11: goldfish 5. Lanes 12 & 13: goldfish 6. Even lanes = *Sac I* digests. Odd lanes = *Eco RI* digests.

A

Sample	G1 *Eco RI* a)	G1 *Eco RI* b)	G2 *Sac I* a)	G2 *Sac I* b)	G1:G2 a)	G1:G2 b)
Control	17	**16**	17	**16**	-	-
GF 1	114	**121**	231	**250**	1:2.0	1:2.0
GF 2	149	**145**	273	**201**	1:1.8	1:1.4
GF 3	69	**63**	124	**121**	1:1.8	1:1.9
GF 4	75	**59**	175	**123**	1:2.3	1:2.1
GF 5	88	**58**	245	**154**	1:2.8	1:2.7
GF 6	166	**136**	222	**195**	1:1.3	1:1.4
Total	661	582	1270	1044	1:1.9	1:1.8

B

	G1 (*Eco RI*)	G2 (*Sac I*)	G1:G2
Mean	104	193	1:1.9

Table 3.6: **A**) Scintillation counts of the two goldfish green opsins. *Eco RI* results are for the relative amounts of green 1, *Sac I* results for green 2. a) represents the first restriction enzyme digest, b), the subsequent digest after further enzyme was added. the last two columns show the ratios of green 1 to green 2 for a) & b). Figures shown in bold were used in the *t* test to calculate the p value. **B**) Mean of scintillation counts for G1 and G2 and the ratio of G1:G2

3.3 Discussion

The expression of a large number of eukaryote genes is transcriptionally regulated by the binding of various sequence-specific proteins to target DNA elements located in the immediate upstream region. A number of such *cis*-acting transcription factors are common to many genes whereas others are limited to either a set of genes with a similar pattern of expression or to a single gene. Studies of the 5' flanking regions of retinally-expressed genes in a number of species have identified the presence of both types of element; such DNA motifs and the transcription factors with which they interact have been conserved therefore during evolution. The targeting of the expression of different opsin genes to either the rod photoreceptors or the different types of cones is an example of cell-specific gene regulation. In the studies so far, sequence conservation and the presence of protein-binding elements has been demonstrated in the 5' flanking regions of rod opsin genes in a number of mammalian species (Morabito *et al.*, 1991; Zack *et al.*, 1991; DesJardin & Hauswirth, 1996), the chicken (Sheshberadaran & Takahashi, 1994) and *Xenopus* (Batni *et al.*, 1996).

As a first step in identifying *cis*-acting elements controlling expression in the goldfish rod and green cone opsin genes, the 5'promoter regions have been cloned and sequenced. Goldfish sequences were compared to each other and also to opsins of other vertebrates (DesJardin & Hauswirth, 1996; Batni *et al.*, 1996; Sheshberadaran & Takahashi, 1994; Bennett *et al.*, 1995; Nathans & Hogness, 1983; Nathans & Hogness, 1984; Baehr *et al.*, 1988). The goldfish green opsins show significant homology with each other for the first 450bp upstream but not with the rod sequence or those of the other vertebrates. An alignment between the human LWS, MWS, and marmoset LWS/MWS opsin gene sequences showed significant homology for the first 236 bp upstream (Dulai *et al.*, 1999) and in transgenic mouse experiments, the targeting of rod and blue cone expression to the appropriate photoreceptor cell type was largely determined by the first 500 base pairs upstream (DesJardin & Hauswirth, 1996; Lem *et al.*, 1991). This proximal promoter region therefore, appears to be critical in the regulation of expression within the same class of opsin genes in closely related species. When studied in greater detail, *cis*-acting elements known to be

essential for transcription of opsins have been identified in this promoter region and it is these elements which have been conserved between many species throughout evolution.

The TATA box, which plays a critical role in the formation of the transcription initiation complex, is found twice in both green opsin promoters in close proximity to the transcription start site. In comparison, other vertebrate opsin promoters including the goldfish rod opsin only have one TATA element. Without DNase I footprinting experiments, we cannot be sure which is the functional TATA box. In chicken, *Xenopus*, bovine (Sheshberadaran & Takahashi, 1994; Batni *et al.*, 1996; Nathans & Hogness, 1983) and goldfish rod, the TATA element is located between approximately -29 and -40bp upstream from the Kozak sequence, ANNATG G (Kozak, 1983). This sequence indicates the transcription start site and suggests that the first TATA box in each green opsin gene, at -29 and -30 upstream, is the functional element. However, when the GC rich flanking regions and their location relative to the CCAAT boxes is taken into account, it is the second group of TATA elements at -51 and -52 which are more similar to the other opsin promoters.

There are numerous motifs present in the goldfish opsin promoters containing the core Ret1/PCE1 sequence, CC/TAAT. The Ret1/PCE1 sites in chicken, *Xenopus*, bovine and human rods are centred at -125, -130, -135 and -135 respectively (Sheshberadaran & Takahashi, 1994; Batni *et al.*, 1996; Nathans & Hogness, 1983; Nathans & Hogness, 1984). In the goldfish rod the motif is present at -123, in green 1 at –199, and in green 2 at –185 and -198. The other sites are much further upstream than the functional motifs of the other photoreceptor genes. Whether they are also able to bind retinal nuclear proteins, including CRX, and influence development of the retina remains to be determined.

Nrl has been shown to bind to both NRE and an extended AP-1-like site (Rehemtulla *et al.*, 1996). The goldfish rod promoter region has an NRE-like site and both the rod and green opsins have AP-1 motifs. The NRE motif is important in increasing activity of the chick rod opsin promoter (Kumar *et al.*, 1996) and it has also been found in a wide range of vertebrate opsin promoters (Zack *et al.*, 1991; Morabito *et al.*, 1991; Gale *et al.*, 1992; Nathans & Hogness, 1983; Nathans & Hogness, 1984;

Batni *et al.*, 1996) so it is not surprising that we also find these sites in the goldfish opsin promoters. Additionally, sites homologous to the ubiquitous Sp1 and cMyc motifs have been identified and also some of the CRS elements, although these are not as widely conserved as the Ret1/PCE1 sites. Of the eight CRS elements (DesJardin & Hauswirth, 1996), only CRS-3 is present in the rod and both green opsins. CRS-6 and 8 can be identified in the upstream region of the goldfish rod opsin and CRS-7 in the goldfish green 1 cone opsin. CRS-8 is the only site other than Ret1/PCE1 which is retinal specific.

Although similar to the mammalian RER, the goldfish rod promoter sequence lacks the core CGNTGG motif of RERs and therefore shows higher homology to the human LCR (figure 3.15). Does this motif function as an LCR? Amongst the mammals, the human, mouse and bovine LCRs are perfectly conserved. The goldfish rod LCR-like region however, shows only around 60 % homology and is located much closer to the transcription initiation site (at -1250 compared to -3430 for the human LCR), in this respect it more closely resembles the RERs. The significance of the LCR in regulating expression of mammalian LWS/MWS opsins would suggest that a similar enhancer is present in the opsins of other species. Despite the evolutionary distance between fish and mammalian opsins, there is a high degree of sequence conservation between the LCRs and RERs of these species, suggesting that they are functionally important. It is probable therefore, that the LCR and RER perform a similar function in the regulation of opsin expression. No regions showing homology to the LCR or RER were found in the promoters of the two rod-like green opsins. It is likely that these analogous regions are further upstream, as in the human LCR, in a region which has yet to be sequenced.

In addition to the known *cis*-acting elements, a comparison between goldfish rod and green opsins and the rod opsins of chicken, bovine and *Xenopus* revealed further motifs which are conserved between species but were not detected by TESS as known transcription factor binding sites. One heptanucleotide sequence 5'-CCCAGGCC-3' was found only in the rods in both forward and reverse orientation. Although its significance is unknown, one can speculate that it might be involved in regulating rod

specific expression. The other motifs identified turned out to be part of the known *cis*-acting elements, the NRL binding site and BAT-1.

Since an orthologous green cone opsin gene is present in birds (Okano *et al.*, 1992) and reptiles (Kojima *et al.*, 1992), the original gene duplication that gave rise to the rod and green cone opsin genes must have occurred before the evolution of the first tetrapods at around 450 mya. The absence of any extensive regions of identity in the upstream regions of these genes in the goldfish is a reflection of the age of this duplication and demonstrates that, except for short transcription factor binding sites, the 5' flanking regions of genes of related function are not tightly conserved.

It has previously been demonstrated by *in vitro* expression (Johnson *et al.*, 1993) that the λ_{max} of the two green cone opsin pigments in the goldfish differ by about 5 nm. Two classes of cone photoreceptors in the intact goldfish retina contain a green-sensitive pigment, a class of single cones and the accessory member of the double cones, suggesting the possibility that each cone class may express only one of the green cone opsin genes. However, the ratio of double to single green cones of 3:1 (Stell & Harosi, 1976) is not directly correlated with the expression levels of green 1 to green 2 of 1:2. MSP analysis has also failed to reveal any difference in the λ_{max} of the pigments present in these two cone classes. Under conditions where 11-*cis* retinal predominates, the λ_{max} for both single green cones and the accessory member of the double cones was 509nm (Tsin *et al.*, 1981). Since both mRNAs are present in the retina, the most likely explanation is that both genes are simultaneously expressed in these two cone classes and the visual pigment present is a mixture of the two types of green opsin.

Contrary to the previously published estimate (Johnson *et al.*, 1993), results from this study show that transcripts from green 2 are present in higher quantities than from green 1, indicating that the promoter of green 2 is more active than green 1. A significant feature of the two promoter regions is that they differ in the number of Ret1/PCE1 elements present: six copies in green 2 but only two in green 1, although spatial relationships suggest that only two elements in green 2 and one in green 1 are sufficiently close to the TATA box to be functional. The Ret1/PCE1 element is a highly conserved DNA motif present in the 5' flanking regions of a number of

mammalian retinally-expressed genes (Morabito *et al.*, 1991; Yu, 1993; DesJardin & Hauswirth, 1996; Kikuchi *et al.*, 1993) that have been shown to bind a 40 kD protein. This element is related to the RCS element originally identified in the upstream region of the *Drosophila Rh*1 opsin gene where it was shown by site-directed mutagenesis experiments to determine the rate of transcription (Mismer & Rubin, 1987; Fortini & Rubin, 1990). The difference in the production of the two green cone opsin mRNAs may be the result therefore of the differing number of Ret1/PCE1 elements present in the promoter regions.

In conclusion, approximately 1.5kb of the 5' flanking region of the goldfish rod and over 2kb of the two goldfish green opsin genes have been sequenced. A number of conserved transcription factor binding sites have been identified and also a motif which has not been previously reported. Green 2 opsin is expressed at a higher level than green 1 and this is paralleled by the presence of an increased number of Ret1-/PCE-1 sites in the promoter of green 2. Extensive homology was seen for the first 450bp upstream of the two green cone opsin genes but there is otherwise little homology between these goldfish genes and the equivalent regions of the rod opsin genes of other species.

CHAPTER 4

CHARACTERISATION OF THE BLUE OPSIN GENES IN THE COTTOID FISH OF LAKE BAIKAL

4.1 Introduction

4.1.1 Short-wave vision in cottoid fishes

The molecular mechanisms of red and green colour vision are relatively well understood and have been studied extensively, especially in mammalian systems. In contrast, the process by which spectral tuning occurs within the short-wave pigments is far less clear. There have been no previous studies of spectral tuning within the blue cone opsin class and therefore the aim of this study is, for the first time, to determine which sites are responsible for a short wave shift within the blue cone group of pigments. A short-wave shift in the λ_{max} values of a group of closely related species, the cottoid fish of Lake Baikal, has been reported previously (Bowmaker *et al.*, 1994) making these fishes an ideal group in which to further investigate spectral tuning at the molecular level.

In this study, the blue cone pigments of four species, *Cottus kesslerei, Cottocomephorus inermis, Limnocottus eurstomias* and *Batrachocottus nicolskii*, were cloned, sequenced and compared in order to identify potential spectral tuning sites. A species of cottoid native to rivers in Britain, *Cottus gobio* was also examined and used as the wild type pigment for subsequent mutation and expression studies. The λ_{max} of the blue pigments in the Lake Baikal species varies from 450nm in *Cottus kesslerei* and *Cottocomephorus inermis* to 428nm in *Limnocottus eurstomias* and *Batrachocottus nicolskii* (Bowmaker *et al.*, 1994). In comparison, recent microspectrophotometric analysis of *Cottus gobio* found its blue pigment to be long wave shifted to a λ_{max} of 470nm (Bowmaker, personal communication 1999).

Plate 4.1: Geographical map showing the location of Lake Baikal in Eastern Siberia.

4.1.2 Lake Baikal

Lake Baikal is situated in Eastern Siberia (plate 4.1) and consists of three deep basins ranging from 900m to 1600m. It is 640km long and contains approximately one fifth of the world's liquid freshwater. A process of deep convection occurs episodically due to effects of temperature and pressure on the water density and shifting downwards of the interface between the deep and surface layers by an external force such as wind (Weiss *et al.*, 1991). As a result, even in the deepest regions, oxygen levels are 75-80% of those at the surface allowing colonisation of the entire water column by a wide range of species. Approximately 2,500 species and 198 sub-species of animals inhabit Baikal, a biodiversity which is not matched by any other lake. Lake Tanganyika in East Africa, its nearest rival, has only half this number of species. On the basis of a complex of data, it has been estimated that colonisation of shallow waters by cottoid fishes occurred at the end of the Miocene approximately 5mya (Timoshkin, 1997; Hunt *et al.*, 1997) and that predation by seals and other fishes may have forced the cottoids to greater depths where vacant ecological niches existed.

4.1.3 Classification of cottoid fishes

The cottoids of Lake Baikal are teleost fishes which are characterised by an externally symmetrical caudal fin (plate 4.2). They are members of the order scorpaeniformes of which there are 25 families although the arrangement of family boundaries is currently the subject of much disagreement (Nelson, 1994; Timoshkin, 1997). Three of these families, cottoidae, comephoridae and abyssocottidae are found in Lake Baikal. Of the 300 cottoid species worldwide, 29 are found in Lake Baikal and of these, 27 are endemic.

The cottidae family, commonly known as sculpins, usually have a large eye high on their head. Baikal species belonging to this family occupy a range of habitats from shallow littoral (1-5m) to abyssal (300-1000m). They are characterised by a glassy, dull, translucent body and both known species are endemic to Lake Baikal. The

A

B

Plate 4.2: Cottoid fishes of Lake Baikal. A) *Comephorus baikalensis* B) *Cottus kesslerei*

abyssocottidae are found primarily in Lake Baikal and also inhabit a range of depths from the sub-littoral to abyssal. Comephoridae are commonly known as Baikal oilfishes and live in deep waters down to 1500m.

4.1.4 Vision in cottoid fishes

Vision studies in Baikal cottoids have primarily focused on relationships between visual pigments and the photic environment (Bowmaker *et al.*, 1994), the structure of rod opsin and spectral tuning of rod visual pigments (Fitzgibbon *et al.*, 1995; Hunt *et al.*, 1996) and corneal colouration (Kondrashev *et al.*, 1986; Gnyubkina & Levin, 1987).

Bowmaker *et al.* (1994) studied 17 species of Baikal fish (figure 4.1). In general, the retinas were found to contain rods, large green-sensitive double cones and small blue-sensitive single cones. Only the littoral species *Cottus kesslerei* and *Paracottus kneri* had additional red-sensitive cones. In the deeper dwelling species, *Cotinella boulengeri* and *Abyssocottus korotneffi*, no blue cones were observed and their presence was also questionable in three other species. *Comephorus dybowskii* and *Comephorus baicalensis* had pure rod retinas. A well ordered cone mosaic is present in the surface species; double cones are arranged in a square with the single cones occupying a central position. A mosaic is also preserved in pelagic species although the double cones are in rows. In comparison, deep-water species have a poorly ordered retina and double cones are often absent (Bowmaker *et al.*, 1994). The size of rod and cone cells and the presence of coloured corneas are also related to depth of habitat (Bowmaker *et al.*, 1994; Kondrashev *et al.*, 1986 Gnyubkina & Levin, 1987). Surface species tend to have thinner, shorter cells and corneas are often coloured yellow or orange; in deeper waters where light is restricted the cells tend to become longer and thicker with a higher density of visual pigment to maximise photon capture and corneas are usually colourless. For both rods and cones, there is a stepwise shift in the λ_{max} of the visual pigment towards shorter wavelengths in fish with increasing depth of habitat (figure 4.1). This cannot be attributed to the use of alternative chromophores because as shown by Bowmaker *et al.* (1994) the oxime

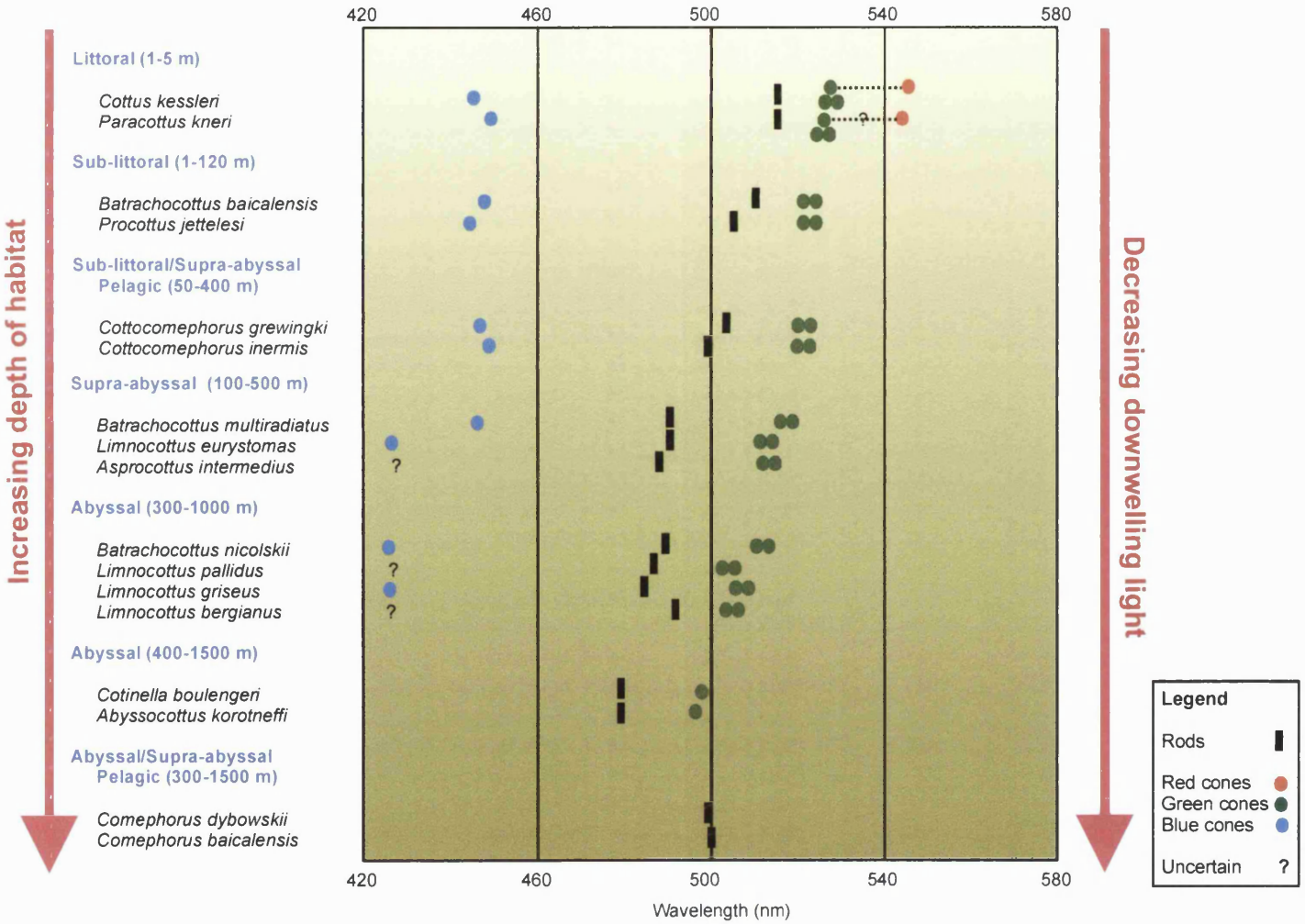


Figure 4.1: Spectral location of rod and cone visual pigments in 17 species of Baikal Cottoids Adapted from Bowmaker *et al.* 1994..

peak formed after pigment bleaching with hydroxylamine consistently corresponded only to that of the A1 chromophore, 11-*cis* retinal. In comparison, most freshwater fish utilise the A2 chromophore, 11-*cis* 3 dehydroretinal, or a mixture of A1 and A2. The Lake Baikal cottoids are unusual therefore in having pure rhodopsin based visual pigments, which may suggest a marine origin (this will be discussed further in the following chapter).

The short wave shift in the λ_{max} of the visual pigments with increasing depth habitat is partially a reflection of adaptation to habitat, however the down-welling light at 400m is not blue. Even in the upper layers of the Lake, between 50-200m, the maximum transmission is around 550-600nm, depending upon the season. There are no bioluminescent sources in the Lake although chemiluminescence below 400m does provide low background light. (BEZRUKOV et al., 1984) This raises the question as to why the blue visual pigments absorb light maximally between 428 and 450nm when the ambient light at this depth is in the yellowish-green part of the spectrum. One possible theory is that pigments are short-wave shifted to reduce background noise and thus enhance the sensitivity of the eye at low levels of ambient illumination (Barlow, 1957). This displacement of the maximum sensitivity of the eye towards the blue end of the spectrum at low light levels occurs in a wide variety of vertebrates and is termed the Purkinje shift. Shifting the spectral sensitivity of a pigment towards shorter wavelengths reduces thermal decomposition since the energy required for activation is greater and thus benefits the sensitivity of the eye. Long wavelength cone pigments have a much higher background "dark" noise and, in comparison to rods and short wave cone pigments, are less stable (Donner, 1992). It has also been proposed that the spontaneous generation of dark noise by photoreceptors is temperature and pH dependent (Barlow *et al.*, 1993; Ashmore & Falk, 1977).

4.1.5 Spectral tuning mechanisms of visual pigments

The absorption maximum of a visual pigment is primarily regulated by the interaction of charges on the chromophore with dipolar residues in the opsin-binding pocket. In a comparison of different visual pigments with identical chromophores, any difference

in spectral properties must therefore arise from differences in the opsin amino acid sequence. Most spectral tuning sites identified so far are situated within the transmembrane helices, face into the lumen of the chromophore binding pocket and involve a non-conservative substitution, i.e. a substitution between amino acids with different properties. A list of amino acids and their relative properties is shown in table 4.1. One exception to these rules is site 100. Here, replacement of Ser with Tyr in the human long wave pigment results in a 4nm short wave shift yet this residue is in the extracellular loop between helices II and III; its effect is therefore indirect as it is not in close proximity to the chromophore but may act by altering the basic structure of the opsin molecule. Chloride ions are also known to indirectly affect the absorption spectra of visual pigments (Crescitelli & Karvaly, 1991). By binding to His and Lys residues at sites 181 and 184 in the extracellular loop between helices IV and V, chloride ions shift the λ_{max} to longer wavelengths. These residues are conserved in all long wave cone pigments but are absent in middle wave sensitive pigments including those of the mouse and blind mole rat (*Spalax ehrenbergi*) (Sun *et al.*, 1997; David-Gray *et al.*, 1998), in rods and in short wave shifted opsins (Wang *et al.*, 1993).

The extent of delocalisation of the electronic charge on the chromophore determines its λ_{max} and is influenced by three factors: the strength of the interaction between the protonated Schiff base and the counterion, the presence of full or partial charges or polarisable groups close to the chromophore, and the degree of planarisation of the polyene chain. An increase in the strength of the interaction between the Glu113 counterion and the protonated Schiff base (PSB) by the placement of polar residues within the PSB region, results in a short wave shift in the λ_{max} due to stabilisation of the ground state (Blatz *et al.*, 1972; Kakitani *et al.*, 1985; Lin *et al.*, 1998). In comparison, the placement of polar residues and an increase in delocalisation at the cyclohexyl-ring end of the chromophore results in a long wave shift in λ_{max}. Delocalisation of π electrons is brought about by photoexcitation of 11-*cis* retinal. By increasing the delocalisation and reducing the energy difference between ground and excited states, the pigment is long wave shifted; conversely a decrease in delocalisation and an increase in the energy difference results in a short wave shift (figure 4.2) (Mathies & Stryer, 1976; Lin & Sakmar, 1999).

Amino acid	1 & 3 letter codes	Properties
Glycine Alanine Valine Leucine Isoleucine	G Gly A Ala V Val L Leu I Ile	Non-polar residues with hydrophobic side chains. Prefer an environment within a protein where they are shielded from H_2O.
Proline	P Pro	Cyclic structure which kinks and folds protein architecture. Non-polar and hydrophobic.
Phenylalanine Tryptophan Tyrosine	F Phe W Trp Y Tyr	Have aromatic side chains. Hydrophobic residues containing localised pi-electron clouds. Tyr has an OH group which makes it less hydrophobic than Phe or Trp.
Cysteine Methionine	C Cys M Met	Have sulphur atoms. In Cys the sulphydryl group allows disulphide bonds. Hydrophobic.
Serine Threonine	S Ser T Thr	Ser & Thr are Ala & Val with OH groups. Hydrophilic residues.
Lysine Arginine Histidine	K Lys R Arg H His	Positively charged residues with polar groups. Highly hydrophilic.
Aspartic acid Glutamic acid	D Asp E Glu	Negatively charged residues with acidic side chains.
Asparagine Glutamine	N Asn Q Gln	Asn and Gln are uncharged Asp and Glu. Have amide (NH_2) groups

Table 4.1: Classification of amino acids grouped according to properties and side chains (Stryer, 1988).

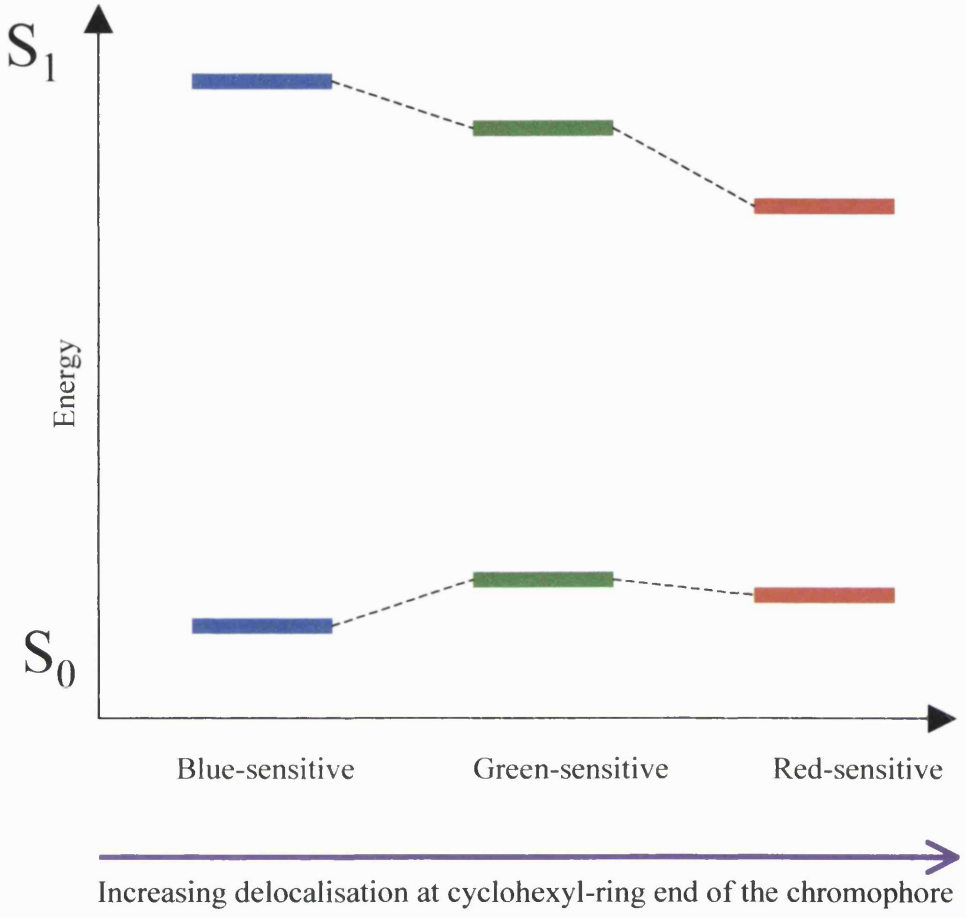


Figure 4.2: The absorption spectrum of rhodopsin represents to a large degree the difference in energy levels between the ground state (S_0) and the excited state (S_1). With a bathochromic shift, delocalisation increases and the interaction between the PSB and the counterion becomes weaker.

Recent advances in spectroscopic techniques have enabled examination of the molecular basis of spectral tuning through measurement of resonance Raman vibrational spectra of the retinal chromophore in recombinant visual pigments. Comparisons of spectra for red, green and blue pigments confirmed that the single most important physical interaction in the opsin shift was that between dipolar residues and the chromophore ground and excited-state charge distributions (Lin *et al.*, 1998; Lin & Sakmar, 1999; Kochendorfer *et al.*, 1999).

4.1.6 Spectral tuning sites

The process of spectral tuning has been extensively studied in the visual pigments of many species. Candidate sites have been proposed on the basis of sequence comparisons (Neitz *et al.*, 1991; Ibbotson *et al.*, 1992; Williams *et al.*, 1992; Archer *et al.*, 1995; Hunt *et al.*, 1996; Hope *et al.*, 1997; Shyue *et al.*, 1998; Yokoyama & Radlwimmer, 1999; Yokoyama & Radlwimmer, 1998) (table 4.2) and a number of these have subsequently been shown by site-directed mutagenisis and *in vitro* expression studies to be involved in spectral tuning (Nathans, 1990; Chan *et al.*, 1992; Yokoyama *et al.*, 1995; Sun *et al.*, 1997; Fasick & Robinson, 1998, Lin *et al.*, 1998; Asenjo *et al.*, 1994; Wilkie *et al.*, 2000). These findings are summarised in table 4.3, and show that the combination of individual mutations causes additive shifts.

In the bovine rod opsin, seven sites have been shown to cause significant spectral shifts of 8 nm or above. These are: 83, 90, 122, 261, 265, 269 and 292 (Nathans, 1990; Chan *et al.*, 1992; Yokoyama *et al.*, 1995; Sun *et al.*, 1997; Fasick & Robinson, 1998; Lin *et al.*, 1998). The mutation Trp 265 Tyr results in the largest single shift in λ_{max} from 500 to 485nm but this is not a natural substitution within the rod opsin group. Instead, this site is one of nine identified by Lin *et al.* (1998), which may be responsible for the difference in λ_{max} between human green and blue opsins. Other spectral tuning sites which have been extensively studied, are those which account for the shift of approximately 30nm between long wave and middle wave sensitive visual pigments in humans and other primates. It has been proposed that

Table 4.2: Candidate amino acid sites for spectral tuning of fish and mammalian opsins. The numbering system for amino acid sites relates to the bovine rod opsin sequence (Nathans *et al.*, 1986).

Candidate site	Shift in λ_{max}	Species/opsin	Reference
Ser 164 Ala	-7nm	Mammalian LWS/MWS	(Yokoyama & Radlwimmer)
His 181 Tyr	-28nm		
Tyr 261 Phe	-7nm		
Thr 269 Ala	-15nm		
Ala 392 Ser	-16nm		
Ala 164 Ser	+5nm	New world monkeys	(Shyue *et al.*, 1998)
Ile 213 Phe	-2nm		
Gly 217 Ser	-1nm		
Phe 261 Tyr	+8nm		
Ala 269 Tyr	+15nm		
Asp 83 Asn/ Ala 118 Thr/ Ala 124 Ser/ Ala 168 Ser/ Ala 292 Ser/ Ser 299 Ala	502 – 460nm	Fish rod shallow sea/ deep sea	(Hope *et al.*, 1997)
Asp 83 Asn/ Tyr 261 Phe/ Ala 292 Ser	516 – 484nm	Cottoid fish rod shallow water/ deep water	(Hunt *et al.*, 1996)
Tyr 261 Phe/ Asp 83 Asn	516 – 495nm		
Tyr 261 Phe/ Ala 292 Ser	505 – 490nm		
Asp 83 Asn/ Ala 124 Ser/ Ala 292 Ser	523 – 501nm	Eel rod freshwater/ deep sea	(Archer *et al.*, 1995)
Ser 164 Ala/ Gly 217 Ser/ Thr 269 Ala	563 – 543nm	Marmoset LWS/MWS	(Williams *et al.*, 1992)
Ile 214 Thr/ Ala 217 Thr/ Tyr 261 Phe/ Thr 269 Ala	565 – 535nm	Old World Monkey LWS/MWS	(Ibbotson *et al.*, 1992)
Ser 164 Ala/ Tyr 261 Phe/ Thr 269 Ala	561 – 530nm	Human deuteranope/protanope	(Neitz *et al.*, 1991)

Candidate site	Species/opsin	Reference
Leu 228 Met/ Ala 232 Thr/	Rabbit LWS/MWS	(Yokoyama & Radlwimmer, 1998)
Ala 282 Ser/ Leu 290 Ile		
Ala 85 Val/ Leu 290 Ile	Deer LWS/MWS	(Yokoyama & Radlwimmer, 1998)
Ile 45 Thr/ Phe 46 Ile/	Guinea pig LWS/MWS	(Yokoyama & Radlwimmer, 1998)
Val 52Ile/ Gln 225 His/		
Ala 282 Ser		
Leu 39 Ile/ Phe 46 Ile/	Squirrel LWS/MWS	(Yokoyama & Radlwimmer, 1998)
Leu 119 Val/ Val 146 Met/		
Ser 195 Tyr/ Val 200 Met		

Table 4.3: Absorption maxima and shifts in λ_{max} of vertebrate opsin mutants. The numbering system for mutation sites relates to the bovine rod opsin sequence (Nathans *et al.*, 1986). A positive shift indicates a shift towards the red end of the visual spectrum, a negative towards the blue.

Mutation	λ_{max} of mutant	Shift from wt	wt	Reference
Ala 86 Ser	362nm	-1nm	Budgerigar UV	(Wilkie *et al.*, 2000)
Cys 90 Ser	398nm	+35nm		
Thr 93 Val	366nm	+3nm		
Ala 118 thr	366nm	+3nm		
Ser 292 Ala	552nm	+28nm	Dolphin LWS	(Fasick & Robinson, 1998)
Asp 83 Asn	495nm	-4nm	Bovine rod	(Fasick & Robinson, 1998)
Ala 292 Ser	489nm	-10nm		
Ala 299 Ser	501nm	+2nm		
Gly 90 Ser	489nm	-11nm	Bovine rod	(Lin *et al.*, 1998)
Ala 117 Gly	496nm	-4nm		
Glu 122 Leu	495nm	-5nm		
Ala 124 Thr	497nm	-3nm		
Trp 265 Tyr	485nm	-15nm		
Ala 292 Ser	491nm	-9nm		
Ala 295 Ser	495nm	-5nm		
Ala 299 Cys	498nm	-2nm		
Ala 292 Ser	488nm	-10nm	Bovine rod	(Sun *et al.*, 1997)
Ser 292 Ala	526nm	+18nm	Mouse LWS	(Sun *et al.*, 1997)
Tyr 261 Phe/ Thr 269 Ala	487nm	-21nm		
Tyr 261 Phe/ Thr 269 Ala/ Ser 292 Ala	506nm	-2nm		
His 181 Tyr	526nm	-28nm	Human LWS	(Sun *et al.*, 1997)
His 181 Tyr/ Ala 292 Ser	508nm	-46nm		
Tyr 261 Phe	496nm	-8nm	Bovine rod	(Yokoyama *et al.*, 1995)
Ser 100 Tyr/ Ser 164 Ala/ Ile 214 Thr/ Ala 217 Ser/ Tyr 261 Phe/ Thr 269 Ala/ Tyr 293 Phe	532nm	-31nm	Human LWS	(Asenjo *et al.*, 1994)
Ala 164 Ser	502nm	+2nm	Bovine rod	(Chan *et al.*, 1992)
Phe 261 Tyr	510nm	+10nm		
Ala 269 Thr	514nm	+14nm		
Phe 261 Tyr/ Ala 269 Thr	520nm	+20nm		
Ala 164 Ser/ Phe 261 Tyr	512nm	+12nm		
Ala 164 Ser/ Ala 269 Thr	514nm	+14nm		
Asp 83 Gly	499.5nm	+1.5nm	Bovine rod	(Nathans, 1990)
Asp 83 Asn	489.5	-8.5nm		
Met 86 Glu	492.5nm	-5.5nm		
Glu 122 Ile	496nm	-2nm		
Glu 122 Gln	481nm	-17nm		
His 211 Phe	493.5nm	-4.5nm		
His 211 Cys	493nm	-5nm		

three are responsible for the principal differences: 164, 261 and 269 (Neitz *et al.*, 1991; Asenjo *et al.*, 1994). These sites have Ala or Thr, Phe or Tyr and Ala or Ser/Thr in the shorter or longer wavelength pigments respectively. More recent studies have shown that the same substitutions account for the major shift in λ_{max} between the long and middle wave sensitive pigments of goldfish (Johnson *et al.*, 1993) and avian species including the chicken, canary and ostrich (Okano *et al.*, 1992; Das *et al.*, 1999; Wright, 2000). Additionally, in humans the introduction of Tyr, Thr, Ser, and Phe at sites 100, 214, 217, and 293 in the human long wave pigment account for further minor short-wave adjustments (Asenjo *et al.*, 1994). Many other candidate sites have been identified in mammals which appear to be species specific as shown in table 4.2. However, whether they are all involved in spectral tuning remains to be determined (Yokoyama & Radlwimmer, 1998). In a comparison of human and dolphin long wave sensitive pigments, the replacement of Ser with Ala at site 292 causes a large red shift of 28nm (Fasick & Robinson, 1998). Interestingly, the mouse long wave sensitive pigment has both Tyr at position 261 and Thr at 269, which would suggest that the λ_{max} should be similar to that of the corresponding human pigment (its closest homologue). However, it also has Ser at 292 which results in a blue shift and hence its λ_{max} of 508nm is closer to that of the rod opsins (Sun *et al.*, 1997).

In the cottoid fishes of Lake Baikal there is a shift in the λ_{max} of the rod visual pigments from 516nm to 484nm with increasing depth of habitat (Bowmaker *et al.*, 1994). The rod opsin gene from a number of these species has been sequenced and comparisons made to identify amino acid substitutions which may account for the shifts (Hunt *et al.*, 1996). As with the Lake Baikal fishes, there are also short wave shifts in the opsins of marine fishes with increasing depth of habitat and between the fresh water and deep sea eel (Hope *et al.*, 1997; Archer *et al.*, 1995). In the rod opsins of fish, site 292 is again thought to be important in short wave shifting by the replacement of Ala in shallow or fresh water species with the polar residue Ser in deeper dwelling species. The replacement of Asp with Asn at site 83 is also common to eels, freshwater cottoids and marine fishes, leading to a blue shift in the deeper living species. In addition, sites 124 in the eel, 261 in the cottoid fishes and 118, 124,

168 and 299 in shallow and deep sea fishes have been proposed as spectral tuning sites (Archer *et al.*, 1995; Hunt *et al.*, 1996; Hope *et al.*, 1997). Site directed mutagenesis studies in the bovine rod opsin have confirmed that the introduction of polar amino acids at positions 124 and 299 lead to a small short wave shift (Lin *et al.*, 1998; Fasick & Robinson, 1998); at sites 118 and 261, replacement with non-polar residues have a similar affect (Wilkie *et al.*, 1998; Chan *et al.*, 1992; Yokoyama *et al.*, 1995; Asenjo *et al.*, 1994; Sun *et al.*, 1997). With the exception of *Astyanax fasciatus* (Yokoyama *et al.*, 1995), two shallow living species of Cottoid fish, *Cottus kesslerei* and *Paracottus kneri* (Hunt *et al.*, 1996) and the deep-sea dragon fish *Malacosteus niger* (Douglas *et al.*, 1999), all other rod opsins known have Phe at site 261. The rods of these fishes are unusual in that they are red shifted and resemble long wave sensitive opsins by having Tyr at 261. To date, position 168 remains a candidate spectral tuning site until its role can be determined by mutagenesis and expression studies.

4.2 Results

4.2.1 Amplification and sequencing of the blue cone pigment of *Cottus gobio*

Unfortunately, only genomic DNA was available with which to study the cottoid fish of Lake Baikal. However, there was a species of cottoid native to Britain, *Cottus gobio,* which was readily available and for which the λ_{max} of 470nm had been determined by Professor Jim Bowmaker as shown in figure 4.3 (unpublished). From *Cottus gobio* it was possible to isolate mRNA. Retinal cDNA was then produced using the 3' RACE system together with the 3' anchor primer and a gene specific primer, GFB1+. When run on an agarose gel, a smear was observed containing a fuzzy band. The band was excised, DNA eluted and further PCR reactions performed using nested primers. The primer pair GFB2+ and GFB4- was eventually successful in giving three discreet bands, one of which was the expected size of approximately 350bp (figure 4.4). This band was excised, DNA eluted and cloned into the pGEM-T easy vector ready for sequencing. Sequence obtained for the cottoid was compared to goldfish opsin sequences and showed highest homology, 74%, to the blue gene suggesting that the amplified fragment was part of the blue opsin gene (figure 4.5).

The cottoid blue opsin sequence was then extended using 3' RACE together with new gene specific primers, BHB2+ and BHB3+. Multiple bands were observed when the samples were run on an agarose gel (figure 4.6); these were excised and the DNA eluted for cloning. Positive clones were sequenced using the pTAg 5' and 3' primers and an internal sequencing primer, BHB internal+. The data obtained from 3' RACE extended the sequence to the poly A tail at the 3' end of the cDNA transcript.

Obtaining the 5' sequence of the cottoid blue cDNA opsin proved to be much more problematic and initial attempts at 5' RACE were abandoned. Using the existing *Cottus gobio* sequence, primers were designed to amplify similar regions in the Lake Baikal cottoids as discussed later. With this information it was then possible to use the walking PCR technique (Dominguez & Lopez-Larrea, 1994; Bellingham *et al.*, 1998) on genomic DNA of *Cottus kesslerei* to amplify the 5' sequence. Cottoid specific primers BHB1+ and BHB2-, were designed from this sequence and used to amplify the 5' sequence of *Cottus gobio*. The full blue cDNA sequence of *Cottus*

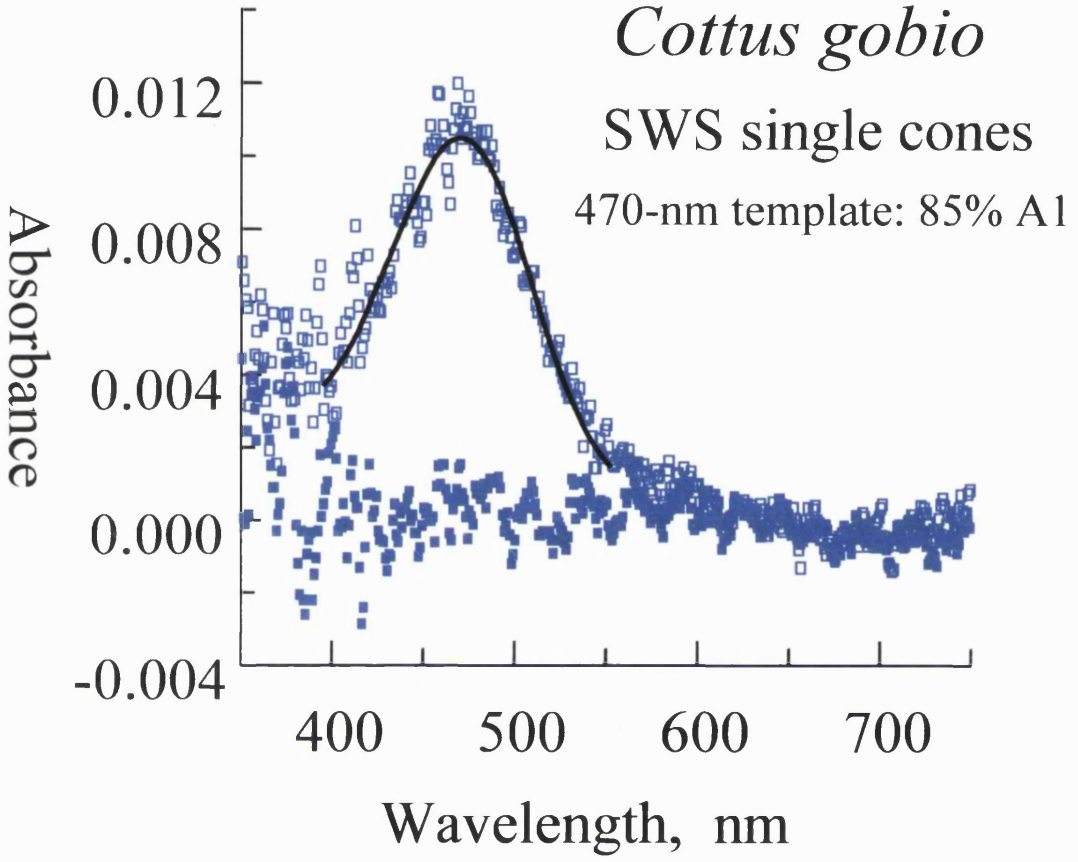


Figure 4.3: Microspectrophotometric absorbance spectra of the short wave sensitive cones in *Cottus gobio*. Results courtesy of J.K. Bowmaker. The spectra are the means of records obtained from 5 individual cone outer segments. Open squares represent the dark spectrum and closed squares, the bleached spectrum.

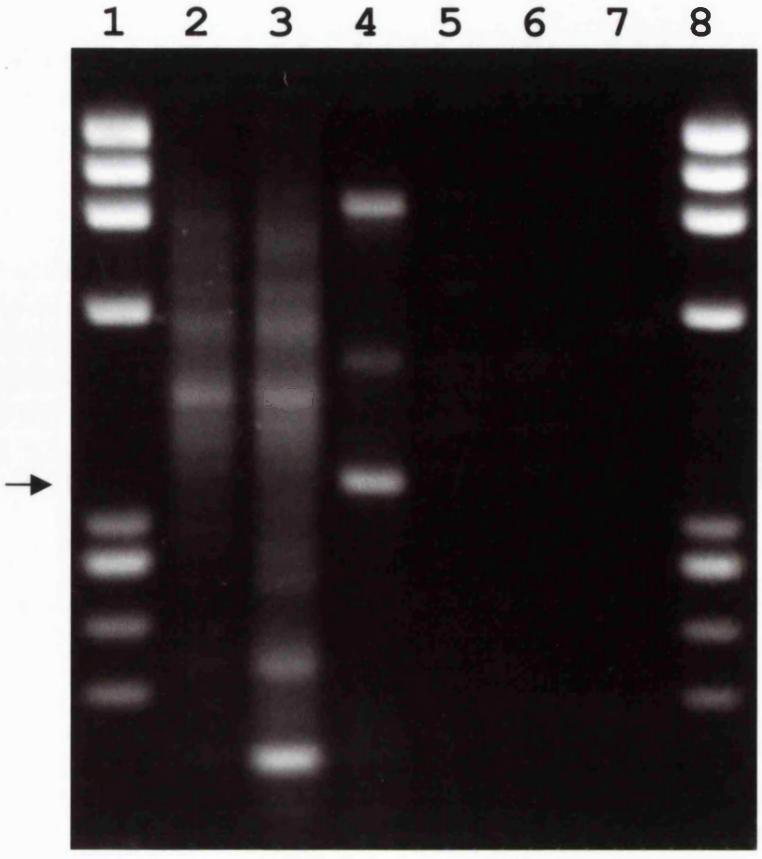


Figure 4.4: Nested PCRs of *Cottus gobio* blue cone opsin 3'RACE. Lanes 1 & 8: ϕX174/Hae III DNA ladder. Lane 2 :GFB1+/GFB4- (55°C). Lane 3:GFB1+/GFB4- (50°C). Lane 4: GFB2+/GFB4-. Lanes 5, 6 & 7: No DNA Controls. The black arrow indicates the 350bp band which was excised from lane 4.

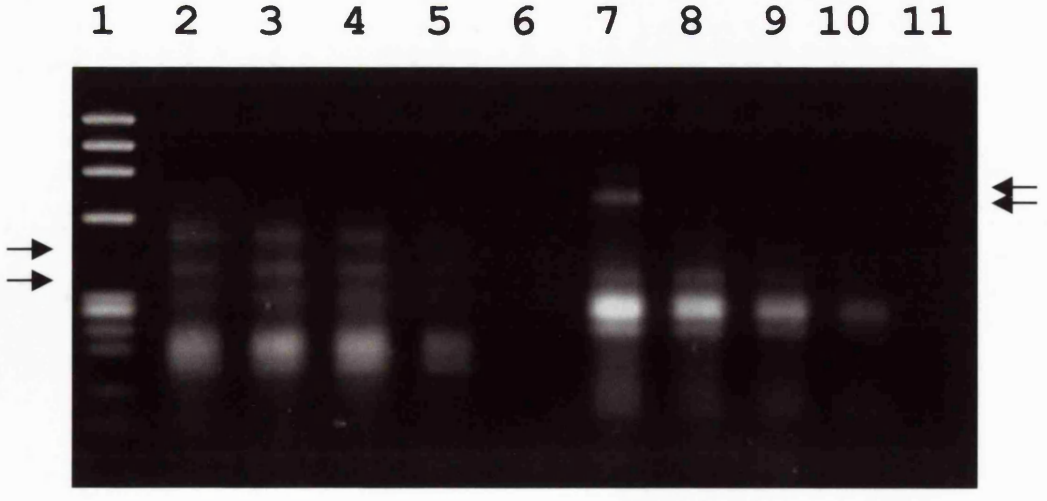


Figure 4.6: PCR amplification of the *Cottus gobio* blue cone opsin gene first strand cDNA. Lane 1: ϕX174/Hae III DNA ladder. Lanes 2,3,4 & 5: BHB2+/PCR anchor primer. Lanes 7,8,9 & 10 BHB3+/PCR anchor primer. Lanes 6 & 11: No DNA controls. Black arrows indicate bands which were excised from lanes 4 and 7.

```
Cottus gobio  ACTTGGTACC TTTGTTTTCA AGCCCGACCA TGCTATGGCC TGCTGCATGG
Goldfish      C*****G*A* ***ACC**** **A**CCT** ***C**A**T G*******AC

Cottus gobio  TGACCTGGGT GTTGGCACTG CTTGCCTCAG GTCCTCCTCT GTTTGGATGG
Goldfish      *TC*T***A* AAGT***T** GCA**T***C TC*****A** ******C***

Cottus gobio  AGCAGGTACA TCCCAGAAGG CCTGCAGTGC TCCTGTGGTC CAGACTGGTA
Goldfish      ***CG***** *A**T***** TT******** ********A* *T********

Cottus gobio  TACCACCAAC AACAAATACA ACAATGAATC CTACGTGATA TATCTCTTCA
Goldfish      ******T*** ********** ********** ******C**G *T*T*G***T

Cottus gobio  CCTGCCACTT TACTGTTCCC CTCTTAATCC TTGTCTTTTG CTANGCCCAA
Goldfish      G**T*TG*** *G*******T T**GGC*C*A *C**G**C** T****G****

Cottus gobio  CTCCTTTTTA CACTGAAAAT GGCGGCAAAG GCCCAAGCAG AATCTGCCTC
Goldfish      **A**CA*C* ****C***T* A**A*****A **T******* *T**A**T**

Cottus gobio  CACCCAGAAA
Goldfish      G*********
```

Figure 4.5: Alignment of the *Cottus gobio* GFB2+/GFB4- fragment with goldfish blue opsin sequence. Identical bases are marked with an asterisk.

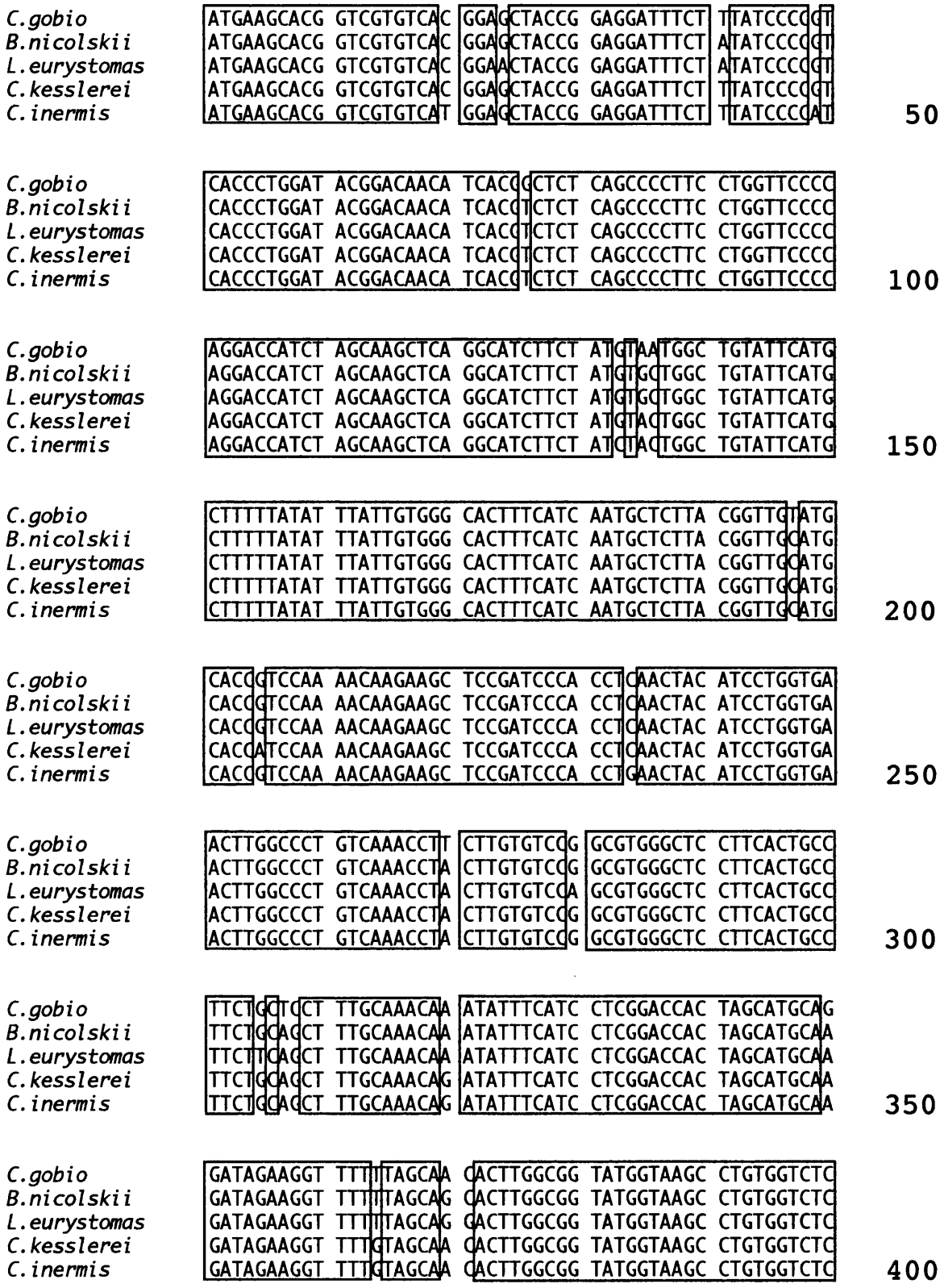
gobio is shown in figure 4.7 aligned with the blue opsin gene sequences of the Lake Baikal cottoids.

4.2.2 Amplification and sequencing of the blue cone pigments of the Lake Baikal Cottoids

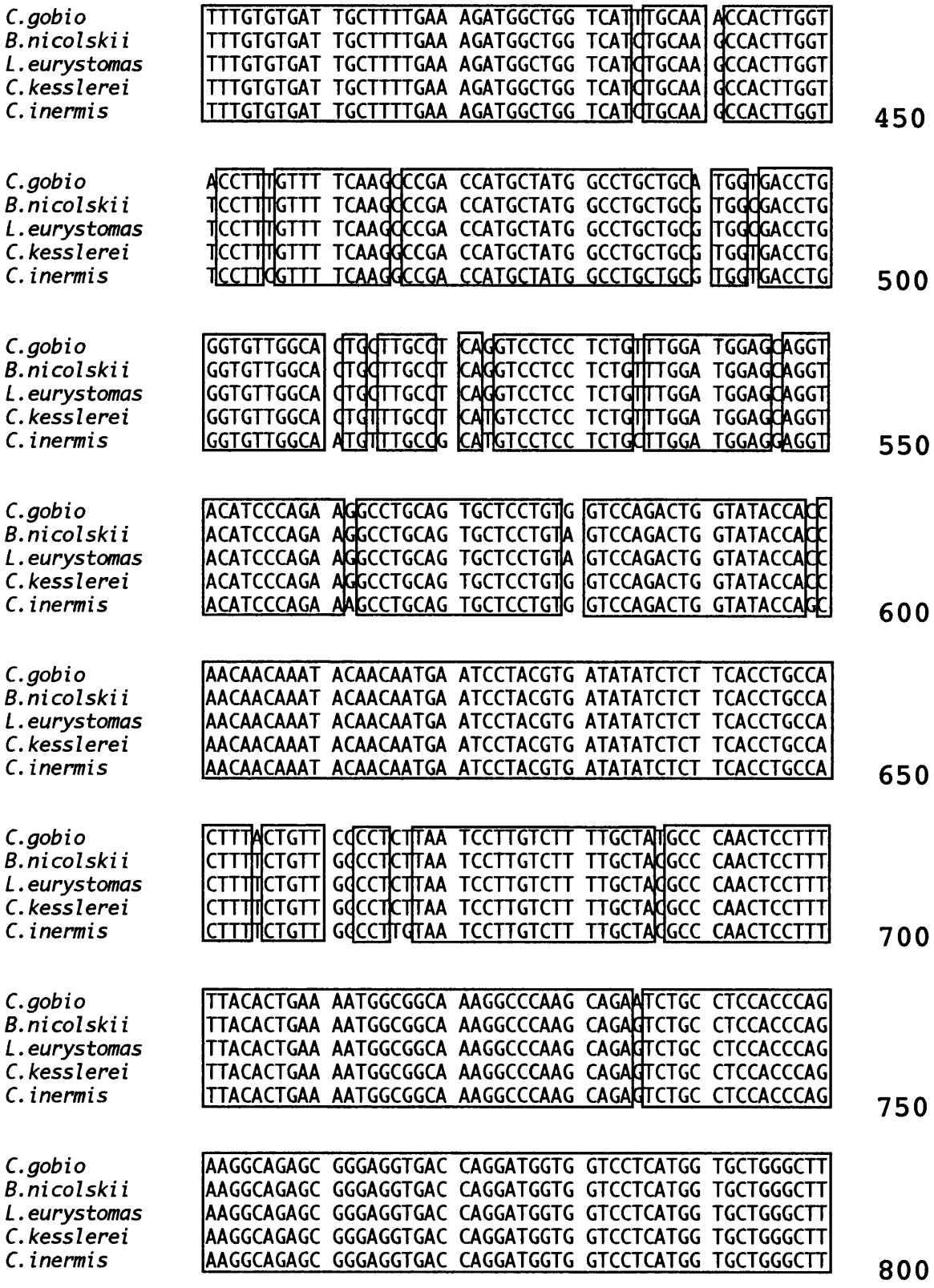
The four species of cottoid from Lake Baikal studied were; *Cottus kesslerei* and *Cottocomephorus inermis* which both have blue cones with a λ_{max} of approximately 450nm, and *Limnocottus eurystomias* and *Batrachocottus nicolskii* with a λ_{max} of approximately 430nm. Using the primers designed from *Cottus gobio*, BHB2+ and BHB4-, a 550bp fragment was amplified from *Cottus kesslerei* (figure 4.8). The exonic sequence between BHB2+ and BHB4- was 370bp, and the combined sizes of introns 2 and 3 was approximately 180bp. The fragment was eluted from an agarose gel and cloned into the pGEM-T easy vector and sequenced. In order to obtain the 5' sequence, the walking PCR method was used (Dominguez & Lopez-Larrea, 1994; Bellingham *et al.*, 1998). Primers UNI 28 and 1st walk CK out, were used in the primary round and gave a smear when viewed on an agarose gel (figure 4.9). The nested primers, UNI 17 and 1st walk CK in, were used for the secondary round of amplification and produced a discreet band of approximately 900bp in size (figure 4.10). This PCR product was cloned into pGEM-T easy and sequenced using pTAg 5' and 3' primers. It was not possible to sequence the whole 900bp using these primers so internal sequencing primers, CK internal+ and CK internal-, were used to fill in the gap. Exon 1 of the blue cone pigment gene was located within the 900bp product by alignment of the sequence with the goldfish blue opsin exon 1. The goldfish and *Cottus kesslerei* exon 1 blue cone opsin sequences showed 54% homology. From this sequence, primers were designed to directly amplify the 5' regions of *Cottus gobio*. The low sequence identity in exon 1 would account for the initial difficulty in amplifying this region from *Cottus gobio* using the goldfish specific primers.

Primers designed from the *Cottus gobio* cDNA sequence were used to amplify the blue cone opsin genes of *Cottocomephorus inermis*, *Limnocottus eurystomias* and

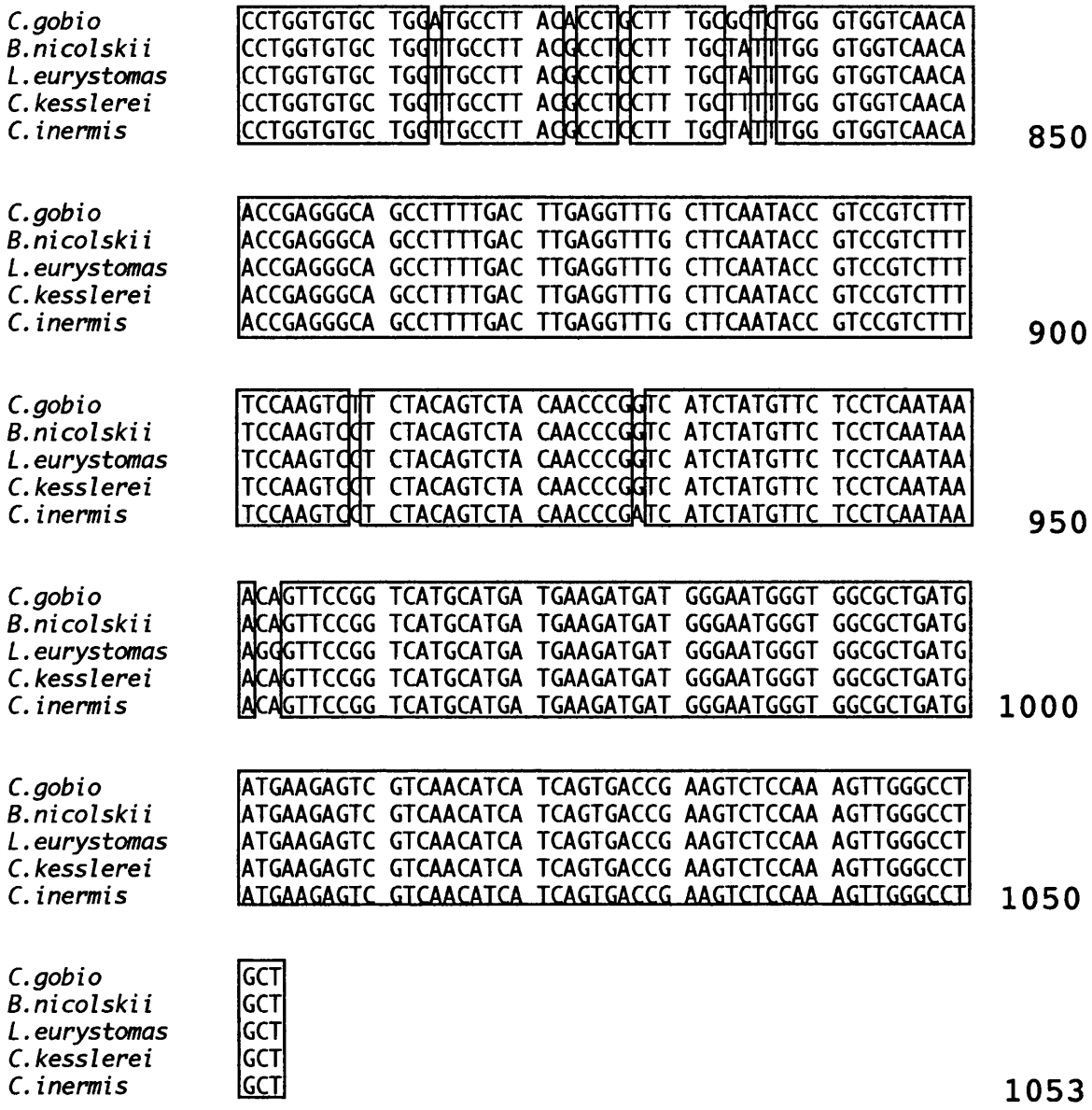
Figure 4.7: Aligned sequences of the blue cone opsin genes of cottoid fishes. Identical sequences are boxed. The numbers to the right represent the position relative to the translation start site.


C.gobio ATGAAGCACG GTCGTGTCAC GGAGCTACCG GAGGATTTCT TTATCCCCGT
B.nicolskii ATGAAGCACG GTCGTGTCAC GGAGCTACCG GAGGATTTCT ATATCCCCGT
L.eurystomas ATGAAGCACG GTCGTGTCAC GGAACTACCG GAGGATTTCT ATATCCCCGT
C.kesslerei ATGAAGCACG GTCGTGTCAC GGAGCTACCG GAGGATTTCT TTATCCCCGT
C.inermis ATGAAGCACG GTCGTGTCAT GGAGCTACCG GAGGATTTCT TTATCCCCAT 50
C.gobio CACCCTGGAT ACGGACAACA TCACCCTCT CAGCCCCTTC CTGGTTCCCC
B.nicolskii CACCCTGGAT ACGGACAACA TCACCTCTCT CAGCCCCTTC CTGGTTCCCC
L.eurystomas CACCCTGGAT ACGGACAACA TCACCTCTCT CAGCCCCTTC CTGGTTCCCC
C.kesslerei CACCCTGGAT ACGGACAACA TCACCTCTCT CAGCCCCTTC CTGGTTCCCC
C.inermis CACCCTGGAT ACGGACAACA TCACCTCTCT CAGCCCCTTC CTGGTTCCCC 100
C.gobio AGGACCATCT AGCAAGCTCA GGCATCTTCT ATGTAATGGC TGTATTCATG
B.nicolskii AGGACCATCT AGCAAGCTCA GGCATCTTCT ATGTGCTGGC TGTATTCATG
L.eurystomas AGGACCATCT AGCAAGCTCA GGCATCTTCT ATGTGCTGGC TGTATTCATG
C.kesslerei AGGACCATCT AGCAAGCTCA GGCATCTTCT ATGTACTGGC TGTATTCATG
C.inermis AGGACCATCT AGCAAGCTCA GGCATCTTCT ATCTACTGGC TGTATTCATG 150
C.gobio CTTTTTATAT TTATTGTGGG CACTTTCATC AATGCTCTTA CGGTTGTATG
B.nicolskii CTTTTTATAT TTATTGTGGG CACTTTCATC AATGCTCTTA CGGTTGCATG
L.eurystomas CTTTTTATAT TTATTGTGGG CACTTTCATC AATGCTCTTA CGGTTGCATG
C.kesslerei CTTTTTATAT TTATTGTGGG CACTTTCATC AATGCTCTTA CGGTTGCATG
C.inermis CTTTTTATAT TTATTGTGGG CACTTTCATC AATGCTCTTA CGGTTGCATG 200
C.gobio CACCGTCCAA AACAAGAAGC TCCGATCCCA CCTCAACTAC ATCCTGGTGA
B.nicolskii CACCGTCCAA AACAAGAAGC TCCGATCCCA CCTCAACTAC ATCCTGGTGA
L.eurystomas CACCGTCCAA AACAAGAAGC TCCGATCCCA CCTCAACTAC ATCCTGGTGA
C.kesslerei CACCATCCAA AACAAGAAGC TCCGATCCCA CCTCAACTAC ATCCTGGTGA
C.inermis CACCGTCCAA AACAAGAAGC TCCGATCCCA CCTGAACTAC ATCCTGGTGA 250
C.gobio ACTTGGCCCT GTCAAACCTT CTTGTGTCCG GCGTGGGCTC CTTCACTGCC
B.nicolskii ACTTGGCCCT GTCAAACCTA CTTGTGTCCG GCGTGGGCTC CTTCACTGCC
L.eurystomas ACTTGGCCCT GTCAAACCTA CTTGTGTCCA GCGTGGGCTC CTTCACTGCC
C.kesslerei ACTTGGCCCT GTCAAACCTA CTTGTGTCCG GCGTGGGCTC CTTCACTGCC
C.inermis ACTTGGCCCT GTCAAACCTA CTTGTGTCCG GCGTGGGCTC CTTCACTGCC 300
C.gobio TTCTGCTCCT TTGCAAACAA ATATTTCATC CTCGGACCAC TAGCATGCAG
B.nicolskii TTCTGCAGCT TTGCAAACAA ATATTTCATC CTCGGACCAC TAGCATGCAA
L.eurystomas TTCTTCAGCT TTGCAAACAA ATATTTCATC CTCGGACCAC TAGCATGCAA
C.kesslerei TTCTGCAGCT TTGCAAACAG ATATTTCATC CTCGGACCAC TAGCATGCAA
C.inermis TTCTGCAGCT TTGCAAACAG ATATTTCATC CTCGGACCAC TAGCATGCAA 350
C.gobio GATAGAAGGT TTTTTTAGCAA CACTTGGCGG TATGGTAAGC CTGTGGTCTC
B.nicolskii GATAGAAGGT TTTTTTAGCAG CACTTGGCGG TATGGTAAGC CTGTGGTCTC
L.eurystomas GATAGAAGGT TTTTTTAGCAG CACTTGGCGG TATGGTAAGC CTGTGGTCTC
C.kesslerei GATAGAAGGT TTTGTAGCAA CACTTGGCGG TATGGTAAGC CTGTGGTCTC
C.inermis GATAGAAGGT TTTGTAGCAA CACTTGGCGG TATGGTAAGC CTGTGGTCTC 400

Continued ....


C.gobio TTTGTGTGAT TGCTTTTGAA AGATGGCTGG TCATTTGCAA ACCACTTGGT
B.nicolskii TTTGTGTGAT TGCTTTTGAA AGATGGCTGG TCATCTGCAA GCCACTTGGT
L.eurystomas TTTGTGTGAT TGCTTTTGAA AGATGGCTGG TCATCTGCAA GCCACTTGGT
C.kesslerei TTTGTGTGAT TGCTTTTGAA AGATGGCTGG TCATCTGCAA GCCACTTGGT
C.inermis TTTGTGTGAT TGCTTTTGAA AGATGGCTGG TCATCTGCAA GCCACTTGGT
450
C.gobio ACCTTTGTTT TCAAGCCCGA CCATGCTATG GCCTGCTGCA TGGTGACCTG
B.nicolskii TCCTTTGTTT TCAAGCCCGA CCATGCTATG GCCTGCTGCG TGGCGACCTG
L.eurystomas TCCTTTGTTT TCAAGCCCGA CCATGCTATG GCCTGCTGCG TGGCGACCTG
C.kesslerei TCCTTTGTTT TCAAGGCCGA CCATGCTATG GCCTGCTGCG TGGTGACCTG
C.inermis TCCTTCGTTT TCAAGGCCGA CCATGCTATG GCCTGCTGCG TGGTGACCTG
500
C.gobio GGTGTTGGCA CTGCTTGCCT CAGGTCCTCC TCTGTTTGGA TGGAGCAGGT
B.nicolskii GGTGTTGGCA CTGCTTGCCT CAGGTCCTCC TCTGTTTGGA TGGAGCAGGT
L.eurystomas GGTGTTGGCA CTGCTTGCCT CAGGTCCTCC TCTGTTTGGA TGGAGCAGGT
C.kesslerei GGTGTTGGCA CTGTTTGCCT CATGTCCTCC TCTGTTTGGA TGGAGCAGGT
C.inermis GGTGTTGGCA ATGTTTGCCG CATGTCCTCC TCTGCTTGGA TGGAGGAGGT
550
C.gobio ACATCCCAGA AGGCCTGCAG TGCTCCTGTG GTCCAGACTG GTATACCACC
B.nicolskii ACATCCCAGA AGGCCTGCAG TGCTCCTGTA GTCCAGACTG GTATACCACC
L.eurystomas ACATCCCAGA AGGCCTGCAG TGCTCCTGTA GTCCAGACTG GTATACCACC
C.kesslerei ACATCCCAGA AGGCCTGCAG TGCTCCTGTG GTCCAGACTG GTATACCACC
C.inermis ACATCCCAGA AAGCCTGCAG TGCTCCTGTG GTCCAGACTG GTATACCAGC
600
C.gobio AACAACAAAT ACAACAATGA ATCCTACGTG ATATATCTCT TCACCTGCCA
B.nicolskii AACAACAAAT ACAACAATGA ATCCTACGTG ATATATCTCT TCACCTGCCA
L.eurystomas AACAACAAAT ACAACAATGA ATCCTACGTG ATATATCTCT TCACCTGCCA
C.kesslerei AACAACAAAT ACAACAATGA ATCCTACGTG ATATATCTCT TCACCTGCCA
C.inermis AACAACAAAT ACAACAATGA ATCCTACGTG ATATATCTCT TCACCTGCCA
650
C.gobio CTTTACTGTT CCCCTCTTAA TCCTTGTCTT TTGCTATGCC CAACTCCTTT
B.nicolskii CTTTTCTGTT GCCCTCTTAA TCCTTGTCTT TTGCTACGCC CAACTCCTTT
L.eurystomas CTTTTCTGTT GCCCTCTTAA TCCTTGTCTT TTGCTACGCC CAACTCCTTT
C.kesslerei CTTTTCTGTT GCCCTCTTAA TCCTTGTCTT TTGCTACGCC CAACTCCTTT
C.inermis CTTTTCTGTT GCCCTTGTAA TCCTTGTCTT TTGCTACGCC CAACTCCTTT
700
C.gobio TTACACTGAA AATGGCGGCA AAGGCCCAAG CAGAATCTGC CTCCACCCAG
B.nicolskii TTACACTGAA AATGGCGGCA AAGGCCCAAG CAGAGTCTGC CTCCACCCAG
L.eurystomas TTACACTGAA AATGGCGGCA AAGGCCCAAG CAGAGTCTGC CTCCACCCAG
C.kesslerei TTACACTGAA AATGGCGGCA AAGGCCCAAG CAGAGTCTGC CTCCACCCAG
C.inermis TTACACTGAA AATGGCGGCA AAGGCCCAAG CAGAGTCTGC CTCCACCCAG
750
C.gobio AAGGCAGAGC GGGAGGTGAC CAGGATGGTG GTCCTCATGG TGCTGGGCTT
B.nicolskii AAGGCAGAGC GGGAGGTGAC CAGGATGGTG GTCCTCATGG TGCTGGGCTT
L.eurystomas AAGGCAGAGC GGGAGGTGAC CAGGATGGTG GTCCTCATGG TGCTGGGCTT
C.kesslerei AAGGCAGAGC GGGAGGTGAC CAGGATGGTG GTCCTCATGG TGCTGGGCTT
C.inermis AAGGCAGAGC GGGAGGTGAC CAGGATGGTG GTCCTCATGG TGCTGGGCTT
800

Continued....


C.gobio CCTGGTGTGC TGGATGCCTT ACACCTGCTT TGCGCTCTGG GTGGTCAACA
B.nicolskii CCTGGTGTGC TGGTTGCCTT ACGCCTCCTT TGCTATTTGG GTGGTCAACA
L.eurystomas CCTGGTGTGC TGGTTGCCTT ACGCCTCCTT TGCTATTTGG GTGGTCAACA
C.kesslerei CCTGGTGTGC TGGTTGCCTT ACGCCTCCTT TGCTTTTTGG GTGGTCAACA
C.inermis CCTGGTGTGC TGGTTGCCTT ACGCCTCCTT TGCTATTTGG GTGGTCAACA
850
C.gobio ACCGAGGGCA GCCTTTTGAC TTGAGGTTTG CTTCAATACC GTCCGTCTTT
B.nicolskii ACCGAGGGCA GCCTTTTGAC TTGAGGTTTG CTTCAATACC GTCCGTCTTT
L.eurystomas ACCGAGGGCA GCCTTTTGAC TTGAGGTTTG CTTCAATACC GTCCGTCTTT
C.kesslerei ACCGAGGGCA GCCTTTTGAC TTGAGGTTTG CTTCAATACC GTCCGTCTTT
C.inermis ACCGAGGGCA GCCTTTTGAC TTGAGGTTTG CTTCAATACC GTCCGTCTTT
900
C.gobio TCCAAGTCTT CTACAGTCTA CAACCCGGTC ATCTATGTTC TCCTCAATAA
B.nicolskii TCCAAGTCCT CTACAGTCTA CAACCCGGTC ATCTATGTTC TCCTCAATAA
L.eurystomas TCCAAGTCCT CTACAGTCTA CAACCCGGTC ATCTATGTTC TCCTCAATAA
C.kesslerei TCCAAGTCCT CTACAGTCTA CAACCCGGTC ATCTATGTTC TCCTCAATAA
C.inermis TCCAAGTCCT CTACAGTCTA CAACCCGATC ATCTATGTTC TCCTCAATAA
950
C.gobio ACAGTTCCGG TCATGCATGA TGAAGATGAT GGGAATGGGT GGCGCTGATG
B.nicolskii ACAGTTCCGG TCATGCATGA TGAAGATGAT GGGAATGGGT GGCGCTGATG
L.eurystomas AGGGTTCCGG TCATGCATGA TGAAGATGAT GGGAATGGGT GGCGCTGATG
C.kesslerei ACAGTTCCGG TCATGCATGA TGAAGATGAT GGGAATGGGT GGCGCTGATG
C.inermis ACAGTTCCGG TCATGCATGA TGAAGATGAT GGGAATGGGT GGCGCTGATG
1000
C.gobio ATGAAGAGTC GTCAACATCA TCAGTGACCG AAGTCTCCAA AGTTGGGCCT
B.nicolskii ATGAAGAGTC GTCAACATCA TCAGTGACCG AAGTCTCCAA AGTTGGGCCT
L.eurystomas ATGAAGAGTC GTCAACATCA TCAGTGACCG AAGTCTCCAA AGTTGGGCCT
C.kesslerei ATGAAGAGTC GTCAACATCA TCAGTGACCG AAGTCTCCAA AGTTGGGCCT
C.inermis ATGAAGAGTC GTCAACATCA TCAGTGACCG AAGTCTCCAA AGTTGGGCCT
1050
C.gobio GCT
B.nicolskii GCT
L.eurystomas GCT
C.kesslerei GCT
C.inermis GCT
1053

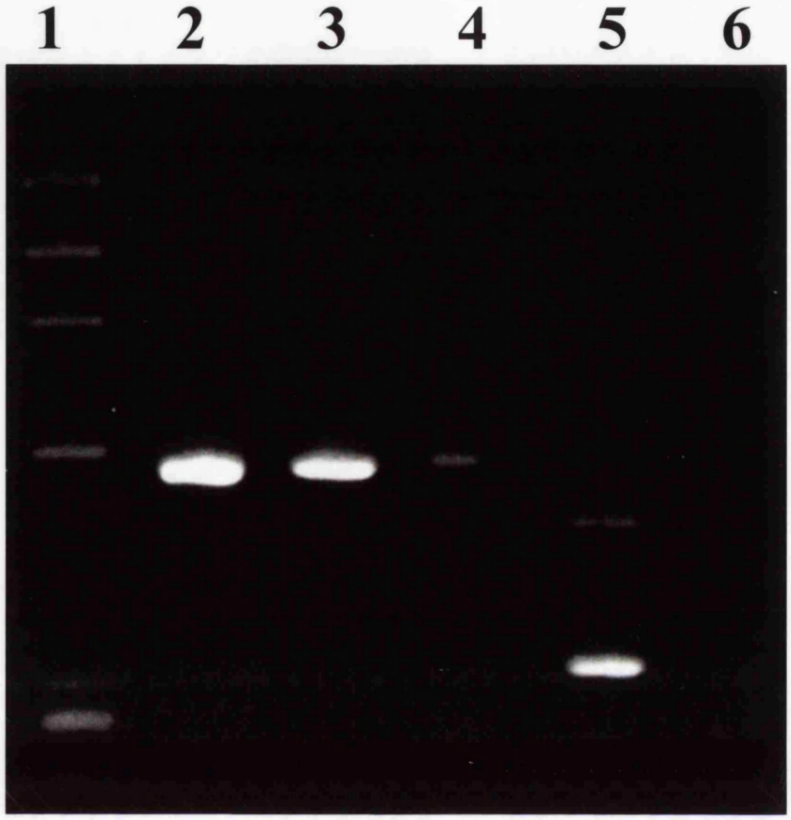


Figure 4.8: PCR of cottoid gDNA using blue opsin primers BHB2+ and BHB4-. Lane 1: ϕX174/Hae III DNA ladder. Lane 2: *C. kesslerei*. Lane 3: *L. pallidus*. Lane 4: *C. boulengerei*. Lane 5: *B. nicolskii*. Lane 6: No DNA control

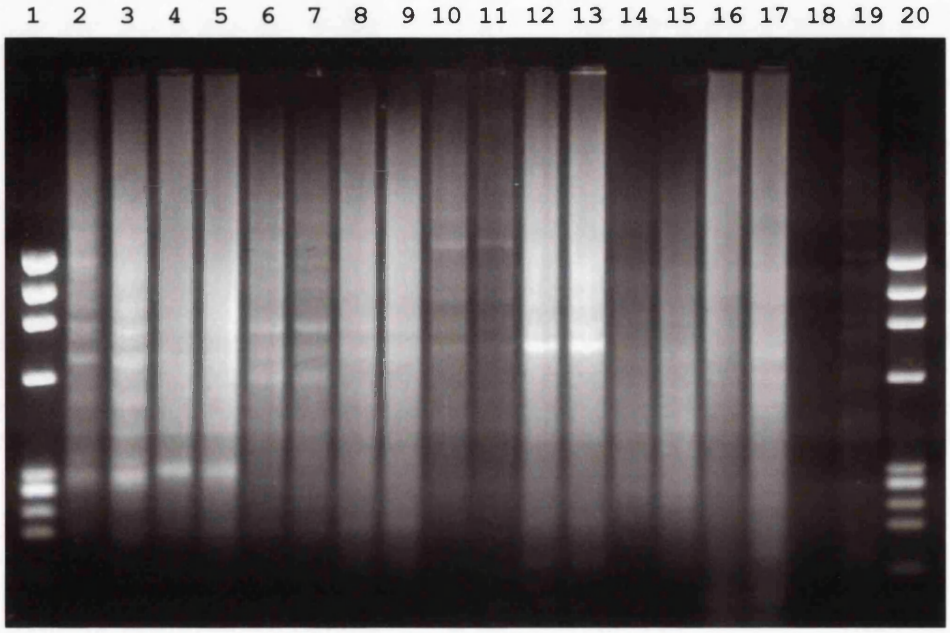


Figure 4.9: Primary walking PCR of *Cottus kesslerei* blue cone opsin. Lanes 1 & 20: ϕX174/Hae III DNA ladder. Lanes 6 & 7 *Cottus kesslerei*, primer Uni28, 1.5mM Mg Cl_2. Lanes 8 & 9: *Cottus kesslerei*, primer UNI 28, 3.0mM $MgCl_2$. Lanes 19 & 20: No DNA Controls. Lanes 2,3,4,5,10,11,12,13,14,15,16 & 17 are walking PCRs of other fish opsins not included in this study.

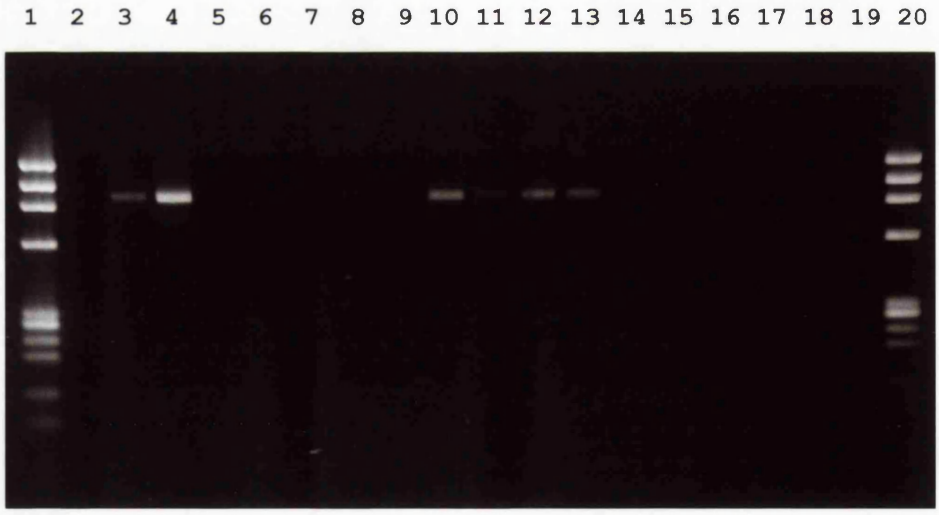


Figure 4.10: Secondary walking PCRs of *Cottus kesslerei* blue cone opsin. Lanes 1 & 20: ϕX174/Hae III DNA ladder. Lanes 10,11,12 & 13: Secondary walking PCRs of samples in lanes 6,7,8 & 9 above respectively. Lanes 18 & 19: No DNA controls. Lanes 2,3,4,5,6 7,8,9,14,15,16 & 17 are secondary walking PCRs of other fish opsins not included in this study.

Batrachocottus nicolskii and the remainder of *Cottus kesslerei*. The primer pairs used and the most successful reaction conditions for each exon are shown in table 4.4 below; images of the agarose gels showing the PCR products are shown in figure 4.11.

EXON	PRIMER PAIR	REACTION BUFFER	ANNEALING TEMPERATURE
1	BHB1+/BHB2-	NH_4	58°C
2	BHB1a+/BHB3-	KCl	60°C
3	BHB2+/BHB4-	NH_4	58°C
4	BHB3+/BHB5-	NH_4	58°C
5	BHB4+/BHB5a-	NH_4	58°C

Table 4.4: Primers and reaction conditions used for the amplification of exons 1 - 5 of the blue cone opsin gene of the Lake Baikal cottoids.

The full blue cone opsin sequences for *Cottus gobio* and the Lake Baikal cottoid fishes have been assembled using MacVector™ (figure 4.7) and amino acid alignments performed using Geneworks™ (figure 4.12). Homology within the cottoid blue opsin genes ranges from 91.7%, between *C. gobio* and *C. inermis*, to 98.6%, between *L. eurystomias* and *B. nicolskii,* as shown in table 4.5. Pairwise comparisons were also made with the blue cone opsin sequence from other species of fish: the goldfish, *Carassius auratus*; killifish, *Oryzias latipes*; zebrafish, *Danio rerio* and the blind cave fish, *Astyanax fasciatus*. The blue cone pigments of birds also belong to the same group as fish blue opsins and so additional comparisons were made with the blue pigment of the chicken, *Gallus gallus*. The comparisons are

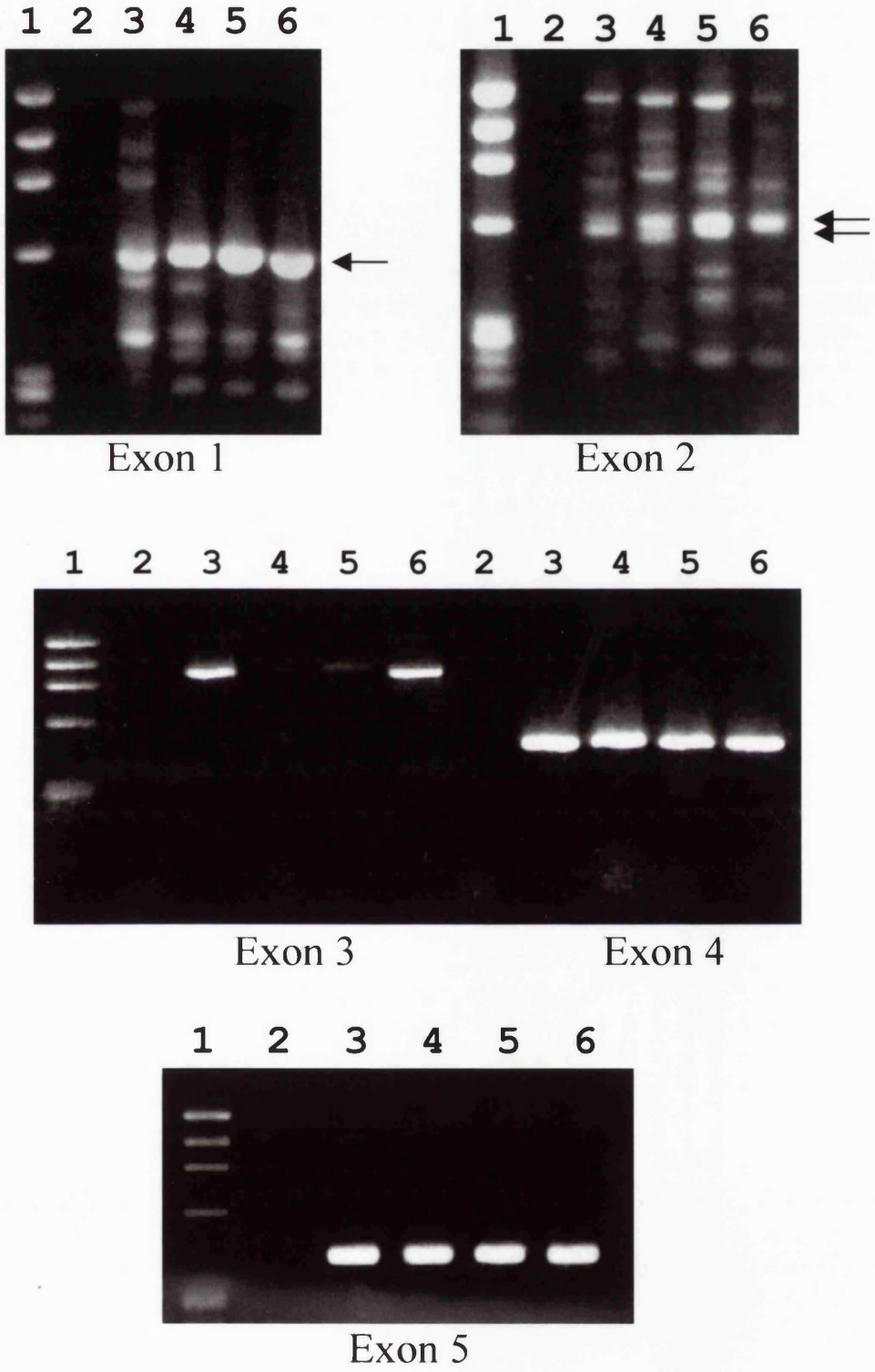


Figure 4.11: PCRs of exons 1-5 of the blue cone opsin gene from the cottoid fish of Lake Baikal. Lane 1: ϕX174/Hae III DNA ladder. Lane 2: No DNA control. Lane 3: *Cottus kesslerei*. Lane 4: *Cottocomephorus inermis*. Lane 5: *Batracocottus nicolskii*. Lane 6: *Limnocottus Eurystomas*. The arrows indicate bands which were excised.

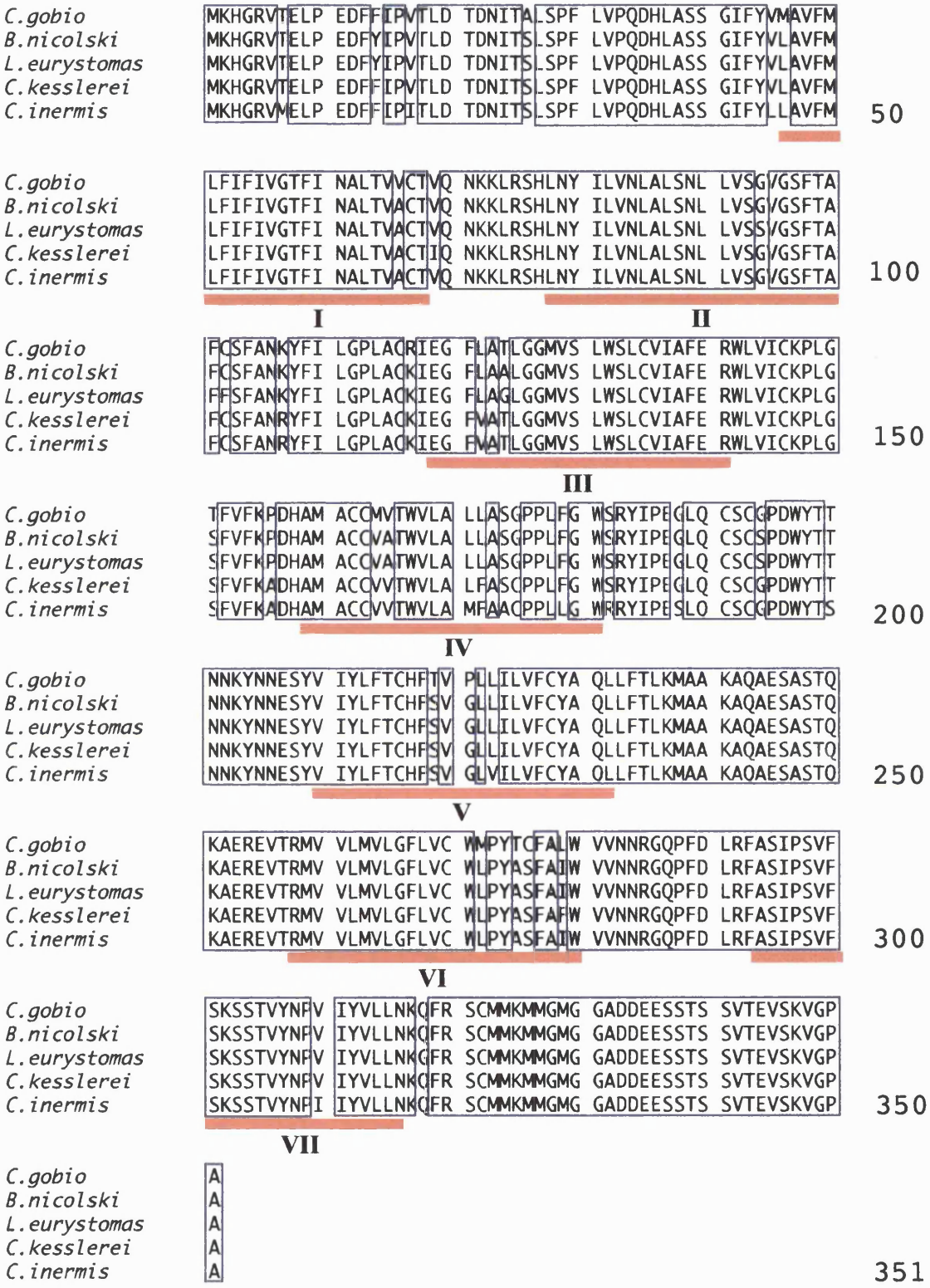


Figure 4.12: Alignment of cottoid blue cone opsin amio acid sequences. Identical amino acids are boxed. The positions of helices I-VII are underlined in red.

shown in table 4.5 below and an amino acid alignment of the helical regions of the blue opsins is shown in table 4.6.

	Cg	Ck	Ci	Le	Bn	Ca	Ol	Dr	Af	Gg
Gg	67.2	66.4	64.4	66.1	66.4	66.4	62.4	64.1	69.8	100.0
Af	74.9	74.6	73.2	74.1	74.3	78.9	71.2	77.2	100.0	
Dr	68.9	68.4	67.5	67.8	68.1	80.3	69.5	100.0		
Ol	70.7	71.2	69.8	71.2	70.7	68.4	100.0			
Ca	68.7	68.7	66.7	67.8	68.4	100.0				
Bn	**95.2**	**96.6**	**94.0**	**98.6**	**100.0**					
Le	**94.3**	**95.7**	**93.2**	**100.0**						
Ci	**91.7**	**96.0**	**100.0**							
Ck	**94.6**	**100.0**								
Cg	**100.0**									

Table 4.5: Percentage amino acid identities when blue cone opsin sequences are aligned with each other. Cg: *Cottus gobio*. Ck: *Cottus kesslerei*. Ci: *Cottocomephorus inermis*. Le: *Limnocottus eurystomias*. Bn: *Batrachocottus nicolskii*. Ca: *Carassius auratus*. Ol: *Oryzias latipes*. Dr: *Danio rerio*. Af: *Astyanax fasciatus*. Gg: *Gallus gallus*. Figures in blue show percentage identity between the cottoid fishes.

Table 4.6: The specific residues contained within the seven transmembrane helices of fish blue (Sb) opsins. Residues which are conserved between all vertebrate opsins are underlined. The helix position is the location of each amino acid residue within the helix and the bovine aa residue no. is the amino acid position relative to bovine rhodopsin (Nathans *et al.*, 1986).

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	M	L	L	L	L	M	M	M	M	M	40
4	A	A	A	A	A	S	S	S	T	A	41
5	V	V	V	V	V	V	A	A	V	A	42
6	F	F	F	F	F	F	L	F	F	F	43
7	M	M	M	M	M	M	M	M	M	M	44
8	L	L	L	L	L	F	F	L	L	F	45
9	F	F	F	F	F	F	V	F	F	L	46
10	I	I	I	I	I	I	L	L	L	L	47
11	F	F	F	F	F	F	F	F	F	I	48
12	I	I	I	I	I	I	V	I	I	A	49
13	V	V	V	V	V	G	A	A	G	L	50
14	G	G	G	G	G	G	G	G	G	G	51
15	T	T	T	T	T	A	T	T	T	V	52
16	F	F	F	F	F	S	A	A	S	P	53
17	I	I	I	I	I	I	I	I	I	I	54
18	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	55
19	A	A	A	A	A	I	L	V	V	T	56
20	L	L	L	L	L	L	L	L	L	L	57
21	T	T	T	T	T	T	T	T	T	T	58
22	V	V	V	V	V	I	I	I	I	I	59
23	V	A	A	A	A	L	A	V	V	F	60
24	C	C	C	C	C	C	C	C	C	C	61
25	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	62

HELIX I

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418	*A.fasciatus* 453	*G.gallus* 455nm	Bovine aa residue no.
3	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	72
4	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	73
5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	74
6	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	75
7	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	76
8	V	V	V	V	V	V	V	V	V	V	77
9	N	N	N	N	N	N	N	N	N	N	78
10	L	L	L	L	L	L	M	L	L	L	79
11	A	A	A	A	A	S	A	A	A	A	80
12	L	L	L	L	L	I	V	I	I	L	81
13	S	S	S	S	S	A	A	S	S	A	82
14	N	N	N	N	N	N	N	N	N	N	83
15	L	L	L	L	L	L	L	L	L	L	84
16	L	L	L	L	L	F	I	W	L	L	85
17	V	V	V	V	V	V	V	V	V	V	86
18	S	S	S	S	S	A	A	S	S	I	87
19	G	G	G	S	G	I	S	V	T	L	88
20	V	V	V	V	V	T	T	F	V	G	89
21	G	G	G	G	G	G	G	G	G	V	90
22	S	S	S	S	S	S	S	S	S	S	91
23	F	F	F	F	F	P	S	S	F	T	92
24	T	T	T	T	T	L	T	V	T	T	93
25	A	A	A	A	A	S	C	A	A	A	94

HELIX II

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	113
4	G	G	G	G	G	G	G	G	G	G	114
5	F	F	F	F	F	F	F	F	F	F	115
6	L	V	V	L	L	L	T	T	V	A	116
7	A	A	A	A	A	A	A	S	A	A	117
8	T	T	T	G	A	T	A	T	T	T	118
9	L	L	L	L	L	L	L	I	L	L	119
10	G	G	G	G	G	G	G	G	G	G	120
11	G	G	G	G	G	G	G	G	G	G	121
12	M	M	M	M	M	M	M	M	M	M	122
13	V	V	V	V	V	V	V	V	V	V	123
14	S	S	S	S	S	G	S	S	S	S	124
15	L	L	L	L	L	L	L	L	L	L	125
16	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	126
17	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	127
18	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	128
19	C	C	C	C	C	A	A	A	S	A	129
20	V	V	V	V	V	V	V	V	V	V	130
21	I	I	I	I	I	V	I	V	V	V	131
22	A	A	A	A	A	A	A	A	A	A	132
23	F	F	F	F	F	F	F	L	F	F	133
24	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	**E**	134
25	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	**R**	135

HELIX III

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	A	A	A	A	A	A	A	A	A	A	153
4	M	M	M	M	M	I	L	I	I	V	154
5	A	A	A	A	A	A	L	A	I	L	155
6	C	C	C	C	C	G	C	G	G	G	156
7	C	C	C	C	C	C	C	C	C	C	157
8	M	V	V	V	V	I	A	I	A	V	158
9	V	V	V	A	A	L	L	L	L	A	159
10	T	T	T	T	T	P	T	P	T	T	160
11	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	161
12	V	V	V	V	V	I	V	C	F	V	162
13	L	L	L	L	L	S	C	M	F	L	163
14	A	A	A	A	A	A	G	A	A	G	164
15	L	L	M	L	L	L	L	L	L	F	165
16	L	F	F	L	L	A	C	A	L	V	166
17	A	A	A	A	A	A	A	A	A	A	167
18	S	S	A	S	S	S	S	G	S	S	168
19	G	C	C	G	G	L	V	L	T	A	169
20	P	P	P	P	P	P	P	P	P	P	170
21	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	171
22	L	L	L	L	L	L	L	L	L	L	172
23	F	F	L	F	F	F	V	L	F	F	173
24	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**	174
25	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	**W**	175

HELIX IV

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	V	V	V	V	V	V	V	V	V	V	204
4	I	I	I	I	I	M	M	M	M	L	205
5	Y	Y	Y	Y	Y	F	F	F	F	F	206
6	L	L	L	L	L	L	L	L	L	L	207
7	F	F	F	F	F	F	F	F	F	F	208
8	T	T	T	T	T	C	C	C	C	T	209
9	C	C	C	C	C	F	F	F	F	F	210
10	H	H	H	H	H	C	C	C	C	C	211
11	F	F	F	F	F	F	F	F	F	F	212
12	T	S	S	S	S	A	A	A	G	G	213
13	V	V	V	V	V	V	V	V	F	V	214
14	P	G	G	G	G	P	P	P	P	P	215
15	L	L	L	L	L	F	F	F	F	L	216
16	L	L	V	L	L	G	S	S	I	A	217
17	I	I	I	I	I	T	I	T	V	I	218
18	L	L	L	L	L	I	I	I	I	I	219
19	V	V	V	V	V	V	V	V	L	V	220
20	F	F	F	F	F	F	F	F	F	F	221
21	C	C	C	C	C	C	C	C	C	S	222
22	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	223
23	A	A	A	A	A	G	S	G	G	G	224
24	Q	Q	Q	Q	Q	Q	Q	Q	Q	R	225
25	L	L	L	L	L	L	L	L	L	L	226

HELIX V

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.inermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	R	R	R	R	R	K	R	K	K	K	252
4	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	253
5	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	254
6	V	V	V	V	V	V	V	V	V	V	255
7	L	L	L	L	L	V	V	V	V	V	256
8	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	**M**	257
9	V	V	V	V	V	V	V	V	V	V	258
10	L	L	L	L	L	L	V	F	M	L	259
11	G	G	G	G	G	G	A	G	G	G	260
12	F	F	F	F	F	F	F	F	F	F	261
13	L	L	L	L	L	L	L	L	L	L	262
14	V	V	V	V	V	V	V	I	V	V	263
15	C	C	C	C	C	C	C	C	C	C	264
16	W	W	W	W	W	W	Y	W	W	W	265
17	M	L	L	L	L	A	V	G	L	A	266
18	P	P	P	P	P	P	P	P	P	P	267
19	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	268
20	T	A	A	A	A	A	A	I	A	T	269
21	C	S	S	S	S	S	S	A	S	A	270
22	F	F	F	F	F	F	F	F	F	F	271
23	A	A	A	A	A	S	A	A	A	A	272
24	L	F	I	I	I	L	L	I	L	L	273
25	W	W	W	W	W	W	W	W	W	W	274

HELIX VI

Helix position	*C.gobio* 470nm	*C.kesslerei* 450nm	*C.lnermis* 450nm	*L.eurystomas* 428nm	*B.nicolskii* 428nm	*C.auratus* 441nm	*O.latipes* 400nm	*D.rerio* 418nm	*A.fasciatus* 453nm	*G.gallus* 455nm	Bovine aa residue no.
3	A	A	A	A	A	A	A	A	G	A	288
4	S	S	S	S	S	T	T	T	T	S	289
5	I	I	I	I	I	I	I	I	I	I	290
6	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	291
7	S	S	S	S	S	S	S	S	S	S	292
8	V	V	V	V	V	C	C	C	C	V	293
9	F	F	F	F	F	F	F	L	F	F	294
10	S	S	S	S	S	S	S	C	S	S	295
11	**K**	**K**	**K**	**K**	**K**	**K**	**K**	**K**	**K**	**K**	296
12	S	S	S	S	S	A	A	A	A	S	297
13	S	S	S	S	S	S	S	S	S	S	298
14	T	T	T	T	T	T	T	T	T	T	299
15	V	V	V	V	V	V	V	V	V	V	300
16	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	301
17	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	302
18	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	**P**	303
19	V	V	I	V	V	V	V	V	V	V	304
20	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	***I***	305
21	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	***Y***	306
22	V	V	V	V	V	V	V	V	V	V	307
23	L	L	L	L	L	L	L	L	F	L	308
24	L	L	L	L	L	M	L	M	M	M	309
25	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	310

HELIX VII

4.2.3 Identification of potential spectral tuning sites

The amino acid sequences deduced were translated and the positions of the seven α-helices identified as shown in figure 4.12. In order to select sites which may be important in tuning of the blue cone pigments in cottoid fish, two basic criteria were used. Firstly, the amino acid should face into the interior of the retinal binding pocket where it can interact with the chromophore and secondly, there should be non-conservative change, i.e. a net change in charge or the gain or loss of an hydroxyl group. A decrease in the electrostatic interaction of the retinal-protonated Schiff's base group with the surrounding protein environment and a decrease in the molecular polarizability of residues lining the chromophore-binding pocket are correlated with a short wave shift in the λ_{max} of the visual pigment (Lin *et al.*, 1994). Each helix was studied in turn and comparisons made between the three classes of blue cone opsin, those with a λ_{max} of 470nm, 450nm and 428nm (table 4.6). There were substitutions between the groups at 13 sites, however only three of these met the criteria above, i.e. were non-conservative substitutions which faced into the retinal binding pocket. These sites are located on helix III at position 8 (equivalent to site 118 in bovine rhodopsin), on helix V at position 14 (equivalent to site 215 in bovine rhodopsin) and on helix VI at position 20 (equivalent to site 269 in bovine rhodopsin) and were considered to be potential spectral tuning sites. This is summarised in the table 4.7. Substitutions at the remaining ten sites were conservative and did not face into the retinal binding pocket so they were not investigated further.

At site 118, threonine in the 470 and 450nm groups, is replaced by either glycine or alanine in the 428nm group. This results in the loss of an hydroxyl group close to the Schiff's base counterion. On helix V at site 215, proline in the 470nm group is substituted with glycine in the 450 and 428nm groups resulting in a conformational change. The final substitution at site 269 on helix VI, involves a loss of an hydroxyl group, with threonine in the 470nm group being replaced by alanine in the 450 and 428nm groups. The nature and location of these substitutions are detailed in figures 4.13, 4.14 and 4.15 respectively.

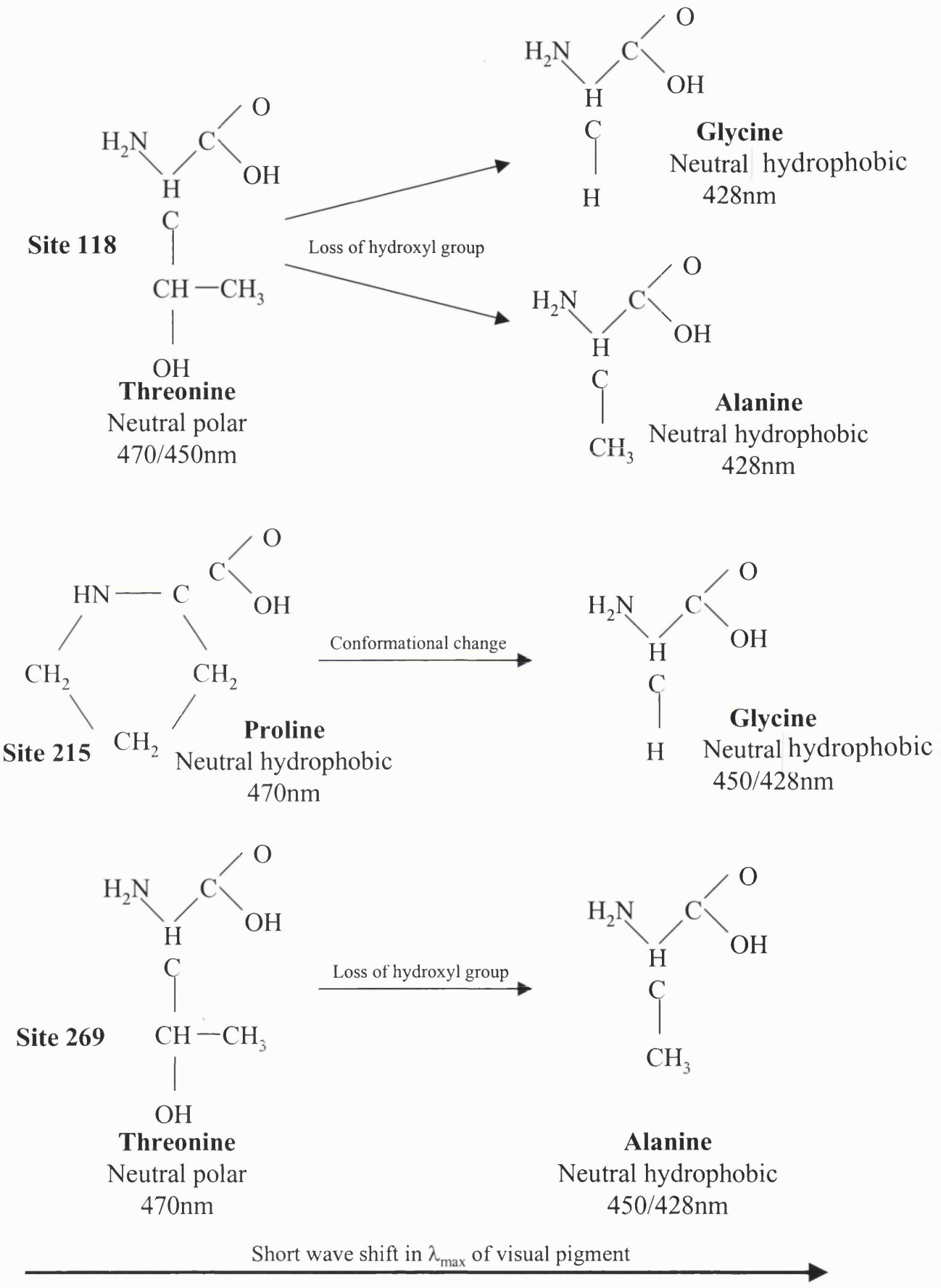


Figure 4.13: Amino acid substitutions at proposed spectral tuning sites

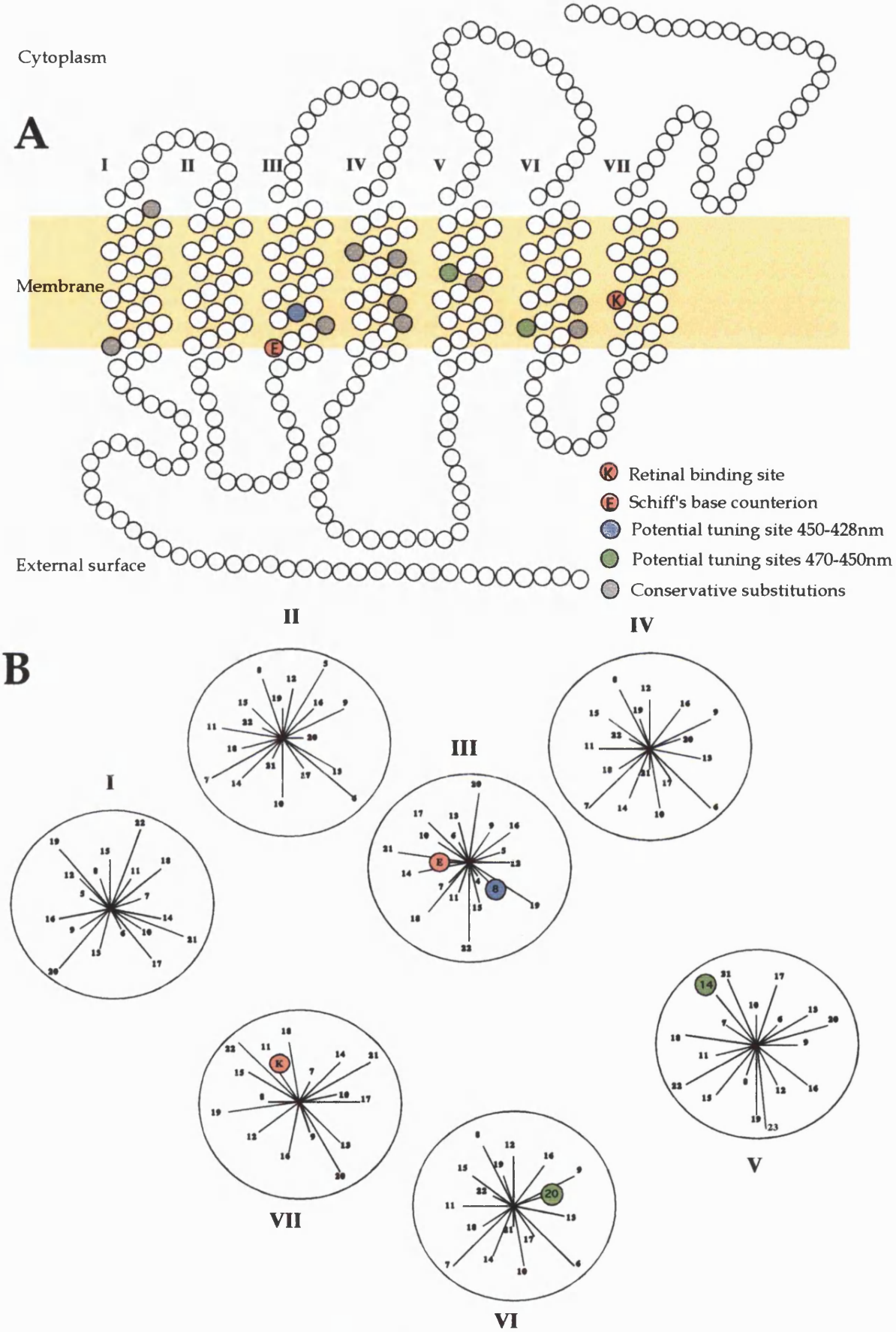


Figure 4.14: Models of the cottoid blue opsin showing positions of potential spectral tuning sites. A) 2D structure showing the seven transmembrane helices. B) 3D structure showing the position of the retinal binding pocket. Each circle represents a helix and each radiating line, an amino acid. The length of the line is proportional to the depth within the helix. Adapted from Baldwin 1998.

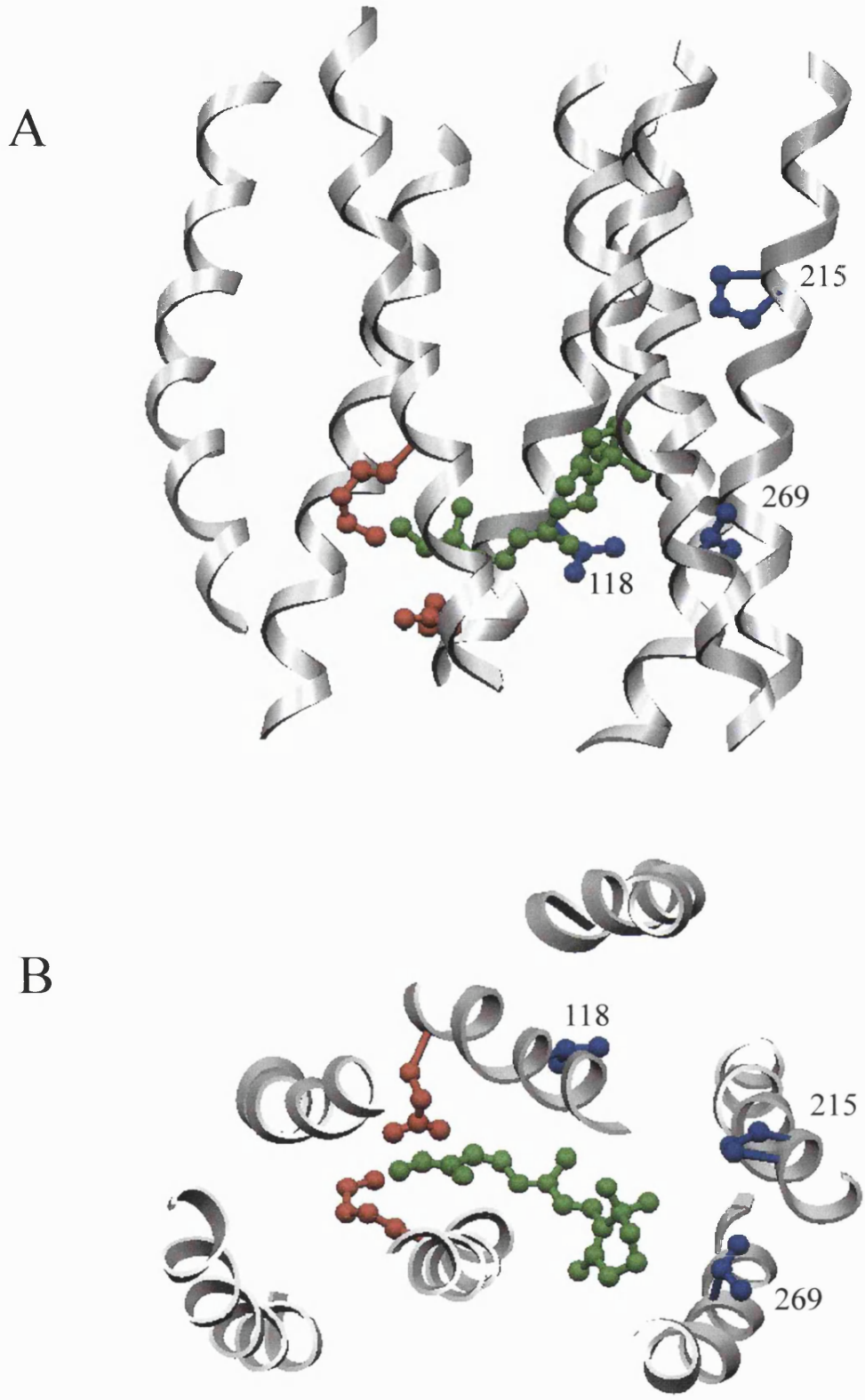


Figure 4.15: Structural models of the *Cottus gobio* blue opsin pigment. A) side view. B) cross section viewed from the cytoplasmic side of the membrane. The three candidate spectral tuning sites are shown in blue. The 11-*cis* retinal chromophore is shown in green. The retinal attachment site, Lys-296, and the Schiffs base counterion, Glu-113, are shown in red. The model was built using Setor and is based on the theoretical model for bovine rhodopsin (Pogozheva 1997).

Species	λmax (MSP)	Site 118	Site 215	Site 269
C.gobio	470nm	Thr	Pro	Thr
C. kesslerei	450nm	Thr	Gly	Ala
C. inermis	450nm	Thr	Gly	Ala
L. eurystomias	428nm	Gly	Gly	Ala
B. nicolskii	428nm	Ala	Gly	Ala

Table 4.7: Amino acid substitutions at potential spectral tuning sites in the blue cone pigments of cottoid fish.

4.2.4 Site-directed mutagenesis to introduce amino acid substitutions

These results suggest that spectral tuning of the blue pigments in cottoid fish may occur in two stages: firstly the shift from 470-450nm by substitutions at either site 215 or 269 or both, and secondly, the shift from 450-428nm by one substitution at site 118. To test this theory the *wild type* cDNA of *Cottus gobio* was altered by site directed mutagenisis introducing each of these substitutions in turn. The two mutations at site 118 where threonine was replaced by either glycine or alanine were created using the Promega Altered Sites II kit. The others, at sites 215 replacing proline for glycine and at 269 replacing threonine for alanine, were constructed using the Stratagene QuikChange site-directed mutagenesis kit.

The *Cottus gobio* full length cDNA was isolated from the pGEM-T easy vector by digestion with *Eco RI* and *Sal I.* The purified fragment was then ligated into the pMT4 vector which had also been digested with the same enzymes to generate the appropriate cloning site. The cloned cDNA was sequenced to confirm that no

mutations had been introduced and that the full cDNA was intact. This clone was used as the *wild type* for subsequent manipulations. In order to generate the mutations, the *Cottus gobio* DNA was isolated from pMT4 using *Eco RI, Sal I* and *Pst I. Pst I* digested the *Cottus gobio* DNA into two fragments of approximately 570 and 480bp (figure 4.16) which were easier to manipulate for mutagenesis. The fragment with *Eco RI* and *Pst I* ends contained site 118 and the *Sal I / Pst I* fragment, sites 215 and 269.

Mutation of the *Eco RI /Pst I* fragment was achieved using the Promega Altered Sites II in vitro mutagenisis system which involves cloning of the desired DNA fragment into a vector pALTER-1, denaturation of the DNA, annealing of phosphorylated oligonucleotide primers containing the mutation to be introduced, synthesis of the mutant strand and two transformation steps, firstly into ES1301 *mutS* cells and then into JM109. The introduction of glycine and alanine at site 118 was successful using this technique and was confirmed by sequencing (figure 4.17). The creation of mutations at sites 215 or 269 in the *Sal I / Pst I* fragment was unsuccessful using this approach. For substitutions at these sites, the Stratagene QuikChange site-directed mutagenisis kit proved successful. This technique was more straightforward and utilises a supercoiled double-stranded DNA vector with the insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The proof reading polymerase, Pfu Turbo, was used in a cycling reaction and the product digested with *Dpn I* which targets methylated DNA. The resulting product was then used to transform JM109 cells. Mutations were confirmed by sequencing and comparison to the wild type as shown in figure 4.18.

The fragments, which had been successfully mutated, were then cloned back into pMT4 along with the unmutated section of wild type DNA to generate plasmids each containing a different substitution. Presence of the full DNA sequence was confirmed by colony PCR using primers BHB2+ and BHB4- (Figure 4.19). Bulk plasmid preparations were then made ready for transfection of 293T cells and expression of the blue opsin protein.

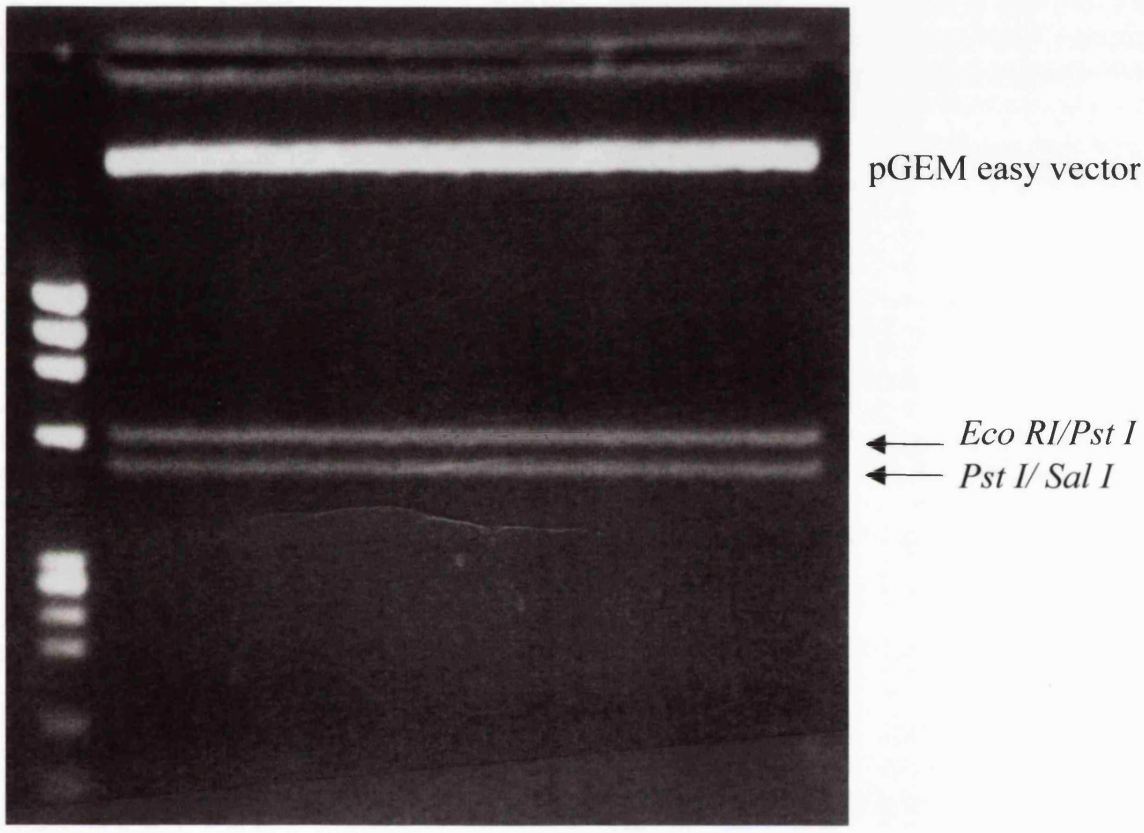


Figure 4.16: *Eco RI/ Sal I/Pst I* digest to release *Cottus gobio* cDNA from the pGEM-T easy vector. The arrows indicate the two fragments of *Cottus gobio* cDNA which were subsequently cloned into the pALTER-1 vector for mutation.

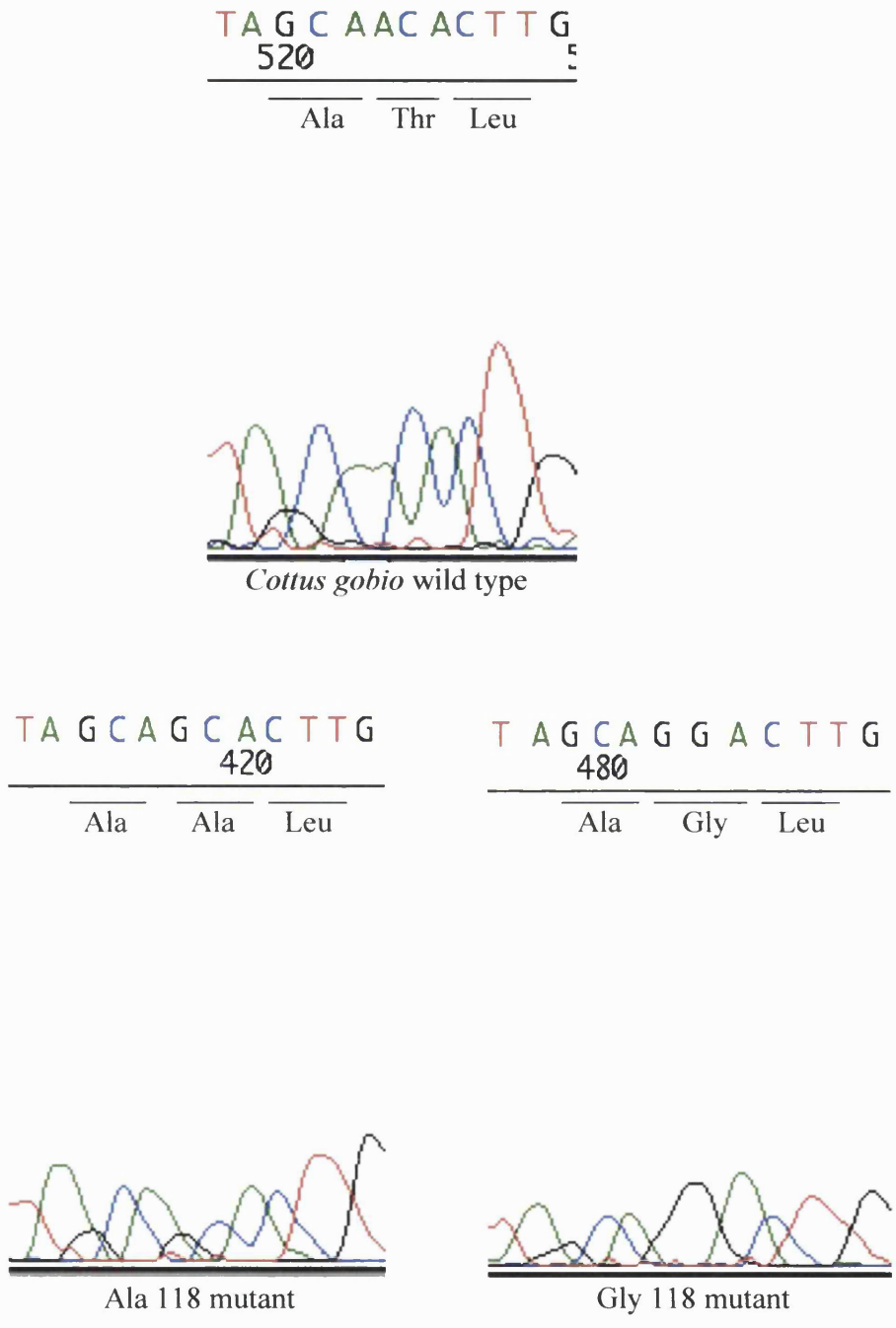


Figure 4.17: Electrophoretograms showing partial sequences of *Cottus gobio* wild type and mutant blue opsin clones

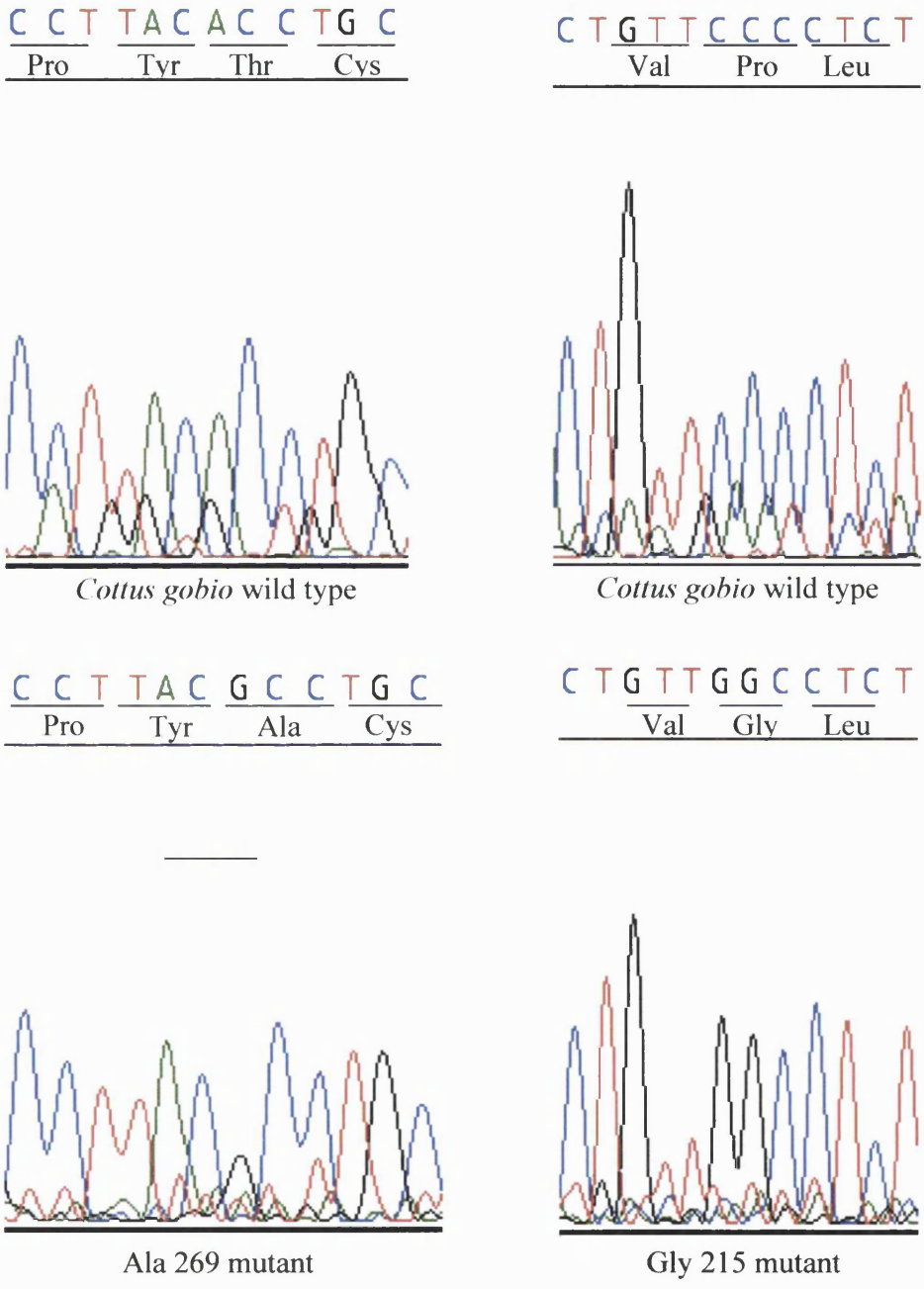


Figure 4.18: Electrophoretograms showing partial sequences of *Cottus gobio* wild type and mutant blue opsin clones.

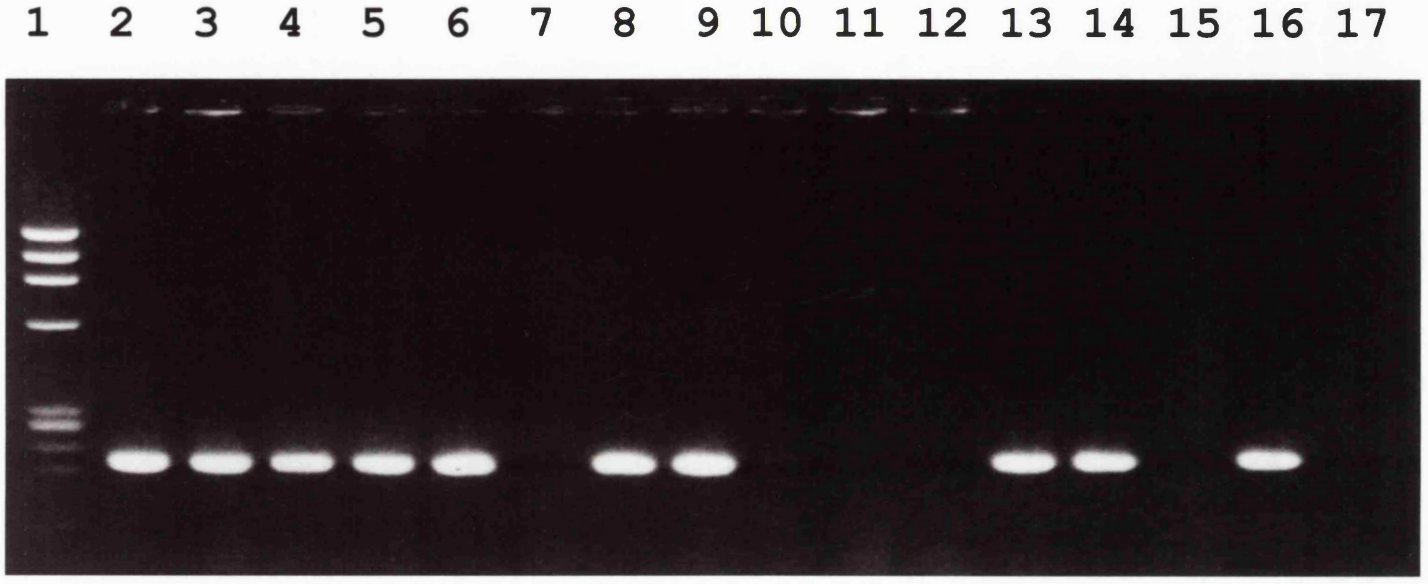


Figure 4.19: Colony PCRs using primers BHB 2+ and BHB 4- to confirm the presence of the full blue opsin DNA sequence. Lane 1 ϕX174/Hae III DNA ladder. Lanes 2,3,4 & 5: 118 Ala mutant. Lanes 6,7,8 & 9: 118 Gly mutant. Lanes 10,1,12 & 13: 215 Gly mutant. Lanes 14, 15 & 16: Ala 269 mutant. Lane 17: No DNA control

4.2.5 Expression of Cottus gobio wild type and mutant blue opsins in 293T cells

293T cells were grown in 90mm petri dishes until they reached between 60-80% confluence. For each wild type or mutant sample, 10 plates were transfected with plasmid DNA (5μg/plate). The cells were incubated for a further 48 hours prior to harvesting. During this time, cell growth was very slow and many even died reducing the confluence to approximately 50% in some cases. The volume of cells collected during harvesting ranged from 300-500μl per 10 plates.

Additionally, two plates were prepared for the bovine rhodopsin, pMT4 vector, wild type *Cottus gobio* and each of the mutant pigments, to determine expression levels by Western blotting.

4.2.6 Determination of expression levels by Western blotting

Western blots were prepared of the wild type and each mutant pigment as described in section 2.23. As seen in figure 4.20, the expression levels were very similar for each sample. Once it had been confirmed that the opsins were actually being expressed, large-scale preps were reconstituted with 11-*cis* retinal and the pigments scanned using a spectrophotometer to determine their $\lambda_{max.}$

4.2.7 Preparation of 11-cis retinal

A crystal of 11-cis retinal was dissolved in 400μl of anhydrous ethanol. 1μl of this stock was diluted with 500μl ethanol to determine its concentration. The absorption spectrum was measured between 190 and 600nm and showed a peak at 380nm with an absorbance reading of 1.16 (Figure 4.21). Using Baer's law, the concentration of the solution was calculated to be 23.2 nmoles/μl.

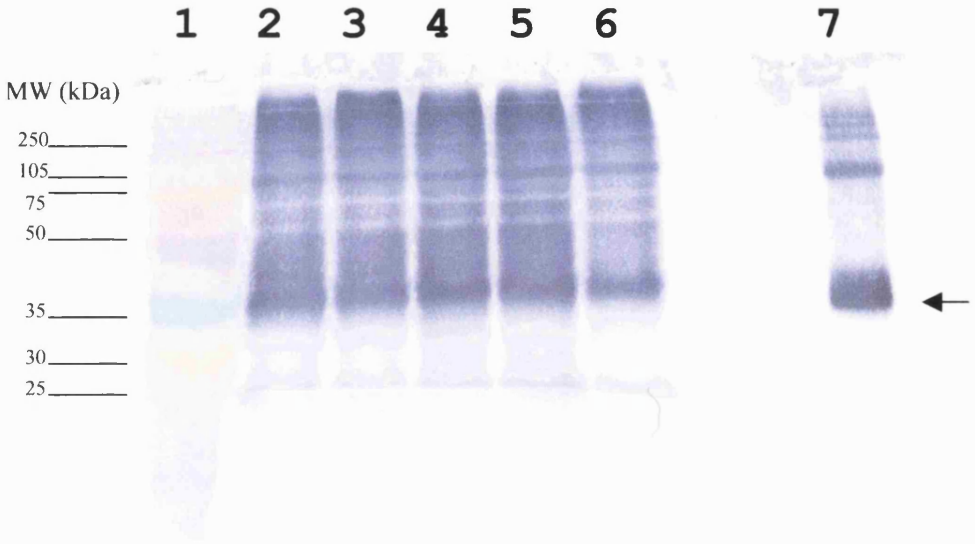


Figure 4.20: Western blot of wild type and mutant opsin pigments. Lane 1: Rainbow marker. Lane 2: wild type. Lane 3: Ala 118 mutant. Lane 4: Gly 118 mutant. Lane 5: Gly 215 mutant. Lane 6: Ala 269 mutant. Lane 7: bovine rod outer segment control. The arrow indicates the position of opsin.

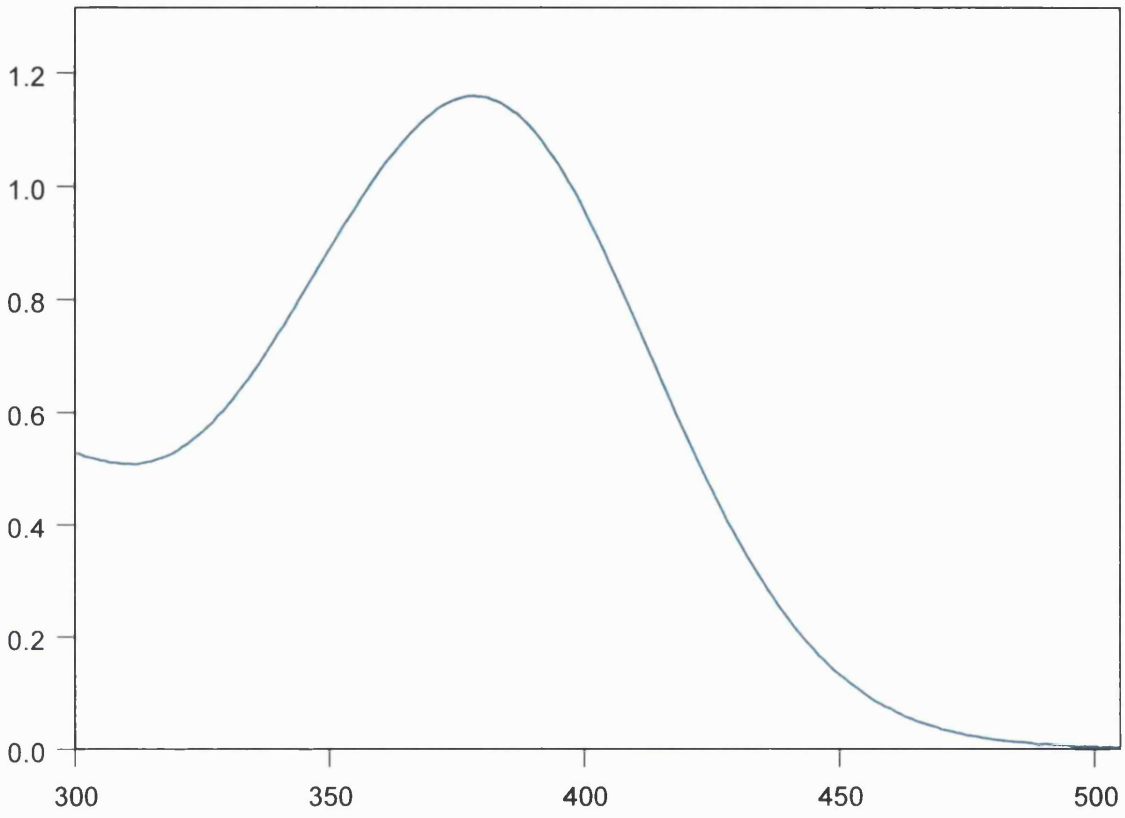


Figure 4.21: Absorbance spectrum of 11-*cis* retinal. The absorbance of 1.16 at 380nm was used to calculate the concentration.

4.2.8 Reconstitution of Cottus gobio wild type and mutant blue opsins with 11-cis retinal

For each reaction, 8.6µl of the 11-*cis* retinal solution was used, giving a final concentration of 40µM. All following procedures were as described in section 2.24. The retinal solution was mixed with the cell pellets to allow its incorporation into the opsin molecules and the cell membranes were digested with a detergent, dodecyl maltoside. PMSF was also added as a protease inhibitor. The opsins were purified by affinity chromatography using a CNBr-activated sepharose column coupled to the 1D4 antibody. Bound opsin was then eluted from the column by the addition of the 1D4 epitope peptide I solution. The λ_{max} of the wild type and each of the mutant opsins was determined by spectrophotometry. Dark spectrums were produced by scanning from 170-700nm. Each sample was then photobleached and re-scanned. The λ_{max} was determined from a difference spectrum, by subtracting the bleached spectrum from the dark spectrum.

Peaks were observed for the wild type, Thr118Gly and Pro215Gly mutants only, showing a λ_{max} of 462nm, 450nm and approximately 463nm respectively. The difference spectra for each showing the raw data is presented in figure 4.22. The smoothed dark, photobleached and difference spectra for the wild type opsin are shown in figure 4.23 and the smoothed difference spectrum for each mutant is shown in figure 4.24.

The data obtained for the Pro215Gly mutant was of poor quality, showing no distinct peak for the λ_{max}. It was however, clear from the spectra that the λ_{max} was certainly not short-wave shifted in comparison to the wild type, as indicated by the position of the ascending arm. In order to estimate the λ_{max}, the spectrum for the wild type pigment was aligned to the 215 mutant spectrum (figure 4.24 B), giving a wavelength of approximately 463nm.

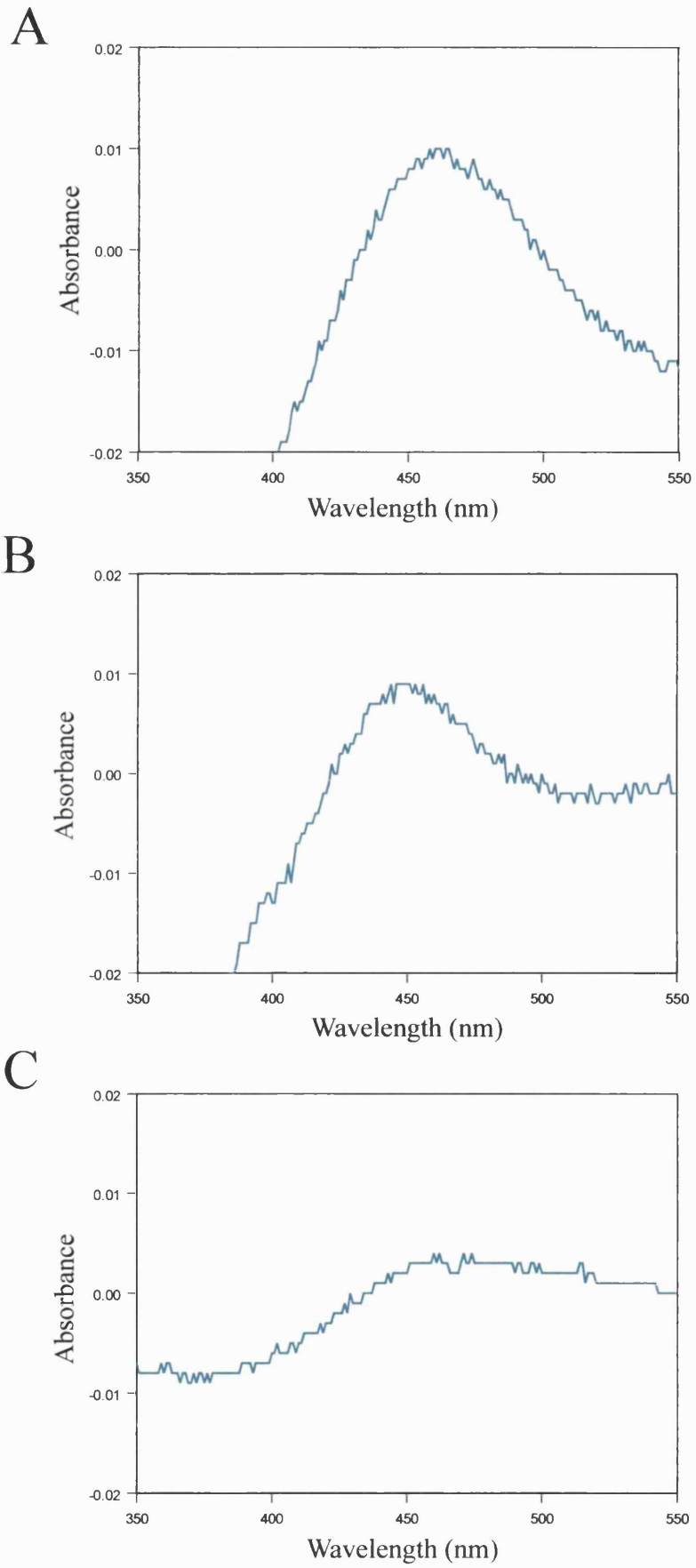


Figure 4.22: Difference spectra for the wild type and mutant blue opsin pigments; raw data. A) Wild type. B) Thr118Gly mutant. C) Pro215Gly mutant.

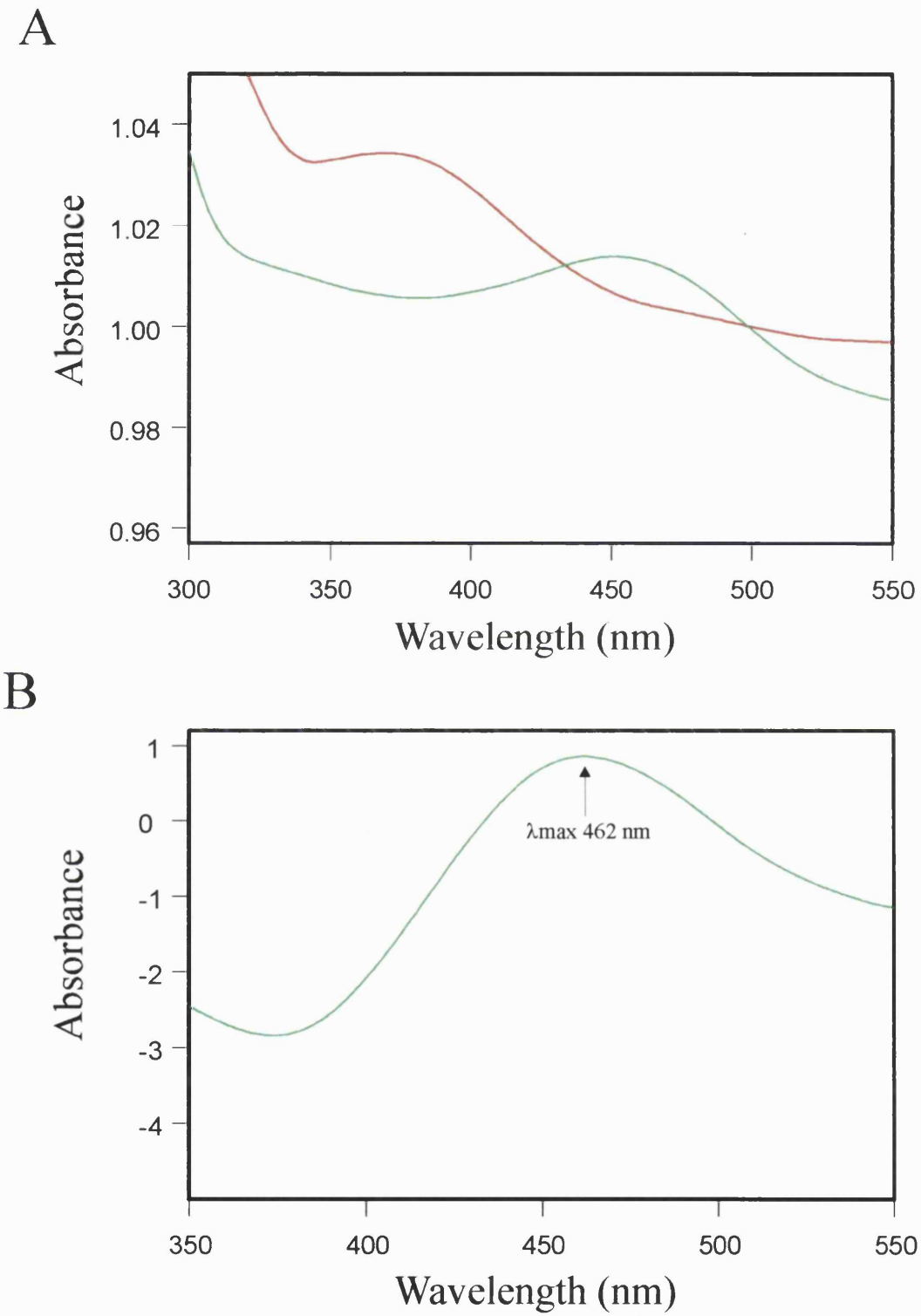


Figure 4.23 : Smoothed spectra of the wild type blue opsin pigment. A) Dark spectrum shown in green, photobleached spectrum shown in red. B) Difference spectrum. Smoothing was performed using the Friedman super algorithm in Axum 6.0

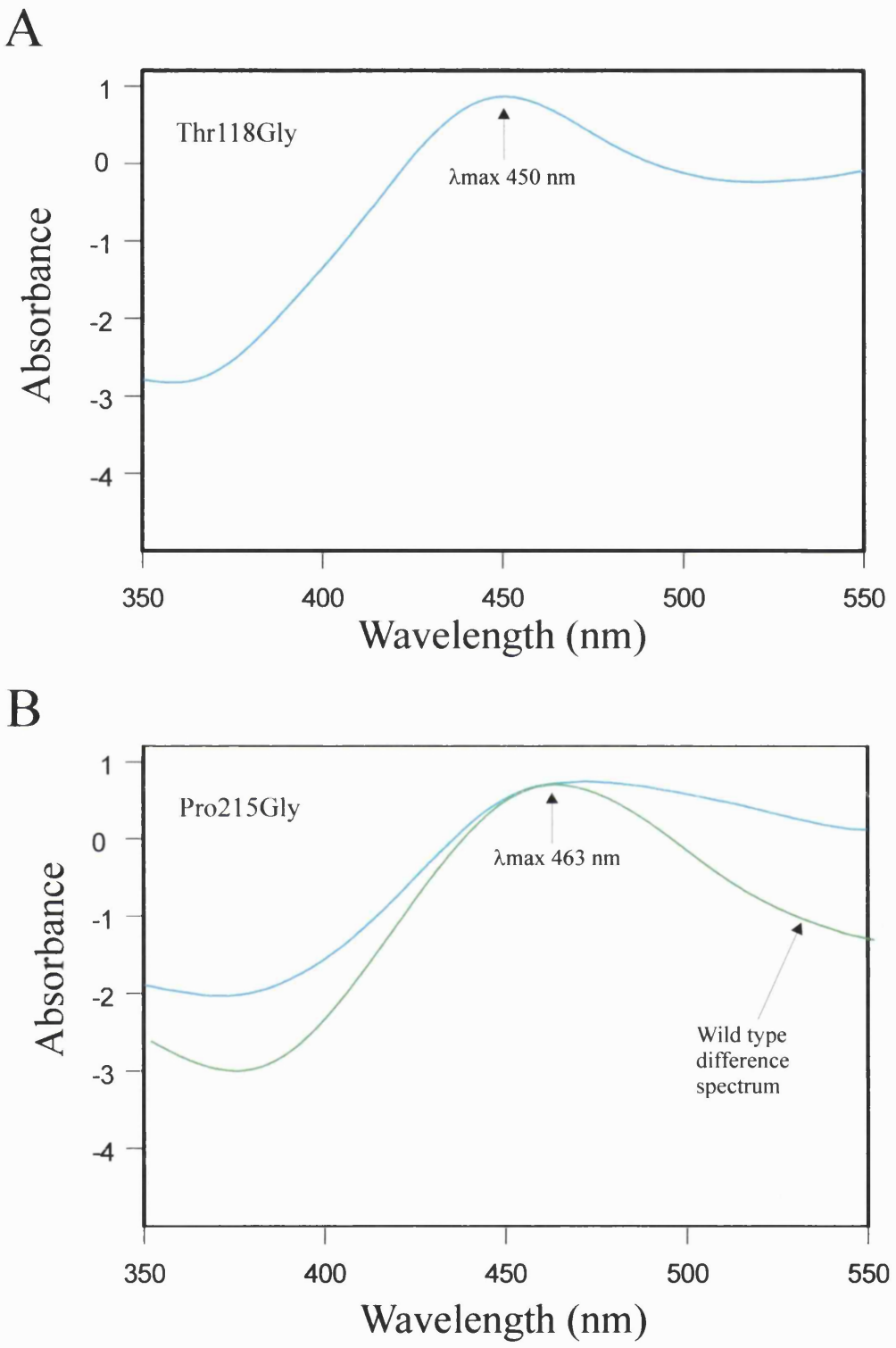


Figure 4.24: Smoothed difference spectra of mutants. A) Thr118Gly. B) Pro215Gly. The wild type difference spectrum has also been included on B In an attempt to clarify the position of the λ_{max} of the Pro215Gly mutant

4.3 Discussion

The data presented here forms the first study of spectral tuning in blue cone opsins of the SWS (Sb) class. In addition to the blue opsins of the cottoid fish of Lake Baikal, the analogous opsin in *Cottus gobio* was also included in the study. The blue cone opsin of *Cottus gobio* has a λ_{max} of 470nm, this is short-wave shifted to 450nm or 428nm in the cottoids of Lake Baikal. This hypsochromic shift is related to increasing depth of habitat but not directly correlated to the wavelength of penetrating light.

Part of the red-shift in λ_{max} of the blue pigment in *Cottus gobio* may be explained by presence of both 11-*cis* retinal and 11-*cis* 3 dehydroretinal acting as chromophores; all Baikal cottoids (Bowmaker *et al.*, 1994) utilise pure 11-*cis* retinal. However, the full 20nm short-wave shift between *Cottus gobio* and the littoral Baikal species, and the further 22nm shift within the Baikal cottoids, cannot be accounted for by this phenomenon so some other process must be responsible for the tuning of these SWS pigments. The substitution of certain amino acids at critical sites within the opsin molecule has previously been shown by many authors to lead to a shift in the λ_{max} of a visual pigment; which sites therefore, are responsible for the tuning of the SWS pigments in these cottoid fishes? In order to answer this question, the blue cone pigment genes were isolated and sequenced from fish representing each wavelength group; *Cottus gobio* (470nm), *Cottus kesslerei* and *Cottocomephorus inermis* (450nm), and *Linmocottus eurystomias* and *Batrachocottus nicolskii* (428nm). Isolation of these genes was not straightforward and involved a combination of 5' and 3' RACE on the cDNA of *Cottus gobio* and walking PCR on genomic DNA of all species. Comparisons with published opsin sequences showed that the cottoid sequences were most similar to the blue opsins of various fish species including the blind cave fish, *Astyanax fasciatus* and the goldfish, *Carassius auratus,* with identities of around 74 and 68 % respectively. Phylogenetic analysis also places the chicken blue cone opsin in the same group as those of fish (Chiu *et al.*, 1994; Yokoyama, 1997; Vihtelic *et al.*, 1999) and shows an identity of approximately 66% with the cottoid fishes. The identities between the cottoid fish in Lake Baikal were generally higher than those between the Baikal fish and their more distant relative, *Cottus*

gobio. One exception to this however is *Limnocottus eurystomias,* which is a member of a separate family, the Abyssocottidae.

By comparing the deduced amino acid sequences of the five blue cone opsins, a number of candidate sites for spectral tuning in the blue region of the spectrum were identified. Out of a total of 13 substitution sites, only three, 118, 215 and 269 were non-conservative and faced into the lumen of the retinal binding pocket. In the pigment with a λ_{max} of 470nm, proline was present at site 215 and threonine at 269; these were substituted by glycine at 215 and alanine at 269 in both the 450 and 428nm pigments. From this one could propose that the shift from 470-450nm was a result of substitutions at these two sites. The shift from 450-428nm would therefore be due to substitutions at position 118; here both the 470 and 450nm pigments have threonine whereas in the 428nm pigments this is replaced by either glycine or alanine.

At residue 118, the wild type ancestral pigment has threonine with the sequence ACA. This is mutated to GCA (alanine) or GGA (glycine) in both of the deep-water Lake Baikal fishes. The logical progression from the wild type would involve two separate mutation steps, firstly from ACA to GCA and then from GCA to GGA. From this one can propose that the ancestral blue pigment of *L. eurystomias* originally had threonine at position 118 which was most probably mutated in two stages, firstly to alanine and then to glycine.

Using the *Cottus gobio* cDNA sequence as the wild type, individual mutations at residues 118, 215 and 269 were introduced by site directed mutagenisis and the recombinant pigments expressed in 293T cells. Purified opsin was then reconstituted with 11-*cis*-retinal and the pigment measured spectrophotometrically to confirm that the substitutions actually caused a shift in the λ_{max} from that of the wild type. Reconstitution of the pigments with 11-*cis* retinal was successful for the wild type, Thr118Gly and Pro215Gly mutants resulting in maximal absorbances at 462nm, 450nm and 463nm respectively. However, no pigment was recovered for either the Thr118Ala or Thr269Ala mutants. This may be due to instability of the pigments *in vitro* when these mutations are introduced, although it is surprising since opsins with

similar mutations have previously been expressed and successfully reconstituted with the chromophore (Wilkie *et al.*, 2000; Sun *et al.*, 1997; Asenjo *et al.*, 1994; Chan *et al.*, 1992).

Site 118 lies towards the luminal side of helix III and is adjacent to the polyene chain of the chromophore. The substitution of threonine with either glycine or alanine results in the loss of an hydroxyl group. Both alanine and glycine are non-polar residues with hydrophobic side chains and so it is to be expected that they would have a similar impact on stabilisation of the opsin molecule. These substitutions may lead to a reduced ability to stabilise the positive charge on the β-ionone ring which is produced upon photoexcitation as a result of destabilisation of the excited state of the chromophore. Although reconstitution of the pigment was only successful for the Thr118Gly mutant, it is very likely that the Thr118Ala mutant would exhibit a similar λ_{max} due to the common properties of these two amino acids. Wilkie *et al.* (2000) also identified a Thr118Ala substitution in avian UV pigments. The replacement of alanine in the native UV pigment with threonine resulted in a red shift of only 3nm. However, in the fish blue pigments, the reverse substitution (but with glycine instead of alanine) resulted in a much larger 12nm blue shift. This difference may be accounted for by the fact that the comparison being made here is between two different classes of visual pigment, the short-wave blue (Sb) and the short-wave UV (Sv). Although this one substitution is at the same site, there are many other amino acid differences between the groups which would act in combination to determine the λ_{max} of the pigment. Not all short-wave pigments have non-polar amino acids at position 118 though (Chang *et al.*, 1995). Hope *et al.* (1997) also identified site 118 as a potential tuning site in fish rod opsins although there was no direct correlation between the λ_{max} of a pigment and the substitution of threonine for alanine; many of the fishes with the blue-shifted rods had threonine at position 118. The alignment of fish and chicken blue opsins in table 4.6 also shows that the polar residue threonine may be present in the short-wave shifted blue pigments such as that of *D. rerio* indicating that, although this site is important in spectral tuning, the presence of an hydrophobic, non-polar residues here, is not essential to all blue pigments. The adjacent site, 117, is also a tuning site identified by Lin *et al.* (1998) and was one of

the residues responsible for the shift in λ_{max} between human rod and blue pigments. Interestingly, *D. rerio* has serine at this position whilst all other fish blue pigments have alanine. Is it possible that a non-conservative substitution at site 117 serves a similar purpose to one at site 118 and that *D. rerio* has a very short wave blue pigment partly because of this Ala 118 Ser substitution?

Although no data was obtained for the Thr269Ala mutant, this site has been proposed as a candidate tuning site in many mammalian opsins (Neitz *et al.*, 1991; Ibbotson *et al.*, 1992; Williams *et al.*, 1992; Shyue *et al.*, 1998; Yokoyama & Radlwimmer, 1999), and avian SWS opsins (Das *et al.*, 1999) and expression studies have confirmed its importance (Chan *et al.*, 1992; Asenjo *et al.*, 1994; Sun *et al.*, 1997). It is close to the β-ionone ring where the loss of polar groups are known cause a short wave shift by destabilising the excited state of the chromophore (Asenjo *et al.*, 1994; Lin *et al.*, 1998; Kochendorfer *et al.*, 1999). However, although the replacement of threonine with alanine does cause a short wave shift, this has only been demonstrated in tuning between LWS and MWS opsins and in rod pigments, never before within the SWS group of opsins. The Thr269Ala substitution found between mammalian LWS and MWS pigments is thought to result in a 15nm blue shift (Yokoyama & Radlwimmer, 1999) and, in expression studies, substitution of alanine with threonine at site 269 in the bovine rod pigment led to a red shift of 14nm (Chan *et al.*, 1992). In numerous other studies, site 269 has been mutated in combination with other spectral tuning sites resulting in a shift in λ_{max} greater than would be expected for any one site alone (Sun *et al.*, 1997; Asenjo *et al.*, 1994; Chan *et al.*, 1992). It would therefore, be logical to conclude that the Thr269Ala substitution is responsible for the 12nm short-wave shift between the wild type blue pigment and those of the lake Baikal cottoid fishes.

In vitro expression studies and reconstitution of the pigment with the Pro215Gly mutation did not show any significant shift in the λ_{max} from that of the wild type. Proline however, is unique among the amino acids in that it has a cyclic structure which kinks and folds the protein architecture. Replacement with glycine could result in a conformational change of the opsin molecule and indirectly affect the interaction

between the protonated Schiff base and the counterion; if these were moved further apart the interaction would be weakened resulting in a short wave shift (Lin *et al.*, 1998; Kochendorfer *et al.*, 1999). However, the failure of this substitution to shift the λ_{max} was not entirely unexpected because of the position of the residue in relation to the chromophore. As seen in fig 4.15A, a side view of the opsin structure, site 215 is high in the molecule and unlikely to interact directly with the 11-*cis* retinal. This site has also never been implicated in the spectral tuning of any other opsins although the adjacent position 214 is thought to be important in tuning the human red and green pigments (Asenjo *et al.*, 1994) and Old World monkey LWS/MWS pigments (Ibbotson *et al.*, 1992). Table 4.6 shows an alignment of other blue opsins from the same class as those of the cottoid fishes. The λ_{max} of the pigments ranges from 418nm-453nm but in all cases proline is present at site 215 except for the cottoid fish of Lake Baikal.

In conclusion, the short-wave shift in the λ_{max} of the blue cone pigments in cottoid fishes is most probably due to substitutions at positions 269 and 118 (summarised in figure 4.25). Site 118 is the critical position distinguishing the two groups of Lake Baikal cottoids, with alanine or glycine replacing threonine in the short-wave shifted, deeper dwelling species. *Cottus gobio* has the longest wavelength blue pigments of the cottoid species studied. This may be accounted for by the presence of the non-polar amino acid, threonine at sites 118 and 269. Although not essential for short-wave pigments, polar residues at site 118 are almost exclusively confined to SWS pigments where as polar residues are found commonly at site 269 in LWS green, SWS blue, violet and UV pigments (Chang *et al.*, 1995; Wilkie *et al.*, 1998; Wilkie *et al.*, 2000).

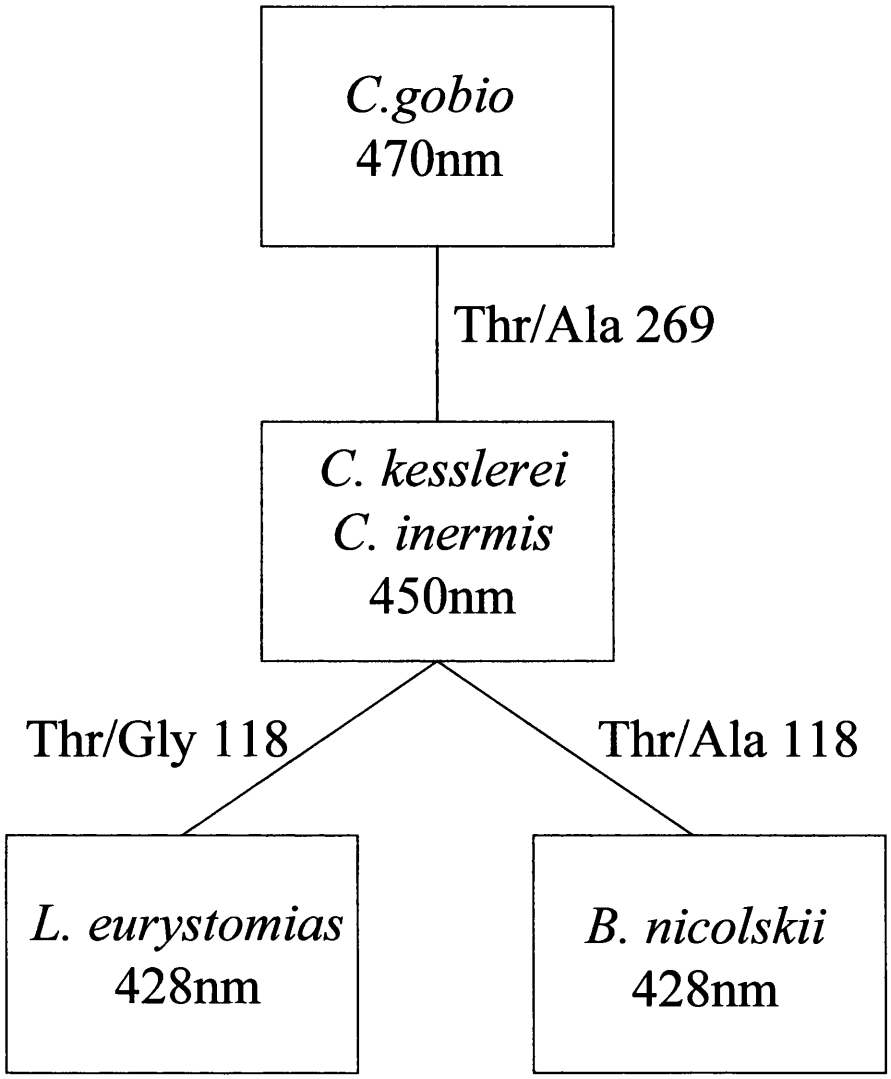


Figure 4.25: Interrelationships between the three classes of blue cone opsins in *Cottus gobio* and the cottoid fish of Lake Baikal. Amino acid substitutions which determine the spectral shifts between classes are shown.

CHAPTER 5

MOLECULAR EVOLUTION OF THE COTTOID FISH OF LAKE BAIKAL

5.1 Introduction

5.1.1 Molecular phylogenetics

Fishes provide an excellent example of how natural selection results in diverse adaptations to common functions. Of the 48,170 recognised living vertebrate species, just over half are fishes (Nelson, 1994). The organisation of this diverse group has been established primarily by studying fossil evidence and comparing external and internal morphologies of living species. The advent of molecular studies has, in most cases, supported this organisation especially at the genus level and has greatly enhanced our understanding of species classification. However, an overall similarity between species and a true evolutionary relationship may not necessarily always be the same so it is useful to compare traditional classifications with those deduced by molecular techniques.

Molecular techniques used for phylogenetic analysis include sequencing regions of mitochondrial and nuclear DNA. Mitochondrial DNA (mtDNA) evolves more quickly than most nuclear DNA and therefore allows identification of informative phylogenetic characters even amongst closely related species. However, the relatively rapid accumulation of mutations makes mtDNA less suitable for studying more ancient evolutionary patterns; in this case nuclear DNA would be more suitable. Knowing the rate of change of a molecule can be crucial in attempts to reconstruct the evolutionary history of a group of organisms. Although molecular sequences generally evolve at a regular clock-like rate in similar species, there are numerous examples to the contrary; a classic case being the α–globin gene in primates which has evolved much more rapidly in baboons than in the rhesus monkey or humans (Hardison & Miller, 1993; Stewart, 1993). Comparisons between a range of

vertebrates have also indicated a strong correlation between substitution rate and body size or generation time; for example rates of nuclear and mtDNA evolution are slower in whales, intermediate in primates and faster in rodents (Martin *et al.*, 1992; Martin & Palumbi, 1993).

The process of unravelling the history of a group of organisms or molecules is complex and numerous methods have been proposed. One of these, maximum parsimony, is commonly used to analyse sequence data and produce molecular phylogenies. It was devised by Eck and Dayhoff in 1966 (Eck & Dayhoff, 1966) and was later adapted to use nucleotide sequences (Fitch, 1977). The principle of maximum parsimony analysis involves producing a tree which requires the minimum number of evolutionary changes to explain differences between the species. Multiple trees may be produced with the minimum number of changes although not all may be informative so trees are often compared before conclusions may be drawn. The neighbour-joining method (Saitou & Nei, 1987) is also commonly used in phylogenetic analysis and is based upon sequential identification of neighbour pairs that minimise the total length of the tree. The length of each branch reflects the evolutionary distance between the species based upon the number of substitutions per site.

Phylogenetic analysis can be used to study the evolution of genes as well as the evolution of species. Although this chapter focuses primarily on the evolution of cottoid fishes, many comparisons are made with opsins of other fishes, birds and mammals; sequence conservation within the opsins is investigated and rates of gene evolution compared.

5.1.2 Classification and evolution of visual pigments

Many phylogenetic trees of vertebrate visual pigments have been constructed on the basis of amino acid sequence identity (Okano *et al.*, 1992; Johnson *et al.*, 1993; Chang *et al.*, 1995; Yokoyama, 1997; Tokunaga *et al.*, 1999). The general

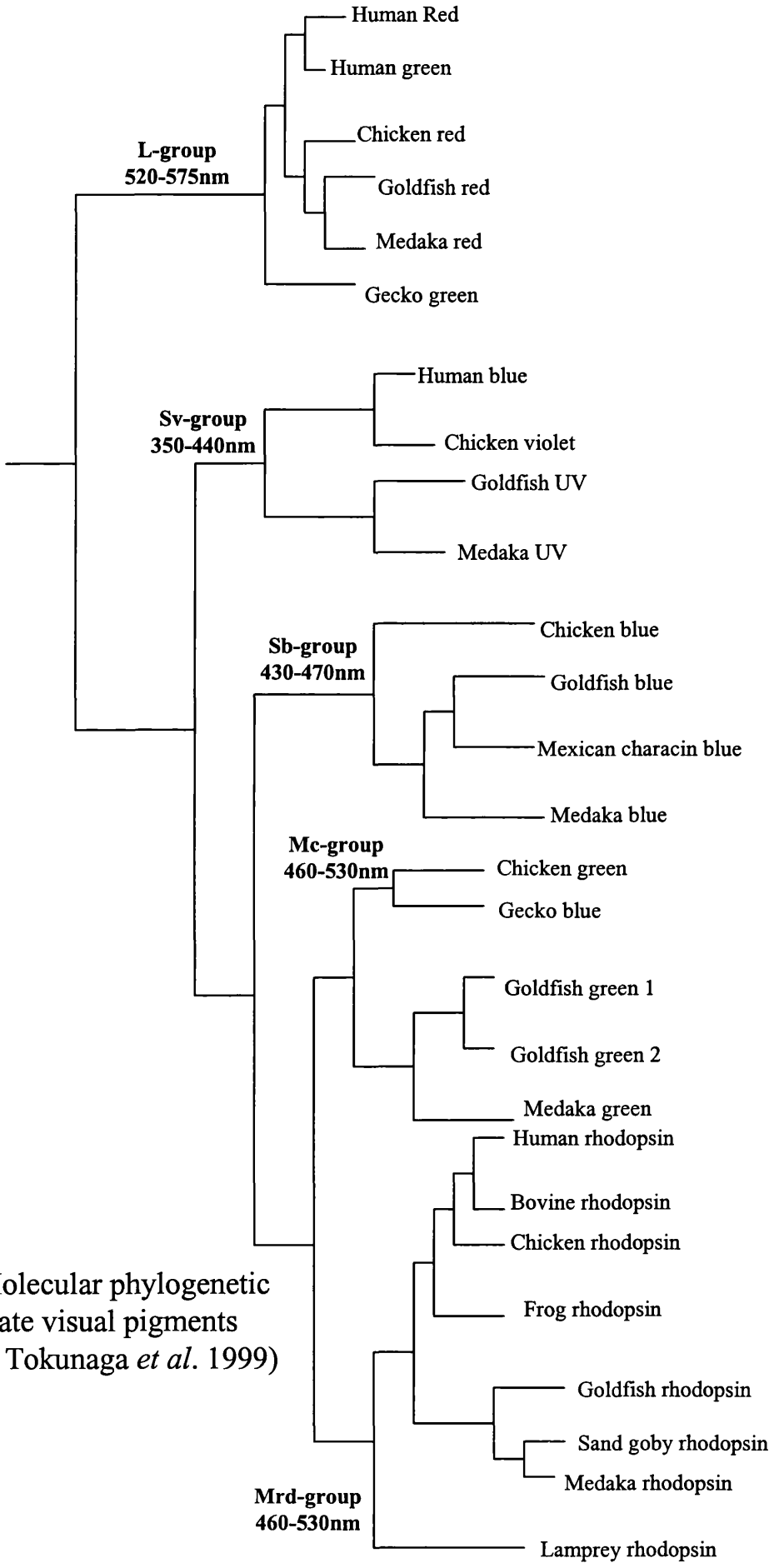


Figure 5.1: Molecular phylogenetic tree of vertebrate visual pigments (adapted from Tokunaga *et al.* 1999)

consensus is that visual pigments can be divided into five major classes as shown in figure 5.1. This tree, which was constructed using the neighbour joining method, is adapted from that of Tokunaga *et al.* 1999 and includes many fish opsins in addition to some mammalian, avian and amphibian pigments. The L-group contains long-wave sensitive cone pigments with a λ_{max} between 520-575nm and includes human red and green, chicken and goldfish red and gecko green. The Mc-group is middle-wave sensitive and includes goldfish and chicken green and gecko blue with a λ_{max} between 460-530nm. Of the short wave pigments, the Sb group includes chicken and goldfish blue, with a λ_{max} between 430-470nm, and the Sv-group, human blue, chicken violet and goldfish UV, all having a λ_{max} between 350-440nm. Finally there is the Mrd-group which consists of the rod opsins of mammals, birds, fish and amphibia with a λ_{max} between 460-530nm (figure 5.1). Many terms are used to classify opsins; those proposed here are from Bowmaker and Hunt 1999.

The first point of divergence of opsin genes is estimated at between 400 (Bowmaker, 1998) and 500 mya (Nathans *et al.*, 1986b), separating a long wave sensitive group from the short wave sensitive pigments and forming the basis for dichromatic colour vision. Subsequent gene duplications within the short wave sensitive group established trichromatic colour vision and a further duplication within the middle wave sensitive group gave rise to the rod opsins. Opsin phylogenetic trees fall into two classes based upon the branch point of the short wave, Sb and Sv, groups. The first class, adopted by most authors, assumes early separation of the L-group followed by subsequent duplications separating the Sv, Sb, Mc and Mrd groups as shown in figure 5.1 (Okano *et al.*, 1992; Johnson *et al.*, 1993; Chang *et al.*, 1995; Yokoyama, 1997; Tokunaga *et al.*, 1999). However, earlier trees produced by Yokoyama (1994) and Hisatomi *et al.* (1994) also propose early separation of the L-group but then a subsequent gene duplication to produce the S and M groups which then divided to produce the Sv and Sb, and then the Mc and Mrd groups. When these early trees were produced, less sequence data were available and bootstrap confidence values were lower than in the first class of trees. Obviously, in the refinement of phylogenetic trees, the more sequence data included in their construction, the better.

Adaptive evolution of different organisms to various photic environments has generated diverse visual systems. In bright open spaces or shallow clear waters the spectrum of light is broad and consequently, animals living in such habitats, tend to have well developed colour vision. The basic vertebrate photoreceptor pattern in these animals consists of rods and four classes of cones, which are spectrally distinct and span the spectrum from near UV to far red. This arrangement is however, noticeably absent in both terrestrial and aquatic mammals, many of which have a rod dominant retina. Of the terrestrial mammals, only Old World primates (including humans) (Nathans *et al.*, 1986b) and the New World primates, *Alouatta seniculus* and *Alouatta caraya* (Jacobs *et al.*, 1996) have full trichromatic colour vision with three distinct photopigments coded for by separate genes. Female New World primates however, may also exhibit a trichromacy based on only two opsin genes, a short-wave gene and different allelic forms of a polymorphic X-linked middle-wave/long-wave gene (Mollon *et al.*, 1984). Most other mammals are dichromatic with a short wave sensitive and either a long or middle wave sensitive pigment (Jacobs, 1993). Two species of aquatic mammal, the bottlenose dolphin (*Tursiops truncates*) and seal (*Phoca hispida*) are unusual in that they completely lack short wave opsins and rely on rod and long-wave sensitive cone pigments for vision (Fasick *et al.*, 1998; Peichl & Moutairou, 1998).

In very dim or dark environments, photopic vision in animals may be lost altogether. This adaptation is common to many deep-water fishes which rely on rod cells for vision and the genes for many of these rod opsins have now been sequenced (Bowmaker *et al.*, 1994; Hunt *et al.*, 1996; Hope *et al.*, 1997; Hunt *et al.*, 2001). The deep-sea dragon fish, *Malacosteus niger,* is unique in terms of visual evolution. It emits far red bioluminescence from suborbital photophores and has developed the ability to perceive this 'light' by enhancing the long wavelength sensitivity of its rod cells by using a derivative of chlorophyll as a photosensitizer (Bowmaker *et al.*, 1988; Douglas *et al.*, 1998; Douglas *et al.*, 1999).

5.1.3 Evolution of visual pigments in cottoid fish

The cottoid fish of Lake Baikal are members of the order Scorpaeniformes which consists of approximately 1,271 species. Of these, 52 species are confined to freshwater (all cottoids) (Nelson, 1994), 29 of them in Lake Baikal. Figure 5.2 shows the hierarchy of cottoid fishes in Lake Baikal based upon morphological classification (Nelson, 1994). The Baikal cottoids form a species flock and are of monophyletic origin (Hunt *et al.*, 1997). It has been suggested that the ancestral species lived in the Pliocene Arctic Ocean and was forced into fresh and salty waters of the lower latitudes by Arctic ice sheets formed during the Pleistocene (Slobodyanyuk *et al.*, 1995). The pure A_1 retinal chromophore also suggests that these fish have a marine origin (Bowmaker *et al.*, 1994); marine species have pure A_1 retinas whereas euryhaline fish generally have A_1/A_2 mixtures, utilising the A_2 form in freshwater (Bowmaker, 1991). The role of glaciation/deglaciation cycles on the speciation of these fish is unknown although it is probable that the ancestral species was littoral or sub-littoral and gradually adapted to deep water during the appearance and disappearance of numerous basins within the lake (Timoshkin, 1997).

It is now generally accepted that gene duplication is the most important mechanism for generating new gene families that have facilitated the evolution of complex organisms. One of these families, the opsins, has been widely studied and the genes for these proteins from many species have been sequenced and used in the construction of phylogenetic trees. A measure of the confidence of a branch point in a tree is given by the bootstrap value. Within the opsin trees this value is generally high (>50%) and therefore the grouping is well supported, especially in gene trees, where there is a clear separation of the various opsin classes eg. (Okano *et al.*, 1992; Johnson *et al.*, 1993; Chang *et al.*, 1995; Yokoyama, 1997; Tokunaga *et al.*, 1999). The use of opsin sequences in the construction of species trees results in more variable bootstrap values. Nevertheless, the use of rod opsin sequences for phylogenetic analysis of Baikal cottoids produced some interesting results. Both neighbour joining and maximum parsimony methods produced evidence for two main lineages (Hunt *et al.*, 1997). The two *Cottocomephorus* species differ from the other Baikal species at eight sites and are placed in one clade on both trees supported by

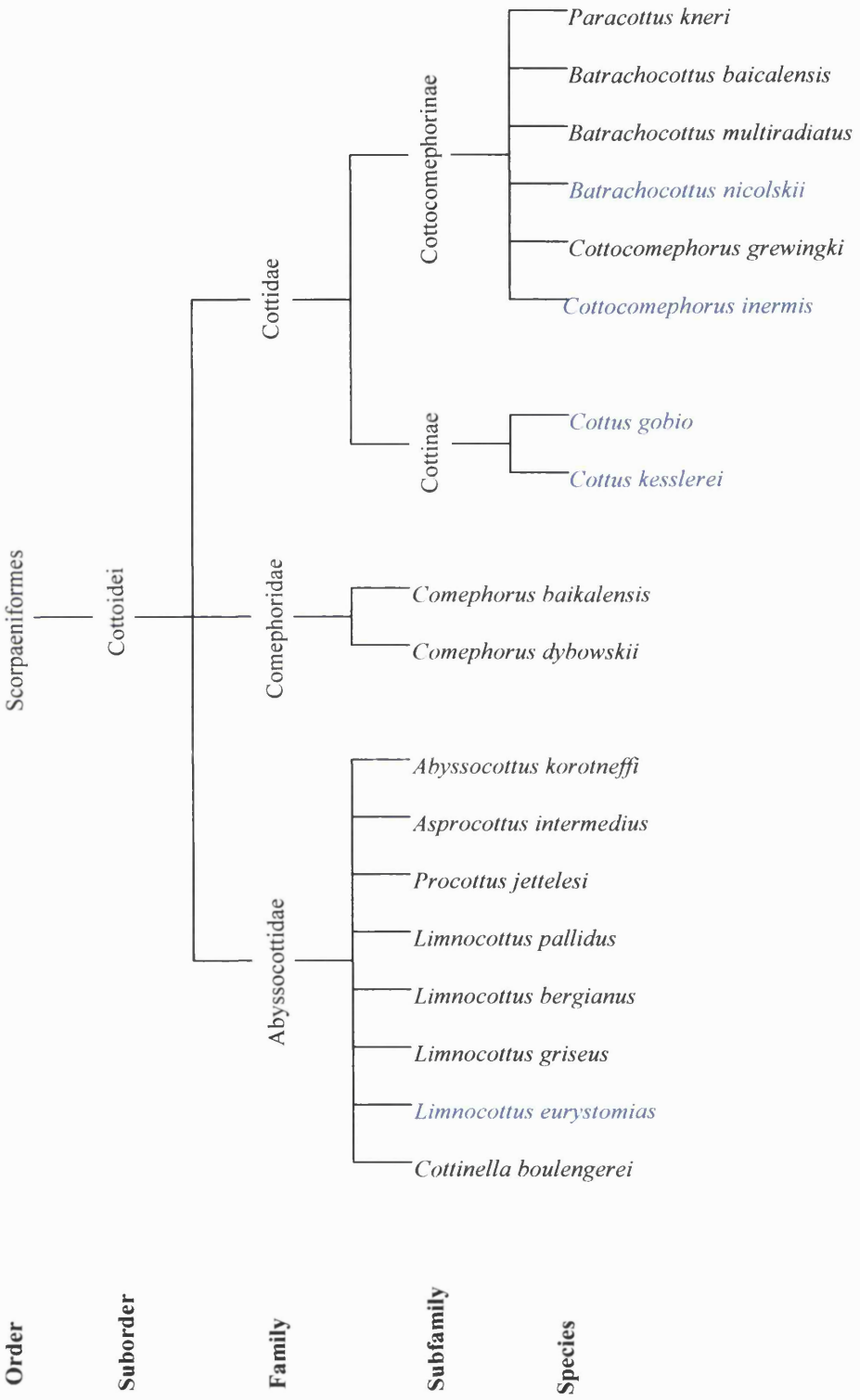


Figure 5.2: Hierarchy of the cottoid fish of Lake Baikal. The blue opsin genes were sequenced from the species shown in blue.

100% confidence values (figure 5.3). In comparison, morphological classification places them in the Cottocomephorinae subfamily along with *Batrachocottus* and *Paracottus* (Nelson, 1994). The two littoral species, *Cottus kesslerei* and *Paracottus kneri* are also placed in a separate clade supported by high bootstrap values of 78%. However, the remaining eight species form a third clade whose sub-divisions are not well supported, with values ranging from only 11 to 56%. Again, morphological classification does not directly correlate with these results where the *Batrachocottus* species are grouped with members of the Abyssocottidae family and the Cottidae are separated into several distinct branches. Slobodyanyuk *et al.* (1995) have constructed neighbour joining trees using mitochondrial DNA data derived from Cytochrome B (CytB) and ATPase 8 and 6 genes (figure 5.4). Six species were compared and different results obtained for each tree. The CytB tree and one produced by restriction analysis, most closely resemble morphological trees, with three of the shallow water Cottidae, *C. kesslerei, C. grewingki and C. inermis,* being placed into one clade. The ATPase tree however, differs considerably from the other trees and may not be so useful for phylogenetic analysis. One reason proposed for this is the higher than expected rate of transitions compared to transversions in ATPase genes (Slobodyanyuk *et al.*, 1995). Bootstrap values to support the branching of neighbour joining trees were not included although some values have been published for the corresponding maximum likelihood trees (Slobodyanyuk *et al.*, 1995). In another paper however, the mtDNA CytB genes were analysed from 16 species of Baikal cottoid and a maximum parsimony tree constructed. Although there was strong support for the clade containing *C. kesslerei, C, grewingki and C. inermis,* confidence values for the other branches were generally low showing that the clustering was not well supported (Kiril'chik *et al.*, 1995) (figure 5.5). Sequence data from the blue opsin genes of five species of cottoid fish have been obtained and will be used to construct further phylogenetic trees. These will be compared to the trees discussed above in an attempt to clarify the classification of cottoid fishes and to determine whether the blue opsin gene is a better system for use in molecular phylogenetic analysis.

The study of 12 Baikal cottoids suggests that their radiation began around 4.9mya (Hunt *et al.*, 1997). This was based on the rate of divergence at synonymous sites within the coding region of the rod opsin genes and differs from mitochondrial DNA data which placed the radiation around 2mya (Slobodyanyuk *et al.*, 1995). A study of the blue opsin sequences of the Lake Baikal cottoid fishes will not only help to clarify the date of species radiation but will also show whether the species grouping obtained with the rod opsin sequence is supported by blue cone opsin data.

5.2 Results

5.2.1 Amplification and sequencing of the blue cone pigment of cottoid fishes

The blue cone pigment gene of five species of cottoid fishes was amplified and sequenced as described previously in sections 2.15. Four of the species are found in Lake Baikal; *Cottus kesslerei, Cottocomephorus inermis, Limnocottus eurstomias* and *Batrachocottus nicolskii* and the fifth, *Cottus gobio,* is native to the rivers of Britain. These sequences together with those of goldfish (*Carassius auratus),* killifish (*Oryzias latipes*), zebrafish (*Danio rerio*), the blind cave fish (*Astyanax fasciatus)* and chicken (*Gallus gallus*), were aligned using Clustal in Geneworks™ and default parameters (figure 5.4). Maximum parsimony (Fitch, 1977) and neighbour joining (Saitou & Nei, 1987) analyses of aligned sequences were carried out using the PAUP 4.0b4a package. An exhaustive search was performed; support for the branching order was assessed using 1000 bootstrap replications and the branch and bound algorithm. The respective trees are shown in figure 5.5. As expected, the chicken blue opsin is clearly separated from the fish opsins. Two of the members of the cottidae family which have the 450nm pigment, *Cottus kesslerei* and *Cottocomephorus inermis*, are placed into one clade with high bootstrap values of either 92 or 81%, depending upon the method used. The remaining two Baikal species, *Limnocottus eurystomias* and *Batrachocottus nicolskii*, fall into another clade with even higher bootstrap values of 97 or 100%. These two species are members of the Abyssocottidae and Cottidae families respectively and both have the shorter-wavelength, 428nm, pigment.

5.2.2 Divergence of cottoid fishes

In order to calculate the rate of divergence between the Lake Baikal Cottoid fishes and set the molecular clock, pairwise comparisons were made across the entire nucleotide sequences and at sites of synonymous substitution only, using MEGA (Molecular Evolutionary Genetic Analysis software) (Kumar *et al.*, 1993). The Jukes and Cantor formula (Jukes & Cantor, 1969), which corrects for multiple substitutions at synonymous and non-synonymous sites, was applied to the calculations. The results

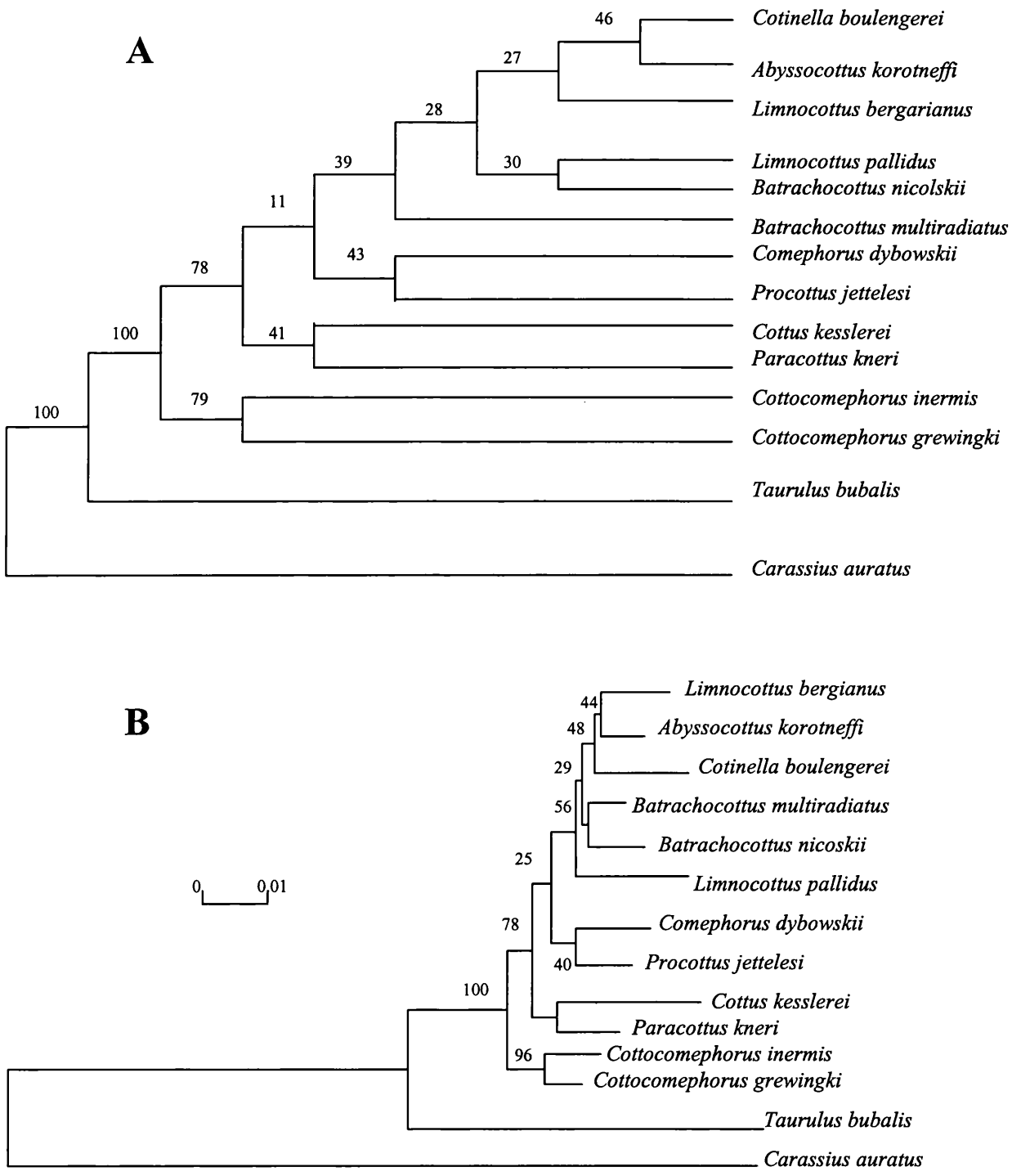


Figure 5.3: Phylogenetic trees based on the total nucleotide sequence of the rod opsin gene of cottoid fishes. A) Maximum parsimony. B) Neighbour joining. The length of each branch reflects the evolutionary distance between species. The scale bar is calculated in substitutions per site. The bootstrap confidence values based on 500 replicates are shown for each branch. Redrawn from Hunt *et al.* 1997

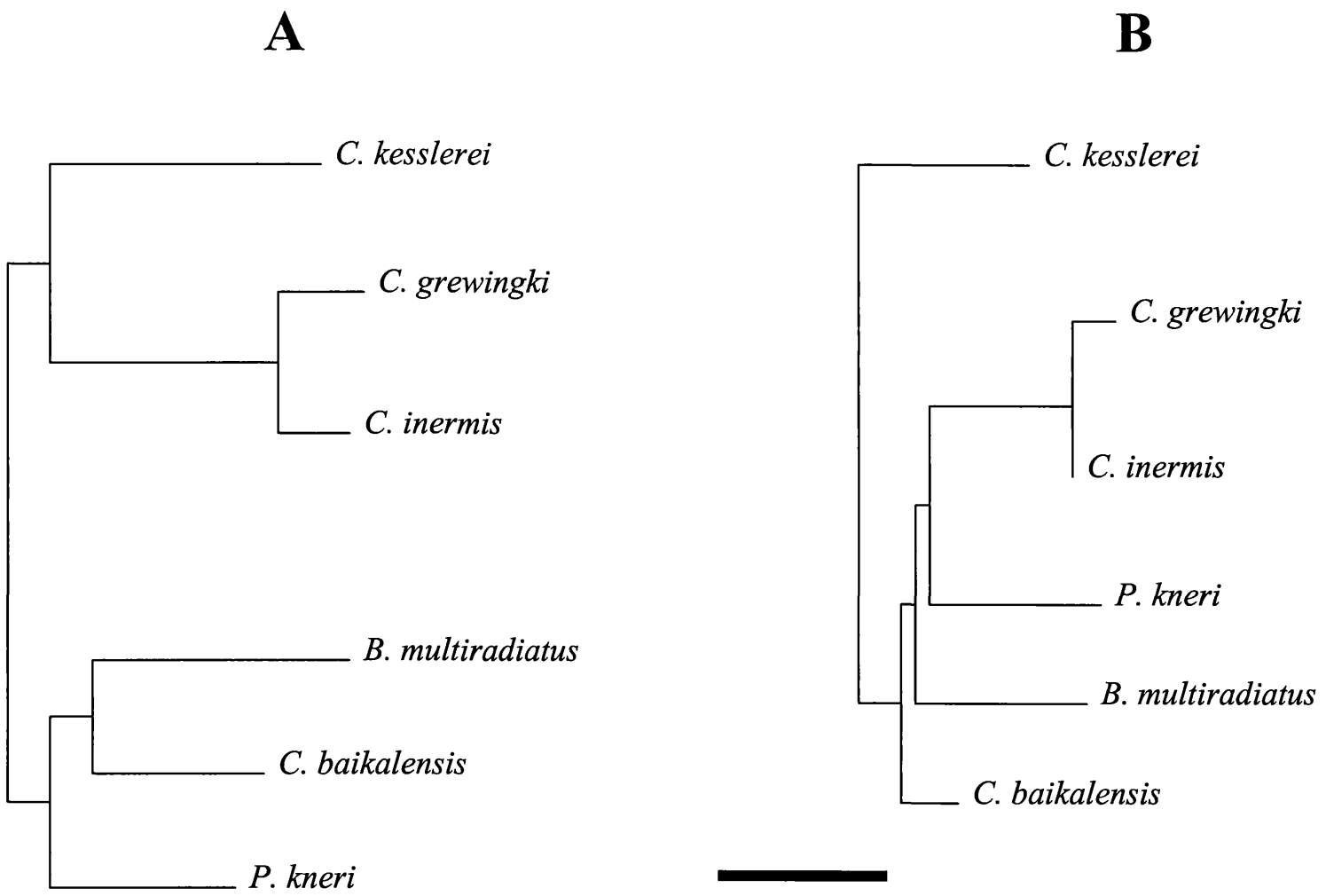


Figure 5.4: Neighbour joining trees for six Lake Baikal cottoid mtDNAs. A) CytB tree. B) ATPase 8+6 tree. The bar under the trees indicates a frequency of 10 nucleotide substitutions. From Slobodyanyuk *et al.* 1995.

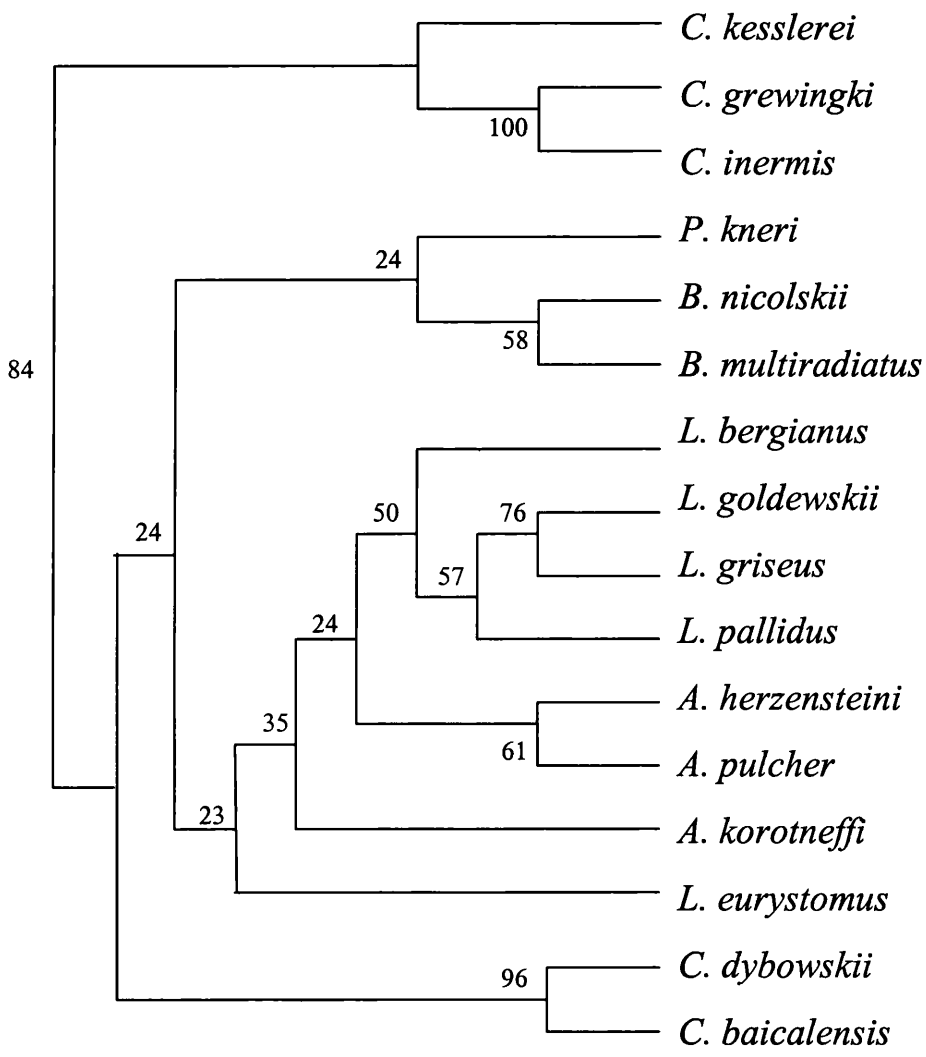
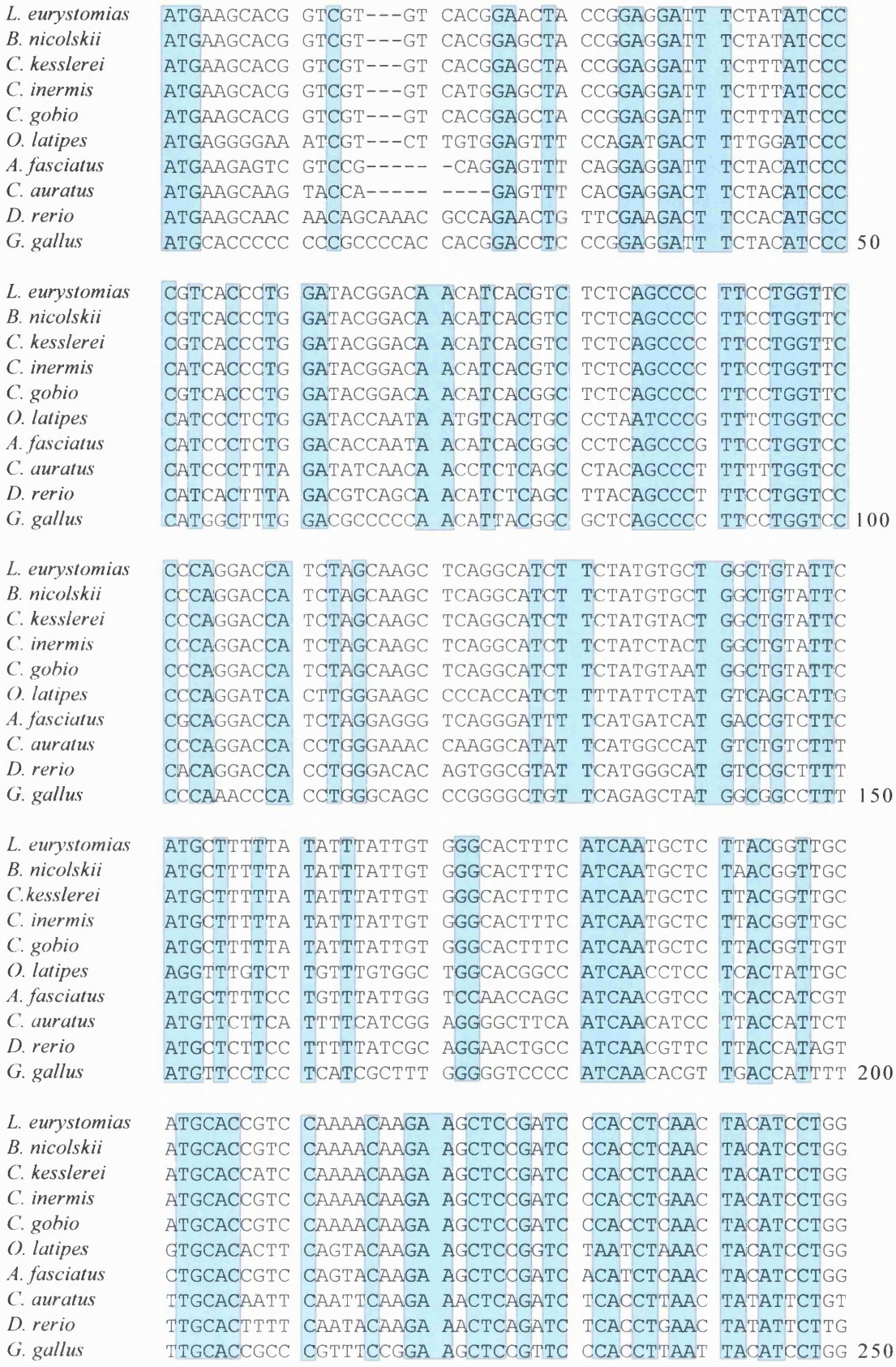
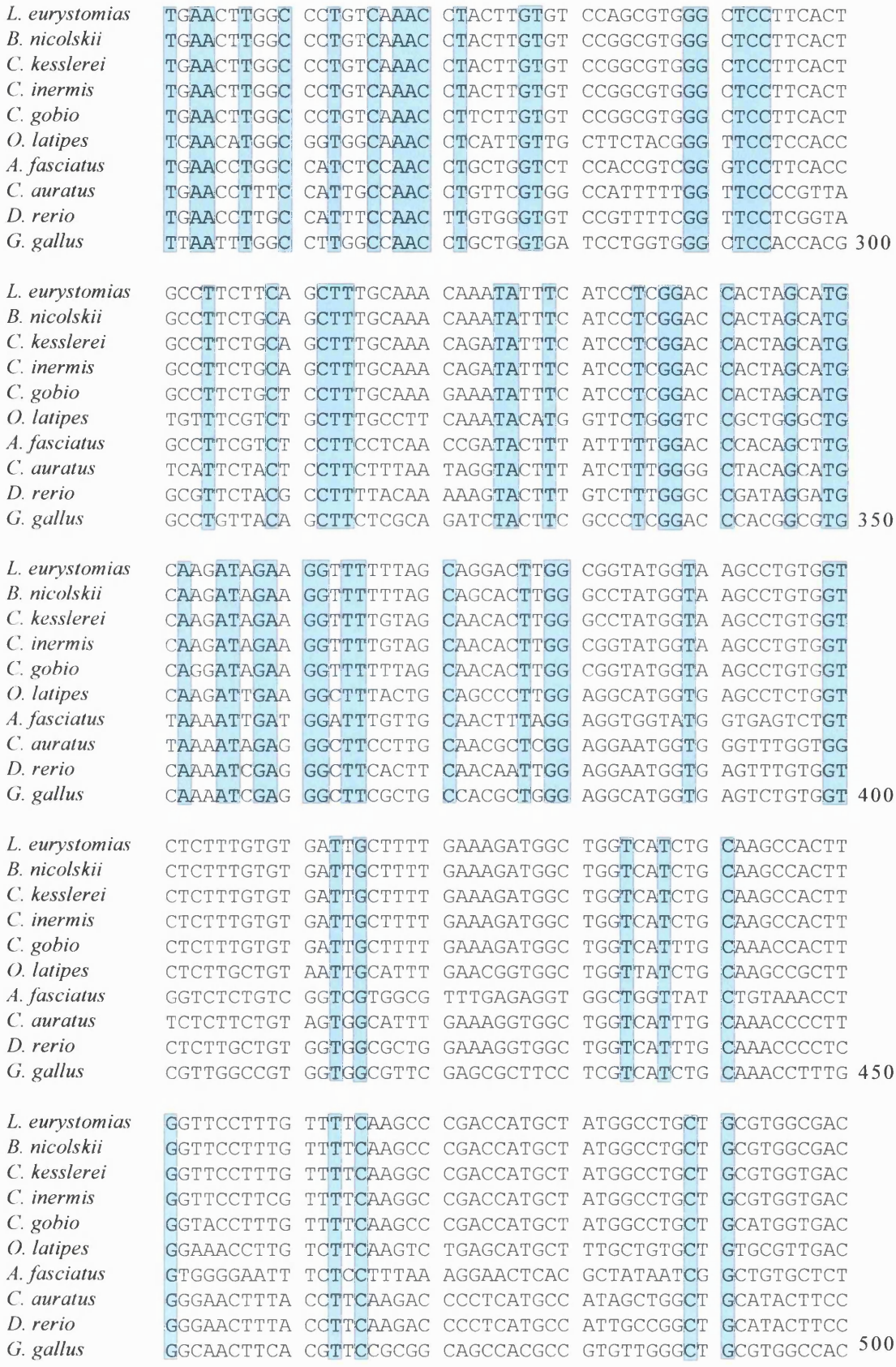


Figure 5.5: PARS tree of 16 Lake Baikal cottoid mtDNAs prepared using CytB sequence data. The bootstrap confidence values based on 100 replicates are shown for each branch. From Kiril'chik *et al.* 1995/

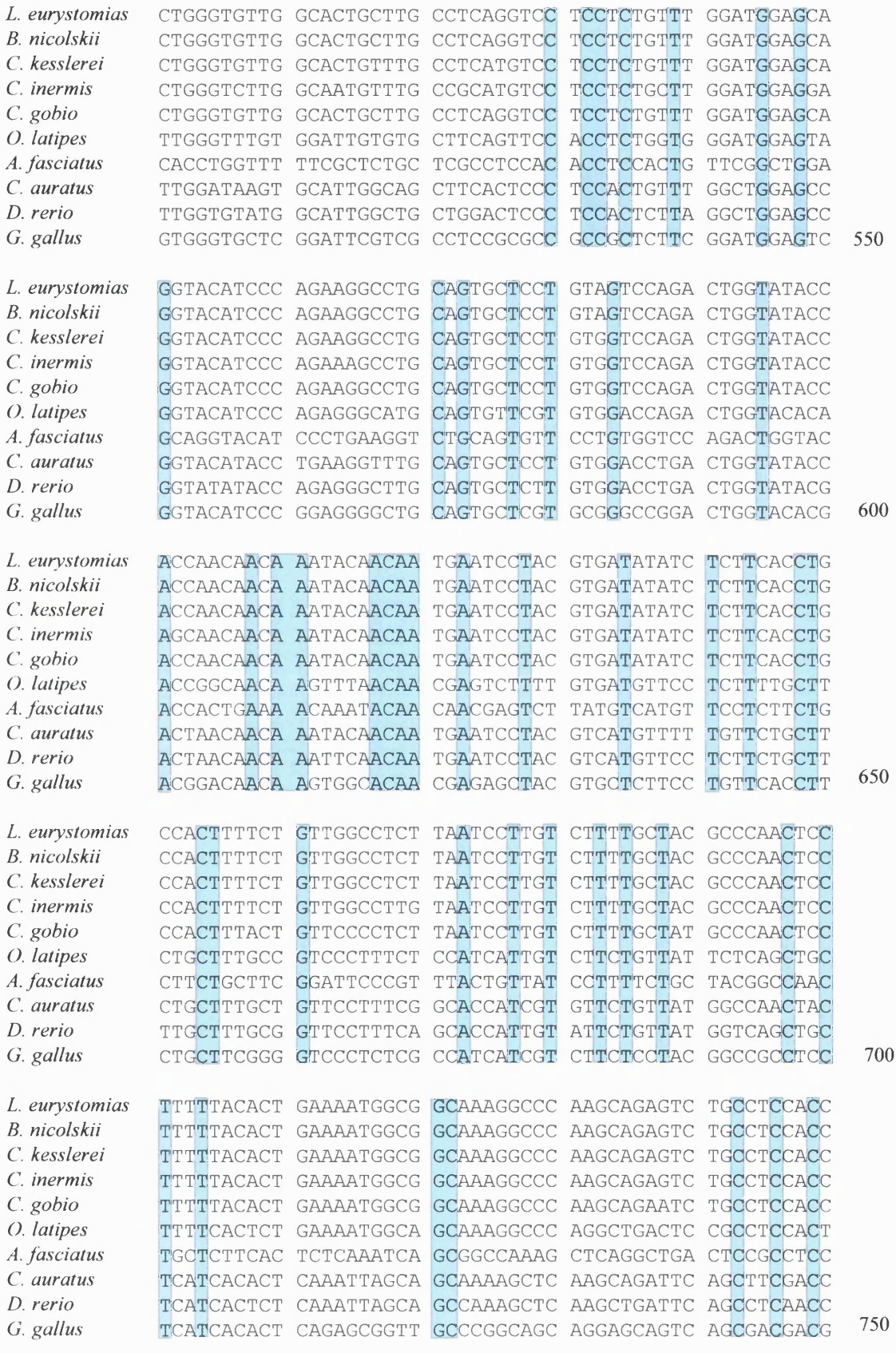
Figure 5.6: Aligned sequences of the blue cone opsin genes of fish and chicken. Identical sequences are boxed. The numbers to the right represent the position relative to the translation start site in chicken (*Gallus gallus*).


L. eurystomias ATGAAGCACG GTCGT---GT CACGGAACTA CCGGAGGATT TCTATATCCC
B. nicolskii ATGAAGCACG GTCGT---GT CACGGAGCTA CCGGAGGATT TCTATATCCC
C. kesslerei ATGAAGCACG GTCGT---GT CACGGAGCTA CCGGAGGATT TCTTTATCCC
C. inermis ATGAAGCACG GTCGT---GT CATGGAGCTA CCGGAGGATT TCTTTATCCC
C. gobio ATGAAGCACG GTCGT---GT CACGGAGCTA CCGGAGGATT TCTTTATCCC
O. latipes ATGAGGGGAA ATCGT---CT TGTGGAGTTT CCAGATGACT TTTGGATCCC
A. fasciatus ATGAAGAGTC GTCCG----- -CAGGAGTTT CAGGAGGATT TCTACATCCC
C. auratus ATGAAGCAAG TACCA----- ----GAGTTT CACGAGGACT TCTACATCCC
D. rerio ATGAAGCAAC AACAGCAAAC GCCAGAACTG TTCGAAGACT TCCACATGCC
G. gallus ATGCACCCCC CCCGCCCCAC CACGGACCTC CCGGAGGATT TCTACATCCC 50
L. eurystomias CGTCACCCTG GATACGGACA ACATCACGTC TCTCAGCCCC TTCCTGGTTC
B. nicolskii CGTCACCCTG GATACGGACA ACATCACGTC TCTCAGCCCC TTCCTGGTTC
C. kesslerei CGTCACCCTG GATACGGACA ACATCACGTC TCTCAGCCCC TTCCTGGTTC
C. inermis CATCACCCTG GATACGGACA ACATCACGTC TCTCAGCCCC TTCCTGGTTC
C. gobio CGTCACCCTG GATACGGACA ACATCACGGC TCTCAGCCCC TTCCTGGTTC
O. latipes CATCCCTCTG GATACCAATA ATGTCACTGC CCTAATCCCG TTTCTGGTCC
A. fasciatus CATCCCTCTG GACACCAATA ACATCACGGC CCTCAGCCCG TTCCTGGTCC
C. auratus CATCCCTTTA GATATCAACA ACCTCTCAGC CTACAGCCCT TTTTTGGTCC
D. rerio CATCACTTTA GACGTCAGCA ACATCTCAGC TTACAGCCCT TTCCTGGTCC
G. gallus CATGGCTTTG GACGCCCCCA ACATTACGGC GCTCAGCCCC TTCCTGGTCC 100
L. eurystomias CCCAGGACCA TCTAGCAAGC TCAGGCATCT TCTATGTGCT GGCTGTATTC
B. nicolskii CCCAGGACCA TCTAGCAAGC TCAGGCATCT TCTATGTGCT GGCTGTATTC
C. kesslerei CCCAGGACCA TCTAGCAAGC TCAGGCATCT TCTATGTACT GGCTGTATTC
C. inermis CCCAGGACCA TCTAGCAAGC TCAGGCATCT TCTATCTACT GGCTGTATTC
C. gobio CCCAGGACCA TCTAGCAAGC TCAGGCATCT TCTATGTAAT GGCTGTATTC
O. latipes CCCAGGATCA CTTGGGAAGC CCCACCATCT TTTATTCTAT GTCAGCATTG
A. fasciatus CGCAGGACCA TCTAGGAGGG TCAGGGATTT TCATGATCAT GACCGTCTTC
C. auratus CCCAGGACCA CCTGGGAAAC CAAGGCATAT TCATGGCCAT GTCTGTCTTT
D. rerio CACAGGACCA CCTGGGACAC AGTGGCGTAT TCATGGGCAT GTCCGCTTTT
G. gallus CCCAAACCCA CCTGGGCAGC CCGGGGCTGT TCAGAGCTAT GGCGGCCTTT 150
L. eurystomias ATGCTTTTTA TATTTATTGT GGGCACTTTC ATCAATGCTC TTACGGTTGC
B. nicolskii ATGCTTTTTA TATTTATTGT GGGCACTTTC ATCAATGCTC TAACGGTTGC
C.kesslerei ATGCTTTTTA TATTTATTGT GGGCACTTTC ATCAATGCTC TTACGGTTGC
C. inermis ATGCTTTTTA TATTTATTGT GGGCACTTTC ATCAATGCTC TTACGGTTGC
C. gobio ATGCTTTTTA TATTTATTGT GGGCACTTTC ATCAATGCTC TTACGGTTGT
O. latipes AGGTTTGTCT TGTTTGTGGC TGGCACGGCC ATCAACCTCC TCACTATTGC
A. fasciatus ATGCTTTTCC TGTTTATTGG TCCAACCAGC ATCAACGTCC TCACCATCGT
C. auratus ATGTTCTTCA TTTTCATCGG AGGGGCTTCA ATCAACATCC TTACCATTCT
D. rerio ATGCTCTTCC TTTTTATCGC AGGAACTGCC ATCAACGTTC TTACCATAGT
G. gallus ATGTTCCTCC TCATCGCTTT GGGGGTCCCC ATCAACACGT TGACCATTTT 200
L. eurystomias ATGCACCGTC CAAAACAAGA AGCTCCGATC CCACCTCAAC TACATCCTGG
B. nicolskii ATGCACCGTC CAAAACAAGA AGCTCCGATC CCACCTCAAC TACATCCTGG
C. kesslerei ATGCACCATC CAAAACAAGA AGCTCCGATC CCACCTCAAC TACATCCTGG
C. inermis ATGCACCGTC CAAAACAAGA AGCTCCGATC CCACCTGAAC TACATCCTGG
C. gobio ATGCACCGTC CAAAACAAGA AGCTCCGATC CCACCTCAAC TACATCCTGG
O. latipes GTGCACACTT CAGTACAAGA AGCTCCGGTC TAATCTAAAC TACATCCTGG
A. fasciatus CTGCACCGTC CAGTACAAGA AGCTCCGATC ACATCTCAAC TACATCCTGG
C. auratus TTGCACAATT CAATTCAAGA AACTCAGATC TCACCTTAAC TATATTCTGT
D. rerio TTGCACTTTT CAATACAAGA AACTCAGATC TCACCTGAAC TATATTCTTG
G. gallus TTGCACCGCC CGTTTCCGGA AGCTCCGTTC CCACCTTAAT TACATCCTGG 250

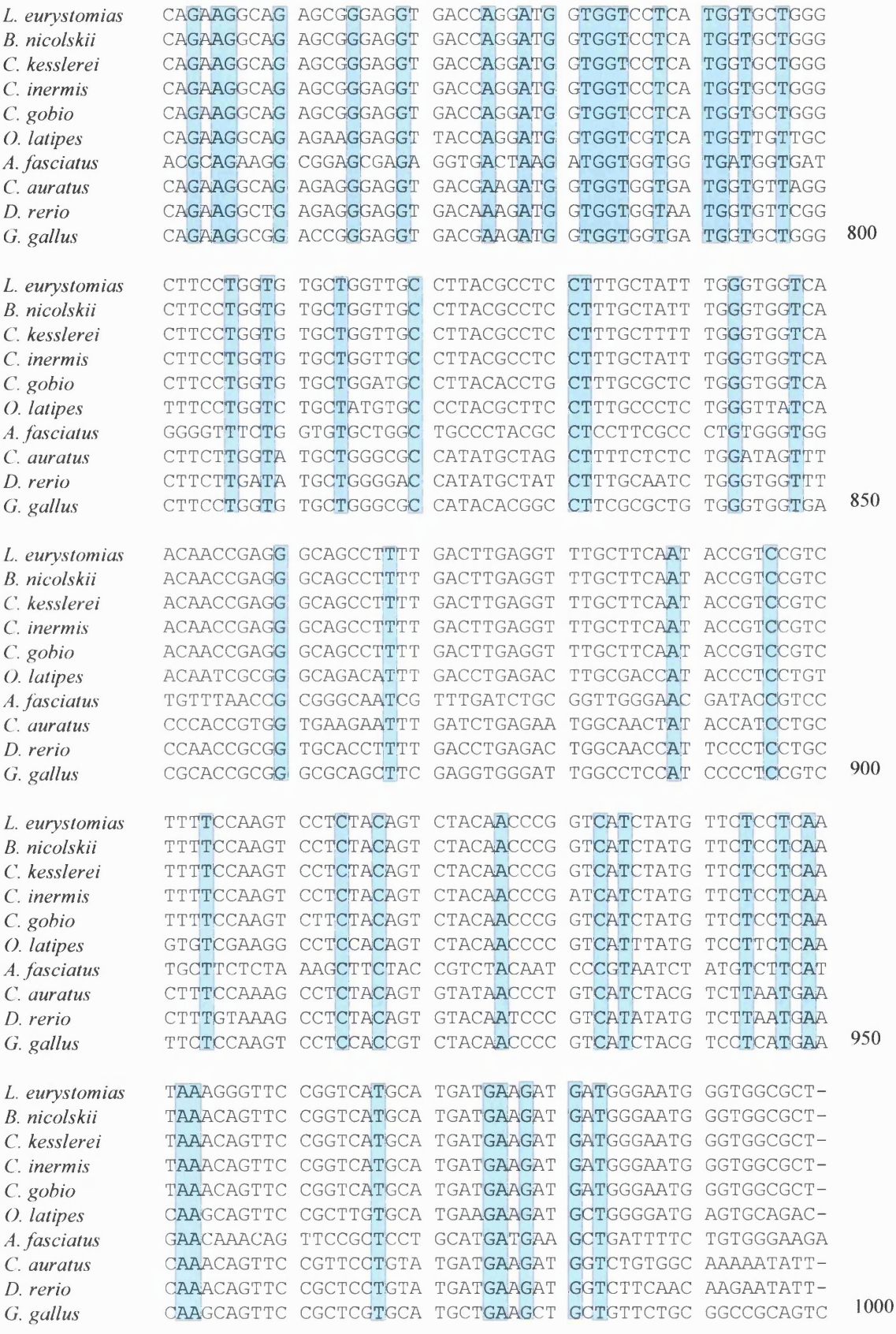
Continued.....


L. eurystomias TGAACTTGGC CCTGTCAAAC CTACTTGTGT CCAGCGTGGG CTCCTTCACT
B. nicolskii TGAACTTGGC CCTGTCAAAC CTACTTGTGT CCGGCGTGGG CTCCTTCACT
C. kesslerei TGAACTTGGC CCTGTCAAAC CTACTTGTGT CCGGCGTGGG CTCCTTCACT
C. inermis TGAACTTGGC CCTGTCAAAC CTACTTGTGT CCGGCGTGGG CTCCTTCACT
C. gobio TGAACTTGGC CCTGTCAAAC CTTCTTGTGT CCGGCGTGGG CTCCTTCACT
O. latipes TCAACATGGC GGTGGCAAAC CTCATTGTTG CTTCTACGGG TTCCTCCACC
A. fasciatus TGAACCTGGC CATCTCCAAC CTGCTGGTCT CCACCGTCGG GTCCTTCACC
C. auratus TGAACCTTTC CATTGCCAAC CTGTTCGTGG CCATTTTTGG TTCCCCGTTA
D. rerio TGAACCTTGC CATTTCCAAC TTGTGGGTGT CCGTTTTCGG TTCCTCGGTA
G. gallus TTAATTTGGC CTTGGCCAAC CTGCTGGTGA TCCTGGTGGG CTCCACCACG 300
L. eurystomias GCCTTCTTCA GCTTTGCAAA CAAATATTTC ATCCTCGGAC CACTAGCATG
B. nicolskii GCCTTCTGCA GCTTTGCAAA CAAATATTTC ATCCTCGGAC CACTAGCATG
C. kesslerei GCCTTCTGCA GCTTTGCAAA CAGATATTTC ATCCTCGGAC CACTAGCATG
C. inermis GCCTTCTGCA GCTTTGCAAA CAGATATTTC ATCCTCGGAC CACTAGCATG
C. gobio GCCTTCTGCT CCTTTGCAAA GAAATATTTC ATCCTCGGAC CACTAGCATG
O. latipes TGTTTCGTCT GCTTTGCCTT CAAATACATG GTTCTGGGTC CGCTGGGCTG
A. fasciatus GCCTTCGTCT CCTTCCTCAA CCGATACTTT ATTTTTGGAC CCACAGCTTG
C. auratus TCATTCTACT CCTTCTTTAA TAGGTACTTT ATCTTTGGGG CTACAGCATG
D. rerio GCGTTCTACG CCTTTTACAA AAAGTACTTT GTCTTTGGGC CGATAGGATG
G. gallus GCCTGTTACA GCTTCTCGCA GATCTACTTC GCCCTCGGAC CCACGGCGTG 350
L. eurystomias CAAGATAGAA GGTTTTTTAG CAGGACTTGG CGGTATGGTA AGCCTGTGGT
B. nicolskii CAAGATAGAA GGTTTTTTAG CAGCACTTGG GCCTATGGTA AGCCTGTGGT
C. kesslerei CAAGATAGAA GGTTTTGTAG CAACACTTGG GCCTATGGTA AGCCTGTGGT
C. inermis CAAGATAGAA GGTTTTGTAG CAACACTTGG CGGTATGGTA AGCCTGTGGT
C. gobio CAGGATAGAA GGTTTTTTAG CAACACTTGG CGGTATGGTA AGCCTGTGGT
O. latipes CAAGATTGAA GGCTTTACTG CAGCCCTTGG AGGCATGGTG AGCCTCTGGT
A. fasciatus TAAAATTGAT GGATTTGTTG CAACTTTAGG AGGTGGTATG GTGAGTCTGT
C. auratus TAAAATAGAG GGCTTCCTTG CAACGCTCGG AGGAATGGTG GGTTTGGTGG
D. rerio CAAAATCGAG GGCTTCACTT CAACAATTGG AGGAATGGTG AGTTTGTGGT
G. gallus CAAAATCGAG GGCTTCGCTG CCACGCTGGG AGGCATGGTG AGTCTGTGGT 400
L. eurystomias CTCTTTGTGT GATTGCTTTT GAAAGATGGC TGGTCATCTG CAAGCCACTT
B. nicolskii CTCTTTGTGT GATTGCTTTT GAAAGATGGC TGGTCATCTG CAAGCCACTT
C. kesslerei CTCTTTGTGT GATTGCTTTT GAAAGATGGC TGGTCATCTG CAAGCCACTT
C. inermis CTCTTTGTGT GATTGCTTTT GAAAGATGGC TGGTCATCTG CAAGCCACTT
C. gobio CTCTTTGTGT GATTGCTTTT GAAAGATGGC TGGTCATTTG CAAACCACTT
O. latipes CTCTTGCTGT AATTGCATTT GAACGGTGGC TGGTTATCTG CAAGCCGCTT
A. fasciatus GGTCTCTGTC GGTCGTGGCG TTTGAGAGGT GGCTGGTTAT CTGTAAACCT
C. auratus TCTCTTCTGT AGTGGCATTT GAAAGGTGGC TGGTCATTTG CAAACCCCTT
D. rerio CTCTTGCTGT GGTGGCGCTG GAAAGGTGGC TGGTCATTTG CAAACCCCTC
G. gallus CGTTGGCCGT GGTGGCGTTC GAGCGCTTCC TCGTCATCTG CAAACCTTTG 450
L. eurystomias GGTTCCTTTG TTTTCAAGCC CGACCATGCT ATGGCCTGCT GCGTGGCGAC
B. nicolskii GGTTCCTTTG TTTTCAAGCC CGACCATGCT ATGGCCTGCT GCGTGGCGAC
C. kesslerei GGTTCCTTTG TTTTCAAGGC CGACCATGCT ATGGCCTGCT GCGTGGTGAC
C. inermis GGTTCCTTCG TTTTCAAGGC CGACCATGCT ATGGCCTGCT GCGTGGTGAC
C. gobio GGTACCTTTG TTTTCAAGCC CGACCATGCT ATGGCCTGCT GCATGGTGAC
O. latipes GGAAACCTTG TCTTCAAGTC TGAGCATGCT TTGCTGTGCT GTGCGTTGAC
A. fasciatus GTGGGGAATT TCTCCTTTAA AGGAACTCAC GCTATAATCG GCTGTGCTCT
C. auratus GGGAACTTTA CCTTCAAGAC CCCTCATGCC ATAGCTGGCT GCATACTTCC
D. rerio GGGAACTTTA CCTTCAAGAC CCCTCATGCC ATTGCCGGCT GCATACTTCC
G. gallus GGCAACTTCA CGTTCCGCGG CAGCCACGCC GTGTTGGGCT GCGTGGCCAC 500

Continued.....


L. eurystomias CTGGGTGTTG GCACTGCTTG CCTCAGGTCC TCCTCTGTTT GGATGGAGCA
B. nicolskii CTGGGTGTTG GCACTGCTTG CCTCAGGTCC TCCTCTGTTT GGATGGAGCA
C. kesslerei CTGGGTGTTG GCACTGTTTG CCTCATGTCC TCCTCTGTTT GGATGGAGCA
C. inermis CTGGGTCTTG GCAATGTTTG CCGCATGTCC TCCTCTGCTT GGATGGAGGA
C. gobio CTGGGTGTTG GCACTGCTTG CCTCAGGTCC TCCTCTGTTT GGATGGAGCA
O. latipes TTGGGTTTGT GGATTGTGTG CTTCAGTTCC ACCTCTGGTG GGATGGAGTA
A. fasciatus CACCTGGTTT TTCGCTCTGC TCGCCTCCAC ACCTCCACTG TTCGGCTGGA
C. auratus TTGGATAAGT GCATTGGCAG CTTCACTCCC TCCACTGTTT GGCTGGAGCC
D. rerio TTGGTGTATG GCATTGGCTG CTGGACTCCC TCCACTCTTA GGCTGGAGCC
G. gallus GTGGGTGCTC GGATTCGTCG CCTCCGCGCC GCCGCTCTTC GGATGGAGTC 550
L. eurystomias GGTACATCCC AGAAGGCCTG CAGTGCTCCT GTAGTCCAGA CTGGTATACC
B. nicolskii GGTACATCCC AGAAGGCCTG CAGTGCTCCT GTAGTCCAGA CTGGTATACC
C. kesslerei GGTACATCCC AGAAGGCCTG CAGTGCTCCT GTGGTCCAGA CTGGTATACC
C. inermis GGTACATCCC AGAAAGCCTG CAGTGCTCCT GTGGTCCAGA CTGGTATACC
C. gobio GGTACATCCC AGAAGGCCTG CAGTGCTCCT GTGGTCCAGA CTGGTATACC
O. latipes GGTACATCCC AGAGGGCATG CAGTGTTCGT GTGGACCAGA CTGGTACACA
A. fasciatus GCAGGTACAT CCCTGAAGGT CTGCAGTGTT CCTGTGGTCC AGACTGGTAC
C. auratus GGTACATACC TGAAGGTTTG CAGTGCTCCT GTGGACCTGA CTGGTATACC
D. rerio GGTATATACC AGAGGGCTTG CAGTGCTCTT GTGGACCTGA CTGGTATACG
G. gallus GGTACATCCC GGAGGGGCTG CAGTGCTCGT GCGGGCCGGA CTGGTACACG 600
L. eurystomias ACCAACAACA AATACAACAA TGAATCCTAC GTGATATATC TCTTCACCTG
B. nicolskii ACCAACAACA AATACAACAA TGAATCCTAC GTGATATATC TCTTCACCTG
C. kesslerei ACCAACAACA AATACAACAA TGAATCCTAC GTGATATATC TCTTCACCTG
C. inermis AGCAACAACA AATACAACAA TGAATCCTAC GTGATATATC TCTTCACCTG
C. gobio ACCAACAACA AATACAACAA TGAATCCTAC GTGATATATC TCTTCACCTG
O. latipes ACCGGCAACA AGTTTAACAA CGAGTCTTTT GTGATGTTCC TCTTTTGCTT
A. fasciatus ACCACTGAAA ACAAATACAA CAACGAGTCT TATGTCATGT TCCTCTTCTG
C. auratus ACTAACAACA AATACAACAA TGAATCCTAC GTCATGTTTT TGTTCTGCTT
D. rerio ACTAACAACA AATTCAACAA TGAATCCTAC GTCATGTTCC TCTTCTGCTT
G. gallus ACGGACAACA AGTGGCACAA CGAGAGCTAC GTGCTCTTCC TGTTCACCTT 650
L. eurystomias CCACTTTTCT GTTGGCCTCT TAATCCTTGT CTTTTGCTAC GCCCAACTCC
B. nicolskii CCACTTTTCT GTTGGCCTCT TAATCCTTGT CTTTTGCTAC GCCCAACTCC
C. kesslerei CCACTTTTCT GTTGGCCTCT TAATCCTTGT CTTTTGCTAC GCCCAACTCC
C. inermis CCACTTTTCT GTTGGCCTTG TAATCCTTGT CTTTTGCTAC GCCCAACTCC
C. gobio CCACTTTACT GTTCCCCTCT TAATCCTTGT CTTTTGCTAT GCCCAACTCC
O. latipes CTGCTTTGCC GTCCCTTTCT CCATCATTGT CTTCTGTTAT TCTCAGCTGC
A. fasciatus CTTCTGCTTC GGATTCCCGT TTACTGTTAT CCTTTTCTGC TACGGCCAAC
C. auratus CTGCTTTGCT GTTCCTTTCG GCACCATCGT GTTCTGTTAT GGCCAACTAC
D. rerio TTGCTTTGCG GTTCCTTTCA GCACCATTGT ATTCTGTTAT GGTCAGCTGC
G. gallus CTGCTTCGGG GTCCCTCTCG CCATCATCGT CTTCTCCTAC GGCCGCCTCC 700
L. eurystomias TTTTTACACT GAAAATGGCG GCAAAGGCCC AAGCAGAGTC TGCCTCCACC
B. nicolskii TTTTTACACT GAAAATGGCG GCAAAGGCCC AAGCAGAGTC TGCCTCCACC
C. kesslerei TTTTTACACT GAAAATGGCG GCAAAGGCCC AAGCAGAGTC TGCCTCCACC
C. inermis TTTTTACACT GAAAATGGCG GCAAAGGCCC AAGCAGAGTC TGCCTCCACC
C. gobio TTTTTACACT GAAAATGGCG GCAAAGGCCC AAGCAGAATC TGCCTCCACC
O. latipes TTTTCACTCT GAAAATGGCA GCAAAGGCCC AGGCTGACTC CGCCTCCACT
A. fasciatus TGCTCTTCAC TCTCAAATCA GCGGCCAAAG CTCAGGCTGA CTCCGCCTCC
C. auratus TCATCACACT CAAATTAGCA GCAAAAGCTC AAGCAGATTC AGCTTCGACC
D. rerio TCATCACTCT CAAATTAGCA GCCAAAGCTC AAGCTGATTC AGCCTCAACC
G. gallus TCATCACACT CAGAGCGGTT GCCCGGCAGC AGGAGCAGTC AGCGACGACG 750

Continued.....


L. eurystomias CAGAAGGCAG AGCGGGAGGT GACCAGGATG GTGGTCCTCA TGGTGCTGGG
B. nicolskii CAGAAGGCAG AGCGGGAGGT GACCAGGATG GTGGTCCTCA TGGTGCTGGG
C. kesslerei CAGAAGGCAG AGCGGGAGGT GACCAGGATG GTGGTCCTCA TGGTGCTGGG
C. inermis CAGAAGGCAG AGCGGGAGGT GACCAGGATG GTGGTCCTCA TGGTGCTGGG
C. gobio CAGAAGGCAG AGCGGGAGGT GACCAGGATG GTGGTCCTCA TGGTGCTGGG
O. latipes CAGAAGGCAG AGAAGGAGGT TACCAGGATG GTGGTCGTCA TGGTTGTTGC
A. fasciatus ACGCAGAAGG CGGAGCGAGA GGTGACTAAG ATGGTGGTGG TGATGGTGAT
C. auratus CAGAAGGCAG AGAGGGAGGT GACGAAGATG GTGGTGGTGA TGGTGTTAGG
D. rerio CAGAAGGCTG AGAGGGAGGT GACAAAGATG GTGGTGGTAA TGGTGTTCGG
G. gallus CAGAAGGCGG ACCGGGAGGT GACGAAGATG GTGGTGGTGA TGGTGCTGGG 800
L. eurystomias CTTCCTGGTG TGCTGGTTGC CTTACGCCTC CTTTGCTATT TGGGTGGTCA
B. nicolskii CTTCCTGGTG TGCTGGTTGC CTTACGCCTC CTTTGCTATT TGGGTGGTCA
C. kesslerei CTTCCTGGTG TGCTGGTTGC CTTACGCCTC CTTTGCTTTT TGGGTGGTCA
C. inermis CTTCCTGGTG TGCTGGTTGC CTTACGCCTC CTTTGCTATT TGGGTGGTCA
C. gobio CTTCCTGGTG TGCTGGATGC CTTACACCTG CTTTGCGCTC TGGGTGGTCA
O. latipes TTTCCTGGTC TGCTATGTGC CCTACGCTTC CTTTGCCCTC TGGGTTATCA
A. fasciatus GGGGTTTCTG GTGTGCTGGC TGCCCTACGC CTCCTTCGCC CTGTGGGTGG
C. auratus CTTCTTGGTA TGCTGGGCGC CATATGCTAG CTTTTCTCTC TGGATAGTTT
D. rerio CTTCTTGATA TGCTGGGGAC CATATGCTAT CTTTGCAATC TGGGTGGTTT
G. gallus CTTCCTGGTG TGCTGGGCGC CATACACGGC CTTCGCGCTG TGGGTGGTGA 850
L. eurystomias ACAACCGAGG GCAGCCTTTT GACTTGAGGT TTGCTTCAAT ACCGTCCGTC
B. nicolskii ACAACCGAGG GCAGCCTTTT GACTTGAGGT TTGCTTCAAT ACCGTCCGTC
C. kesslerei ACAACCGAGG GCAGCCTTTT GACTTGAGGT TTGCTTCAAT ACCGTCCGTC
C. inermis ACAACCGAGG GCAGCCTTTT GACTTGAGGT TTGCTTCAAT ACCGTCCGTC
C. gobio ACAACCGAGG GCAGCCTTTT GACTTGAGGT TTGCTTCAAT ACCGTCCGTC
O. latipes ACAATCGCGG GCAGACATTT GACCTGAGAC TTGCGACCAT ACCCTCCTGT
A. fasciatus TGTTTAACCG CGGGCAATCG TTTGATCTGC GGTTGGGAAC GATACCGTCC
C. auratus CCCACCGTGG TGAAGAATTT GATCTGAGAA TGGCAACTAT ACCATCCTGC
D. rerio CCAACCGCGG TGCACCTTTT GACCTGAGAC TGGCAACCAT TCCCTCCTGC
G. gallus CGCACCGCGG GCGCAGCTTC GAGGTGGGAT TGGCCTCCAT CCCCTCCGTC 900
L. eurystomias TTTTCCAAGT CCTCTACAGT CTACAACCCG GTCATCTATG TTCTCCTCAA
B. nicolskii TTTTCCAAGT CCTCTACAGT CTACAACCCG GTCATCTATG TTCTCCTCAA
C. kesslerei TTTTCCAAGT CCTCTACAGT CTACAACCCG GTCATCTATG TTCTCCTCAA
C. inermis TTTTCCAAGT CCTCTACAGT CTACAACCCG ATCATCTATG TTCTCCTCAA
C. gobio TTTTCCAAGT CTTCTACAGT CTACAACCCG GTCATCTATG TTCTCCTCAA
O. latipes GTGTCGAAGG CCTCCACAGT CTACAACCCC GTCATTTATG TCCTTCTCAA
A. fasciatus TGCTTCTCTA AAGCTTCTAC CGTCTACAAT CCCGTAATCT ATGTCTTCAT
C. auratus CTTTCCAAAG CCTCTACAGT GTATAACCCT GTCATCTACG TCTTAATGAA
D. rerio CTTTGTAAAG CCTCTACAGT GTACAATCCC GTCATATATG TCTTAATGAA
G. gallus TTCTCCAAGT CCTCCACCGT CTACAACCCC GTCATCTACG TCCTCATGAA 950
L. eurystomias TAAAGGGTTC CGGTCATGCA TGATGAAGAT GATGGGAATG GGTGGCGCT-
B. nicolskii TAAACAGTTC CGGTCATGCA TGATGAAGAT GATGGGAATG GGTGGCGCT-
C. kesslerei TAAACAGTTC CGGTCATGCA TGATGAAGAT GATGGGAATG GGTGGCGCT-
C. inermis TAAACAGTTC CGGTCATGCA TGATGAAGAT GATGGGAATG GGTGGCGCT-
C. gobio TAAACAGTTC CGGTCATGCA TGATGAAGAT GATGGGAATG GGTGGCGCT-
O. latipes CAAGCAGTTC CGCTTGTGCA TGAAGAAGAT GCTGGGGATG AGTGCAGAC-
A. fasciatus GAACAAACAG TTCCGCTCCT GCATGATGAA GCTGATTTTC TGTGGGAAGA
C. auratus CAAACAGTTC CGTTCCTGTA TGATGAAGAT GGTCTGTGGC AAAAATATT-
D. rerio CAAACAGTTC CGCTCCTGTA TGATGAAGAT GGTCTTCAAC AAGAATATT-
G. gallus CAAGCAGTTC CGCTCGTGCA TGCTGAAGCT GCTGTTCTGC GGCCGCAGTC 1000

Continued.....


L. eurystomias ---------- -GATGATGAA GAGTCGTCAA CA---TCATC AGTGACCGAA
B. nicolskii ---------- -GATGATGAA GAGTCGTCAA CA---TCATC AGTGACCGAA
C. kesslerei ---------- -GATGATGAA GAGTCGTCAA CA---TCATC AGTGACCGAA
C. inermis ---------- -GATGATGAA GAGTCGTCAA CA---TCATC AGTGACCGAA
C. gobio ---------- -GATGATGAA GAGTCGTCAA CA---TCATC AGTGACCGAA
O. latipes ---------- -GAAGATGAA GAGTCCTCAA CCAGTCAATC AACCACTGAA
A. fasciatus GTCCGTTTGG AGATGATGAA GAAGCCTCCT CCTCCTCTCA GGTGACCCAG
C. auratus ---------- -GAGGAAGAT GAGGCCTCTA CTTCATCTCA GGTCACCCAG
D. rerio ---------- -GAGGAAGAT GAGGCTTCTT CTTCATCTCA GGTCACCCAG
G. gallus CGTTCGGGGA CGACGAGGAC GTGTCGGGCT CCTCGCAGGC CACGCAGGTC
1050
L. eurystomias GTCTCCAAAG TTGGGCCTGC T
B. nicolskii GTCTCCAAAG TTGGGCCTGC T
C. kesslerei GTCTCCAAAG TTGGGCCTGC T
C. inermis GTCTCCAAAG TTGGGCCTGC T
C. gobio GTCTCCAAAG TTGGGCCTGC T
O. latipes GTCTCAAAAG TCGGCCCTTC C
A. fasciatus GTGTCTTCTG TAGGACCGGA GAAA
C. auratus GTCTCCTCCG TTGCACCAGA GAAA
D. rerio GTCTCCTCTG TTGCGCCAGA GAAA
G. gallus TCCTCCGTCT CCTCCAGCCA CGTCGCCCCC GCC
1083

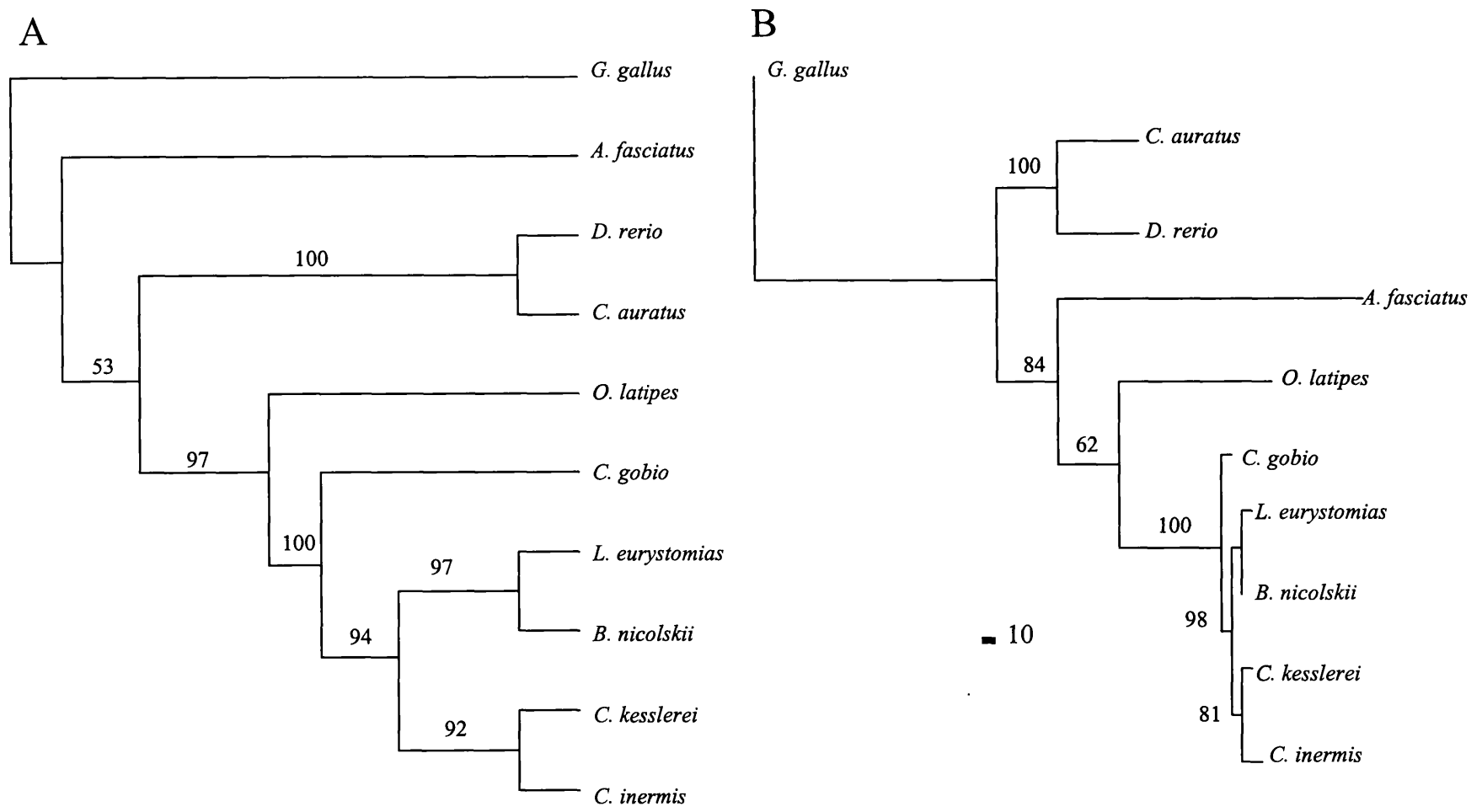


Figure 5.7: Phylogenetic trees based on the total nucleotide sequence of the blue opsin gene. A) Maximum parsimony. B) Neighbour joining – the length of each branch represents the evolutionary distance between species. The bar indicates a frequency of 10 nucleotide substitutions. The bootstrap confidence values based on 1000 replicates are shown on each branch.

A Pairwise comparisons between the blue cone opsin genes of fishes

	Bn	Ck	Ci	Cg	Ol	Af	Ca	Dr
Le	0.012	0.012	0.024	0.040	1.176	1.298	1.717	1.868
Bn		0.008	0.028	0.044	1.155	1.277	1.710	1.961
Ck			0.020	0.036	1.178	1.319	1.696	1.977
Ci				0.048	1.210	1.386	1.786	2.109
Cg					1.199	1.267	1.599	1.798
Ol						1.587	2.502	1.956
Af							1.727	1.838
Ca								0.447

B Divergence of cottoids from goldfish

Limnocottus eurystomias	1.717
Batrachocottus nicolskii	1.710
Cottus kesslerei	1.700
Cottocomephorus inermis	1.786
Cottus gobio	1.599
Mean ± SEM	1.703 ± 0.030

C Divergence of Lake Baikal cottoids from *Cottus gobio*

Limnocottus eurystomias	0.040
Batrachocottus nicolskii	0.044
Cottus kesslerei	0.036
Cottocomephorus inermis	0.048
Mean ± SEM	0.042 ± 0.003

D Divergence of *L.eurystomias* and *B. nicolskii* from *C. kesslerei*

Limnocottus eurystomias	0.012
Batrachocottus nicolskii	0.008

Divergence of *L. eurystomias* and *B. nicolskii* from *C. inermis*

Limnocottus eurystomias	0.024
Batrachocottus nicolskii	0.028

Mean ± SEM	0.018 ± 0.005

Table 5.1: Estimation of divergence at synonymous sites between the blue cone opsin genes of fishes. Substitutions per synonymous site were corrected for multiple replacements by the Jukes and Cantor method (Jukes & Cantor, 1969).

A Pairwise comparisons between the blue cone opsin genes of fishes

	Bn	Ck	Ci	Cg	Ol	Af	Ca	Dr
Le	0.010	0.020	0.030	0.034	0.349	0.880	0.425	0.431
Bn		0.012	0.028	0.032	0.348	0.883	0.425	0.431
Ck			0.019	0.032	0.348	0.883	0.417	0.429
Ci				0.043	0.352	0.880	0.427	0.434
Cg					0.345	0.877	0.399	0.409
Ol						1.027	0.460	0.449
Af							0.922	0.949
Ca								0.183

B Divergence of cottoids from goldfish

Limnocottus eurystomias	0.425
Batrachocottus nicolskii	0.425
Cottus kesslerei	0.417
Cottocomephorus inermis	0.427
Cottus gobio	0.399
Mean ± SEM	0.419 ± 0.012

C Divergence of Lake Baikal cottoids from *Cottus gobio*

Limnocottus eurystomias	0.034
Batrachocottus nicolskii	0.032
Cottus kesslerei	0.032
Cottocomephorus inermis	0.043
Mean ± SEM	0.032 ± 0.005

D Divergence of *L.eurystomias* and *B. nicolskii* from *C. kesslerei*

Limnocottus eurystomias	0.020
Batrachocottus nicolskii	0.012

Divergence of *L. eurystomias* and *B. nicolskii* from *C. inermis*

Limnocottus eurystomias	0.030
Batrachocottus nicolskii	0.028

Mean ± SEM	0.023 ± 0.008

Table 5.2: Estimation of divergence at all nucleotide sites, between the blue cone opsin genes of fishes. The rate of substitution was corrected for multiple replacements by the Jukes and Cantor method (Jukes & Cantor, 1969).

	Pk	Pj	Cg	Ci	Bm	Bn	Lb	Lp	Ak	Cb	Cd	Ca
Ck	0.056	0.051	0.044	0.044	0.034	0.054	0.049	0.051	0.049	0.075	0.044	**0.672**
Pk		0.034	0.027	0.027	0.019	0.029	0.031	0.031	0.032	0.049	0.027	**0.703**
Pj			0.024	0.024	0.014	0.014	0.017	0.027	0.017	0.039	0.012	**0.688**
Cg				0.000	0.005	0.022	0.015	0.027	0.015	0.045	0.012	**0.690**
Ci					0.005	0.024	0.015	0.027	0.015	0.045	0.012	**0.698**
Bm						0.019	0.010	0.022	0.010	0.039	0.005	**0.665**
Bn							0.019	0.032	0.019	0.044	0.015	**0.667**
Lb								0.022	0.010	0.039	0.005	**0.690**
Lp									0.022	0.047	0.017	**0.701**
Ak										0.039	0.005	**0.695**
Cb											0.034	**0.678**
Cd												**0.685**
										Mean ± SEM		**0.686** ± 0.013

Table 5.3: Substitutions per synonymous site between the rod opsins of Lake Baikal cottoid fishes corrected for multiple substitutions by the Jukes and Cantor method (Jukes & Cantor, 1969). Ck, *Cottus kesslerei*: Pk, *Paracottus kneri*; Pj, *Procottus jettelesi*; Cg, *Cottus gobio*; Ci, *Cottocomephorus inermis*; Bm, *Batrachocottus multiradiatus*; Bn, *Batrachocottus nicolskii*; Lb, *Limnocottus bergianus*; Lp, *Limnocottus pallidus*; Ak, *Abyssocottus korotneffi*; Cb, *Cotinella boulengerei*; Cd, *Comephorus dybowskii*; Ca, *Carassius auratus*. Sequences used for the pairwise analysis were those from Hunt *et al.* 1997. Figures in bold were used to calculate the mean.

	Pk	Pj	Cg	Ci	Bm	Bn	Lb	Lp	Ak	Cb	Cd	Ca
Ck	0.034	0.040	0.048	0.049	0.037	0.042	0.048	0.042	0.045	0.053	0.049	**0.230**
Pk		0.018	0.035	0.039	0.026	0.028	0.039	0.033	0.034	0.039	0.029	**0.226**
Pj			0.029	0.033	0.020	0.025	0.031	0.033	0.026	0.032	0.019	**0.221**
Cg				0.018	0.028	0.038	0.037	0.044	0.032	0.042	0.034	**0.223**
Ci					0.029	0.039	0.038	0.045	0.028	0.040	0.035	**0.215**
Bm						0.014	0.021	0.025	0.016	0.022	0.025	**0.220**
Bn							0.026	0.026	0.021	0.023	0.031	**0.226**
Lb								0.031	0.019	0.029	0.028	**0.229**
Lp									0.028	0.033	0.033	**0.240**
Ak										0.020	0.025	**0.223**
Cb											0.034	**0.227**
Cd												**0.230**
										Mean ± SEM		**0.226**
												± 0.006

Table 5.4: Total nucleotide substitutions between the rod opsins of lake Baikal Cottoid fishes, corrected for multiple replacements by the Jukes and Cantor method (Jukes & Cantor, 1969). Ck, *Cottus kesslerei*; Pk, *Paracottus kneri*; Pj, *Procottus jettelesi*; Cg, *Cottus gobio*; Ci, *Cottocomephorus inermis*; Bm, *Batrachocottus multiradiatus*; Bn, *Batrachocottus nicolskii*; Lb, *Limnocottus bergianus*; Lp, *Limnocottus pallidus*; Ak, *Abyssocottus korotneffi*; Cb, *Cotinella boulengerei*; Cd, *Comephorus dybowskii*; Ca, *Carassius auratus*. Sequences used for the pairwise analysis were those from Hunt *et al.* 1997. Figures in bold were used to calculate the mean.

A

	B.rod	Ch.rod	Ch.r	D.g	B.g	Ch.g	Ch.b	Ch.v	B.uv	C.uv
D.rod	**0.254**	**0.291**	0.981	0.923	1.004	1.028	1.135	1.008	0.856	0.883
B.rod		**0.311**	1.060	0.897	1.012	1.077	1.210	1.257	1.063	0.965
Ch.rod			0.998	0.909	1.028	1.115	1.147	0.858	0.771	0.817
Ch.r				0.805	0.945	0.970	0.671	0.705	0.741	0.584
D.g					**0.166**	**0.268**	0.734	0.676	0.630	0.614
B.g						**0.318**	0.837	0.877	0.715	0.679
Ch.g							0.978	0.832	0.846	0.767
Ch.b								0.639	0.778	0.593
Ch.v									**0.593**	**0.548**
B.uv										**0.312**

Opsin divergence

Rod mean ± SEM	0.285 ± 0.029
Green mean ± SEM	0.251 ± 0.078
V/UV mean ± SEM	0.485 ± 0.151

B

	M.rod	H.r	H.g	M.g	H.b	M.b
H.rod	**0.563**	1.133	1.159	1.522	1.250	1.238
M.rod		1.265	1.210	1.606	1.026	1.354
H.r			0.046	0.467	1.373	1.655
H.g				**0.461**	1.332	1.603
M.g					1.429	1.777
H.b						**0.422**

Rod	0.563
Green	0.461
Blue	0.422

Table 5.5: Substitutions per synonymous site corrected for multiple replacements by the Jukes and Cantor method (Jukes & Cantor, 1969). A) between avian opsins. D.rod, duck rod; B.rod, budgerigar rod; Ch.rod, chicken rod; Ch.r, chicken red; D.g, duck green; B.g, budgerigar green; Ch.g, chicken green; Ch.b, chicken blue; Ch.v, chicken violet; B.uv, budgerigar UV; C.uv, canary UV. B) between mammalian opsins. H.rod, human rod; M.rod, mouse rod; H.r, human red; H.g, human green; M.g, mouse green; H.b, human blue; M.b, mouse blue. Figures used to calculate opsin divergences are shown in bold.

A

	B.rod	Ch.rod	Ch.r	D.g	B.g	Ch.g	Ch.b	Ch.v	B.uv	C.uv
D.rod	**0.069**	**0.075**	0.731	0.318	0.313	0.319	0.575	0.630	0.595	0.577
B.rod		**0.088**	0.743	0.323	0.322	0.331	0.613	0.688	0.647	0.605
Ch.rod			0.733	0.327	0.321	0.334	0.610	0.625	0.599	0.575
Ch.r				0.659	0.698	0.696	0.664	0.667	0.661	0.671
D.g					**0.055**	**0.069**	0.505	0.528	0.506	0.510
B.g						**0.083**	0.566	0.620	0.571	0.549
Ch.g							0.583	0.609	0.587	0.563
Ch.b								0.515	0.537	0.468
Ch.v									**0.192**	**0.191**
B.uv										**0.134**

Opsin divergence

Rod mean ± SEM	0.077 ± 0.009
Green mean ± SEM	0.069 ± 0.014
V/UV mean ± SEM	0.172 ± 0.033

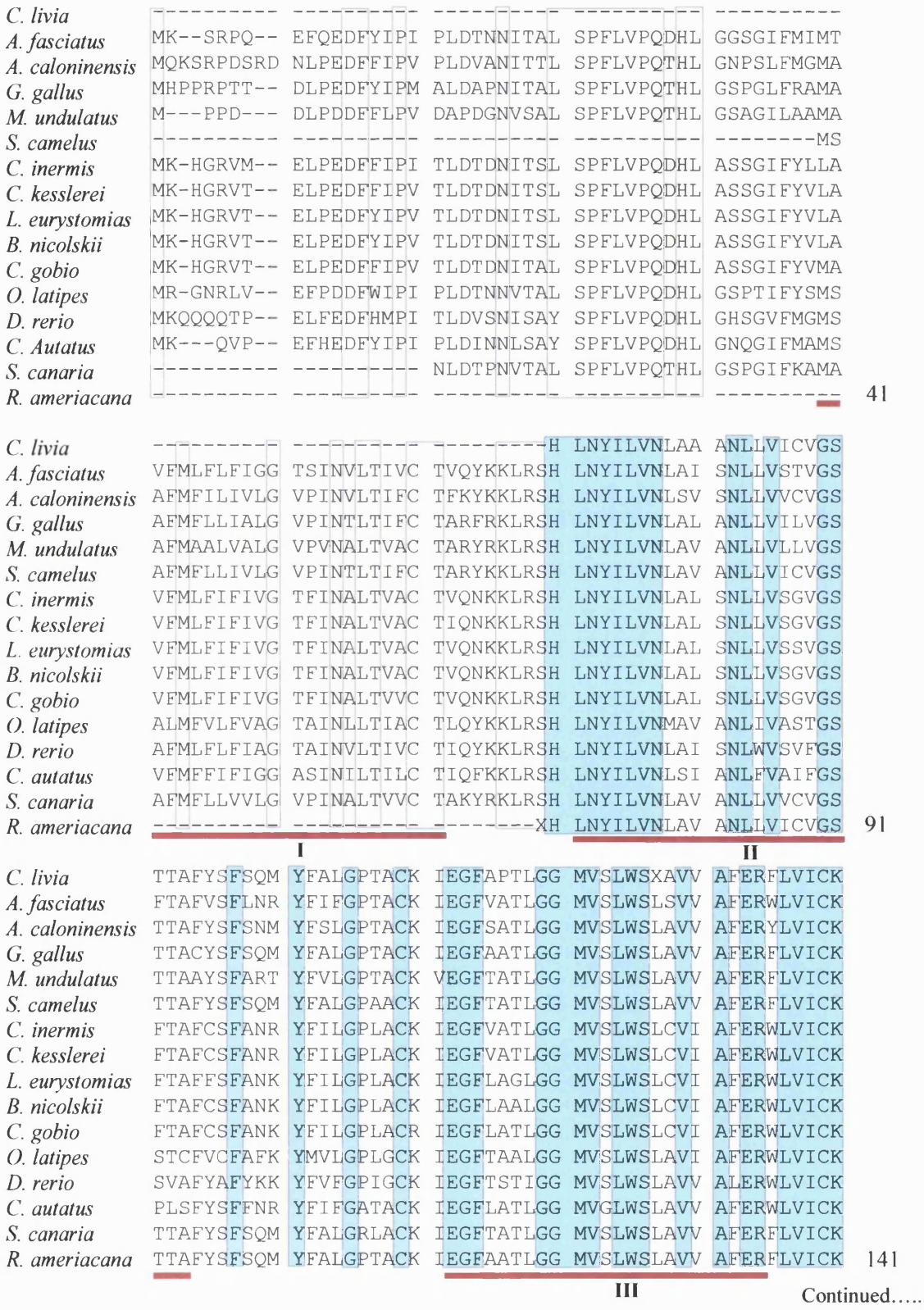
B

	M.rod	H.r	H.g	M.g	H.b	M.b
H.rod	**0.132**	0.690	0.673	0.728	0.633	0.628
M.rod		0.713	0.690	0.745	0.598	0.624
H.r			0.029	0.145	0.723	0.731
H.g				**0.148**	0.711	0.718
M.g					0.727	0.749
H.b						**0.149**

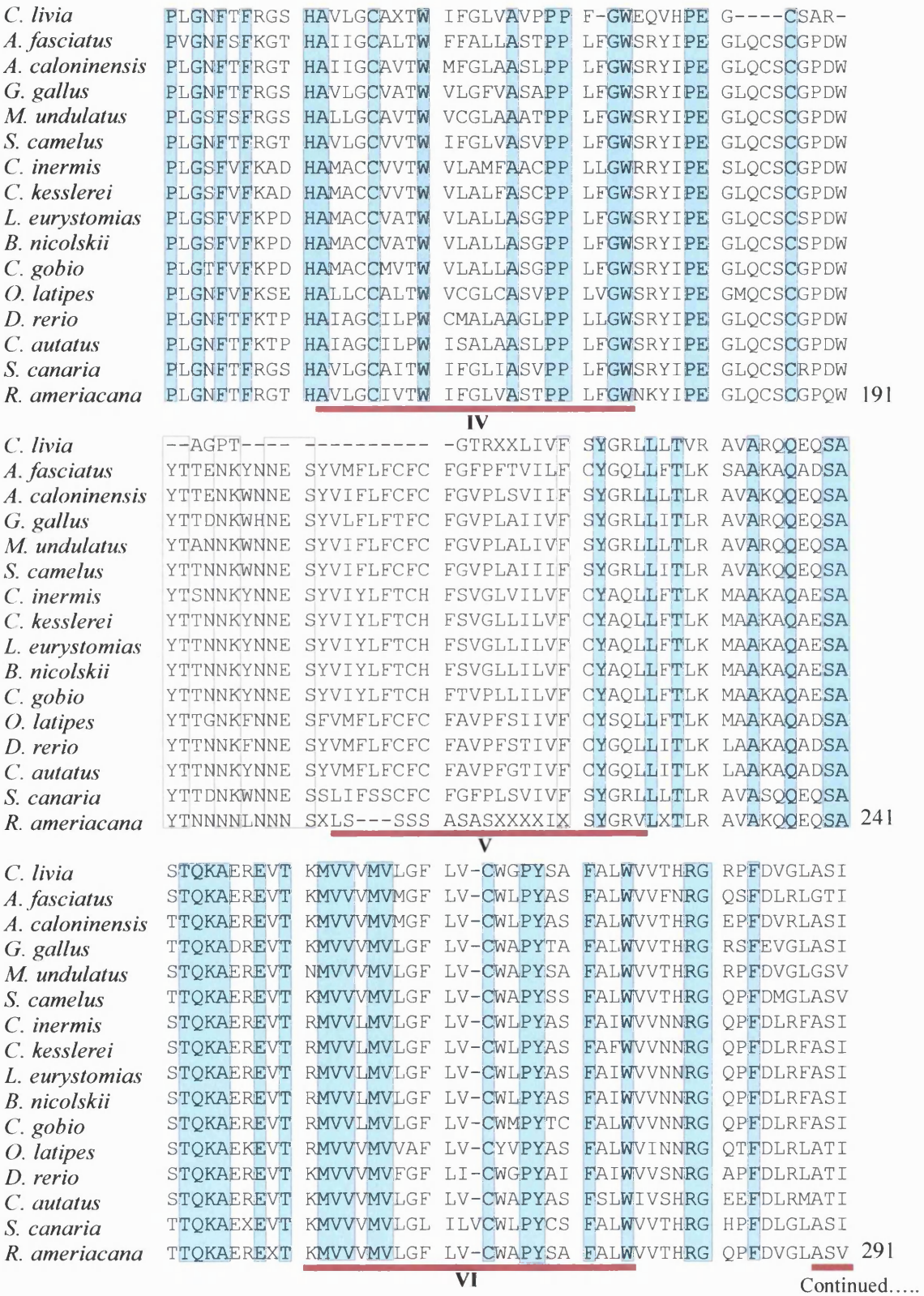
Rod	0.132
Green	0.148
Blue	0.149

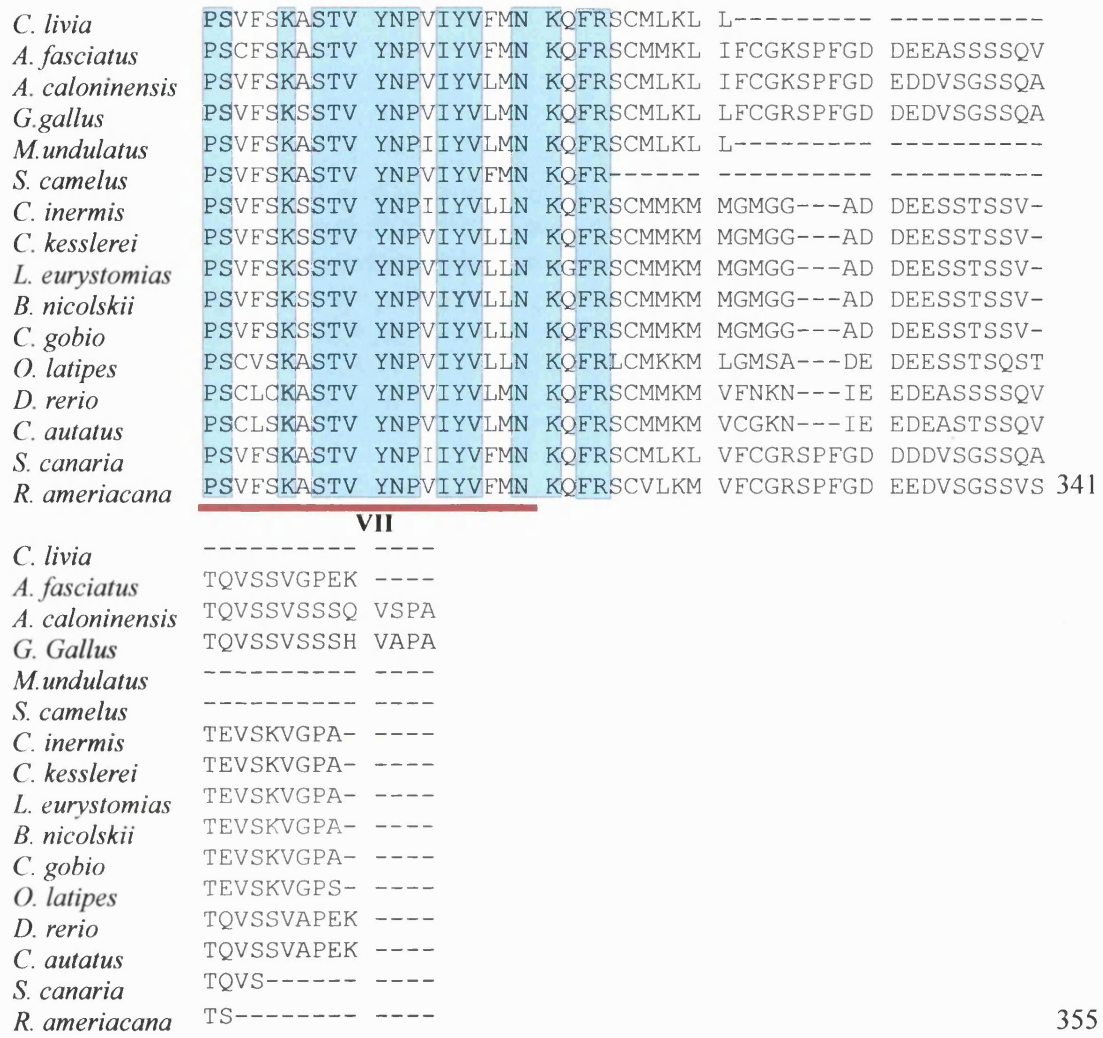
Table 5.6: Total nucleotide substitutions corrected for multiple replacements by the Jukes and Cantor method (Jukes & Cantor, 1969). A) between avian opsins. D.rod, duck rod; B.rod, budgerigar rod; Ch.rod, chicken rod; Ch.r, chicken red; D.g, duck green; B.g, budgerigar green; Ch.g, chicken green; Ch.b, chicken blue; Ch.v, chicken violet; B.uv, budgerigar UV; C.uv, canary UV. B) between mammalian opsins. H.rod, human rod; M.rod, mouse rod; H.r, human red; H.g, human green; M.g, mouse green; H.b, human blue; M.b, mouse blue. Figures used to calculate opsin divergences are shown in bold.

Figure 5.8: Aligned amino acid sequences of the blue cone opsin genes of pigeon (*C. livia*), chameleon (*A. carolonensis*), chicken (*G. gallus*), budgerigar (*M. undulates*), ostrich (*S. camelus*), Baikal bullheads (*C. inermis, C. kesslerei, L. eurystomias* and *B. nicolskii*), bullhead (*C. gobio*), killifish (*O. latipes*), zebrafish (*D. rerio*), goldfish (*C. auratus*), canary (*S. canaria*), rhea (*R. Americana*). Identical sequences are boxed. The positions of the helices are marked with red bars. The numbers to the right relate to the amino acid numbering system of bovine rhodopsin (Nathans *et al.* 1986).


C. livia ---------- ---------- ---------- ---------- ----------
A. fasciatus MK--SRPQ-- EFQEDFYIPI PLDTNNITAL SPFLVPQDHL GGSGIFMIMT
A. caloninensis MQKSRPDSRD NLPEDFFIPV PLDVANITTL SPFLVPQTHL GNPSLFMGMA
G. gallus MHPPRPTT-- DLPEDFYIPM ALDAPNITAL SPFLVPQTHL GSPGLFRAMA
M. undulatus M---PPD--- DLPDDFFLPV DAPDGNVSAL SPFLVPQTHL GSAGILAAMA
S. camelus ---------- ---------- ---------- ---------- --------MS
C. inermis MK-HGRVM-- ELPEDFFIPI TLDTDNITSL SPFLVPQDHL ASSGIFYLLA
C. kesslerei MK-HGRVT-- ELPEDFFIPV TLDTDNITSL SPFLVPQDHL ASSGIFYVLA
L. eurystomias MK-HGRVT-- ELPEDFYIPV TLDTDNITSL SPFLVPQDHL ASSGIFYVLA
B. nicolskii MK-HGRVT-- ELPEDFYIPV TLDTDNITSL SPFLVPQDHL ASSGIFYVLA
C. gobio MK-HGRVT-- ELPEDFFIPV TLDTDNITAL SPFLVPQDHL ASSGIFYVMA
O. latipes MR-GNRLV-- EFPDDFWIPI PLDTNNVTAL SPFLVPQDHL GSPTIFYSMS
D. rerio MKQQQQTP-- ELFEDFHMPI TLDVSNISAY SPFLVPQDHL GHSGVFMGMS
C. Autatus MK---QVP-- EFHEDFYIPI PLDINNLSAY SPFLVPQDHL GNQGIFMAMS
S. canaria ---------- ---------- NLDTPNVTAL SPFLVPQTHL GSPGIFKAMA
R. ameriacana ---------- ---------- ---------- ---------- ---------- 41
C. livia ---------- ---------- ---------H LNYILVNLAA ANLLVICVGS
A. fasciatus VFMLFLFIGG TSINVLTIVC TVQYKKLRSH LNYILVNLAI SNLLVSTVGS
A. caloninensis AFMFILIVLG VPINVLTIFC TFKYKKLRSH LNYILVNLSV SNLLVVCVGS
G. gallus AFMFLLIALG VPINTLTIFC TARFRKLRSH LNYILVNLAL ANLLVILVGS
M. undulatus AFMAALVALG VPVNALTVAC TARYRKLRSH LNYILVNLAV ANLLVLLVGS
S. camelus AFMFLLIVLG VPINTLTIFC TARYKKLRSH LNYILVNLAV ANLLVICVGS
C. inermis VFMLFIFIVG TFINALTVAC TVQNKKLRSH LNYILVNLAL SNLLVSGVGS
C. kesslerei VFMLFIFIVG TFINALTVAC TIQNKKLRSH LNYILVNLAL SNLLVSGVGS
L. eurystomias VFMLFIFIVG TFINALTVAC TVQNKKLRSH LNYILVNLAL SNLLVSSVGS
B. nicolskii VFMLFIFIVG TFINALTVAC TVQNKKLRSH LNYILVNLAL SNLLVSGVGS
C. gobio VFMLFIFIVG TFINALTVVC TVQNKKLRSH LNYILVNLAL SNLLVSGVGS
O. latipes ALMFVLFVAG TAINLLTIAC TLQYKKLRSH LNYILVNMAV ANLIVASTGS
D. rerio AFMLFLFIAG TAINVLTIVC TIQYKKLRSH LNYILVNLAI SNLWVSVFGS
C. autatus VFMFFIFIGG ASINILTILC TIQFKKLRSH LNYILVNLSI ANLFVAIFGS
S. canaria AFMFLLVVLG VPINALTVVC TAKYRKLRSH LNYILVNLAV ANLLVVCVGS
R. ameriacana ---------- ---------- --------XH LNYILVNLAV ANLLVICVGS 91
I
II
C. livia TTAFYSFSQM YFALGPTACK IEGFAPTLGG MVSLWSXAVV AFERFLVICK
A. fasciatus FTAFVSFLNR YFIFGPTACK IEGFVATLGG MVSLWSLSVV AFERWLVICK
A. caloninensis TTAFYSFSNM YFSLGPTACK IEGFSATLGG MVSLWSLAVV AFERYLVICK
G. gallus TTACYSFSQM YFALGPTACK IEGFAATLGG MVSLWSLAVV AFERFLVICK
M. undulatus TTAAYSFART YFVLGPTACK VEGFTATLGG MVSLWSLAVV AFERFLVICK
S. camelus TTAFYSFSQM YFALGPAACK IEGFTATLGG MVSLWSLAVV AFERFLVICK
C. inermis FTAFCSFANR YFILGPLACK IEGFVATLGG MVSLWSLCVI AFERWLVICK
C. kesslerei FTAFCSFANR YFILGPLACK IEGFVATLGG MVSLWSLCVI AFERWLVICK
L. eurystomias FTAFFSFANK YFILGPLACK IEGFLAGLGG MVSLWSLCVI AFERWLVICK
B. nicolskii FTAFCSFANK YFILGPLACK IEGFLAALGG MVSLWSLCVI AFERWLVICK
C. gobio FTAFCSFANK YFILGPLACR IEGFLATLGG MVSLWSLCVI AFERWLVICK
O. latipes STCFVCFAFK YMVLGPLGCK IEGFTAALGG MVSLWSLAVI AFERWLVICK
D. rerio SVAFYAFYKK YFVFGPIGCK IEGFTSTIGG MVSLWSLAVV ALERWLVICK
C. autatus PLSFYSFFNR YFIFGATACK IEGFLATLGG MVGLWSLAVV AFERWLVICK
S. canaria TTAFYSFSQM YFALGRLACK IEGFTATLGG MVSLWSLAVV AFERFLVICK
R. ameriacana TTAFYSFSQM YFALGPTACK IEGFAATLGG MVSLWSLAVV AFERFLVICK 141
III

Continued.....


C. livia PLGNFTFRGS HAVLGCAXTW IFGLVAVPPP F-GWEQVHPE G----CSAR-
A. fasciatus PVGNFSFKGT HAIIGCALTW FFALLASTPP LFGWSRYIPE GLQCSCGPDW
A. caloninensis PLGNFTFRGT HAIIGCAVTW MFGLAASLPP LFGWSRYIPE GLQCSCGPDW
G. gallus PLGNFTFRGS HAVLGCVATW VLGFVASAPP LFGWSRYIPE GLQCSCGPDW
M. undulatus PLGSFSFRGS HALLGCAVTW VCGLAAATPP LFGWSRYIPE GLQCSCGPDW
S. camelus PLGNFTFRGT HAVLGCVVTW IFGLVASVPP LFGWSRYIPE GLQCSCGPDW
C. inermis PLGSFVFKAD HAMACCVVTW VLAMFAACPP LLGWRRYIPE SLQCSCGPDW
C. kesslerei PLGSFVFKAD HAMACCVVTW VLALFASCPP LFGWSRYIPE GLQCSCGPDW
L. eurystomias PLGSFVFKPD HAMACCVATW VLALLASGPP LFGWSRYIPE GLQCSCSPDW
B. nicolskii PLGSFVFKPD HAMACCVATW VLALLASGPP LFGWSRYIPE GLQCSCSPDW
C. gobio PLGTFVFKPD HAMACCMVTW VLALLASGPP LFGWSRYIPE GLQCSCGPDW
O. latipes PLGNFVFKSE HALLCCALTW VCGLCASVPP LVGWSRYIPE GMQCSCGPDW
D. rerio PLGNFTFKTP HAIAGCILPW CMALAAGLPP LLGWSRYIPE GLQCSCGPDW
C. autatus PLGNFTFKTP HAIAGCILPW ISALAASLPP LFGWSRYIPE GLQCSCGPDW
S. canaria PLGNFTFRGS HAVLGCAITW IFGLIASVPP LFGWSRYIPE GLQCSCRPDW
R. ameriacana PLGNFTFRGT HAVLGCIVTW IFGLVASTPP LFGWNKYIPE GLQCSCGPQW 191
IV
C. livia --AGPT---- ---------- -GTRXXLIVF SYGRLLLTVR AVARQQEQSA
A. fasciatus YTTENKYNNE SYVMFLFCFC FGFPFTVILF CYGQLLFTLK SAAKAQADSA
A. caloninensis YTTENKWNNE SYVIFLFCFC FGVPLSVIIF SYGRLLLTLR AVAKQQEQSA
G. gallus YTTDNKWHNE SYVLFLFTFC FGVPLAIIVF SYGRLLITLR AVARQQEQSA
M. undulatus YTANNKWNNE SYVIFLFCFC FGVPLALIVF SYGRLLLTLR AVARQQEQSA
S. camelus YTTNNKWNNE SYVIFLFCFC FGVPLAIIIF SYGRLLITLR AVAKQQEQSA
C. inermis YTSNNKYNNE SYVIYLFTCH FSVGLVILVF CYAQLLFTLK MAAKAQAESA
C. kesslerei YTTNNKYNNE SYVIYLFTCH FSVGLLILVF CYAQLLFTLK MAAKAQAESA
L. eurystomias YTTNNKYNNE SYVIYLFTCH FSVGLLILVF CYAQLLFTLK MAAKAQAESA
B. nicolskii YTTNNKYNNE SYVIYLFTCH FSVGLLILVF CYAQLLFTLK MAAKAQAESA
C. gobio YTTNNKYNNE SYVIYLFTCH FTVPLLILVF CYAQLLFTLK MAAKAQAESA
O. latipes YTTGNKFNNE SFVMFLFCFC FAVPFSIIVF CYSQLLFTLK MAAKAQADSA
D. rerio YTTNNKFNNE SYVMFLFCFC FAVPFSTIVF CYGQLLITLK LAAKAQADSA
C. autatus YTTNNKYNNE SYVMFLFCFC FAVPFGTIVF CYGQLLITLK LAAKAQADSA
S. canaria YTTDNKWNNE SSLIFSSCFC FGFPLSVIVF SYGRLLLTLR AVASQQEQSA
R. ameriacana YTNNNNLNNN SXLS---SSS ASASXXXXIX SYGRVLXTLR AVAKQQEQSA 241
V
C. livia STQKAEREVT KMVVVMVLGF LV-CWGPYSA FALWVVTHRG RPFDVGLASI
A. fasciatus STQKAEREVT KMVVVMVMGF LV-CWLPYAS FALWVVFNRG QSFDLRLGTI
A. caloninensis TTQKAEREVT KMVVVMVMGF LV-CWLPYAS FALWVVTHRG EPFDVRLASI
G. gallus TTQKADREVT KMVVVMVLGF LV-CWAPYTA FALWVVTHRG RSFEVGLASI
M. undulatus STQKAEREVT NMVVVMVLGF LV-CWAPYSA FALWVVTHRG RPFDVGLGSV
S. camelus TTQKAEREVT KMVVVMVLGF LV-CWAPYSS FALWVVTHRG QPFDMGLASV
C. inermis STQKAEREVT RMVVLMVLGF LV-CWLPYAS FAIWVVNNRG QPFDLRFASI
C. kesslerei STQKAEREVT RMVVLMVLGF LV-CWLPYAS FAFWVVNNRG QPFDLRFASI
L. eurystomias STQKAEREVT RMVVLMVLGF LV-CWLPYAS FAIWVVNNRG QPFDLRFASI
B. nicolskii STQKAEREVT RMVVLMVLGF LV-CWLPYAS FAIWVVNNRG QPFDLRFASI
C. gobio STQKAEREVT RMVVLMVLGF LV-CWMPYTC FALWVVNNRG QPFDLRFASI
O. latipes STQKAEKEVT RMVVVMVVAF LV-CYVPYAS FALWVINNRG QTFDLRLATI
D. rerio STQKAEREVT KMVVVMVFGF LI-CWGPYAI FAIWVVSNRG APFDLRLATI
C. autatus STQKAEREVT KMVVVMVLGF LV-CWAPYAS FSLWIVSHRG EEFDLRMATI
S. canaria TTQKAEXEVT KMVVVMVLGL ILVCWLPYCS FALWVVTHRG HPFDLGLASI
R. ameriacana TTQKAEREXT KMVVVMVLGF LV-CWAPYSA FALWVVTHRG QPFDVGLASV 291
VI
Continued.....


C. livia PSVFSKASTV YNPVIYVFMN KQFRSCMLKL L--------- ---------- 341
A. fasciatus PSCFSKASTV YNPVIYVFMN KQFRSCMMKL IFCGKSPFGD DEEASSSSQV
A. caloninensis PSVFSKASTV YNPVIYVLMN KQFRSCMLKL IFCGKSPFGD EDDVSGSSQA
G.gallus PSVFSKSSTV YNPVIYVLMN KQFRSCMLKL LFCGRSPFGD DEDVSGSSQA
M.undulatus PSVFSKASTV YNPIIYVLMN KQFRSCMLKL L--------- ----------
S. camelus PSVFSKASTV YNPVIYVFMN KQFR------ ---------- ----------
C. inermis PSVFSKSSTV YNPIIYVLLN KQFRSCMMKM MGMGG---AD DEESSTSSV-
C. kesslerei PSVFSKSSTV YNPVIYVLLN KQFRSCMMKM MGMGG---AD DEESSTSSV-
L. eurystomias PSVFSKSSTV YNPVIYVLLN KGFRSCMMKM MGMGG---AD DEESSTSSV-
B. nicolskii PSVFSKSSTV YNPVIYVLLN KQFRSCMMKM MGMGG---AD DEESSTSSV-
C. gobio PSVFSKSSTV YNPVIYVLLN KQFRSCMMKM MGMGG---AD DEESSTSSV-
O. latipes PSCVSKASTV YNPVIYVLLN KQFRLCMKKM LGMSA---DE DEESSTSQST
D. rerio PSCLCKASTV YNPVIYVLMN KQFRSCMMKM VFNKN---IE EDEASSSSQV
C. autatus PSCLSKASTV YNPVIYVLMN KQFRSCMMKM VCGKN---IE EDEASTSSQV
S. canaria PSVFSKASTV YNPIIYVFMN KQFRSCMLKL VFCGRSPFGD DDDVSGSSQA
R. ameriacana PSVFSKASTV YNPVIYVFMN KQFRSCVLKM VFCGRSPFGD EEDVSGSSVS
VII
C. livia ---------- ----
A. fasciatus TQVSSVGPEK ----
A. caloninensis TQVSSVSSSQ VSPA
G. Gallus TQVSSVSSSH VAPA
M.undulatus ---------- ----
S. camelus ---------- ----
C. inermis TEVSKVGPA- ----
C. kesslerei TEVSKVGPA- ----
L. eurystomias TEVSKVGPA- ----
B. nicolskii TEVSKVGPA- ----
C. gobio TEVSKVGPA- ----
O. latipes TEVSKVGPS- ----
D. rerio TQVSSVAPEK ----
C. autatus TQVSSVAPEK ----
S. canaria TQVS------ ----
R. ameriacana TS-------- ---- 355

from pairwise comparisons are presented in tables 5.1 and 5.2. In order to estimate the divergence between both cottoid fishes and goldfish, and the Lake Baikal cottoids and *Cottus gobio*, the mean of the group pairs was determined (tables 5.1 B, C & D and 5.2 B, C & D). The approximate age of the emergence of the goldfish lineage has been calculated using evidence from the fossil record and, because this age is known, the goldfish was used as a comparison to calculate divergence rates.

Goldfish are part of the superorder Ostariophysi, the oldest fossils of which have been dated at around 140 mya by Poyato-Ariza (cited in Hunt *et al.*, 1997). This figure was used to calibrate the molecular clock and determine the rate of divergence between cottoid fishes and goldfish. The rate of nucleotide substitution is defined as the number of substitutions per site per year and is calculated using the formula:

$$r = K/(2T)$$

Where K is the number of substitutions between two homologous sequences and T is the time of divergence between the two sequences.

Using figures from the pairwise comparisons of total nucleotide sequences and those at synonymous sites (tables 5.1B & 5.2B), the rate of divergence of cottoid fishes from goldfish was calculated to be:

for **total nucleotide** substitutions: $r = 0.418 / 280 \times 10^{6} = \mathbf{1.50 \times 10^{-9}}$ subs/site/year

for **synonymous** site substitutions: $r = 1.703 / 280 \times 10^{6} = \mathbf{6.08 \times 10^{-9}}$ subs/site/year

The molecular clock is then set using the value r in the formula:

$$K/(r \times 2)$$

The age of divergence between the Lake Baikal cottoids and *Cottus gobio* was calculated using the above formula, where K = mean divergence of Lake Baikal cottoids from *Cottus gobio* (tables 5.1C & 5.2C), to be:

using **total nucleotide** substitutions, age of divergence =

$0.032 / (1.5^{-9} \times 2) =$ **10.8 ± 1.7** mya

using **synonymous** site substitutions, age of divergence =
$0.042 / (6.08^{-9} \times 2) =$ **3.4 ± 0.2** mya

The age of divergence between *Limnocottus eurystomias/Bhatrachocottus nicolskii,* with the 428nm pigment, and *Cottus kesslerei/Cottocomephorus inermis,* with the 450nm pigment (tables 5.1D & 5.2D), was calculated to be:

using **total nucleotide** substitutions , age of divergence =
$0.023 / (1.5^{-9} \times 2) =$ **7.6 ± 2.6** mya

using **synonymous** site substitutions, age of divergence =
$0.018 / (6.08^{-9} \times 2) =$ **1.4 ± 0.4** mya

The corrected frequency of substitutions at synonymous sites between the rod opsins of the Lake Baikal cottoid fishes and goldfish is less than half that of the blue opsins, 0.686 ± 0.013 (table 5.3) compared to 1.703 ± 0.030. This raises the question as to whether substitution rates also vary between other opsins in different species. As a comparison, estimates of divergence between avian opsins and mammalian opsins have also been calculated using pairwise comparisons as shown in tables 5.5 and 5.6. In the avian opsins, the mean substitution rate at synonymous sites within the rods is 0.285 (table 5.5A). This figure is similar to the rate within the green opsins of 0.251 (table 5.5A). However, the short wave violet and UV opsins show a mean substitution rate of 0.485, almost twice that of the other groups (table 5.5A). A similar pattern was observed when total nucleotide substitution rates were compared (table 5.6A). Mammalian opsins showed very little difference between rod, green and blue pigments; these had synonymous substitution rates of 0.563, 0.461 and 0.422 respectively (table 5.5B). Again, when total nucleotide substitutions were evaluated, similar figures were obtained (table 5.6B).

5.2.3 Sequence conservation within the blue opsin gene

Given that there is divergence between the blue opsin genes of various species, are some regions more highly conserved than others? A comparison was made between amino acid sequences of the blue genes of 16 species including 9 fishes, 6 birds and the chameleon (figure 5.8), using Geneworks™. Percentage amino acid identity was calculated for each of the transmembrane helices and extracellular loops; the results are summarised in table 5.7. Helices II, III and VII show greater than 50% amino acid identity. Helix III, which contains the Schiff's base counterion is the most highly conserved, followed by helix VII in which the retinal binding site is located. The least conserved helix was V, where only 13% of residues were identical. Interestingly, the loop between two poorly conserved helices, IV and V, was the most highly conserved at 68%.

Helices	% identity	Loops	% identity
I	30	I-II	56
II	52	II-III	23
III	61	III-IV	59
IV	30	IV-V	68
V	13	V-VI	48
VI	43	VI-VII	23
VII	57		

Table 5.7: Percentage amino acid identities in the helical and extracellular loop regions between the blue cone opsin genes of fishes and chicken.

5.3 Discussion

5.3.1 Evolutionary relatedness of Baikal fishes from phylogenetic analysis

Morphological classification of the Lake Baikal cottoids places *Cottocomephorus inermis* and *Batrachocottus nicolskii* in the same sub-family of Cottidae, the Cottocomephorinae. *Cottus kesslerei* and *Cottus gobio* are grouped together in another subfamily, the Cottinae. *Limnocottus eurystomias* is placed in a separate family, the Abyssocottidae (figure 5.2) (Nelson, 1994). Results from phylogenetic analysis of blue cone opsin genes however, place *Limnocottus eurystomias* and *Batrachocottus nicolskii* in the same clade, separate and distinct from *Cottus kesslerei* and *Cottocomephorus inermis*. This grouping is also supported by phylogenetic analysis of the rod opsin genes of the Baikal cottoids (Hunt *et al.*, 1997). Although *Limnocottus eurystomias* was not included in their study, *Batrachocottus nicolskii* was placed in a clade with other members of the Abyssocottidae family, supported by a bootstrap value of 100. Kiril'chik *et al.* studied the phylogenetic relatedness of Lake Baikal cottoids by sequencing and comparing partial nucleotide sequences of mitochondrial DNA from cytochrome *b* genes (Kiril'chik *et al.*, 1995). Their data also place *Batrachocottus nicolskii* and *Cottocomephorus inermis* into separate clades although the grouping of *Batrachocottus nicolskii* with other members of the Abyssocottidae family is not supported (figure 5.5).

The arrangement of species into separate clades in the phylogenetic trees shown in figure 5.5, closely mirrors morphological classification at the order level (Nelson, 1994). As expected, all cottoid fishes are grouped together in one clade. The two Cypriniformes, *Carassius auratus* and *Danio rerio* are placed in a separate clade with *Astyanax fasciatus* and *Oryzias latipes*. The topology of both trees is very similar, differing only in the position of the *Astyanax fasciatus* branch. Inconsistencies with the original taxonomy only arise at the species level and therefore, based upon phylogenetic data from the blue cone opsin, rod opsin (Hunt *et al.*, 1997) and cytochrome *b* genes (Kiril'chik *et al.*, 1995), it seems that the classification of Lake Baikal cottoid fishes needs to be re-evaluated.

5.3.2 Divergence time of cottoid fishes

The substitution rate at non-synonymous sites is extremely variable between different kinds of genes and reflects the rate of protein evolution; synonymous rates may also vary but by not as much. Mutations at non-synonymous sites alter the amino acids and often cause deleterious effects on protein function resulting in their elimination from the population by purifying selection. In contrast, synonymous site substitutions are more likely to become fixed in a population giving a more accurate reflection of the general rate of evolution. This discussion therefore, focuses on the rates at synonymous sites rather than those at non-synonymous sites.

Based on the rate of divergence at synonymous sites within the coding region of the blue opsin gene, it is estimated that the radiation between the shallow and deep water Lake Baikal cottoids began around 1.4 mya. Studies by a group in Siberia on the cytochrome *b*, ATPase 8 and ATPase 6 mitochondrial genes, place radiation of the species flock at between 1 – 2.5 mya (Grachev *et al.*, 1992; Slobodyanyuk *et al.*, 1995; Kiril'chik *et al.*, 1995). However, their calculations were based on the rate of evolution of primate mtDNA which may not be directly applicable to fish mtDNA as rates of gene evolution are known to vary between species (Martin *et al.*, 1992; Martin & Palumbi, 1993). Hunt *et al.* placed the evolution of the Baikal species flock at approximately 4.9 mya (Hunt *et al.*, 1997). This figure is based on evolution of the rod opsin gene and is significantly different to the 1.4 mya predicted from the blue cone opsin genes.

What can be concluded from these various rates of evolution? It is clear that molecular divergence is roughly correlated with divergence time. However, the constancy of rate of divergence and the extent of the margin of error involved in these predictions is unclear. Also, the molecular clock is set using the fossil record which itself is often incomplete and unreliable. Comparisons between the dates obtained for the opsins, cytochrome *b* and ATPases do however suggest that the species flock radiation is relatively recent in geological terms and probably occurred during the

Pliocene or late Miocene, in the Cenozoic era. Results obtained for the divergence between *Cottus gobio* and the Lake Baikal species pre-date the Lake Baikal radiation as expected.

5.3.3 Evolution rates of opsin genes

Members of a gene family can evolve at different rates both within and between species (Sheldon, 1987; Hardison & Miller, 1993; Carulli & Hartl, 1992). Carulli and Hartl have shown that opsin genes in *Drosophila* evolve at significantly different rates (Carulli & Hartl, 1992). Based on comparisons between *D. melanogaster* and *D. obscura,* which were thought to have diverged 30 million years ago, evolutionary rates at synonymous sites for the rhodopsin 2, 3 and 4 genes of *Drosophila* are estimated to be 17.18×10^{-9}, 15.38×10^{-9} and 15.58×10^{-9} substitutions per site per year respectively (Li, 1997). Different structural regions have also evolved at different rates; membrane-spanning regions showing the lowest substitution level. Other proteins such as the histones are extremely conserved and show a substitution rate at non-synonymous sites of virtually zero. In comparison, interferon γ has approximately 3×10^{-9} substitutions per non-synonymous site per year (Li, 1997).

The rate of substitution depends upon mutation rate and the probability that the mutation will become fixed in a population. It is known that the rate of substitution at synonymous sites is generally higher than at non-synonymous sites because active genes are subjected to functional constraints. Also, the rate of synonymous substitution varies less between genes than the rate at non-synonymous sites (Li & Graur, 1991). Therefore, the rate of amino acid replacement is not always correlated with the rate of synonymous nucleotide substitution. For example the two UV receptors in *Drosophila*, *Rh* 3 and *Rh* 4 have similar rates of synonymous site substitution but significantly different rates of amino acid evolution (Carulli & Hartl, 1992). It is unclear why substitution rates vary from gene to gene although some possible explanations may be chromosomal location, codon usage bias or local base composition. The rate of synonymous site substitution may be reflected in the chromosomal position of a gene, for example its location in a particular isochore or on

a sex chromosome. In mammalian males, the number of germ cell divisions per generation is much larger than in the female and hence it is hypothesized that there is a higher mutation rate in the male germ line. It is unknown whether any opsin genes are sex linked in cottoid fishes because to date, studies have been unable to identify any sex chromosomes in these fishes (Amores & Postlethwait, 1999; Mandrioli *et al.*, 2000). The non-random usage of synonymous codons is termed codon usage bias. Although most amino acids are encoded by more than one codon, studies have shown that related species often favour a particular codon, particularly those ending in G or C (Shields *et al.*, 1988; Li, 1997). The rate of synonymous substitution is negatively correlated with codon usage bias (Shields *et al.*, 1988) and is probably affected to some extent by natural selection. The relationship between codon usage bias and rate of evolution of opsin genes is however, rather inconsistent (Carulli & Hartl, 1992).

The rate of substitution at synonymous sites in the rod opsins of Lake Baikal cottoid fishes was less than half that of the blue opsin genes. This corresponds to observations by Yokoyama and Yokoyama (1989) that human cone opsin genes are also evolving at a significantly higher rate than rod opsin genes. Rod opsins are common to a wide range of species living in very different environments suggesting that they are important for basic vision and would therefore be subjected to functional constraints. Colour pigments, in comparison, are less constrained and have become finely tuned to suit the environment inhabited by a particular species (this process of spectral tuning by the introduction of selected mutations at critical sites was discussed in chapter 4). The substitution rate at synonymous sites in *Drosophila* also varies between the four rhodopsin genes from 26 % in *Rh*1, to 36% in *Rh* 3 and 4, and 39% in *Rh*2 (Carulli & Hartl, 1992).

Is there a correlation between the number of synonymous substitutions and wavelength of a cone pigment? Pairwise comparisons between the MWS and violet/UV pigments of avian species did show a significant increase in substitution rate with decreasing λ_{max}. However, a similar comparison of substitution rates amongst mammalian opsins showed very little variation as expected. In conclusion

therefore, rod opsins appear to have evolved more slowly that cone opsins and the rate of evolution within cone opsins seems to vary between species. The differential rate of evolution between the rod and blue opsin genes may be a combined result of codon usage bias and local base composition as a result of different selection pressures.

5.3.4 Sequence conservation within opsins

Regions of a molecule, which are functionally important, are likely to be conserved during evolution. In a comparison of 38 vertebrate visual opsins, Wright (2000) identified 30 residues, which were conserved within the α-helices (table 5.8). However, when other fish blue (Sb) opsins were included, it was found that three of these sites were no longer conserved. The blue opsins of the cottoids, *L. eurystomias, B. nicolskii, C. kesslerei, C. inermis* and *C. gobio* all had substitutions at position 219 and 309; Ile-219 was replaced with Leu-219 and Met-309 replaced with Leu-309. At position 294, Phe was replaced with Val in the Beloniform, *O. latipes*, and Leu in the Cypriniformes, *D. rerio* and *C. auratus*. These substitutions are all conservative and unlikely to have any affect on the λ_{max} and functioning of the opsin molecule. Hunt *et al.* (1997) sequenced the rod opsins from many of the Lake Baikal cottoid fishes. These opsins did not contain the substitutions at sites 219 and 309 suggesting that this replacement is unique to the blue class of opsins in the cottoid fishes and occurred relatively recently but before the split between *C. gobio* and the Lake Baikal cottoids. The substitutions at position 294 probably occurred independently in the Beloniformes and Cypriniformes, the variation at this site suggesting that it is not functionally important.

Comparisons were made between the blue opsin genes of 9 species of fish, *C. inermis, C. kesslerei, L. eurystomias, B. nicolskii, C. gobio, O. latipes, D. rerio, A fasciatus* and *C, auratus*; 6 species of bird, *C. livia, G. gallus, M. undulates, S. camelus, S. canaria* and *R. ameriacana*; and the chameleon, *A. caloninensis*. The comparison revealed relatively high sequence conservation in helices II, III and VII and in cytoplasmic loops I-II, III-IV and IV-V (table 5.7). Helix III contains the Schiff's

base counterion at position 113 (Sakmar *et al.*, 1989) and transducin binding residues at 135 and 136 (Franke *et al.*, 1990), all of which are critically important in the phototransduction process. The retinal binding site is at position 296 in helix VII which is again, highly conserved. A disulphide bridge is formed between cysteine residues at 110, in the cytoplasmic loop between helices II and III, and 187, in the loop between helices IV and V (Karnik & Khorana, 1990; Karnik *et al.*, 1988). These residues are conserved in all of the blue opsins and, around 187 in particular, there is a high degree of general sequence conservation, including sites 181 and 184. In long-wavelength cone pigments, these two sites are responsible for binding chloride ions which are critical to achieve the long-wavelength sensitivity of a pigment (Crescitelli, 1977; Kleinschmidt & Harosi, 1992; Wang *et al.*, 1993). His-181 and Lys-184, in long-wave cone pigments (except mouse), are replaced by Glu-181 and Gln-184 in all of the blue (Sb) opsins, resulting in a loss of charge and inability to bind chloride ions; this in turn leads to a short-wave shift in the λ_{max} of the pigment. This suggests that long-wave pigments were already established when an ancestral pigment gained the ability to bind chloride ions. Why there should be such a high degree of sequence conservation at these sites within the short-wave pigments is unclear, although one possibility is that a negative charge may be required to prevent binding of the chloride ions.

The other regions, which show a high degree of conservation, are helix II and the transmembrane regions between loops I and II, and III and IV. Interestingly, the loop regions are more highly conserved than the helices in the blue opsins studied. Although these do not directly interact with 11-*cis* retinal or other members of the phototransduction cascade they may be important in maintaining the structural integrity of blue opsin molecules.

Helix	aa number	aa residue
I	55	N
II	72	L
	73	N
	75	I
	76	L
III	113	E
	126	W
	127	S
	128	L
	134	E
	135	R
IV	161	W
	171	P
	174	G
	175	W
V	219	I *
	223	Y
VI	253	M
	254	V
	257	M
VII	291	P
	294	F **
	296	K
	301	Y
	302	N
	303	P
	305	I
	306	Y
	309	M *
	310	N

Table 5.8: Amino acid residues conserved within transmembrane helices of all vertebrate visual opsins, as identified by Wright (2000). With the exception of those marked by an asterisk, these sites are also conserved within fish blue (Sb) opsins. * - Residues not conserved in Cottoid fishes. ** - Residue not conserved in the Cypriniformes *D. rerio* and *C. auratus*, and the Beloniform, *O. latipes*. The numbers relate to bovine rod (Nathans *et al.*, 1986).

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

The aim of this project was to study the regulation, spectral tuning and evolution of visual pigment genes in fish to enhance our understanding of the complex world of colour vision. A great deal is already known about visual systems of primates but much less is known about fish. This study offered an opportunity to investigate gene regulation of a group of rod-like green cone opsins which are not found in mammals and, for the first time, to analyse the process of spectral tuning within the blue cone opsin class.

The 5' promoter regions of the goldfish rod and two green opsins have been successfully isolated by the walking PCR method (Dominguez & Lopez-Larrea, 1994) and sequenced. Comparisons between the two green gene promoters showed them to be highly conserved for the first 450bp upstream, however comparisons with the rod promoter and regulatory regions of other opsin genes showed very little homology. A number of known transcription factor binding sites were identified and also a novel 8bp sequence which was common to fish, mammalian and avian opsin promoters. A next step would be to determine if these motifs are involved in binding *trans*-acting factors. Crude studies can be made using the electrophoretic band shift assay (EBSA) to determine if any proteins bind to regions of the promoter. A more detailed assessment would involve DNaseI footprinting, which would highlight any specific DNA sequences to which a transcription factor is binding. Once these sequences have been identified, the yeast one-cell hybrid system could be used to identify transcription factors binding to particular DNA motifs (Li & Herskowitz, 1993;Wang & Reed, 1993). This method involves using yeast cells, containing plasmids with the relevant DNA insert, to screen a retinal cDNA expression library. Positive clones isolated by this method can then be sequenced and the data used in a BLAST search to identify the class of protein. Complementary studies could also involve the use of transgenic models to help determine the importance of *cis*-acting binding sites and the

effect of any mutations in the promoter region on retinal development and expression of opsin proteins.

The cottoid fish of Lake Baikal offered a unique opportunity to study evolution and spectral tuning within the blue cone opsins of a group of closely related species. The original aim of isolating, sequencing and comparing the blue opsins of fishes was successfully achieved. Potential spectral tuning sites were identified at amino acid residues 118, 215 and 269, and individual mutations at these positions were introduced into the wild type, longer wavelength blue opsin to determine their affect on the λ_{max}. However, problems were experienced with the expressed mutant pigments during the purification procedure following reconstitution of the visual pigments with 11-*cis* retinal. Results were obtained for the Thr118Gly and Pro215Gly mutants but not for the Thr118Ala or Thr269Ala mutants. The main problem appears to be failure of the opsin to bind to the 1D4 antibody column. It is possible that blue photopigments are less stable than longer wavelength pigments and that the protocols used by other groups for purifying the opsins, needs to be optimised further. One possibility would be to alter the buffering system from PBS to HEPES, which is a more efficient buffer and is stable when frozen. Another alternative is to replace the 1D4 epitope with a c-myc tag and to purify using a nickel, c-myc coupled, antibody column or to use a His tag and purify using commercially available His-bind resin.

Further mutation and expression studies would be desirable to improve the data for the Pro215Gly mutant and to confirm the role of site 269 in short-wave shifting, blue pigments. It would also be interesting to determine if, by introducing polar amino acids at both sites 118 and 269 together, the wild type pigment could be induced to short-wave shift its λ_{max} by the full 42nm. Site 117 as a spectral tuning site in blue pigments also warrants further investigation.

In order to understand the physical basis underlying the short-wave shift in λ_{max} of these blue pigments, the process of resonance Raman spectroscopy could be applied. An electric field is used to induce dipole moment; the process relies on there being a change in the polarizability of the system of vibrating atoms during a vibrational cycle. This technique would provide chemical and structural information about the

interaction of opsin mutants with the chromophore and has been successfully applied by Lin *et al.* (1998) to study differences between bovine rhodopsin and artificial short-wave shifted mutants.

In the final chapter, the evolution of cottoid fish was investigated. The divergence rate at synonymous sites showed that this species flock radiated relatively recently, in the Cenozoic era. Phylogenetic trees were produced using the blue cone opsin sequence data and these supported morphological classification at the order level, although there were some inconsistencies at the species level. The evolution rate of the blue opsin genes was found to be faster than that of the rod opsins, as is the case in humans. Comparisons were also made with other blue (sb) opsins showing that the degree of sequence conservation varied significantly between different regions of the molecule.

ABBREVIATIONS

μg	microgram
μl	microlitre
λ_{max}	wavelength of maximum absorbance
A_1	11-*cis* retinal
A_2	11-*cis* 3 dehydro retinal
amp	ampicillin
bp	base pairs
cDNA	complementary deoxyribonucleic acid
cGMP	cyclic guanosine monophosphate
CNGCs	cyclic nucleotide gated channels
dATP	deoxy-adenosine-5' triphosphate
dCTP	deoxy-cytosine-5' triphosphate
DEPC	diethyl pyrocarbonate
dGTP	deoxy-guanosine-5' triphosphate
DM	dodecyl maltoside
DMED	Dulbecco's modified eagle medium
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleotide-5' triphosphate
dTTP	deoxy-thymidine-5' triphosphate
E	phosphodiesterase
EDTA	ethylene-diamine-tetra acetic acid
gDNA	genomic deoxyribonucleic acid
GDP	guanosine mono phosphate
GTP	guanosine triphosphate
IPTG	isopropyl-β-D-thiogalactoside
Kb	kilobases
L	long-wavelength
LCR	locus control region
LWS	long wave sensitive

M	molar
Mc	middle-wavelength rod-like cone
ml	millilitre
Mrd	middle-wavelength rod
MSP	microspectrophotometry
mtDNA	mitochondrial deoxyribonucleic acid
MWS	middle wave sensitive
mya	million years ago
ng	nanogram
nm	nanometer
P	probability
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenyl methyl sulfonyl fluoride
PNK	polynucleotide kinase
RACE	rapid amplification of cDNA ends
RER	rhodopsin enhancer region
RNA	ribonucleic acid
rpm	revolutions per minute
Sb	short-wavelength blue
SEM	standard error of the mean
Sv	short-wavelength violet
SWS	short wave sensitive
Ta	annealing temperature
TAE	Tris acetate EDTA buffer
TBE	Tris borate EDTA buffer
TESS	transcription element search system
tet	tetracycline
TF	transcription factor
Tm	melting temperature
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside

REFERENCES

7.1 List of references

Ahmad, I. (1995). Mash-1 is expressed during ROD photoreceptor differentiation and binds an E-box, E(opsin)-1 in the rat opsin gene. *Brain Res Dev Brain Res, 90*, 184-9.

Ahmad, I., Yu, X., & 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*, 396-9.

Ahnelt, P. K., Hokoc, J. N., & Rohlich, P. (1995). Photoreceptors in a primitive mammal, the South American opossum, *Didelphis marsupialis aurita*: characterization with anti-opsin immunolabeling. *Vis Neurosci, 12*, 793-804.

Ali, M. A., & Anctil, M. (1976). *Retinas of fishes: an atlas*. Berlin: Springer-Verlag.

Amores, A., & Postlethwait, J. H. (1999). Banded chromosomes and the zebrafish karyotype. *Methods Cell Biol, 60*, 323-38.

Archer, S., Hope, A., & Partridge, J. C. (1995). The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. *Proc R Soc Lond B Biol Sci, 262*, 289-95.

Asenjo, A. B., Rim, J., & Oprian, D. D. (1994). Molecular determinants of human red/green color discrimination. *Neuron, 12*, 1131-8.

Ashmore, J. F., & Falk, G. (1977). Dark noise in retinal bipolar cells and stability of rhodopsin in rods. *Nature, 270*, 69-71.

Baehr, W., Falk, J. D., Bugra, K., Triantafyllos, J. T., & McGinnis, J. F. (1988). Isolation and analysis of the mouse opsin gene. *FEBS Lett, 238*, 253-6.

Baldwin, J. M. (1993). The probable arrangement of the helices in G protein-coupled receptors. *Embo J, 12*, 1693-703.

Baldwin, J. M., Schertler, G. F., & Unger, V. M. (1997). An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol, 272*, 144-64.

Barlow, H. B. (1957). Purkinje Shift and Retinal Noise. *Nature, 179*, 255-256.

Barlow, R. B., Birge, R. R., Kaplan, E., & Tallent, J. R. (1993). On the molecular origin of photoreceptor noise. *Nature, 366*, 64-66.

Batni, S., Scalzetti, L., Moody, S. A., & Knox, B. E. (1996). Characterization of the Xenopus rhodopsin gene. *J Biol Chem, 271*, 3179-86.

Bellingham, J., Morris, A. G., & Hunt, D. M. (1998). The rhodopsin gene of the cuttlefish Sepia officinalis: sequence and spectral tuning. *J Exp Biol, 201*, 2299-306.

Bennett, J., Sun, D., & Kariko, K. (1995). Sequence analysis of the 5.34-kb 5' flanking region of the human rhodopsin-encoding gene. *Gene, 167*, 317-20.

Blatz, P. E., Mohler, J. H., & Navangul, H. V. (1972). Anion-induced wavelength regulation of absorption maxima of Schiff bases of retinal. *Biochemistry, 11*, 848-55.

Boatright, J. H., Borst, D. E., Peoples, J. W., Bruno, J., Edwards, C. L., Si, J. S., & Nickerson, J. M. (1997a). A major cis activator of the IRBP gene contains CRX-binding and Ret- 1/PCE-I elements. *Mol Vis, 3*, 15.

Boatright, J. H., Buono, R., Bruno, J., Lang, R. K., Si, J. S., Shinohara, T., Peoples, J. W., & Nickerson, J. M. (1997b). The 5' flanking regions of IRBP and arrestin have promoter activity in primary embryonic chicken retina cell cultures. *Exp Eye Res, 64*, 269-77.

Bowmaker, J. K. (1991). The evolution of visual pigments in photoreceptors. London: Macmillan.

Bowmaker, J. K. (1998). Evolution of colour vision in vertebrates. *Eye, 12*, 541-7.

Bowmaker, J. K., Astell, S., Hunt, D. M., & Mollon, J. D. (1991a). Photosensitive and photostable pigments in the retinae of Old World monkeys. *J Exp Biol, 156*, 1-19.

Bowmaker, J. K., Dartnall, H. J. A., & Herring, P. J. (1988). Longwave-sensitive visual pigments in some deep-sea fishes: segregation of 'paired' rhodopsins and porphyropsins. *Journal of Comparative Physiology A, 163*, 688-698.

Bowmaker, J. K., Govardovskii, V. I., Shukolyukov, S. A., Zueva, L. V., Hunt, D. M., Sideleva, V. G., & Smirnova, O. G. (1994). Visual pigments and the photic environment: the cottoid fish of Lake Baikal. *Vision Res, 34*, 591-605.

Bowmaker, J. K., & Hunt, D. M. (1999). *Molecular biology of photoreceptor spectral sensitivity*. (1st ed.): Kluwer Academic Publishers.

Bowmaker, J. K., Thorpe, A., & Douglas, R. H. (1991b). Ultraviolet-sensitive cones in the goldfish. *Vision Res, 31*, 349-52.

Bownds, D. (1967). Site of attachment of retinal in rhodopsin. *Nature, 216*, 1178-81.

Buratowski, S., Hahn, S., Guarente, L., & Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell, 56*, 549-61.

Cameron, D. A., & Easter, S. S., Jr. (1995). Cone photoreceptor regeneration in adult fish retina: phenotypic determination and mosaic pattern formation. *J Neurosci, 15*, 2255-71.

Carter-Dawson, L. D., & LaVail, M. M. (1979). Rods and cones in the mouse retina. II. Autoradiographic analysis of cell generation using tritiated thymidine. *J Comp Neurol, 188*, 263-72.

Carulli, J. P., & Hartl, D. L. (1992). Variable rates of evolution among Drosophila opsin genes. *Genetics, 132*, 193-204.

Chan, T., Lee, M., & Sakmar, T. P. (1992). Introduction of hydroxyl-bearing amino acids causes bathochromic spectral shifts in rhodopsin. *The Journal of Biological Chemistry, 267*, 9478-9480.

Chang, B. S., Crandall, K. A., Carulli, J. P., & Hartl, D. L. (1995). Opsin phylogeny and evolution: a model for blue shifts in wavelength regulation. *Mol Phylogenet Evol, 4*, 31-43.

Chen, J., Tucker, C. L., Woodford, B., Szel, A., Lem, J., Gianella-Borradori, A., Simon, M. I., & 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. USA, 91*, 2611-2615.

Chen, S., Wang, Q., Nie, Z., Sun, H., Lennon, G., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., & Zack, D. J. (1997). Crx, a novel Otx-like paired-homeodomain protein, binds to and transactivates photoreceptor cell-specific genes. *Neuron, 19*, 1017-1030.

Chen, S., & Zack, D. J. (1996). Ret 4, a positive acting rhodopsin regulatory element identified using a bovine retina in vitro transcription system. *Journal of Biological chemistry, 271*, 28549-28557.

Chiu, M. I., & 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*, 3426-36.

Chiu, M. I., & 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*, 773-80.

Chiu, M. I., Zack, D. J., Wang, Y., & 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*, 440-3.

Collin, S. P. (1999). *Behavioural ecology and retinal cell topography.* (1st ed.): Kluwer academic Publishers.

Courey, A. J., & Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell, 55*, 887-98.

Crescitelli, F. (1972). Handbook of sensory physiology (Dartnall, H.J.A. Ed.). *Springer, New York, VII/1*, 245-363.

Crescitelli, F. (1977). *Handbook of sensory physiology*. (Vol. VII/5). Berlin: Springer-Verlag.

Crescitelli, F., & Karvaly, B. (1991). The gecko visual pigment: the anion hypsochromic effect. *Vision Res, 31*, 945-50.

Cvekl, A., & Paces, V. (1992). Interactions between proteins bound to the duck beta A-globin gene promoter and enhancer detected by the DNaseI footprinting. *Gene, 110*, 225-8.

Dartnall, H. J. A. (1957). *The Visual Pigments*. New LondonYork: Methuen & Co. Ltd.

Das, D., Wilkie, S. E., Hunt, D. M., & Bowmaker, J. K. (1999). Visual pigments and oil droplets in the retina of a passerine bird, the canary *Serinus canaria*: microspectrophtometry and opsin sequences. *Vision Research*.

David-Gray, Z. K., Janssen, J., deGrip, W., Nevo, E., & Foster, R. G. (1998). Isolation and functional expression of a middle wavelength sensitive cone opsin from the eye of the blind mole rat Spalax ehrenbergi. *Investigative Ophthalmology and Visual Science, 39*, S957.

DesJardin, L. E., & Hauswirth, W. W. (1996). Developmentally important DNA elements within the bovine opsin upstream region. *Invest Ophthalmol Vis Sci, 37*, 154-65.

Dominguez, O., & Lopez-Larrea, C. (1994). Gene walking by unpredictably primed PCR. *Nucleic Acids Res, 22*, 3247-8.

Donner, K. (1992). Noise and the absolute thresholds of cone and rod vision. *Vision research, 32*, 853-866.

Douglas, R. H. (1982). The function of the photomechanical movements in the retina of the rainbow trout (*Salmo gairdneri*). *Journal of Experimental Biology, 96*, 389-403.

Douglas, R. H., & Marshall, N. J. (1999). *A review of vertebrate and invertebrate ocular filters*: Kluwer Academic Publishers.

Douglas, R. H., Partridge, J. C., Dulai, K. S., Hunt, D. M., Mullineaux, C. W., & Hynninen, P. H. (1999). Enhanced retinal longwave sensitivity using a chlorophyll-derived photosensitiser in Malacosteus niger, a deep-sea dragon fish with far red bioluminescence. *Vision Res, 39*, 2817-32.

Douglas, R. H., Partridge, J. C., Dulai, K. S., Hunt, D. M., Mullineaux, C. W., Tauber, A. Y., & Hynninen, P. H. (1998). Dragon fish see using chlorophyll. *Nature, 393*, 423-424.

Douglas, R. H., & Wagner, H. J. (1982). Endogenous patterns of photomechanical movements in teleosts and their relation to activity rhythms. *Cell Tissue Res, 226*, 133-44.

Dulai, K. S., von Dornum, M., Mollon, J. D., & Hunt, D. M. (1999). The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates. *Genome Res, 9*, 629-38.

Eck, R. V., & Dayhoff, M. O. (1966). *Atlas of protein Sequence and Structure.* Silver Spring MD: National Biomedical Research Foundation.

Endler, J. A. (1993). Some general comments on the evolution and design of animal communication systems. *Philos Trans R Soc Lond B Biol Sci, 340*, 215-25.

Fasick, J. I., Cronin, T. W., Hunt, D. M., & Robinson, P. R. (1998). The visual pigments of the bottlenose dolphin (*Tursiops truncatus*). *Vis Neurosci, 15*, 643-51.

Fasick, J. I., & Robinson, P. R. (1998). Mechanism of spectral tuning in the dolphin visual pigments. *Biochemistry, 37*, 433-438.

Fernald, R. D., & Wright, S. E. (1985). Growth of the visual system in the African cichlid fish, Haplochromis burtoni. Accommodation. *Vision Res, 25*, 163-70.

Fitch, W. M. (1977). On the problem of discovering the most parsimonious tree. *Am. Nat., 111*, 223-257.

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

Fitzgibbon, J., Hope, A., Slobodyanyuk, S. J., Bellingham, J., Bowmaker, J. K., & Hunt, D. M. (1995). The rhodopsin-encoding gene of bony fish lacks introns. *Gene, 164*, 273-7.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics, 151*, 1531-45.

Fortini, M. E., & Rubin, G. M. (1990). Analysis of cis-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organization to rhodopsin promoters in Drosophila melanogaster. *Genes Dev, 4*, 444-63.

Franke, R. R., Konig, B., Sakmar, T. P., Khorana, H. G., & Hofmann, K. P. (1990). Rhodopsin mutants that bind but fail to activate transducin. *Science, 250*, 123-5.

Franke, R. R., Sakmar, T. P., Oprian, D. D., & Khorana, H. G. (1988). A single amino acid substitution in rhodopsin (lysine 248----leucine) prevents activation of transducin. *J Biol Chem, 263*, 2119-22.

Freund, C. L., Gregory-Evans, C. Y., Furukawa, T., Papaioannou, M., Looser, J., Ploder, L., Bellingham, J., Ng, D., Herbrick, J. A., Duncan, A., Scherer, S. W., Tsui, L. C., Loutradis-Anagnostou, A., Jacobson, S. G., Cepko, C. L., Bhattacharya, S. S., & McInnes, R. R. (1997). Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell, 91*, 543-53.

Fukada, M. N., Papermaster, D. S., & Hargrave, P. A. (1979). Rhodopsin carbohydrate: structure of small oligosaccharides attached at two sites near the NH_2 terminus. *J. Biol. Chem*, 8201-8207.

Furukawa, T., Morrow, E. M., & Cepko, C. L. (1997). Crx, a novel otx-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell, 91*, 531-41.

Gale, J. M., Tobey, R. A., & D'Anna, J. A. (1992). Localization and DNA sequence of a replication origin in the rhodopsin gene locus of Chinese hamster cells. *J Mol Biol, 224*, 343-58.

Gnyubkina, V. P., & Levin, A. V. (1987). Changeable Corneal Coloration in some Baikalian and River Sculpins (Pisces: Cottoidei). *Copeia, 3*, 758-762.

Goldsmith, T. H. (1990). Optimization, constraint, and history in the evolution of eyes. *Q Rev Biol, 65*, 281-322.

Grachev, M. A., Slobodyanyuk, S., Kholodilov, N. G., Fyodorov, S. P., Belikov, S. I., Sherbakov, D., Sideleva, V. G., Zubin, A. A., & Kharchenko, V. V. (1992). Comparative study of two protein-coding regions of mitochondrial DNA from three endemic sculpins (Cottoidei) of Lake Baikal. *J Mol Evol, 34*, 85-90.

Hailman, J. P. (1976). Oildroplets in the eyes of adult anuran amphibians: a comparative survey. *J Morphol, 148*, 453-68.

Hardison, R., & Miller, W. (1993). Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol Biol Evol, 10*, 73-102.

Hashimoto, Y., Harosi, F. I., Ueki, K., & Fukurotani, K. (1988). Ultra-violet sensitive cones in the colour-coding systems of cyprinid retinas. *Neurosci Res Suppl, 8*, 81-96.

Henderson, R., & Schertler, G. F. (1990). The structure of bacteriorhodopsin and its relevance to the visual opsins and other seven-helix G-protein coupled receptors. *Philos Trans R Soc Lond B Biol Sci, 326*, 379-89.

Higgins, D. G., Bleasby, A. J., & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Comput Appl Biosci, 8*, 189-91.

Hisatomi, O., Kayada, S., Aoki, Y., Iwasa, T., & Tokunaga, F. (1994). Phylogenetic relationships among vertebrate visual pigments. *Vision Res, 34*, 3097-102.

Hope, A. J., Partridge, J. C., Dulai, K. S., & Hunt, D. M. (1997). Mechanisms of wavelength tuning in the rod opsins of deep-sea fishes. *Proc R Soc Lond B Biol Sci, 264*, 155-63.

Hunt D.M., Dulai, K.S., Cottrill, P., Bowmaker, J.K., & Partridge, J.C. (2001). Spectral tuning of rod visual pigments in deep-sea fish. *Submitted for publication.*

Hunt, D. M., Fitzgibbon, J., Slobodyanyuk, S. J., & Bowmaker, J. K. (1996). Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in Lake Baikal. *Vision Res, 36*, 1217-24.

Hunt, D. M., Fitzgibbon, J., Slobodyanyuk, S. J., Bowmaker, J. K., & Dulai, K. S. (1997). Molecular evolution of the cottoid fish endemic to Lake Baikal deduced from nuclear DNA evidence. *Mol Phylogenet Evol, 8*, 415-22.

Ibbotson, R. E., Hunt, D. M., Bowmaker, J. K., & 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*, 145-54.

Imai, H., Kojima, D., Oura, T., Tachibanaki, S., Terakita, A., & Shichida, Y. (1997). Single amino acid residue as a functional determinant of rod and cone visual pigments. *Proc Natl Acad Sci U S A, 94*, 2322-6.

Jacobs, G. H. (1993). The distribution and nature of colour vision among the mammals. *Biol. Rev., 68*, 413-471.

Jacobs, G. H., Neitz, M., Deegan, J. F., & Neitz, J. (1996). Trichromatic colour vision in New World monkeys. *Nature, 382*, 156-8.

Johnson, R. L., Grant, K. B., Zankel, T. C., Boehm, M. F., Merbs, S. L., Nathans, J., & Nakanishi, K. (1993). Cloning and expression of goldfish opsin sequences. *Biochemistry, 32*, 208-14.

Jukes, T. H., & Cantor, C. R. (1969). *Evolution of protein molecules.* (Vol. 3). New york: Academic Press.

Kakitani, H., Kakitani, T., Rodman, H., & Honig, B. (1985). On the mechanism of wavelength regulation in visual pigments. *Photochem Photobiol, 41*, 471-9.

Karnik, S. S., & Khorana, H. G. (1990). Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *J Biol Chem, 265*, 17520-4.

Karnik, S. S., Sakmar, T. P., Chen, H. B., & Khorana, H. G. (1988). Cysteine residues 110 and 187 are essential for the formation of correct structure in bovine rhodopsin. *Proc Natl Acad Sci U S A, 85*, 8459-63.

Kikuchi, T., Raju, K., Breitman, M. L., & Shinohara, T. (1993). The proximal promoter of the mouse arrestin gene directs gene expression in photoreceptor cells and contains an evolutionarily conserved retinal factor-binding site. *Mol Cell Biol, 13*, 4400-8.

Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T., & Fujiyoshi, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature, 389*, 206-11.

Kiril'chik, S. V., Slobodianiuk, S., Belikov, S. I., & Pavlova, M. E. (1995). Phylogenetic relationships between 16 species of Lake Baikal sculpin fish based on a nucleotide sequence analysis of a mitochondrial DNA fragment of the cytochrome b gene. *Mol Biol (Mosk), 29*, 817-25.

Kleinschmidt, J., & Harosi, F. I. (1992). Anion sensitivity and spectral tuning of cone visual pigments in situ. *Proc Natl Acad Sci U S A, 89*, 9181-5.

Knowles, A. D., H.J.A. (1977). The photobiology of vision. *Davson, H. (Ed.) The Eye, 2B*, 1-689.

Kochendorfer, G. G., Lin, S. W., Sakmar, T. P., & Mathies, R. A. (1999). How colour visual pigments are tuned. *TIBS, 24*, 300-305.

Kojima, D., Okano, T., Fukada, Y., Shichida, Y., Yoshizawa, T., & Ebrey, T. G. (1992). Cone visual pigments are present in gecko rod cells. *Proc Natl Acad Sci U S A, 89*, 6841-5.

Kondrashev, S. L., Gamburtzeva, A. G., Gnjubkina, V. P., Orlov, O., & Pam Thi, M. Y. (1986). Coloration of corneas in fish. A list of species. *Vision Res, 26*, 287-90.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol Rev, 47*, 1-45.

Kumar, R., Chen, S., Scheurer, D., Wang, Q. L., Duh, E., Sung, C. H., Rehemtulla, A., Swaroop, A., Adler, R., & Zack, D. J. (1996). The bZIP transcription factor Nrl stimulates rhodopsin promoter activity in primary retinal cell cultures. *J Biol Chem, 271*, 29612-8.

Kumar, S. K., Tamura, K., & Nei, M. (1993). MEGA molecular evolutionary genetic analysis (Version 1.0). Pennsylvania: Pennsylvania State University.

Kunz, Y. W., & Ennis, S. (1983). Ultrastructural diurnal changes of the retinal photoreceptors in the embryo of a viviparous teleost (Poecilia reticulata P.). *Cell Differ, 13*, 115-23.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*, 680-5.

Lamb, T. D., & Pugh, E. N. J. (1990). Physiology of transduction and adaptation in rod and cone photoreceptors. *Seminars in the Neurosciences, 2*, 3-13.

Larhammar, D., & Risinger, C. (1994). Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Molecular phylogenetics and evolution, 3*, 59-68.

Lem, J., Applebury, M. L., Falk, J. D., Flannery, J. G., & Simon, M. I. (1991). Tissue-specific and developmental regulation of rod opsin chimeric genes in transgenic mice. *Neuron, 6*, 201-10.

Levine, J. S., & MacNichol, E. F. (1982). Colour Vision in fishes. *Sci Am, 246*.

Levine, M. W. (1982). Retinal processing of intrinsic and extrinsic noise. *J Neurophysiol, 48*, 992-1010.

Lewin, B. (1990). *Genes IV*. Oxford University Press, Oxford.

Li, J. J., & Herskowitz, I. (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system [see comments]. *Science, 262*, 1870-4.

Li, W.-H. (1997). *Molecular Evolution*. (1st ed.): Sinauer Associates Inc.

Li, W.-H., & Graur, D. (1991). *Fundamentals of Molecular evolution*. Sunderland, Mass.: Sinauer Associates, Inc.

Lin, S. W., Imamoto, Y., Fukada, Y., Shichida, Y., Yoshizawa, T., & Mathies, R. A. (1994). *Biochemistry, 33*, 2151-2160.

Lin, S. W., Kochendoerfer, G. G., Carroll, K. S., Wang, D., Mathies, R. A., & Sakmar, T. P. (1998). Mechanisms of spectral tuning in blue cone visual pigments. Visible and raman spectroscopy of blue-shifted rhodopsin mutants. *J Biol Chem, 273*, 24583-91.

Lin, S. W., & Sakmar, T. P. (1999). Colour tuning mechanisms of visual pigments`, *Rhodopsins and phototransduction* (224 ed., pp. 124-141). Chichester: John Wiley & Sons Ltd.

Lipetz, L. E. (1984). Pigment types, densities and concentrations in cone oil droplets of Emydoidea blandingii. *Vision Res, 24*, 605-12.

Locket, N. A. (1999). *Vertebrate photoreceptors*: Kluwer Academic Publishers.

Lyall, A. H. (1957). The growth of trout retina. *Quaterly Journal of Microscopical Science, 98*.

Lythgoe, J. N. (1979). *The Ecology of Vision.* Oxford: Clarendon press.

Mandrioli, M., Cuoghi, B., Marini, M., & Manicardi, G. C. (2000). Cytogenetic analysis of the pufferfish Tetraodon fluviatilis (Osteichthyes) [In Process Citation]. *Chromosome Res, 8*, 237-42.

Marc, R. E., & Sperling, H. G. (1976). Colour receptor identities of goldfish cones. *Science, 191*, 487-488.

Martin, A. P., Naylor, G. J., & Palumbi, S. R. (1992). Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature, 357*, 153-5.

Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A, 90*, 4087-91.

Mastrangelo, I. A., Courey, A. J., Wall, J. S., Jackson, S. P., & Hough, P. V. (1991). DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. *Proc Natl Acad Sci U S A, 88*, 5670-4.

Mathies, R., & Stryer, L. (1976). Retinal has a highly dipolar vertically excited singlet state: implications for vision. *Proc Natl Acad Sci U S A, 73*, 2169-73.

MathSoft. (1999). Axum (Version 6.0). Seattle: Data Analysis Products Division, MathSoft Inc.

McKnight, S., & Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. *Cell, 46*, 795-805.

Mednick, A. S., Berk, M. F., & Springer, A. D. (1988). Asymmetric distribution of cells in the inner nuclear and cone mosaic layers of the goldfish retina. *Neurosci Lett, 94*, 241-6.

Mismer, D., & Rubin, G. M. (1987). Analysis of the promoter of the ninaE opsin gene in Drosophila melanogaster. *Genetics, 116*, 565-78.

Mollon, J. D., Bowmaker, J. K., & Jacobs, G. H. (1984). Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proc R Soc Lond B Biol Sci, 222*, 373-99.

Morabito, M. A., Yu, X., & Barnstable, C. J. (1991). Characterization of developmentally regulated and retina-specific nuclear protein binding to a site in the upstream region of the rat opsin gene. *J Biol Chem, 266*, 9667-72.

Moses, K., Ellis, M. C., & Rubin, G. M. (1989). The glass gene encodes a zinc-finger protein required by Drosophila photoreceptor cells. *Nature, 340*, 531-6.

Moses, K., & Rubin, G. M. (1991). Glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing Drosophila eye. *Genes Dev, 5*, 583-93.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol, 51*, 263-73.

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol, 155*, 335-50.

Munk, O. (1966). Ocular anatomy of some deep-sea teleosts. *Dana Rep, 70*, 1-62.

Nathans, J. (1987). Molecular biology of visual pigments. *Annu Rev Neurosci, 10*, 163-94.

Nathans, J. (1990). Determinants of visual pigment absorbance: role of charged amino acids in the putative transmembrane segments. *Biochemistry, 29*, 937-42.

Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., & et al. (1989). Molecular genetics of human blue cone monochromacy. *Science, 245*, 831-8.

Nathans, J., & Hogness, D. S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell, 34*, 807-14.

Nathans, J., & Hogness, D. S. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc Natl Acad Sci U S A, 81*, 4851-5.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B., & Hogness, D. S. (1986a). Molecular genetics of inherited variation in human color vision. *Science, 232*, 203-10.

Nathans, J., Thomas, D., & Hogness, D. S. (1986b). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science, 232*, 193-202.

Neitz, M., Neitz, J., & Jacobs, G. H. (1991). Spectral tuning of pigments underlying red-green color vision. *Science, 252*, 971-4.

Nelson, G. J. (1973a-a). Notes on the structure and relationships of certain Cretaceous and Eocene teleostean fishes. *Am. Mus. Novit., 2524*, 31.

Nelson, G. J. (1973a-b). Notes on the structure and relationships of certain Cretaceous and eocene teleostean fishes. *Am. Mus. Novit., 2524*, 31.

Nelson, J. S. (1994). *Fishes of the World.* (3rd ed.): John Wiley & Sons, Inc.

Nie, Z., Chen, S., Kumar, R., & Zack, D. J. (1996). RER, an evolutionarily conserved sequence upstream of the rhodopsin gene, has enhancer activity. *J Biol Chem, 271*, 2667-75.

Ohtsuka, T. (1985). Relation of spectral types to oil droplets in cones of turtle retina. *Science, 229*, 874-7.

Okano, T., Kojima, D., Fukada, Y., Shichida, Y., & Yoshizawa, T. (1992). Primary structures of chicken cone visual pigments: vertebrate rhodopsins have evolved out of cone visual pigments. *Proc Natl Acad Sci U S A, 89*, 5932-6.

Oprian, D. D., Molday, R. S., Kaufman, R. J., & Khorana, H. G. (1987). Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc Natl Acad Sci U S A, 84*, 8874-8.

Oxford Molecular Group. (1996). MacVector.

Oxford Molecular Group. (1997). GeneWorks 2.5.1N.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science, 289*, 739-45.

Pankhurst, N. W., & Lythgoe, J. N. (1983). Changes in vision and olfaction during sexual maturation in the European eel *Anguilla anguilla. J. Fish Biol., 23*, 229-240.

Partridge, J. C. (1986). Microspectrophotometry of vertebrate photoreceptors. *University of Bristol.*

Partridge, J. C. (1989). The visual ecology of avian cone oil droplets. *Journal of Comparative Physiology A, 165.*

Partridge, J. C. (1999). *Adaptive Mechanisms in the Ecology of Vision: Biophysical adaptations.* (1st ed.). (Vol. 1): Kluwer Academic Publishers.

Partridge, J. C., Archer, S. N., & Lythgoe, J. N. (1988). Visual pigments in the individual rods of deep-sea fishes. *J. Comp. Physiol. A, 162.*

Patterson, C. (1968). The caudal skeleton in Lower Liassic pholidophoroid fishes. *Bull. Br. Mus. Nat. Hist. (Geol.), 16*, 203-239.

Patterson, C., & Rosen, D. E. (1977). Review of ichthyodectiform and other Mesozoic teleost fishes and the theory and practice od classyfying fossils. *Bull. Amer. Mus. Nat. Hist., 158*, 81-172.

Peichl, L., & Moutairou, K. (1998). Absence of short-wavelength sensitive cones in the retinae of seals (Carnivora) and African giant rats (Rodentia). *Eur J Neurosci, 10*, 2586-94.

Pogozheva, I.D., Lomize, A.L., & Mosberg, H.I. (1997). The transmembrane 7-alpha bundle of rhodopsin; distance geometry calculations with hydrogen bonding constraints. *Biophys J, 72*, 1963-85.

Pugh, E. N., Jr. (1999). Variability in single photon responses: a cut in the Gordian knot of rod phototransduction? *Neuron, 23*, 205-8.

Raymond, P. A. (1985). Cytodifferentiation of photoreceptors in larval goldfish: delayed maturation of rods. *J Comp Neurol, 236*, 90-105.

Raymond, P. A., Barthel, L. K., Rounsifer, M. E., Sullivan, S. A., & Knight, J. K. (1993). Expression of rod and cone visual pigments in goldfish and zebrafish: a rhodopsin-like gene is expressed in cones. *Neuron, 10*, 1161-74.

Register, E. A., Yokoyama, R., & Yokoyama, S. (1994). Multiple origins of the green-sensitive opsin genes in fish. *J Mol Evol, 39*, 268-73.

Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D. J., & Swaroop, A. (1996). The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc Natl Acad Sci U S A, 93*, 191-5.

Robinson, S. R. (1994). Early vertebrate colour vision. *Nature, 367*, 121.

Roorda, A. W., D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature, 397*, 520-522.

Rosenfeld, P. J., & Kelly, T. J. (1986). Purification of nuclear factor I by DNA recognition site affinity chromatography [published erratum appears in J Biol Chem 1986 Jul 25;261(21):10015]. *J Biol Chem, 261*, 1398-408.

Rushton, W. (1963). A cone pigment in the protanope. *J. Physiol, 168*, 345-359.

Saitou, N., & Nei, M. (1987). The neighbour joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol., 4*, 406-425.

Sakmar, T. P., Franke, R. R., & 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*, 8309-13.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*, 5463-7.

Schertler, G. F., & Hargrave, P. A. (1995). Projection structure of frog rhodopsin in two crystal forms. *Proc Natl Acad Sci U S A, 92*, 11578-82.

Schertler, G. F., Villa, C., & Henderson, R. (1993). Projection structure of rhodopsin. *Nature, 362*, 770-2.

Schug, J., & Overton, G. C. (1997). TESS: Transcription Element Search Software (Version 1.2.2). Pennsylvania: Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania.

Shaaban, S. A., & Deeb, S. S. (1998). Functional analysis of the promoters of the human red and green visual pigment genes. *Invest Ophthalmol Vis Sci, 39*, 885-96.

Sheldon, F. H. (1987). Rates of single-copy DNA evolution in herons. *Mol Biol Evol, 4*, 56-69.

Sheshberadaran, H., & Takahashi, J. S. (1994). Characterization of the chicken rhodopsin promoter: identification of retina-specific and glass-like protein binding domains. *Mol Cell Neurosci, 5*, 309-18.

Shichida, Y., Imai, H., Imamoto, Y., Fukada, Y., & Yoshizawa, T. (1994). Is chicken green-sensitive cone visual pigment a rhodopsin-like pigment? A comparative study of the molecular properties between chicken green and rhodopsin. *Biochemistry, 33*, 9040-4.

Shields, D. C., Sharp, P. M., Higgins, D. G., & Wright, F. (1988). "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol, 5*, 704-16.

Shyue, S. K., Boissinot, S., Schneider, H., Sampaio, I., Schneider, M. P., Abee, C. R., Williams, L., Hewett-Emmett, D., Sperling, H. G., Cowing, J. A., Dulai, K. S., Hunt, D. M., & Li, W. H. (1998). Molecular genetics of spectral tuning in New World monkey color vision. *J Mol Evol, 46*, 697-702.

Slobodyanyuk, S., Kirilchik, S. V., Pavlova, M. E., Belikov, S. I., & Novitsky, A. L. (1995). The evolutionary relationships of two families of cottoid fishes of Lake Baikal (east Siberia) as suggested by analysis of mitochondrial DNA. *J Mol Evol, 40*, 392-9.

Snell, R., & Lemp, M. (1989). *Clinical anatomy of the eye*: Blackwell Scientific.

Stavenga, D. G., Smits, R. P., & Hoenders, B. J. (1993). Simple exponential functions describing the absorbance bands of visual pigment spectra. *Vision Res, 33*, 1011-7.

Stell, W. K., & Harosi, F. I. (1975). Cone structure and visual pigment content in the retina of the goldfish. *Vision Res, 16*, 647-657.

Stell, W. K., & Harosi, F. I. (1976). Cone structure and visual pigment content in the retina of the goldfish. *Vision Res, 16*, 647-57.

Stenkamp, D. L., Barthel, L. K., & Raymond, P. A. (1997). Spatiotemporal coordination of rod and cone photoreceptor differentiation in goldfish retina. *J Comp Neurol, 382*, 272-84.

Stewart, C. B. (1993). The powers and pitfalls of parsimony [see comments]. *Nature, 361*, 603-7.

Stryer, L. (1988). *Biochemistry*. (3rd ed.): WH Freeman and Company.

Su, C. Y., Lim, J., & Tsai, H. J. (2000). Structural characterization and transcriptional pattern of two types of carp rhodopsin gene [In Process Citation]. *Comp Biochem Physiol B Biochem Mol Biol, 125*, 37-45.

Sun, H., Macke, J. P., & Nathans, J. (1997). Mechanisms of spectral tuning in the mouse green cone pigment. *Proc Natl Acad Sci U S A, 94*, 8860-5.

Swaroop, A., Xu, J. Z., Pawar, H., Jackson, A., Skolnick, C., & Agarwal, N. (1992). A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc Natl Acad Sci U S A, 89*, 266-70.

Timoshkin, O. A. (1997, 1997). *Biodiversity of Baikal Fauna: state-of-the-art.* Paper presented at the New Scope on Boreal Ecosystems in East Siberia. Proceedings of the International Workshop, Kyoto, Japan.

Tokunaga, F., Hisatomi, O., Satoh, T., Taniguchi, Y., Matsuda, S., Imanishi, Y., Honkawa, H., Takahashi, Y., Kobayashi, Y., Yoshida, M., & Tsukahara, Y.

(1999). Evolution of visual pigments and related molecules. *Novartis Found Symp, 224*, 44-52.

Tsin, A. T., Liebman, P. A., Beatty, D. D., & Drzymala, R. (1981). Rod and cone visual pigments in the goldfish. *Vision Res, 21*, 943-6.

Unger, V. M., Hargrave, P. A., Baldwin, J. M., & Schertler, G. F. (1997). Arrangement of rhodopsin transmembrane alpha-helices. *Nature, 389*, 203-6.

Vihtelic, T. S., Doro, C. J., & Hyde, D. R. (1999). Cloning and characterization of six zebrafish photoreceptor opsin cDNAs and immunolocalization of their corresponding proteins. *Vis Neurosci, 16*, 571-85.

Walls, G.L. (1942). *The vertebrate eye and its adaptive radiation.* Cranbrook Institute of Science. Bloomfield hills, Michigan.

Wang, Y., Smallwood, P. M., Cowan, M., Blesh, D., Lawler, A., & Nathans, J. (1999). Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors [see comments]. *Proc Natl Acad Sci U S A, 96*, 5251-6

Wang, M. M., & Reed, R. R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature, 364*, 121-6.

Wang, S. Z., Adler, R., & Nathans, J. (1992a). A visual pigment from chicken that resembles rhodopsin: amino acid sequence, gene structure, and functional expression. *Biochemistry, 31*, 3309-15.

Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., Gearhart, J., & Nathans, J. (1992b). A locus control region adjacent to the human red and green visual pigment genes. *Neuron, 9*, 429-40.

Wang, Z., Asenjo, A. B., & Oprian, D. D. (1993). Identification of the Cl(-)-binding site in the human red and green color vision pigments. *Biochemistry, 32*, 2125-30.

Weiss, R. F., Carmack, E. C., & Koropalov, V. M. (1991). Deep-water renewal and biological production in Lake Baikal. *Nature, 349*, 665-669.

Weitz, C. J., & Nathans, J. (1992). Histidine residues regulate the transition of photoexcited rhodopsin to its active conformation, metarhodopsin II. *Neuron, 8*, 465-72.

Welkowitz, J., Ewen, R. B., & Cohen, J. (2000). *Introductory statistics for the behavioural sciences*. (5th ed.): Harcourt Brace College Publications.

Wetts, R., & Fraser, S. E. (1988). Multipotent precursors can give rise to all major cell types of the frog retina. *Science, 239*, 1142-5.

Whitmore, A. V., & Bowmaker, J. K. (1989). Seasonal variation in cone sensitivity and shortwave absorbing visual pigments in the rudd, *Scardinius erythrophthalmus*. *J. Comp. Physiol. A, 166*, 103-115.

Wilkie, S. E., Robinson, P. R., Cronin, T. W., Poopalasundaram, S., Bowmaker, J. K., & Hunt, D. M. (2000). Spectral tuning of avian violet and ultraviolet-sensitive visual pigments. *Biochemistry, 39*, 7895-901.

Wilkie, S. E., Vissers, P. M., Das, D., Degrip, W. J., Bowmaker, J. K., & Hunt, D. M. (1998). The molecular basis for UV vision in birds: spectral characteristics, cDNA sequence and retinal localization of the UV-sensitive visual pigment of the budgerigar (Melopsittacus undulatus). *Biochem J, 330*, 541-7.

Williams, A. J., Hunt, D. M., Bowmaker, J. K., & Mollon, J. D. (1992). The polymorphic photopigments of the marmoset: spectral tuning and genetic basis. *Embo J, 11*, 2039-45.

Williams, D. R. (1986). Seeing through the photoreceptor mosaic. *Trends in neurosciences, 9*, 193-198.

Wood, P., & Partridge, J. C. (1993). Opsin substitution induced in retinal rods of the eel (*Anguilla anguilla* (L.): a model for G-protein linked receptors. *Proc. R. Soc. Lond. B, 254*, 227-232.

Wright, M. W. (2000). *Retinal photoreceptor complement of Paleognathous birds.* , University College london, London.

Yau, K. W. (1994). Phototransduction mechanism in retinal rods and cones. The Friedenwald Lecture. *Invest Ophthalmol Vis Sci, 35*, 9-32.

Yokoyama, R., Knox, B. E., & Yokoyama, S. (1995). Rhodopsin from the fish, Astyanax: role of tyrosine 261 in the red shift. *Invest Ophthalmol Vis Sci, 36*, 939-45.

Yokoyama, S. (1994). Gene duplications and evolution of the short wavelength-sensitive visual pigments in vertebrates. *Mol Biol Evol, 11*, 32-9.

Yokoyama, S. (1997). Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. *Annu Rev Genet, 31*, 315-36.

Yokoyama, S., & Radlwimmer, F. B. (1998). The "five-sites" rule and the evolution of red and green color vision in mammals. *Mol Biol Evol, 15*, 560-7.

Yokoyama, S., & Radlwimmer, F. B. (1999). The molecular genetics of red and green color vision in mammals. *Genetics, 153*, 919-32.

Yokoyama, S., & Yokoyama, R. (1989). Molecular evolution of human visual pigment genes. *Mol Biol Evol, 6*, 186-97.

Young, R. W. (1971). Shedding of discs from rod outer segments in the rhesus monkey. *J Ultrastruct Res, 34*, 190-203.

Yu, C.-Y., Motamed, K., Chen, J., Bailey, A. D., & Shen, C.-K. J. (1991). The CACC box upstream of human embryonic globin gene binds Sp1 and is a functional promoter element *in vitro* and *in vivo. The journal of Biological Chemistry, 266*, 8907-8915.

Yu, X., Chung M., Morabito M. A., and Barnstable C. J. (1993). Shared nuclear protein binding sites in the upstream region of the rat opsin gene. *Biochem. Biophys. Res. Commun,, 191*, 76-82.

Yu, X., Leconte, L., Martinez, J. A., & Barnstable, J. B. (1996). Ret 1, a *cis*-acting element of the rat opsin promoter can direct gene expression in rod photoreceptors. *Journal of Neurochemistry, 67*, 2494-2504.

Zack, D. J., Bennett, J., Wang, Y., Davenport, C., Klaunberg, B., Gearhart, J., & Nathans, J. (1991). Unusual topography of bovine rhodopsin promoter-lacZ fusion gene expression in transgenic mouse retinas. *Neuron, 6*, 187-99.

Bezrukov, L. B., Budnev, N. M., Galperin, M. D., Dobrynin, V. I., Dudkin, G. N., Zurbanov, V. L., Kokhomskij, A. G., Nikiforov, S. A., Poleschuk, V. A., Sherstjankin, P. P., & Sherstakov, A. A. (1984). On the glow of deep waters of Lake Baikal. *Proceedings of the USSR Academy of Sciences, 227*, 1240-1224. (In Russian.).