Molecular Genetics of Mammalian Blue Cone Pigment Genes

Binoy Appukuttan

Institute of Ophthalmology
University College, London
University of London

Thesis submitted to the University of London for the degree of Doctor of Philosophy

1997

ProQuest Number: 10055423

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10055423

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

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

Abstract

Since the isolation of the human cone opsin genes 10 years ago, the advance of molecular biology techniques has played a significant role in increasing our knowledge on the genetics of visual pigment genes from a wide variety of vertebrate and invertebrate species. However, only two non-primate mammalian blue cone pigment (BCP) genes have been characterised to date, and the amino acid residues which are responsible for maintaining the short wavelength sensitivity of this particular class of pigment (the S group) have yet to be determined. Which *cis*-acting elements confer photoreceptor or cone-specific gene expression? The identification of cone and rod-specific *cis*-acting elements will not only increase our understanding of photoreceptor-specific gene expression but also contribute to gene therapy applications which target inherited retinal disorders.

The BCP gene from the domestic pig (*Sus scrofa*) was sequenced and characterised, and analysis of the deduced amino acid sequence showed that this opsin is a new member of the S group of visual pigments. Comparison of the S group of opsins with other vertebrate visual pigments resulted in the identification of amino acids which are specific to this family of opsins. Amino acids at 16 positions were identified as putative spectral tuning sites which may contribute to maintaining the short λ_{max} observed in the S group of opsins.

1194 bp, 949 bp, 573 bp and 570 bp of the 5′ flanking regions from the BCP genes of human, porcine, talapoin monkey and capuchin monkey were sequenced, respectively. A comparison of 5′ flanking BCP gene sequences, between species, revealed several evolutionary conserved motifs, one of which is the photoreceptor conserved element (PCE1). The cells of a retinoblastoma cell line, Weri-Rb1, were shown to express cone-specific transcripts by RT-PCR. It was determined that nuclear proteins isolated from the Weri-Rb1 cells could be used for electrophoretic band shift assay (EBSA) experiments on fragments of the 5′ flanking region of the human BCP gene. Competitive EBSA, using short double-stranded oligonucleotides designed to sequences within the human BCP gene 5' flanking region, showed that the specific regions within the BCP proximal promoter to which proteins were binding could be identified.

The yeast, *Pichia pastoris* was tested as an alternative to mammalian cells for the *in vitro* expression of opsins. The human BCP gene was found to be associated with two *Alu* repeats, one is situated within intron 3, and the other is between 1194 and 916 bp upstream, of the gene.

Acknowledgments

I wish to thank Professor David M Hunt, my supervisor, for giving me the opportunity to study for a PhD and for the support and advice which he has provided throughout the course of this study, especially during the latter stages for his criticism and corrections of this thesis. I would like to thank the Wellcome Trust for awarding the grant which allowed this research to be undertaken.

I also wish to thank Dr Vaiman Marcel (Laboratoire de Radiobiolgie Appliquée, 78352 Jouy-en-Josas, France) for the kind donation of a pig genomic cosmid library and Dr Nick Occleston (Dept. of Pathology) for the sample of fibroblast cells which were used in this study.

Throughout the course of this study I have had the pleasure of working and associating with a wide variety of wonderful people who have helped me at various stages of this PhD and have made my stay at the Institute one of the happiest and most memorable of experiences. The following is a feeble attempt to thank these wonderful people, who are too numerous to remember all and so please forgive me if I have forgotten your name. I should start with Ro (Ro Ro Ro your boat, 'I love my McCoys/Brannigans') Kelsell one of the kindest and nicest persons I have met, who kindly gave up her time to read and correct drafts of this thesis, for all the encouragement and 'Carry On' talks; Neil (mutant Tamil, who has kept me sane (well almost), for all his assistance, especially during the latter stages of writing up), and other members of his family (eg Priya, Mr Drippy Candles,...) for making me so welcome; Kamal (Mango Milkshake Maker, 58 E) Dulai, for all his advice, time and enormous knowledge on computers, expert help on the ABI sequencer, thesis formatting, diet tips, all the fun times and for introducing me to the USA (and his gracious family); The USA FC: Reshma (honorary AA person, 'I love Art!'), a very special person/scientist who has been there since the very beginning (marmoset BCP) through all the fun, laughter and tears, thank you and Alison (Miss Style Challenge, 'I love Minnesota') Fab Gromit friend and Su (I love/hate my Ball & Chain, i.e. sister!) Wijjy for a most excellent time in the USA; The Proof Readers: Karen ('I'll just have a fag') the Elder, for help with upstream refs etc (keep up the good work), Naheed (what can I say?!) Kangaroo, for invaluable advice and assistance with the band shift assays (& the hair!) and Jane ('Transcription Factor person', I wish you were here 10 months ago!) Sowden, for discussions on TFs and EBSA;

Sarah ('Dear Lump', 'I love Jarvis'), for protein discussions, the CDs, the fun and of course the peas; The Veggies: Dr Zoeiki ('Fidel for the president of USA, Viva le proletarians'), Jilly (Beano organiser, 'Captain Jean-Luc is gorgeous') Willy and Dawny ('X is wonderful') for all the good times and veggie stuff (please note that good times and veggie stuff are not necessarily synonymous!); The Mutilinguals: Evelyne (Bonjour Bonjour), Fabiana (Tesoro, Bellisama) and Zara (the giggler/snorter) for increasing my knowledge of European languages by 10-20 words; Jack Daniels buddies: Dr David Kelsell and Dr Ian; The Schizo-group: Alexander ('I can still taste THAT wine!') the Great Bohemian Posh One, Gorgonzola (Godzilla) and Sinthuja (Sicky, Shiatsu) Terrorist I; The Pensioner: Sue (3 legged hopalong), for spectral tuning discussions; The Illegal Immigrants: Debi ('Lets take her to Ikea', spinster chicken gobbler) who still has not escaped, Erangutan (Terrorist II), Mani ('AllWind, I don't beeelieve it!'), Mai (Terrorist III), Sana (big hair) and Vinny (stop speaking Hindi!); Football Junkies: Jim (Boat-Crusher) Bellingham, for his informative advice on UPW-PCR, Emma (Le grande Gluteus Maximus) Tortilla, for the rod opsin primers; The Protein Fiddlers: Ed (a good footballer) the Bolt, Richard ('Don't tease me I'm very sensitive') the Lionheart, Seraphina (baby of the lab) and Woodcock (Wife I) for help with SDS-PAGE, Natalie (young doctor), Trasher ('that malarkey'), Jenny (nice line in jumpers) and Dr Warren (Tetrapyrrole-Man, 'Are you still writing your thesis!?'); Cataract-group: Donna (X-Phile) Kebab, Alfonso ('Trust me I'm a medic'), Marcela (Nachos) V and Alan (Chelsea reject) Sheils who is getting on a bit, but is still on the ball!; Lesley (Fly-Catcher) Heath; A (17th century) Payne; Chris ('I want to WIN at football!') Inglehearn, for discussions and refs on polymorphic repeats; RACHEL!! (answer that phone!!!); Jeffrey (I have a candidate gene but I'm NOT telling anybody) Keen, for the 90 gms paper; Cheryl (overdue bun in the oven) Gregory; Kevin ('I'm Welsh & I'm proud of it'); Martin (always on the go/Auf weidersien) Reichel; Robin (the perfect politician?) Ali; Stephanie (a smart cookie) of Islington; Lloyd and Sunil (the weight lifters); Soraya (the South African); Myrto (Yasoo Yasoo); Annika ('I am from Sweeden') Soderpalm; Dr Andy (Pandy Pudding & Pie) Hope; Jude (I don't have a TV so I'll act like Siggy Freud) Fitzgibbon, for all the side-line projects and teaching me the technique of cosmid screening for which I am forever grateful; Texas Tim ('pop round ure ouse') Stout, for a 'Tickety Boo' time and the lecture on the advantages & disadvantages of side-line projects; and Prof Shomi (Demon spin bowler, you will get that wicket/gene one day soon!) who has

always found me amusing for some odd reason. If I may quote a dear rotund friend "I love you all".

I would like to thank members of my family, here and abroad, and family friends, too numerous to mention all by name, who have supported and cared for me throughout this study. Thank you to Mr N Surendran and family for all their help with various things that are too numerous to account. Thank you to my sisters with a very big thank you to Biju, who has been very helpful and understanding and for putting up with me during the last few months. Finally a very huge and special thank you to my mother, who has been very patient, supportive, encouraging and understanding, especially during the final stages of this research. I am extremely grateful to my mother for providing the sustenance and the love that has kept me going, (and of course for all the biryanis, appums, idi appums, dosais etc etc).

Dedication

For my mum
and to the memory of my father
(who believed that education was the key to success)

Table of Contents

Abstract 2

Acknowledgments 3

Dedication 6

Table of Contents 8

List of Figures 13

List of Tables 16

Chapter 1 17

Introduction 17

1.1 Introduction 17
1.2 Eye structure and function 17
1.3 The retina 20
1.4 Photoreceptor structure 20
1.5 Photoreceptor function 23
1.6 Rod and cone cells 27
1.7 Visual pigments 29
1.8 Phototransduction 31
1.9 Rod opsin 36
1.10 Cone pigments 37
1.11 Cone opsins 37
1.12 Spectral sensitivity 41
1.13 Colour vision 42
1.14 Classification and evolution of visual pigments 43
1.15 Molecular genetics of visual pigments 45
1.16 Gene expression 47
1.16.1 Transcription 47
1.16.2 Enhancer elements 50
1.16.3 Tissue-specific expression 50
1.16.4 Photoreceptor-specific expression 51
1.17 Aims 55

Chapter 2 56

Materials and Methods 56

2.1 DNA extraction 56
2.2 Polymerase chain reaction (PCR) 58
2.3 DNA digestion with restriction endonucleases 64
2.4 Electrophoresis and visualisation of DNA 65
2.5 Ligation of DNA inserts into plasmid vectors 65
2.6 Transformation of competent *E coli* 66
2.7 Growth media and selection of recombinants 66
2.8 Screening recombinant *E coli* for inserts 67
2.9 Screening human and porcine cosmid genomic libraries 67
2.10 Southern blotting 70
2.11 Autoradiography 71
2.12 DNA sequencing 71
2.13 Denaturing polyacrylamide gel electrophoresis 73
2.14 Subcloning of cosmid DNA into plasmids 74
2.15 Routine cell culture 75
2.16 Isolation of poly-A^+ RNA 75
2.17 cDNA synthesis 76
2.18 Protein extractions and quantification 77
2.19 DNA mobility shift assay 79
2.20 The *Pichia pastoris* yeast expression system 81
2.21 Computing and software 87
2.22 Reagents and suppliers 88

Chapter 3 91

Characterisation of the Porcine Blue Cone Pigment Gene and Identification of Spectral Tuning Sites

3.1 Introduction 91
3.1.1 Porcine visual pigments 91
3.1.2 Terminology 92
3.1.3 Spectral tuning 92
3.1.4 Predicting spectral tuning sites 95
3.2 Results 99

3.2.1 Isolation and subcloning of a genomic cosmid clone containing the porcine BCP gene 99
3.2.2 The porcine BCP is a member of the S group of opsins 99
3.2.3 Genomic structure of the porcine BCP gene 104
3.2.4 Comparison of the S group proteins 109
3.2.41 Identification of amino acids which are exclusive to individual S opsins 113
3.2.42 Identification of S group-specific amino acids ... 115
3.2.5 Identification of spectral tuning sites 119
3.2.51 S group-specific spectral tuning sites: 122
(i) Non-conserved amino acid substitutions 122
(ii) Conservative amino acid substitution 123
3.2.52 Other spectral tuning sites 123
3.2.53 Spectral tuning sites in extracellular loops 125
3.2.54 Ultraviolet spectral tuning sites 126
3.3 Discussion 128
3.3.1 The porcine BCP: a new S group opsin 128
3.3.2 Analysis of S group proteins 130
3.3.3 Evaluation of the putative spectral tuning sites 132
3.4 Summary 134

Chapter 4 **136**

Expression of the Human Blue Cone Pigment Gene in the *Pichia pastoris* Yeast Expression System

4.1 Introduction 136
4.1.1 *In vitro* expression of opsin genes 136
4.1.2 The *Pichia pastoris* yeast expression system 137
4.2 Results 139
4.2.1 Isolation and cloning of a human BCP complementary DNA into a yeast expression vector 139
4.2.2 Transformation and selection of yeast recombinants containing the human BCP cDNA .. 146
4.2.3 Pilot expression studies of H^+/M^- yeast clones 148
4.2.4 Further analysis of H^+/M^- yeast clones 148
4.3 Discussion 152

Chapter 5 156

Characterisation of the 5′ proximal promoter region of the mammalian Blue Cone Pigment Gene

5.1 Introduction 156
5.1.1 Aims 156
5.1.2 Photoreceptor gene-specific enhancer elements 157
5.1.3 Retinoblastoma 158
5.2 Results 159
5.2.1 Isolation and sequencing of the human and porcine blue cone pigment (BCP) gene 5′ flanking region 159
5.2.2 Isolation and sequencing of the 5′ flanking region of the blue cone pigment gene from an Old World and a New World primate 167
5.2.3 Analysis of the promoter regions of the porcine and primate BCP genes 172
5.2.4 Evolutionary conservation of upstream sequences in mammalian BCP genes 175
5.2.5 Identification of putative protein binding sites within the extended porcine and human BCP gene 5′ flanking regions 180
5.2.6 Characterisation of the Weri-Rb1 cell line 185
5.2.61 Isolation of nuclear proteins from Weri-Rb1 and fibroblast cell lines 187
5.2.7 Electrophoretic band shift assays (EBSA) on the human BCP gene promoter 187
5.2.71 Competitive electrophoretic band shift assays ... 197
5.3 Discussion 202
5.3.1 Evaluation of the primate products obtained by UPW-PCR 202
5.3.2 The photoreceptor conserved element (PCE1) 203
5.3.3 Putative protein binding sites 205
5.3.4 Weri-Rb1 cells do express BCP transcripts 208
5.3.5 Analysis of the electrophoretic band shift assays 208

Chapter 6 **211**

Characterisation and Localisation of the Human Blue Cone Pigment Genomic Cosmid Clone

6.1 Introduction 211
6.1.1 Chromosomal assignment of the human cone opsin genes 211
6.1.2 Repetitive DNA sequences 211
6.2 Results 212
6.2.1 Localisation of the human blue cone pigment gene to chromosome band 7q31.3-32 212
6.2.2 Dinucleotide and trinucleotide repeats found within the human BCP cosmid 212
6.2.3 *Alu* repeat elements situated within the BCP cosmid .. 215
6.3 Discussion 219

Chapter 7 **226**

Conclusions and Future Prospects **226**

Abbreviations **229**

Appendix **231**

Publications **232**

References **233**

INSTITUTE OF OPHTHALMOLOGY LIBRARY BATH STREET EC1V 9EL

List of Figures

Figure No.		Page
1.1	A cross-section of the human eye	18
1.2	Cross-section of the primate retina	21
1.3	Mammalian rod and cone photoreceptors	22
1.4	Turnover of rod cell discs	24
1.5	The dark current	26
1.6a	Photoreceptor outer segment	28
1.6b	Two-dimensional depiction of the opsin protein	30
1.7	Transverse section of a visual pigment	32
1.8	The retinal cycle of rod opsin	33
1.9	The phototransduction cascade	34
1.10	Absorption spectra curves	39
1.11	Phylogenetic tree of the visual pigment families	46
1.12	Schematic representation of two zinc finger motifs	49
1.13	Octamer motifs of the immunoglobulin genes	52
2.1	Vectorette protocol	63
2.2	Identification of recombinant yeast clones	82-83
3.1	Baldwin's model of the seven helical transmembranes	98
3.2A	Digestion of a porcine BCP gene cosmid clone	100
3.2B	Southern blot of a porcine BCP gene cosmid clone	101
3.3	Identification of *Bam*HI subclones	102
3.4	Nucleotide sequence of the coding region of the porcine BCP gene	103
3.5	Alignment of the porcine and human BCP amino acid sequences	105
3.6	Alignment of the porcine and human BCP gene sequences	107
3.7	Amplification of human and porcine BCP gene introns	110
3.8	Amino acid sequence comparison of the S group of opsins	111-112
3.9	Classification of the primate suborder Anthropoidea	116
3.10	Baldwin's model of the helical transmembrane regions	124
4.1A	First round amplification of a human BCP cDNA	140
4.1B	Second round amplification of a human BCP cDNA	140
4.2	Sequence of the full length human BCP cDNA	141

4.3	The yeast pPIC9 expression vector	142
4.4	*Bam*HI digests of pPIC9/BCP cDNA clones	144
4.5	The pPIC9 expression vector multi-cloning site with attached αF secretion signal sequence	145
4.6	*Bgl*II digests of pPIC9/BCP cDNA clones	147
4.7a	Secreted proteins from *P pastoris* pPIC9/BCP cDNA clones	149
4.7b	Secreted proteins from *P pastoris* pPIC9/BCP cDNA clones	149
4.7c	Secreted and cell membrane proteins from *P pastoris* pPIC9/BCP cDNA clones	150
4.8	PCR of yeast genomic DNA from 46 colonies identified as H^+/M^- clones	151
4.9	Enhanced expression of secreted proteins	153
4.10	Cell membrane protein preps	153
4.11	PCR amplification of cDNA from H^+/M^- clones	154
5.1	PCR of a human BCP gene cosmid clone	161
5.2	PCR of the human BCP gene 5' flanking region	162
5.3a	962 bp of the human BCP gene 5' flanking region	163
5.3b	949 bp of the porcine BCP gene 5′ flanking region	164
5.4	Alignment of the human and porcine BCP gene 5′ flanking sequences	165-166
5.5	Alignment of exon one of primate BCP genes	168
5.6	Unpredictably Primed Walking PCR products	169
5.7	Alignment of primate sequences obtained by UPW-PCR	170
5.8	Alignment of the first 44-46 amino acids of primate BCP proteins	171
5.9a	573 bp of the BCP gene 5′ flanking region from talapoin monkey	173
5.9b	570 bp of the BCP gene 5′ flanking region from capuchin monkey	173
5.10	Alignment of BCP gene upstream sequences	176-177
5.11	TESS results for the human BCP gene 5′ flanking region	181
5.12	TESS results for the porcine BCP gene 5' flanking region	182

5.13	Alignment of the human and porcine BCP gene 5′ flanking sequences	183-184
5.14	PCR of cDNA isolated from Weri-Rb1 cells	186
5.15	Nuclear protein extracts	189
5.16	DNA fragments spanning the human BCP gene promoter	190
5.17	Autoradiograph showing equal amounts of radioactively labelled fragments	192
5.18	Autoradiograph of electrophoretic band shift assay (EBSA) on fragments of the human BCP gene 5′ flanking region	194
5.19	Autoradiograph of an electrophoretic band shift assay	195
5.20	Autoradiograph of an electrophoretic band shift assay	196
5.21	Autoradiograph of an electrophoretic band shift assay	198
5.22	Autoradiograph of a competitive electrophoretic band shift assay (CEBSA)	200
5.23	-747 to -629 bp upstream of the BCP gene	201
5.24a	Sequence of the 33mer UPW-PCR primer	204
5.24b	Alignment of primate BCP gene 5′ flanking sequence	204
6.1	Localisation of the human BCP gene	213
6.2	An ideogram of human chromosome 7	214
6.3	Vectorette PCR of a triplet repeat	216
6.4	507 bp Vectorette PCR product	217
6.5	PCR products obtained after amplification with the vectorette primer and BCP Repeat+ primer	218
6.6	Sequence of the triplet repeat within the human BCP cosmid	218
6.7	962 bp of 5' flanking region of the human BCP gene	220
6.8	Alignment of the *Alu* repeat found upstream of the BCP gene to the complement of a consensus *Alu* sequence	221
6.9	Intron 3 of the human BCP gene (606 bp)	222
6.10	Alignment of the *Alu* repeat found within intron 3 of the BCP gene to a consensus *Alu* sequence	223
6.11	Alignment of the two *Alu* repeats found in the human BCP gene cosmid	224

List of Tables

Table No.		Page
1.1	The λ_{max} of some vertebrate opsins	38
1.2	The colours and wavelengths of the spectrum	40
2.1	BCP gene-specific PCR primers	59
2.2	BCP gene 5' flanking sequence-specific PCR primers	61
2.3	Opsin primers used to amplify from Weri-Rb1 cDNA	61
2.4	Unpredictably primed walking PCR primers	62
2.5	Triplet and dinucleotide repeat oligonucleotides	62
3.1	The human RCP and GCP spectral tuning sites	94
3.2	Identified and predicted spectral tuning sites	96
3.3	Percentage amino acid identity between opsins	106
3.4	The porcine BCP gene exon-intron junctions	108
3.5	24 non-S group vertebrate opsins	114
3.6	Amino acids which are specific to the S group opsins	117-118
3.7	Comparison of S group-specific residues	120-121
3.8	Comparison of the mouse and goldfish S opsins	127
3.9	Potential spectral tuning sites	129
4.1	Heterologous proteins expressed in *Pichia pastoris*	138
4.2	Human BCP gene PCR primers with linkers	143
5.1	Percentage identity of BCP gene promoter sequences	174
5.2a	Bradford assay for proteins	188
5.2b	Nuclear protein concentrations	188

Chapter 1

1.1 Introduction

All animals rely on senses to receive information and to comprehend the world around them. Vision may be considered as one of the most precious senses within the animal kingdom. Certain animals may consider other senses as their most important, such as hearing in the case of dolphins and bats whom utilise echolocation for everyday activities. The majority of animals however rely heavily on vision to perceive and interpret their surrounding environment, especially in the case of human beings since our senses of hearing and smell are considered to be far inferior to most other mammals.

The human eye is the primary receptor organ of light, that part of the electromagnetic spectrum visible to humans (usually defined as 400-700 nm). Other animals such as certain fish and birds are able to perceive wavelengths of light within the ultraviolet region. What structures in the eye receive the light stimulus and convert this energy to a form which is understood? Which proteins are involved in producing the sensation of sight and are any of these involved in discriminating colour? Which genes within the genome encode for these proteins? What factors regulate the genes which are exclusively expressed in the eye? Due to the recent advances of new techniques in molecular biology, the answers to some of these questions are being unravelled. The study of genes which are normally responsible for specific tasks within the eye, but when mutated result in visual defects, has also advanced our understanding of the processes involved in vision.

The following chapter reviews the structure of the eye and in particular that of the retina. The biochemical and genetic factors involved in the process of vision are also discussed. The molecular characteristics and the regulation of the genes which are involved in colour vision is also reviewed. The following discussion is based mainly on mammalian systems unless otherwise stated.

1.2 Eye structure and function

The eye functions to receive light energy and convert it into electrical signals. These electrical signals are then transported from the eye to the visual cortex in the brain, via the lateral geniculate nucleus. The typical structure of a mammalian eye is given in Figure 1.1. Most of the

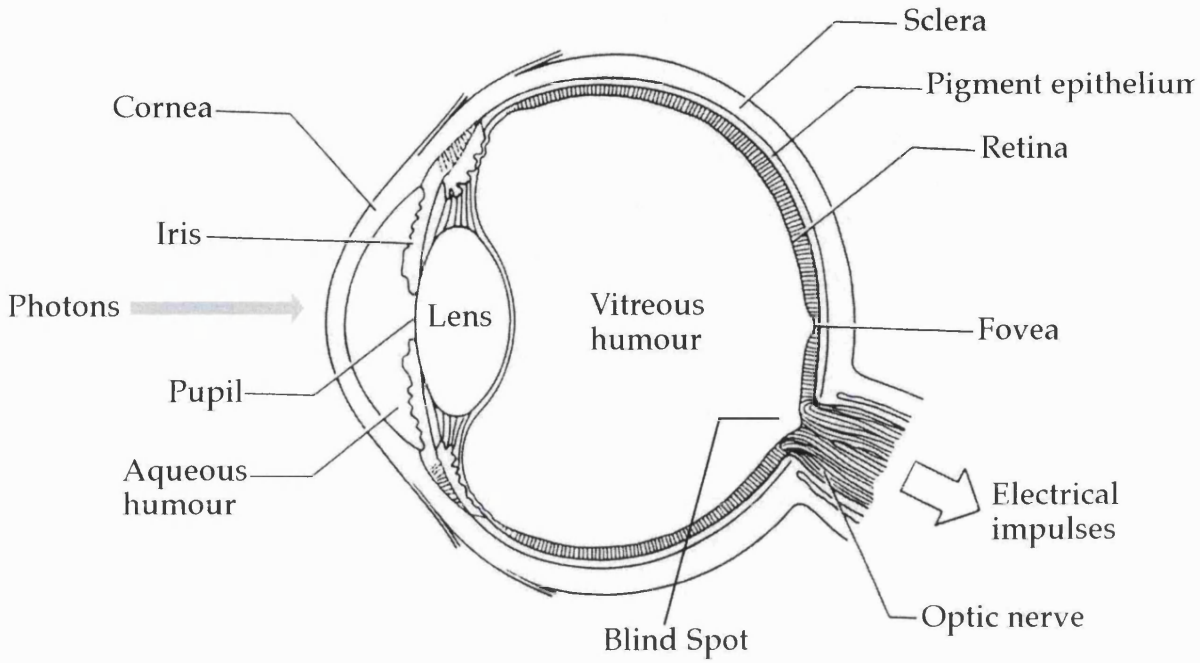


Figure 1.1 A cross section of the human eye (adapted from Goldstein, 1989).

light captured by the eye comes from reflected light from surrounding objects. The first step of visual perception requires the focusing of incoming light on to the retina. The retina is a light sensitive layer of cells lining the back of the eye (Figure 1.1).

Light enters the eye through the cornea and then passes through the aqueous humour, the pupil (the hole at the centre of the iris), the lens and the vitreous humour before finally reaching the retina (Figure 1.1). The iris consists of pigmented circular and radial smooth muscle cells, which inter-link together to form a doughnut shape. The hole in the centre of the 'doughnut' is the pupil, through which light enters the eyeball. The iris functions to control the amount of light entering the eyeball through the lens by regulating the size of the pupil. Contraction of the circular muscles of the iris decreases the size of the pupil, and hence reduces the amount of light entering the eye, in bright light conditions. Whereas, contraction of the radial muscles of the iris increases pupil size and hence the amount of light entering the eye, during dim light conditions.

The lens and cornea act as focusing bodies to focus the light entering the eye precisely on to the retina. Light rays entering the eye from distant objects are generally in parallel. The cornea acts to refract these parallel rays to focus the image onto the retina. Light rays entering the eye from a nearby object are not parallel and the lens changes shape to bring into focus the image formed by non parallel light rays onto the retina. The shape of the lens is controlled by the action of cillary muscles, which surround the lens, found at the front of the eye. When the cillary muscles contract the lens becomes fatter, and thus increases its surface curvature which in turn increases the refraction of light rays to bring about the sharp focusing of a nearby image on the retina. This process of focusing of images onto the retina is known as accommodation (Goldstein, 1989).

The lens consists largely of soluble structural proteins known as crystallins (Piatigorsky and Wistow, 1989). These crystallins are arranged in layers within the lens, similar to the layers found in an onion. The lens is perfectly transparent and any loss of this transparency is known as a cataract.

The aqueous and vitreous humours both contribute to the pressure in the eye, known as the intraocular pressure. The intraocular pressure maintains the shape of the eyeball and holds the retina smoothly against the internal layers of the eye, so the retina can form even focused images. Excessive intraocular pressure causes glaucoma, which results in optic nerve atrophy and partial or complete loss of vision.

1.3 The retina

The retina is a highly specialised layer of neuronal cells (depicted in Figure 1.2). It contains five main classes of cells: photoreceptors, horizontal cells, bipolar cells, amacrine cells and ganglion cells. These neurons are arranged such that the photoreceptors (the cells responsible for initiating the conversion of light energy to electrical energy) actually face away from the source of light. The photoreceptors synapse with the bipolar cells which in turn synapse with the ganglion cells. The axons of the ganglion cells extend around the retina to a small area known as the optic disc. These axons exit the retina through this optic disc region to form the optic nerve which transmits the signals, originally initiated in the photoreceptors, to the brain. The optic disc region contains no photoreceptors and an image focused this region cannot be perceived (Goldstein, 1989). For this reason the optic disc region is also referred to as the blind spot.

There is a great deal of convergence of signal transmission from the photoreceptors to the ganglion cells, which is not clearly obvious from Figure 1.2. In each human retina approximately 126 million photoreceptors transmit their signals, via bipolar cells, to only 1 million ganglion cells (Goldstein, 1989).

Signals are also transmitted across the retina via the amacrine and horizontal cells (Figure 1.2). Horizontal cells allow communication between photoreceptors and bipolar cells which are not directly connected to each other. Amacrine cells increase the number of bipolar cells and ganglion cells which can interact with each other. Signals which travel along the ganglion and bipolar cells can be modulated by the action of these amacrine and horizontal cells (Goldstein, 1989). Thus, these amacrine and horizontal cells act to fine tune the visual signals transmitted to the brain.

1.4 Photoreceptor structure

All photoreceptors possess a similar body plan, all contain an outer segment, an inner segment and a synaptic terminal which transmits impulses to the bipolar cells (Figure 1.3). The majority of vertebrate retinas contain two types of photoreceptors, the rod and cone cells. Both types of photoreceptors are found in mammalian retinas (Jacobs, 1993). The rod and cone cells were originally distinguished by the structure of their outer segments. The outer segments of rods are rod shaped,

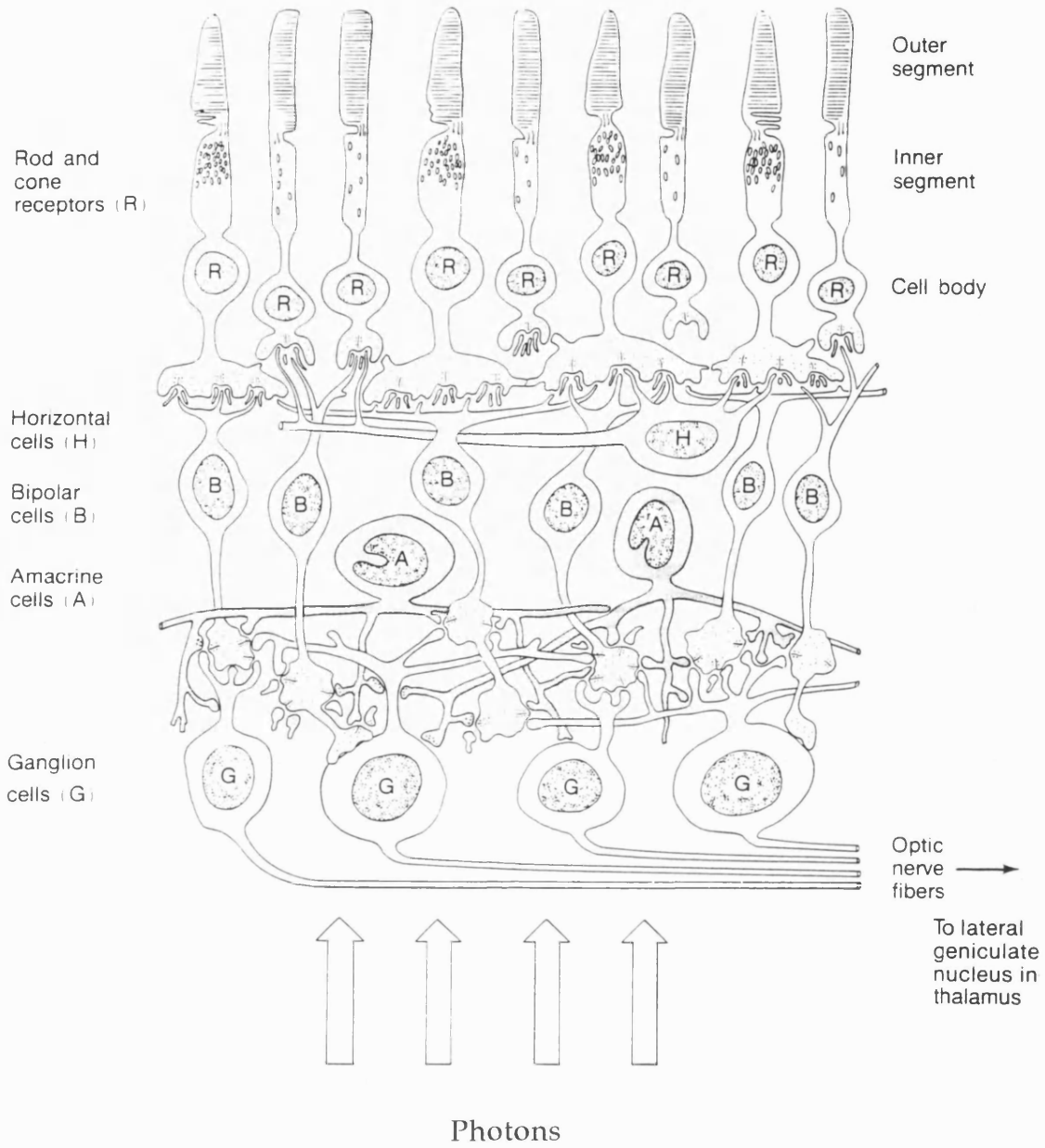


Figure 1.2 Cross section of the primate retina (adapted from Goldstein, 1989).

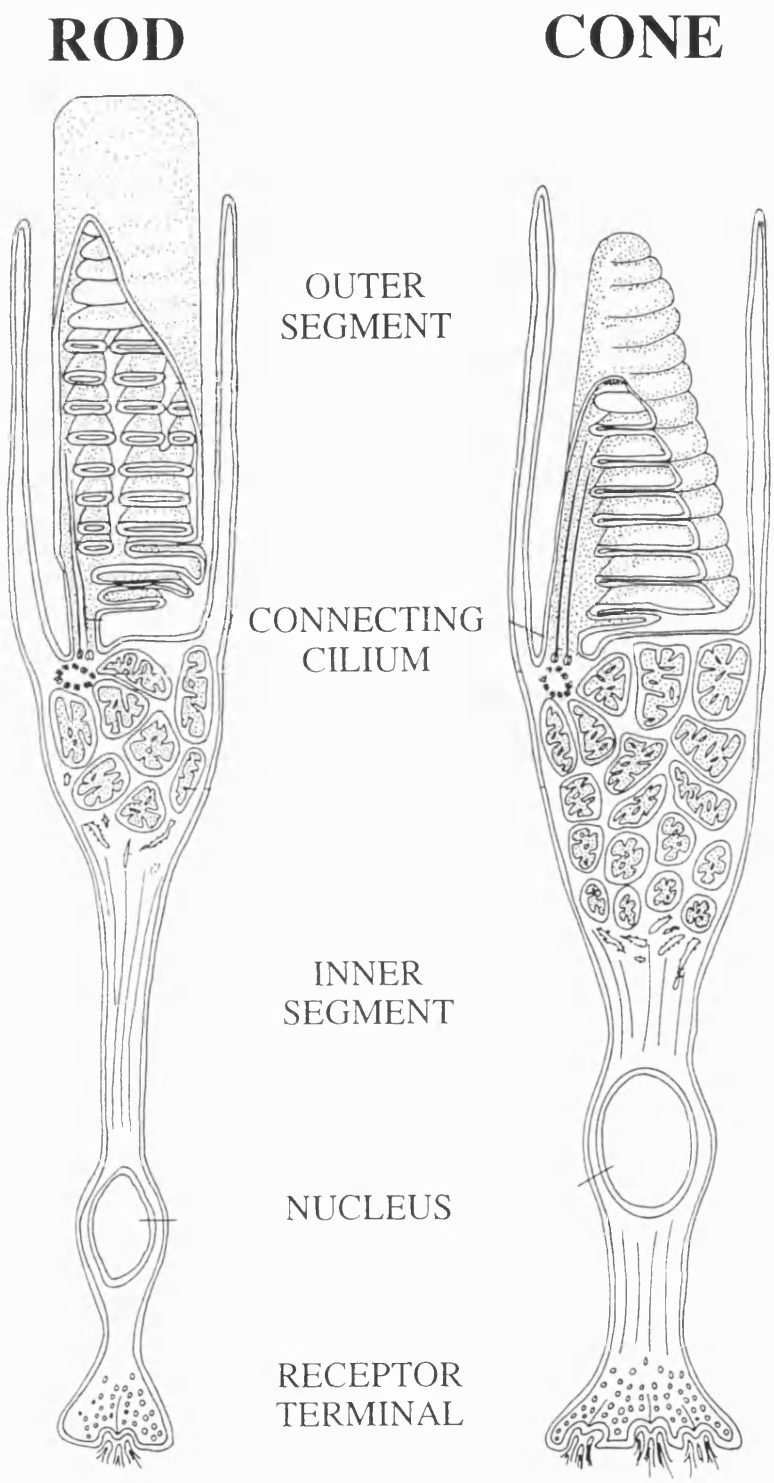


Figure 1.3 Mammalian rod and cone photoreceptors. (Adapted from Ali and Klyne, 1985).

possessing a uniform cylindrical shape and are longer than the relatively short and conical outer segments of cone cells (Figure 1.3). Subsequently, rods and cones have also been distinguished by their physiological functions (Section 1.6 and 1.7).

The photoreceptor outer segments contain stacks of flattened membrane vesicles (discs). These discs stem from outgrowths of the plasma membrane. Light sensitive proteins, known as visual pigments, reside within the membranes of these discs. These visual pigments react to photons of light to initiate the visual transduction process. They are the most abundant protein within the outer segment, accounting for approximately 80% of all the proteins present (Nathans, 1987). The structure of the outer segments is a dynamic one. Work on Rhesus monkey rod cells has demonstrated that about 90 discs, each containing the components of visual transduction, are continually being assembled at the base of the outer segment, near the inner segment (Young, 1971). 10 days later these particular discs, after traversing up the complete length of the outer segment, are engulfed and digested by the pigment epithelial cells that surround the tips of the outer segments (see Figure 1.4).

The inner segment contains the nucleus of the cell, which is spherical and larger in size in cones when compared to the oval shaped nuclei of rod cells (Figure 1.3). The biosynthetic machinery for the production of the proteins and other components of visual transduction are situated within the inner segment. The inner segment is connected to the outer segment by a short cilia, but the actual mechanism of transport of these components from the inner to the outer segment is not fully understood.

The synaptic terminal of both rods and cones contain numerous vesicles. These vesicles contain the neurotransmitters, which when released stimulate bipolar cells. The synaptic terminals of rods are smaller and more localised in comparison to the wider and flat terminals of cones (Figure 1.2 and 1.3). Photoreceptor terminals also contain processes which synapse with the horizontal cells.

1.5 Photoreceptor function

The photoreceptors function to house the light-sensitive visual pigments and to propagate the converted light energy to electrical energy, initiated within themselves, to the brain via the bipolar and ganglion cells. Unlike other cells that undergo depolarisation to propagate action potentials, vertebrate photoreceptors actually hyperpolarise.

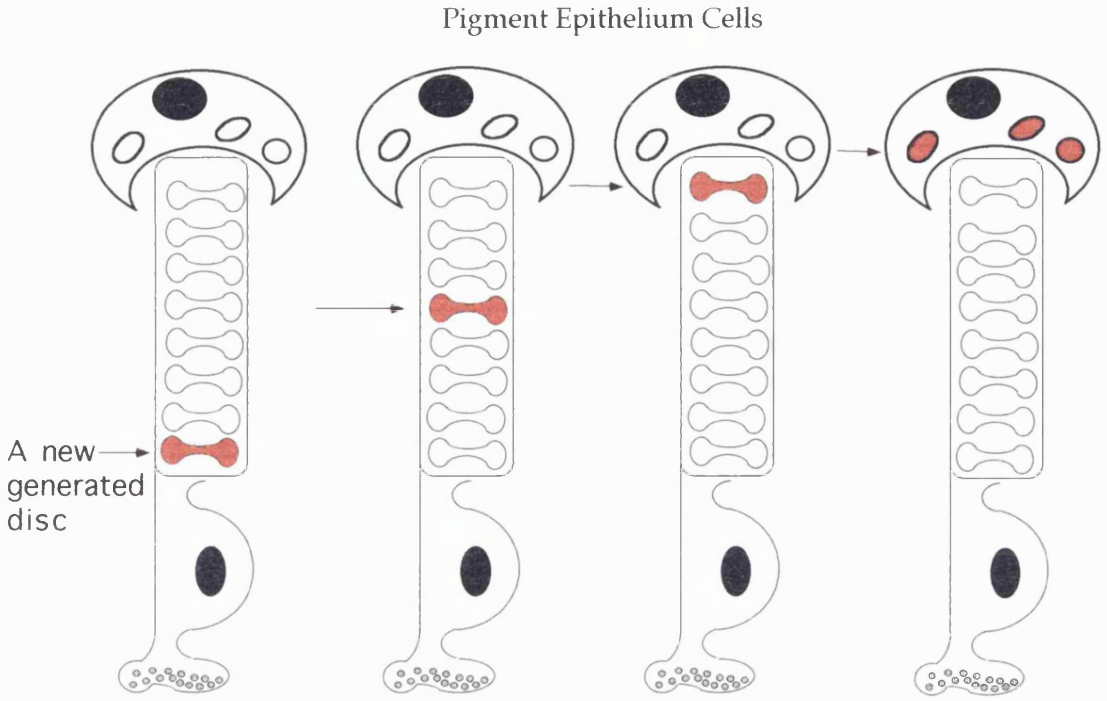


Figure 1.4 Turnover of rod cell discs (adapted from Levitan and Kaczmarek, 1991). New discs are assembled at the base of the outer segment and move up to the tip of the photoreceptor where they are engulfed and degraded by pigment epithelium cells.

When a typical neuron is not conducting nerve impulses it has a membrane potential of -40 to -90 mV, which is termed the resting potential (Levitan and Kaczmarek, 1991). Membrane potentials are expressed relative to the extracellular fluid. A negative membrane potential refers to the fact that the inside of the cell membrane is more negative than the outside. A cell is depolarised when its membrane potential becomes higher than the resting potential. The cell is said to be hyperpolarised when the membrane potential becomes more negative then the resting potential.

A rod cell is considered at rest when it is not stimulated by light (in the dark). The membrane potential of a rod cell at rest is about -30 mV which is higher than the resting potential of the average neuron, as mentioned above (Levitan and Kaczmarek, 1991). In the dark a rod cell is actually in a depolarised state relative to the resting potentials of most neurons, which results in perpetual release of neurotransmitter from the synaptic terminal. This in turn causes a continuous stimulation of bipolar cells. A light stimulus results in the rod cell becoming hyperpolarised, that is the membrane potential becomes more negative than -30 mV. This hyperpolarisation produces a reduction of neurotransmitter release (Rayer *et al.*, 1990).

What causes the membrane potential to change? In the dark the sodium (Na^+) and calcium (Ca^{2+}) pass through cation channels in the plasma membrane of the outer segment into the photoreceptor. Due to the abundance of Na^+ ions in the extracellular fluid it is mainly Na^+ ions which enter the cell (Levitan and Kaczmarek, 1991). With such a high influx of Na^+ ions the membrane potential would be expected to be very positive. However, this influx is counterbalanced by open potassium channels in the inner segment of the rod cell (Figure 1.5). Due to good electrical continuity between the outer and inner segments and the spatial distribution of the ion channels this creates a circulating current of cations through the cell, known as the dark current (Figure 1.5). The dark current, which flows in through the outer segment and out through the inner segment cation channels, maintains the average membrane potential within the negative range of -30mV. A light stimulus results in the closure of cation channels in the outer segment (Figure 1.5). Since the ion channels remain open within the inner segment, there is a significant decrease in the dark current, which reduces the concentration of cations within the cell (Figure 1.5). Thus, the decrease in the dark current results in hyperpolarisation of the cell and a subsequent reduction of

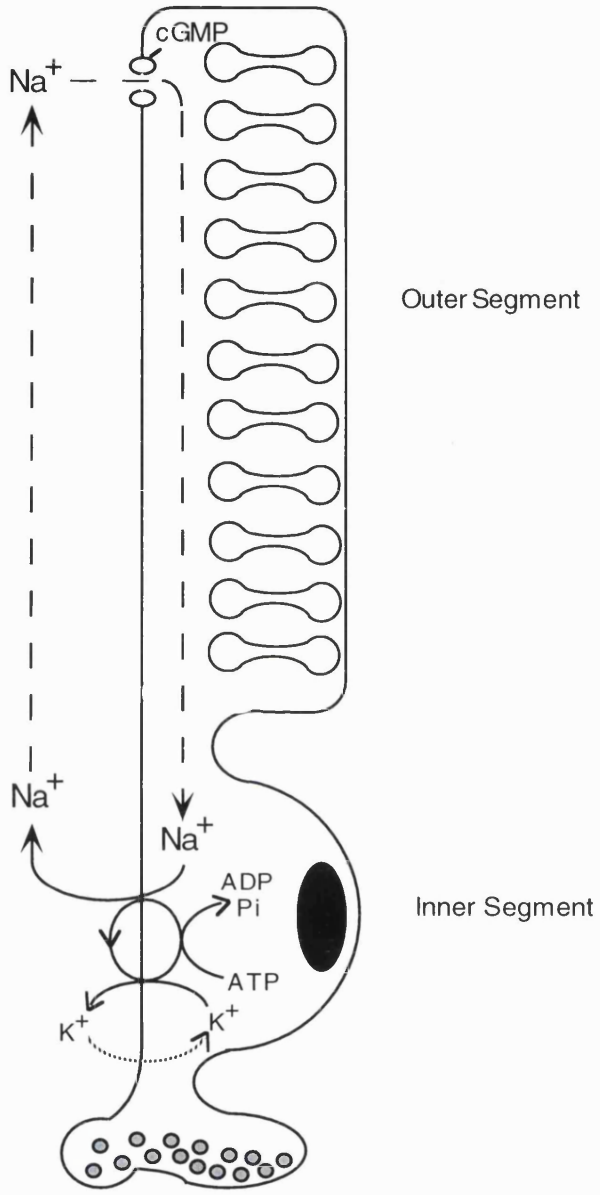


Figure 1.5 A schematic diagram of a photoreceptor showing the passage of cations during the dark current (adapted from Zubay, 1993). In vertebrate rods the cation channels are held open in the dark and they close in the light, resulting in hyperpolarisation of the cell. In invertebrates the cation channels open in the light.

neurotransmitter release. The more photons of light that stimulate the outer segment the more cation channels become closed and the more hyperpolarised the cell becomes. This increased hyperpolarisation results in even greater reduction of neurotransmitter release.

The rod photoreceptors are more sensitive to photons of light then the cones. To attain an equivalent level of hyperpolarisation in rods and cones, the cones require a much greater number of photons than the rods. The nucleotide cyclic GMP (cGMP) is involved in keeping ion channels open. Thus, any reduction in cGMP levels in the outer segment results in closure of the channels and hyperpolarisation of the cell. The biochemical events leading to closure of the ion channels in the outer segments in response to a light stimulus are discussed in Section 1.8.

1.6 Rod and cone cells

As mentioned in Section 1.4 the structure of rods differs from cones. In rod cells the flattened membrane vesicles (discs) are found completely enclosed within the outer segment cell membrane (Figure 1.4). These discs are arranged like a pile of stacked up coins within the rod cell outer segment. Cone cell outer segments possess fewer membranous vesicles than rod cells (Levitan and Kaczmarek, 1991). The membranous vesicles of cone cells do not form separate disc like structures, as observed in rod cells, but remain as outgrowths of the outer segment plasma membrane. (Compare the enclosed discs of a rod cell Figures 1.4 to the finger-like projections of the outer segment in cones 1.6a).

Rod cells show synaptic convergence where several rod cells synapse with only one bipolar cell. Generally, cone cells do not show synaptic convergence, instead there is a one to one relationship of cone to bipolar to ganglion cell. The non-convergent arrangement of cells increases visual acuity, relative to a convergent arrangement. This non-convergent set up for cone cells is particularly true for the region of the retina known as the fovea. The fovea is found in the centre of the posterior portion of the retina, corresponding to the visual axis of the eye, which places it directly on the line of sight (Figure 1.1). Since cones are tightly packed together at a very high density within the fovea, this area of the retina gives the sharpest vision. The greater the density of cones in the fovea the sharper the visual acuity. An animal that relies heavily on vision to catch prey, such as a bird of prey, possesses a high cone dense fovea.

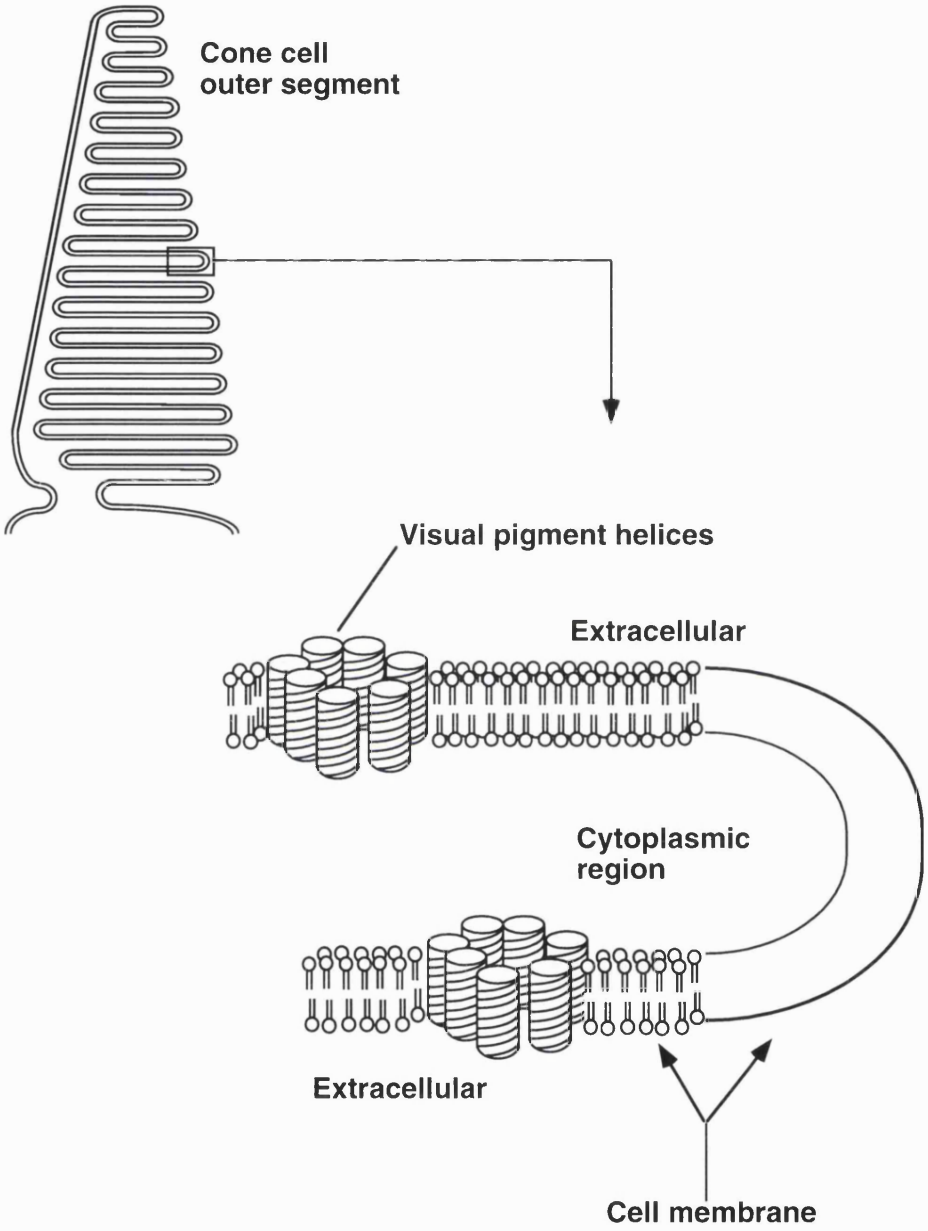


Figure 1.6a Photoreceptor outer segment with a close up of an individual flattened stack, showing the palisade structure of the seven helices of the membrane bound visual pigments (adapted from Mollon, 1991).

Conversely, the rod cells show lower visual acuity as a result of synaptic convergence within the retina. This arrangement of rod cells functions to increase their sensitivity to light, and hence rods are utilised for vision in dim-light environments (scotopic vision). Nocturnal animals such as fruit bats (*Pteropus giganteus*) or the bush baby (*Galago senegalensis*) (Jacobs, 1993) and other species that live in dim-light level environments such as deep water fish (Partridge *et al.*, 1989; Hunt *et al.*, 1996) tend to possess rod-rich retinas. Whereas, cones require higher light levels than rods for stimulation and are used principally in daylight (photopic vision).

The distribution of cones and rods in the retina varies considerably from species to species (reviewed Jacobs, 1993). In humans the rod to cone ratio is 20:1, there being approximately 120 million rods to 6 million cones. Other animals, such as certain squirrels, possess cone rich retinas with 60 to 90% of their photoreceptors being cones (Long and Fisher, 1983).

1.7 Visual pigments

Rods usually only contain one visual pigment, rod opsin, contained within their outer segments. Cones are classified into various subtypes according to the properties of the visual pigment (cone opsin) they contain, (described later in Section 1.10). Visual pigments, found in abundance within the outer segments of the photoreceptors, are the initial responders to the light stimulus (Figure 1.6a). Visual pigments function as light-absorbing molecules and when activated by the absorption of a photon of light they initiate the phototransduction cycle. The phototransduction cycle consists of a series of biochemical events, which lead to the closure of the sodium ion channels and the subsequent hyperpolarisation of the photoreceptor (described in Section 1.5).

Visual pigments are members of a large family of proteins, the G protein-coupled receptors. The first G protein-coupled receptor to be sequenced at the amino acid level was the visual pigment, bovine rod opsin (Hargrave *et al.*, 1983). The bovine rod opsin sequence contains seven interspersed stretches of amino acid residues with hydrophobic properties. These hydrophobic stretches, each of 20-28 amino acids in length, represent regions of the protein which span the cell membrane (Figure 1.6a). Hence, this protein spans the membrane seven times. The structure of the opsin protein spanning the cell membrane is shown in Figure 1.6b. All G protein-coupled receptors possess seven cell membrane spanning stretches, which form α-helical structures (Baldwin, 1993;

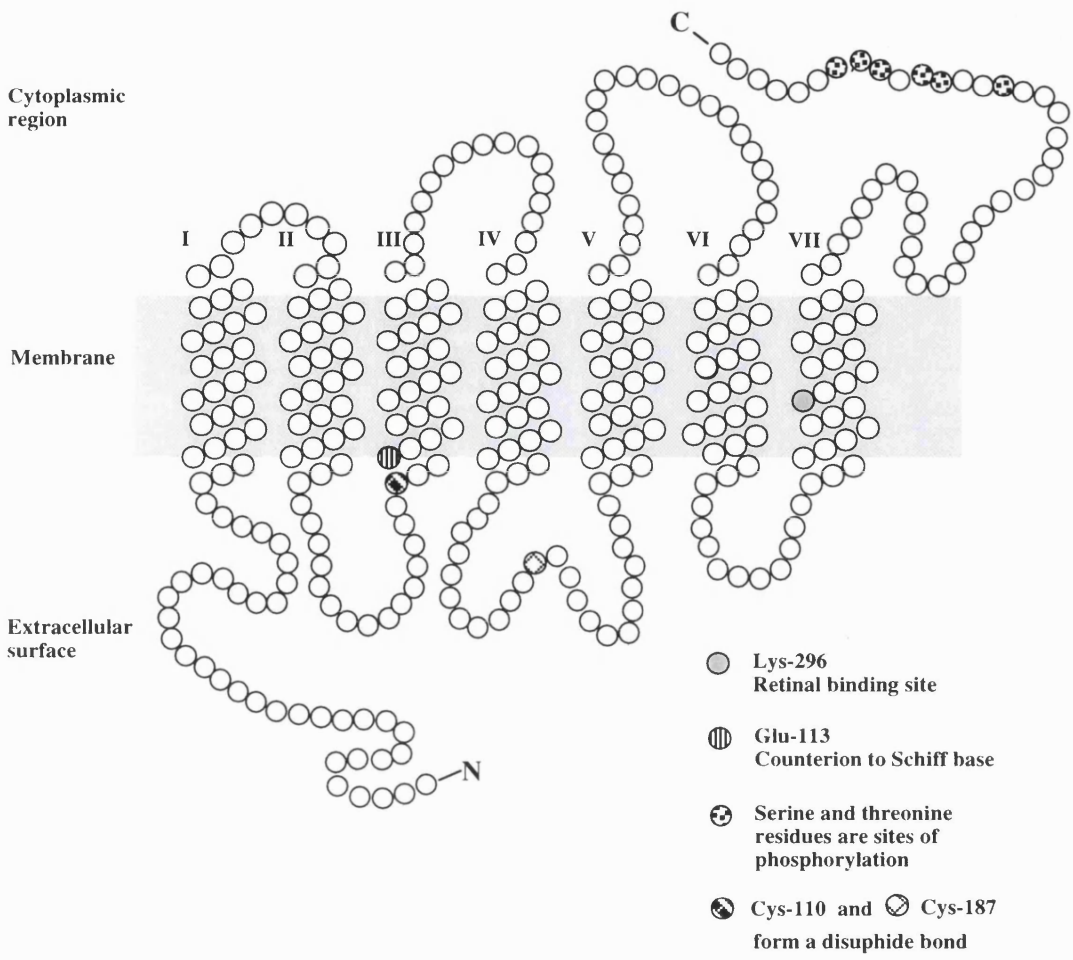


Figure 1.6b Two dimensional depiction of the opsin protein, indicating the 7 transmembrane helices. The N-terminal end and C-terminal lie within the extracellular and cytoplasmic regions of the photoreceptor, respectively. Some of the essential amino acid residues that are conserved in all vertebrate opsins studied to date are also indicated (adapted from Dulai, 1996).

Schertler *et al.*, 1993; Alkorta and Du, 1994). Each helix is linked to the next by an extramembrane loop. Three of the seven α-helices in rod opsin are predicted to be perpendicular to the plane of the membrane, whereas the other 4 are tilted (Schertler and Hargrave, 1995). The seven α-helices of visual pigments are connected by three cytoplasmic loops and 3 extracellular loops. The amino-terminal (N-terminal) and the carboxyl-terminal (C-terminal) of the protein lie within the extracellular and cytoplasmic regions of the cell membrane, respectively (Figure 1.6b).

All visual pigments share the same characteristic features. They consist of a protein moiety known as an opsin (either a rod opsin or cone opsin depending on the photoreceptor type), covalently joined via a protonated Schiff base to a chromophore. The chromophore is usually the 11-*cis* isomer of vitamin A aldehyde (11-*cis* -retinal). However, some vertebrates utilise 11-*cis* -3-dehydroretinal, and some invertebrates use 11-*cis*-3-hydroxy-retinaldehyde or 4-hydroxyretinal, as their chromophore (Vogt and Kirschfeld, 1984; Kito *et al.*, 1992). The seven transmembrane helices (helices I to VII) form a pocket within which the chromophore is situated (Figure 1.7). The chromophore, 11-*cis* -retinal, is attached to the amino group of a conserved lysine residue, found in all opsins, in transmembrane helix VII (Figure 1.7). It is the chromophore which causes the visual pigment to be light sensitive.

All G protein-coupled receptors are activated by an external stimulus which results in the binding and interaction of a guanine nucleotide-binding protein (G protein) to the receptor. In the case of visual pigments, photons of light induce the isomerisation of the covalently bound chromophore 11-*cis*-retinal to all-*trans*-retinal (Figure 1.8). This isomerisation causes the opsin to change shape, exposing certain regions of its cytoplasmic loops, which allows the binding and activation of the G protein transducin, a trimeric protein consisting of α, β and γ units (reviewed in Farber, 1995). The phototransduction cascade which leads on from this conformational change of the visual pigment is described in Section 1.8.

1.8 Phototransduction

The biochemical events involved in the phototransduction cascade have been particularly well researched for rod photoreceptors, which contain the visual pigment rod opsin (reviewed in Yau, 1994 and Farber, 1995). The following description refers to the events which have been elucidated for rod opsin. The initial steps of the phototransduction cycle is shown schematically in Figure 1.9.

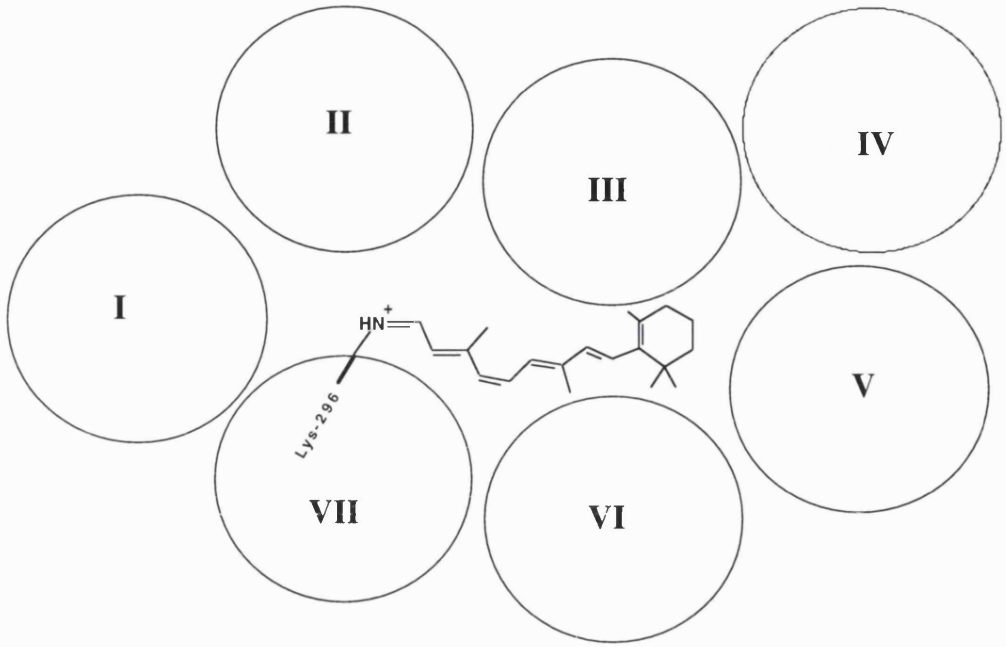


Figure 1.7 Schematic representation of a transverse section through the seven helices of a visual pigment (adapted from Nakayama and Khorana, 1991). The seven helices form a pocket within which the chromophore, 11-*cis*-retinal, sits bound to Lysine-296 in helix VII.

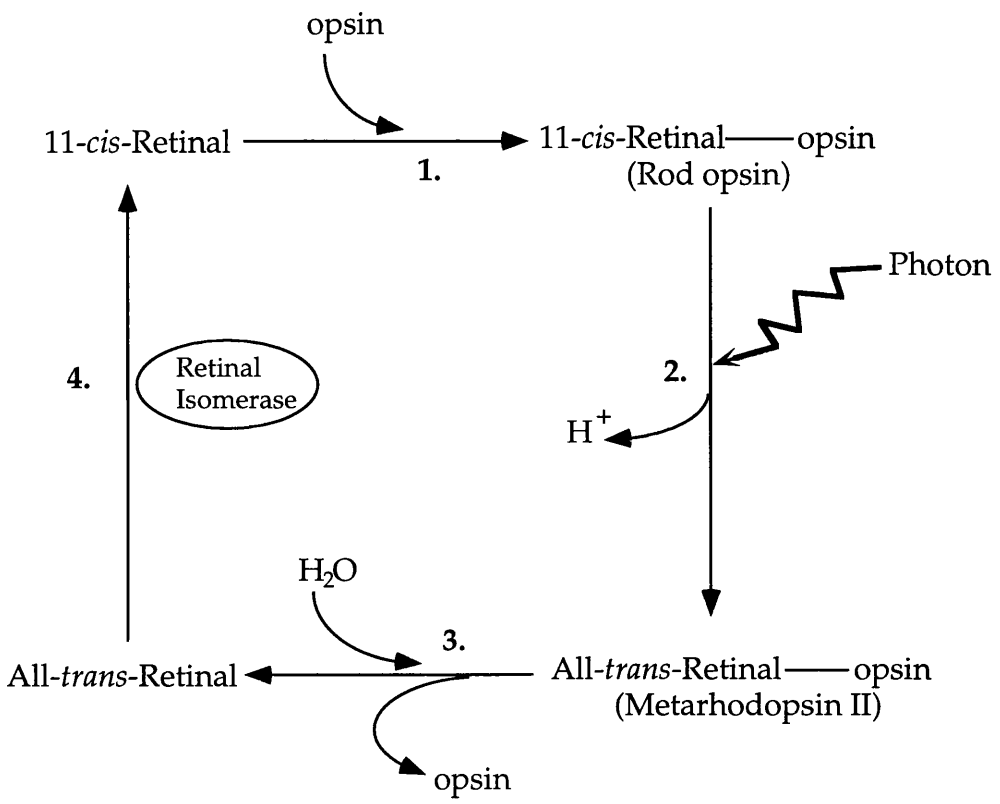


Figure 1.8 The retinal cycle of rod opsin (adapted from Maden, 1995).
The following steps are also described in Section 1.8.
1. 11-*cis*-retinal is attached to opsin via a protonated Schiff base, to form rod opsin.
2. Absorption of a photon of light by rod opsin causes the isomerisation of 11-*cis*-retinal to all-*trans*-retinal.
3. All-*trans*-retinal is released from metarhodopsin II by hydrolysis at the Schiff base linkage.
4. All-*trans*-retinal is converted back to 11-*cis*-retinal by an enzyme, retinal isomerase, found in the pigment epithelium.

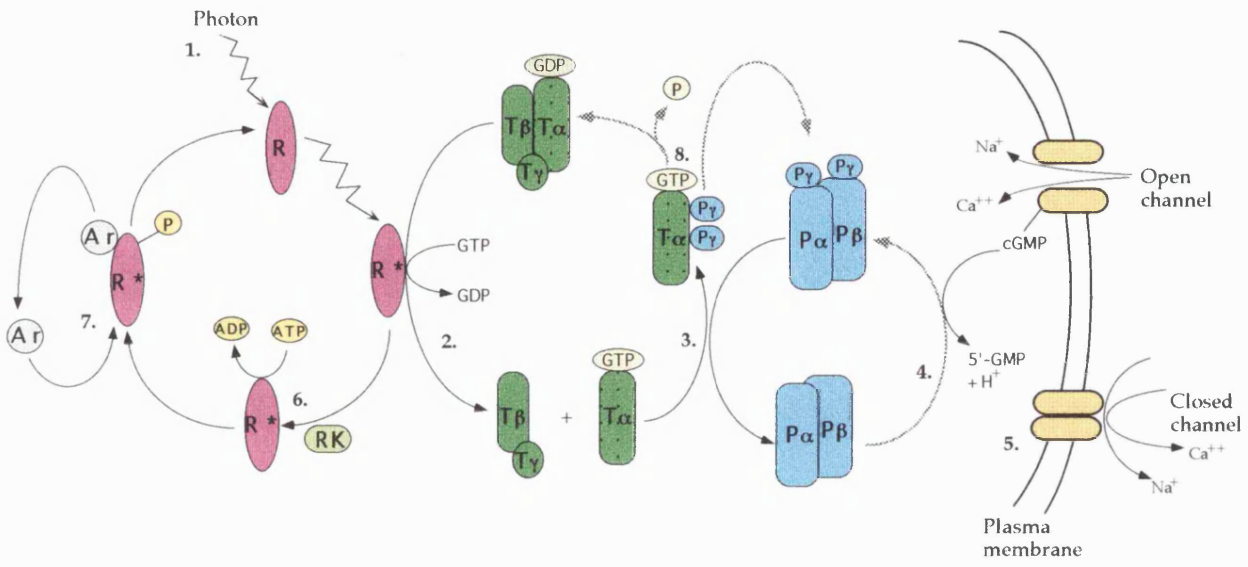


Figure 1.9 Schematic diagram depicting the biochemical events, in 8 steps, within the phototransduction cascade (adapted from Farber, 1995).

The 8 steps indicated in the diagram are also explained in Section 1.8.

1. Rod opsin (R) is activated by the absorption of a photon of light.

2. The activated rod opsin (R*) can now bind and activate transducin. Binding of transducin results in the exchange of GDP for GTP and the subsequent release of the Tα subunit with bound GTP from the Tβ and Tγ subunits.

3. The Tα–GTP molecules activate PDE (P) by binding to the inhibitry Pγ subunits and thereby releasing the catalytic Pα and Pβ subunits.

4. The catalytic Pα and Pβ subunits now hydrolyse molecules of cGMP.

5. A depletion of intracellular cGMP results in closure of the cGMP-gated cation channels.

6. Inactivation of R* is initiated by the phosphorylation of the rod opsin by a kinase (RK).

7. Arrestin (Ar) binds to the rod opsin, which blocks the binding site on R* for tranducin.

8. Inactivation of Tα–GTP occurs by hydrolysation of its GTP molecule to GDP, with the subsequent release of the Pγ subunits. The Pγ subunits are free to inhibit the Pα and Pβ subunits of PDE again.

The photon activated rod opsin (R) with attached all-*trans*-retinal is called metarhodopsin II (labelled as R* in Figure 1.9). The initial light induced activation process is labelled as step 1 in Figure 1.9. Metarhodopsin II in turn activates transducin (T), a trimeric protein consisting of α (Tα), β (Tβ) and γ (Tγ) subunits. In the dark, GDP is bound to Tα within the trimeric transducin. When transducin is activated the Tα binds to metarhodopsin II, resulting in exchange of GDP for GTP and its disassociation from the Tβ/Tγ subunits (step 2). The Tα-GTP then dissociates from metarhodopsin II, allowing the latter to activate more molecules of transducin (step 2). The free Tα-GTP molecules can now activate the enzyme cyclic GMP-phosphodiesterase (PDE). PDE consists of four subunits, two inhibitory γ (Pγ) and two catalytic α (Pα) and β (Pβ). The Tα-GTP molecules bind to the inhibitory Pγ subunits of PDE thereby freeing the catalytic subunits (Pα and Pβ) of PDE (step 3). The PDE enzyme is now able to convert cyclic GMP (cGMP) to 5′-GMP (step 4). One molecule of metarhodopsin II activates many transducins and each of these in turn activate many PDEs, thereby amplifying the response to the initial light stimulus. The resulting depletion of cGMP concentration causes the cGMP-gated cation ion channels of the outer segment to close (step 5), which in turn leads to hyperpolarisation and reduction of transmitter release (described in Section 1.5).

Inactivation of metarhodopsin II (R*) involves the initial step of phosphorylation of the protein component (rod opsin) by a kinase (rhodopsin kinase (RK), step 6), and then subsequent binding of the protein arrestin (Ar) which blocks the binding of Tα-GTP (step 7). An endogenous GTPase activity of the Tα subunit inactivates Tα-GTP, which releases the Pγ subunits of PDE ready to inhibit the action of PDE again (step 8). The cGMP levels return to normal and the ion channels reopen, until another light stimulus repeats the whole cascade again.

The chromophore attached to rod opsin undergoes a series of changes during the phototransduction cycle (Figure 1.8). The initial photon activates rod opsin by the isomerisation of its chromophore to all-*trans*-retinal (step 2). The all-*trans*-retinal is hydrolysed at its Schiff base link to metarhodopsin II which releases it from rod opsin (step 3). All-*trans*-retinal is then converted back to 11-*cis*-retinal by retinal isomerase which can now bind to opsin to regenerate functional rod opsin (step 4).

The phototransduction cascade in cone photoreceptors follows the same plan given above, although the protein components are, in many cases, cone-specific (Cobbs *et al.*, 1985). The cone PDE catalytic region

consist of two α subunits compared to the α and β subunits found in rod PDE (Hurwitz *et al.*, 1985). Recently, the genes encoding the cone-specific α subunit (Piriev *et al.*, 1995) and the γ subunit (Ong *et al.*, 1995) of PDE have been cloned and characterised. Chicken rod and cone photoreceptor cells have been shown to express distinct genes for the cGMP-gated channels (Bönigk *et al.*, 1993). The phototransduction proteins may even be specific to particular sub-types of cones (Lerea *et al.*, 1986). Genetic data indicates that there is more than one gene within the human genome that codes for cone-specific α subunit of transducin (Lerea *et al.*, 1989). Thus, there are still components of the cone-specific phototransduction pathway to be elucidated.

1.9 Rod opsin

Bovine rod opsin was the first vertebrate opsin to be sequenced at the amino acid level (Hargrave *et al.*, 1983). The gene encoding bovine rod opsin was also the first opsin gene to be sequenced (Nathans and Hogness, 1983). Soon after, the gene for human rod opsin was also sequenced (Nathans and Hogness, 1984) which paved the way for the sequencing of rod opsin genes from various other species. Both the rod opsin proteins from human and bovine are 348 amino acids in length and share 94% identity. The structural and functional properties of rod opsin have been reviewed extensively (see Dratz and Hargrave, 1983; Nathans, 1987; Saibal, 1990; Nathans, 1992; Maden, 1995)

Functional and sequence analysis of rod opsins from different species has revealed certain amino acid residues which are conserved in all rod opsins. In some cases, these amino acids are conserved in all opsins, indicating that they are essential for the proper activity of these proteins (some examples are given below). The chromophore 11-*cis* -retinal binds to rod opsin via a protonated Schiff base at Lys-296 (the numbering for amino acids refers to the system applied to bovine rod opsin (Wang *et al.*, 1980)), which sits within helix VII. To counter act the positive charge from the Schiff base the negatively charged residue Glu-113 was found to serve as the counter ion (Sakmar *et al.*, 1989; Zhukovsky and Oprian, 1989). Two cysteine residues, Cys-110 and Cys-187, situated within extracellular loops I and II respectively, form a disulphide bond (Karnik and Khorana, 1990). Cys-110 and Cys-187 are essential for the proper structural formation of rod opsin (Karnik *et al.*, 1988). Rod opsin is covalently modified by the addition of oligosaccharides at two asparagine residues, Asn-2 and Asn-15, within the N-terminal region of the protein

(Fukuda *et al.*, 1979). Multiple serine and threonine residues found in the C-terminal region undergo reversible light-dependent phosphorylation during the phototransduction cycle as mentioned in Section 1.8 (Wilden and Khun, 1982). The cytoplasmic loop II, that connects helices III and IV, and cytoplasmic loop III, which connects helices V and VI (Figure 1.6b), and specific regions within the C-terminal region of rod opsin are essential for the activation of the G protein transducin (Konig *et al.*, 1989).

As described previously, rod opsin and its chromophore, found in rod cells, are responsible for scotopic vision (Section 1.6). Rod opsin plays no role in perceiving colour and absorbs light maximally at about 500nm.

1.10 Cone pigments

There are many different cone pigments, and cone cells are characterised according to the type of cone pigment they contain. Cone pigments in turn are classified according to the particular wavelength of light they are maximally sensitive to (the λ_{max}). The λ_{max} of cone pigments varies from the ultraviolet range of about 360nm to the yellow region of the visible spectrum at about 570nm. A list of some vertebrate visual pigments with their corresponding λ_{max} is given in Table 1.1. The classification of cone pigments into specific groups is described in Section 1.13.

1.11 Cone opsins

Many cone opsin amino acid sequences have been predicted from the sequence of the genes which encode them (Yokoyama, 1994). Human beings were the first vertebrate species to have all their cone opsin genes sequenced (Nathans *et al.*, 1986a,b). These pigments were termed red, green and blue and have a λ_{max} of 563, 532 and 426nm, respectively (Oprian *et al.*, 1991; Merbs and Nathans, 1992). The typical absorption spectra curves observed for the λ_{max} of these three pigments are given in Figure 1.10.

The red/green/blue terminology is misleading for the classification of cone pigments. For instance, the λ_{max} for the human red and blue cone pigments actually correspond to the green-yellow and violet regions of the visible spectrum, respectively (see Table 1.2). Thus, cone opsins termed red, green and blue do not necessarily absorb maximally within the red, green and blue region of the visible spectrum, respectively. An alternative way of referring to the human cone pigments is as follows:

Visual Pigment	Species	λ_{max} (nm)	Phylogenetic Group
Rod Opsin	Human	495	Rh
Rod Opsin	Bovine	500	Rh
Rod Opsin	Murine	498	Rh
Rod Opsin	Chicken	503	Rh
Rod Opsin	Goldfish	492	Rh
Red Cone	Human	563	L
Green Cone	Human	535	L
Red Cone	Chicken	571	L
Red Cone	Goldfish	525	L
Green Cone	Gecko	521	L
Red Cone	Cavefish	563	L
Green Cone	Cavefish	533	L
Green Cone	Chicken	508	M2
Green Cone 1	Goldfish	511	M2
Green Cone 2	Goldfish	506	M2
Blue Cone	Chicken	455	M1
Blue Cone	Goldfish	441	M1
Blue Cone	Human	420	S
Blue Cone	Talapoin	429	S
Blue Cone	Marmoset	423	S
Blue Cone	Murine	360	S
Blue Cone	Bovine	?	S
Violet Cone	Chicken	415	S
Ultraviolet Cone	Goldfish	355-360	S

Table 1.1 A list of some vertebrate visual pigments and their maximum wavelength of absorption (λ_{max}). Pigments are also grouped into their specific families obtained from phylogenetic tree analysis, see Section 1.13. The original references quoting the λ_{max} for most of these pigments can be obtained from Yokoyama (1994), marmoset BCP (Tovee *et al.*, 1992), talapoin BCP (Bowmaker *et al.*, 1991a) murine BCP (Jacobs *et al.*, 1991) goldfish ultraviolet (Bowmaker *et al.*, 1991b). Currently there is no data for the λ_{max} of the bovine BCP.

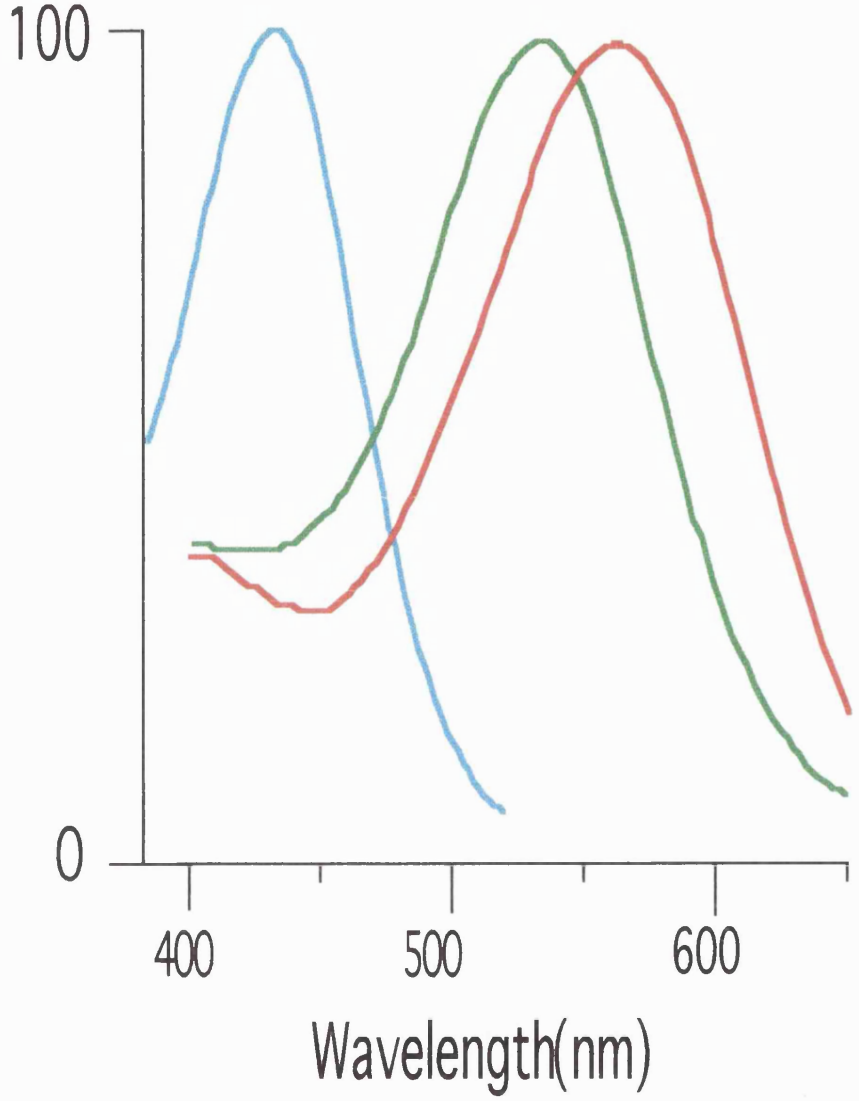


Figure 1.10 Absorption spectra curves for human red, green and blue cone pigments. Curves are indicated in their respective colours: red (563 nm); green (530 nm) and blue (420 nm).

Colour	Wavelength (nm)
Red	620-700
Orange	590-620
Yellow	570-590
Green	500-570
Blue	440-500
Violet	400-440
Ultraviolet	< 400

Table 1.2 Colours of the visible spectrum and the ultraviolet and the approximate ranges of their wavelength.

long-wavelength sensitive (LW) for the red, middle-wavelength sensitive (MW) for the green and short-wavelength sensitive (SW) for the blue (Bowmaker, 1983). However, even this terminology is not adequate for the complete classification of the plethora of vertebrate cone pigments that have been identified to date (see Section 1.13).

Cone opsins show similar structural properties to rod opsins, as would be expected since both are functional components of light-sensitive pigments. Lys-296, for the binding of the chromophore, Cys-110, Cys-187, counter-ion Glu-113 and multiple serine and threonine residues in the C-terminal end are also present within cone opsins. The human red and green opsins share 96% amino acid identity. The blue opsin shares about 40% identity with the green and rod opsin, while the green opsin shows a 38.2% identity with rod opsin (Nathans *et al.*, 1986a). Both human rod opsin and blue cone opsin contain 348 amino acids whereas the red and green opsins contain 364 residues. These extra 16 amino acids in the red and green opsins are found at the N-terminal region of the protein (Nathans *et al.*, 1986a).

1.12 Spectral sensitivity

The maximum sensitivity of each class of cone cell is determined by the wavelength of maximal absorption (λ_{max}) of the visual pigment it contains. What determines the λ_{max} of a visual pigment?

11-*cis*-retinal in a solution of aqueous detergents absorbs maximally at 380nm in solution (Hubbard and Sperling, 1973). The three classes of human cone opsins have been expressed *in vitro* and reconstituted with 11-*cis*-retinal have a λ_{max} of 424nm for the blue cone pigment (BCP), 530nm for the green cone pigment (GCP) and 552 and 557nm for two polymorphic forms of the red cone pigment (RCP) (Oprian *et al.*, 1991; Merbs & Nathans 1992). It would seem that sequence differences between the classes of opsin determine the λ_{max} of cone pigments, which range in mammals from the ultraviolet ($\approx$ 360nm) to the green-yellow ($\approx$ 570nm) end of the visible spectrum. An opsin (opsin A) with a particular protein sequence, attached to its chromophore, will posses a unique λ_{max}. Another opsin (opsin B) which has a different amino sequence to opsin A, but contains the same chromophore, will possess a different λ_{max}. Thus, through evolution a whole range opsins, each with their own unique λ_{max}, have become adapted to the needs of a particular species.

An alternative method to alter the λ_{max} of cone pigments without changing the class of opsin is to utilise a different chromophore. Many

fresh water fish use 11-*cis*-3,4-dehydroretinal in their pigments which results in a red shift in λ_{max}. The American chameleon (*Anolis carolinensis*) also utilises 11-*cis*-3,4-dehydroretinal to red shift the λ_{max} of their visual pigments (Provencio *et al.*, 1992). However, changing the chromophore to alter the λ_{max} is a rare occurrence in nature, the visual pigments (cone and rod) from all mammals and most land vertebrates analysed to date utilise 11-*cis*-retinal as the chromophore. Another method for fine tuning the λ_{max} is the use of coloured oil droplets found within some classes of cone cells that act to filter the incident light (Nathans, 1987). Oil droplets are usually found within photoreceptors of avian species, reptiles, lungfish, amphibians (in which they are clear) and marsupial mammals (in which they may be coloured) but have yet to be seen in rod or cone cells of eutherian mammals (Crescitelli, 1972).

If sequence differences between opsins mediate λ_{max}, the next questions to arise are how, what and where are these differences in the opsin that affect the λ_{max}? This question is discussed further in Section 4.1.7.

1.13 Colour vision

To be able to perceive colour, an animal must possess at least two spectrally different cone photoreceptors to enable a comparison of stimulatory signals. Most mammals are dichromats, that is they compare visual stimuli from only two different classes of cone cells to see colour, with each class containing a spectrally-distinct visual pigment. Cone cells are classified according to the wavelength of light they are maximally sensitive to. In dichromatic mammals one class is sensitive to short wavelengths, usually between 400-450nm, and the other cone class is sensitive to long wavelengths, within the 500-570nm range.

The theory that human colour vision is mediated by three receptors, each with different spectral sensitivities, was originally presented in 1801 by Thomas Young. Young proposed the "trichromatic theory of colour vision" on the basis that test subjects required at least three lights of different wavelengths to match the wavelength of a test light (Young, 1801). Electrophysiological, microspectrophotometrical and genetical data have proved this trichromatic theory to be correct. Trichromatic colour vision, where stimuli from three classes of cones are compared, is seen only in primates (Bowmaker *et al.*, 1991b). The three classes of cones in Old World primates have maximum sensitivity near 420-430nm, 535nm and 565nm, and these are termed the blue, green and red cones,

respectively. Human red and green cone pigments are coded for by separate genes on the X-chromosome and the BCP gene is found on chromosome 7 (Nathans *et al.*, 1986b).

Until recently, all New World primates were thought to possess only two cone pigment genes, an autosomal BCP gene and a polymorphic red/green cone pigment gene on the X chromosome. The howler monkey (*Alouatta* sp) has recently been shown, by electrophysiological and genetic data, to possess trichromatic vision similar to Old World Primates (Jacobs *et al.*, 1996). In other New World primates trichromacy can be achieved in females due to the presence of distinct variant forms of the polymorphic cone pigment gene on each of her X chromosomes (Mollon *et al.*, 1984; Jacobs and Neitz, 1987; Williams *et al.*, 1992). Each of the three pigment variants on the X chromosome of the marmoset, a New World monkey, possess different peak sensitivities at 543nm, 556nm or 563nm (Williams *et al.*, 1992). Thus, after random X-inactivation in females with two different alleles each cone cell will express only one pigment.

Although primates are privileged with trichromatic vision rather than dichromatic vision found in most other mammals, other vertebrates can possess additional cone classes. A fourth class of cone pigment is found in some birds, teleost fish, amphibians and reptiles, which results in tetrachromatic vision. Some birds, such as pigeon, may even possess five classes of cone pigments (Bowmaker, 1991). Interestingly, an exception to the general rule that each class of cone cell expresses only one pigment has been found in some rodents, where two different spectral classes of pigment are expressed during development in one cone cell (Röhlich, 1994).

1.14 Classification and evolution of visual pigments

Vertebrate visual pigments can be classified into 5 major families based on amino acid identity and functional analysis. These visual pigment families and some of their individual members with their corresponding λ_{max} is given in Table 1.1. Any new vertebrate pigments isolated in the future can be compared with these 5 groups and characterised accordingly. The following section describes how these visual pigments were defined into 5 groups.

Each visual pigment can be grouped according to its maximum wavelength of absorption (λ_{max}). Rod opsins generally have a λ_{max} of around 500nm. Due to the similarity between rod opsins from various species they can all be grouped together in one family. Individual

members of this group are easy to subdivide according to λ_{max} and species type. However the subdivision of cone pigments according to λ_{max} is more difficult. The cone pigments were originally grouped into three classes red, green and blue based on λ_{max} similarity to the human pigments (mentioned in Section 1.11). λ_{max} data gathered from various vertebrate photoreceptors/pigments using methods such as microspectrophotometry (msp), electroretinography (ERG) and direct measurements from pigments expressed *in vitro* have shown that there are a vast variety of cone pigments (Bowmaker *et al.*, 1991a; Nathans *et al.*, 1989; Oprian *et al.*, 1991; Jacobs *et al.*, 1991). These data indicate that the original classification of cone pigments into three groups was a gross simplification. The λ_{max} for cone pigments vary from the ultraviolet to the yellow region of the visible spectrum (about 570nm).

Other criteria need to be considered in order to divide cone pigments into subtypes. By comparing the protein sequences of visual pigments from various organisms, individual members of the cone opsin group can be categorised in relation to their amino acid identity. The sequence data can be related back to the λ_{max} of the pigments to classify the cone pigment subtypes and study the evolutionary history of vertebrate visual pigments.

The human red and green cone pigments contain two residues, histidine-181 and lysine-184 (using the bovine rod opsin numbering system), situated within extracellular loop II which bind chloride (Cl^-) ions. Binding of chloride ion results in the modulation of the λ_{max} of the visual pigment (Wang *et al.*, 1993). On the basis that His-181 and Lys-184 are conserved in all cone pigments with a λ_{max} higher than 520nm, but are replaced by Glu-181 and Gln-184 in all other opsins with a λ_{max} less than 520nm (including the rod opsins), all vertebrate opsins can be subdivided into two groups (Wang *et al.*, 1993). Thus, the human red and green cone pigments are grouped together with other vertebrate red and green cone pigments, (which contain His-181 and Lys-184), as members of the long-wavelength (L) pigments. Vertebrate blue cone pigments and rod opsins form the short-wavelength (SW) pigment group.

It is important to note that not all pigments termed green automatically belong to the L group of pigments. For example, the green pigments from chicken (Okano *et al.*, 1992) and goldfish (Johnson *et al.*, 1993) are actually members of the SW pigments based on the above criteria. These green pigments actually form a subclass of their own, the

M2 class, within the SW group (Table 1.1), which are described further below.

Numerous phylogenetic trees have been generated by comparing protein or gene sequences to determine the evolutionary history of vertebrate opsins (Yokoyama, 1994; Okano *et al.*, 1992; Chiu *et al.*, 1994; Chang *et al.*, 1995). The generation of these phylogenetic trees has resulted in the further subdivision of the SW pigments into 4 separate subclasses, the Rh, M1, M2 and S group of pigments. An example of a phylogenetic tree showing the evolution of these subclasses of visual pigments is given in Figure 1.11. Examples of individual members of each subclass are also listed in Table 1.1.

The Rh group represents the rod opsin family of pigments. The S group consists of the human BCP, chicken violet (Okano *et al.*, 1992), bovine and murine BCPs (Chiu *et al.*, 1994), marmoset and talapoin monkey BCPs (Hunt DM *et al.*, 1995) and the goldfish ultraviolet pigment (Hisatomi *et al.*, 1996). The S group of pigments share over 60% amino acid identity with each other and have a λ_{max} of less than 440nm (as indicated in Table 1.1). The M2 group consists of the green visual pigments within the SW group. These M2 pigments are actually rod-like green pigments. M2 members are more related to the rod opsin family, by amino acid identity, than to any of the other cone opsins. The M1 group contains the cone visual pigments which have a λ_{max} between 440 and 500nm (Table 1.1). 440-500nm corresponds to the blue region of the visual spectrum (Table 1.2), thus the M1 members are 'true' BCPs.

1.15 Molecular genetics of visual pigments

The original sequencing of the bovine rod opsin gene (Nathans and Hogness, 1983) has paved the way for the characterisation of rod and cone opsin genes from a variety of different vertebrates and invertebrates, (a list of some of these genes which have been characterised can be found elsewhere (Yokoyama, 1994)). Any new vertebrate visual pigment gene sequenced is generally compared with the human visual pigment genes for analysis. The genetic structure of the four human opsin genes is described below.

The human rod opsin gene consists of 1044 bp of coding region, interrupted by 4 introns (Nathans and Hogness, 1984). The human rod opsin gene has been localised to chromosome 3q21-q24 (Nathans *et al.*, 1986b). The bovine rod opsin gene possesses the similar genomic structure to the human gene (Nathans and Hogness, 1983). The coding region for

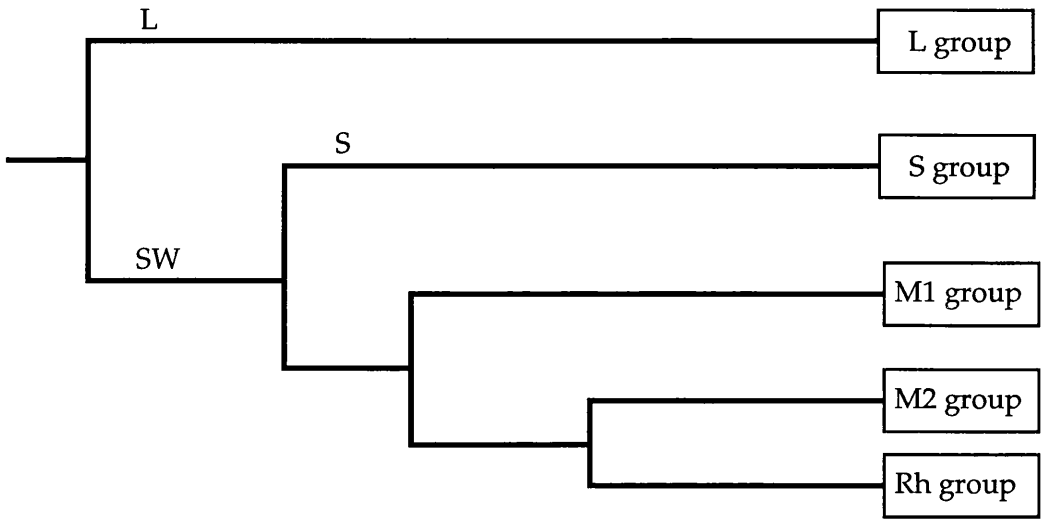


Figure 1.11 Schematic diagram of a typical phylogenetic tree indicating the origin of the 5 main families of visual pigments. (Adapted from Okano *et al.*,1992 and Chiu *et al.*, 1994). Some individual members of each group of pigments are given in Table 1.1.

the human BCP gene is also 1044 bp long and contains 5 exons, interrupted by 4 introns (Nathans *et al.*, 1986a). The BCP gene has a very short 5′ untranslated region of 7 bp and its total length is 3.3 kb (Nathans *et al.*, 1986a). The BCP gene has been mapped to the long arm of chromosome 7 (Nathans *et al.*, 1986b). The human red cone pigment (RCP) and green cone pigment (GCP) genes are organised in a tandem array with the RCP gene 5′ to at least one copy of the GCP gene, on the X chromosome (Nathans *et al.*, 1986a,b; Vollrath *et al.*, 1988). The RCP and GCP genes are 15.2 kb and 13.3 in length, respectively (Nathans *et al.*, 1986a). This difference in size is due to an extra 2 kb of sequence within intron 1 of the RCP gene (Nathans *et al.*, 1986a). Both RCP and GCP genes contain 6 exons, interrupted by 5 introns, which code for 16 extra amino acids when compared to the other human opsins.

1.16 Gene expression

The central dogma of molecular biology is 'DNA-RNA-Protein'. This dogma states that DNA is transcribed into RNA and RNA is then translated into protein. The transcription of DNA into RNA is mediated by the enzyme RNA polymerase. In eukaryotic organisms there are three classes of RNA polymerases, namely RNA polymerase I, II and III. Transcription of mRNA from genes which encode proteins are mediated by RNA polymerase II. The regulation of transcription of mRNA from a particular gene requires the coordinated action of specific proteins or transcription factors (discussed below).

1.16.1 Transcription

The initiation of RNA polymerase II mediated transcription of eukaryotic genes requires the interaction of transcription factors that bind DNA. These transcription factors bind DNA in a sequence-specific manner. The specific DNA sequences (elements or motifs) to which proteins bind can be in close proximity or extremely distant to a particular gene. In most cases the DNA elements involved in these interactions are clustered together within the 5′ flanking region of the gene. The region of the DNA where these transcription factor binding sites are found is termed the promoter. Many different transcription factors have been identified, which are involved in the process of constitutive or enhanced transcription. The properties and binding sequences of these protein factors have been reviewed extensively (Mitchell and Tjian, 1989; Johnson

and McKnight, 1989). The following sections describe the function and properties of a few of the known transcription factors.

The TATA box, an AT-rich sequence (consensus TATAA/TAA/T), is found about 30 base pairs upstream of the transcription start site of most eukaryotic genes. The TATA box is necessary for the correct positioning of RNA polymerase at the start site for the initiation of basal levels of transcription (Buratowski *et al.*, 1988). The rate of transcription can be enhanced considerably from its initial basal level by the binding of other transcription factors to specific sites upstream of the TATA box. These factors may be tissue-specific or be induced by a particular stimulus or act in a constitutive manner. Constitutive factor binding sites are found in a diverse number of genes. Two of the well characterised constitutive transcription factor binding sites are the CCAAT and GC (or Sp1) boxes (Santoro *et al.*, 1988; Rupp *et al.*, 1990; Dynan and Tjian, 1983; Gidoni *et al.*, 1984). The transcription factors CTF (CCAAT box transcription factor) and Sp1 bind to the CCAAT and GC boxes, respectively.

The Sp1 protein binds to the consensus sequence GGGCGG. The Sp1 protein is found in all cell types. Sp1 protein contains two glutamine-rich domains which interact with the RNA polymerase II and associated proteins (Latchman, 1995). These glutamine rich domains are essential for the activation of transcription of genes which contain the Sp1-binding site. The DNA-binding ability of the Sp1 protein is dependent on three zinc finger motifs, situated within the C-terminal region of this transcription factor (Berg, 1992). Zinc finger motifs are found in other transcription factors. The zinc finger is a 30 amino acid repeat unit of which 12 residues form a loop, which projects out from the surface of the protein (Figure 1.12). The amino acid loop is anchored at its base by two cysteine and two histidine residues, which bind a zinc atom (Figure 1.12). These two Cys and His residues are conserved in all zinc finger motifs. Zinc fingers facilitate protein to DNA binding. The Sp1 protein requires the presence of zinc for its sequence-specific binding activity.

There are several different proteins which bind to the CCAAT box (Rajmondjean *et al.*, 1988). Some of the proteins which bind to the CCAAT box can be expressed in a tissue-specific manner, whereas others are expressed in all tissues (reviewed Johnson and McKnight, 1989). One of the first CCAAT box binding proteins to be identified was the transcription factor CTF (Rosenfeld and Kelley, 1986). In contrast to the glutamine-rich activation regions of Sp1, the factor CTF contains proline-rich activation domains (Latchman, 1995). The proline-rich domain of

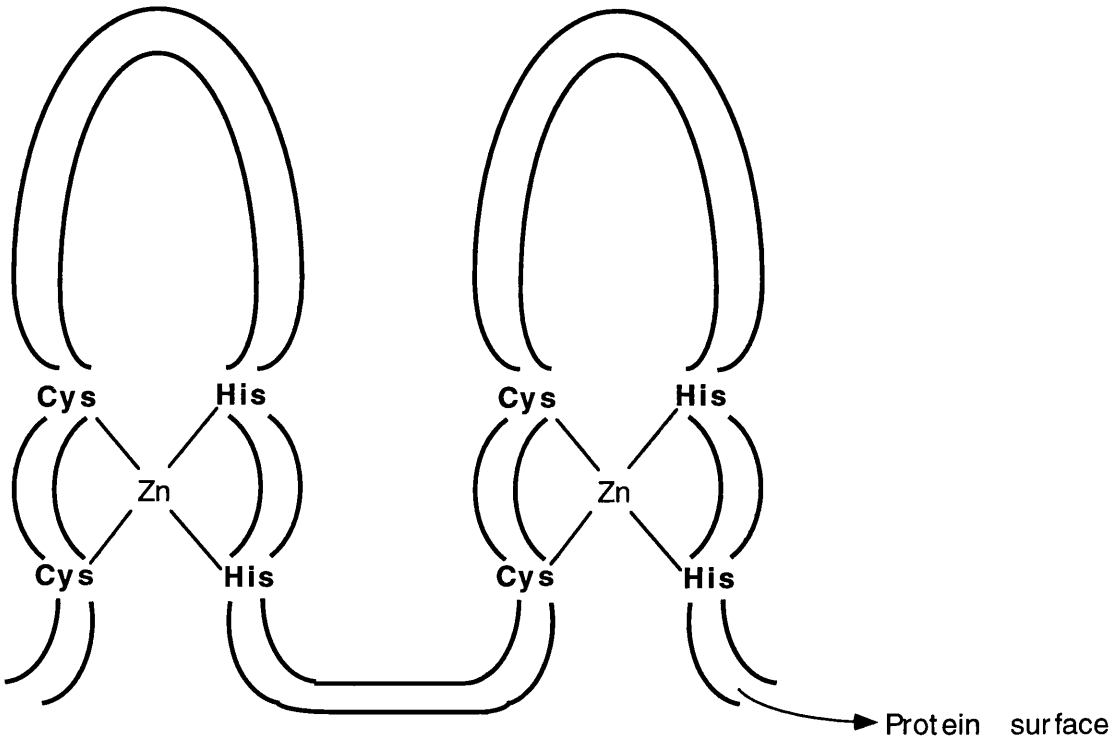


Figure 1.12 Schematic representation of two zinc finger motifs (adapted from Latchman, 1995). The finger motifs project out from the surface of the protein. A zinc atom is bound at the base of each motif by conserved cysteine and histidine residues.

CTF is situated within the C-terminal region of its protein. Proline residues constitute about 25% of the amino acids found within the C-terminal region of the CTF protein. Proline-rich domains are found in other transcription factors, such as AP-2, *c-Jun* and Oct-2 (Mitchell and Tjian, 1989). This proline-rich domain is required for the activation of transcription (Gerber *et al.*, 1994).

Some genes may have both of these Sp1 and CCAAT box DNA elements, whereas others have single or multiple copies of one or the other, working in a concerted fashion to regulate the rate of transcription. Thus, these constitutive factors play an essential role for efficient transcription of the gene to occur.

1.16.2 Enhancer elements

Other DNA sequences located at a distance from a gene (100 bp to several 1000 bp) which influence the activation of transcription are termed enhancer elements. Enhancer elements bind transcription factors which in turn influence the rate of transcription from a particular gene. Enhancer elements are usually found within the 5′ flanking region of a gene. Where enhancer elements are in close proximity to the TATA box, within the 5′ flanking DNA, this region is termed the proximal promoter. Various workers studying enhancer elements have established that these DNA sequences possess three basic properties. These properties are 1; enhancer elements can influence a promoter regardless of the orientation of its sequence to the gene, 2; enhancer elements can influence transcription equally whether it is within the proximal promoter or several 1000 bp away from the promoter, 3; the enhancer element can affect activation of transcription whether situated within the 5′ or 3′ end of the transcribed region (Latchman, 1995). In some cases enhancer elements can also be found within the intron of the gene it influences (Singh *et al.*, 1986). Many enhancer elements act in a tissue-specific manner. That is the enhancer element of a particular gene which is expressed only in a specific tissue-type will only have a stimulatory effect on genes within that tissue and not on any other tissue (see Section 1.16.3).

1.16.3 Tissue-specific expression

In eukaryotes the principal means of controlling tissue-specific production of proteins is by the regulation of the initiation of transcription (Latchman D, 1995). Regulation of transcriptional initiation is achieved by the production of certain protein factors which are exclusive to a particular

tissue or even to a specific cell-type. How protein factors can confer tissue-specificity is described below.

One of the best examples of tissue-specific gene expression is for the immunoglobulin genes. There are two immunoglobulin genes, one encodes for a heavy chain protein and the other for a light chain. Both of these genes are expressed exclusively in B lymphocyte cells. Both the heavy- and light-chain genes attain their B cell specificity via the interaction of B cell-specific factors, which bind to the promoter region. Both of the immunoglobulin genes contain a conserved octamer sequence 5′-ATTTGCAT-3′ within their 5′ flanking promoter regions, which binds the protein factor Oct-2 (Singh *et al.*, 1986). Sequence specific binding of Oct-2 confers B cell-specific enhancement of expression of the light- and heavy-chain genes (Wirth *et al.*, 1987). The tissue-specificity of expression arises from the fact that Oct-2 is transcribed and translated, at high levels, in B lymphocytes but not in most other cell types.

It is important to note that the octamer sequence in the heavy-chain gene promoter is in an inverted orientation (5′-ATGCAAAT-3′) to the octamer sequence in the promoter of the light chain gene (Figure 1.13). This example perfectly illustrates one of the properties of a DNA motif. That is, the function of the DNA motif to act as a protein binding site can be independent of its orientation. Mutation or removal of this octamer motif from the promoter region of the light or heavy genes eliminates B-cell-specific expression (Gerster *et al.*, 1987). Indeed, placing this octamer sequence within a non-immunoglobulin gene promoter results in B cell-specific expression of this altered gene (Wirth *et al.*, 1987). Thus, the Oct-2 sequence acts as a tissue-specific enhancer element to confer B lymphocyte-specific expression.

1.16.4 Photoreceptor-specific expression

Visual pigments are found exclusively in photoreceptor cells (see Section 1.7) and expression studies of photoreceptor genes have shown that they are regulated in a tissue/cell-specific manner (Lem *et al.*, 1991). Thus, it would be logical to assume that the promoter regions of visual pigment genes, and other photoreceptor-specific genes, contain DNA motifs that confer this specificity.

The use of reporter constructs in transgenic mice has allowed the determination of the minimum 5′ flanking region of a gene required for cell-specific expression in the retina. A 1.3 kb fragment from the 5′ end of the human interphotoreceptor retinoid-binding protein gene promoter

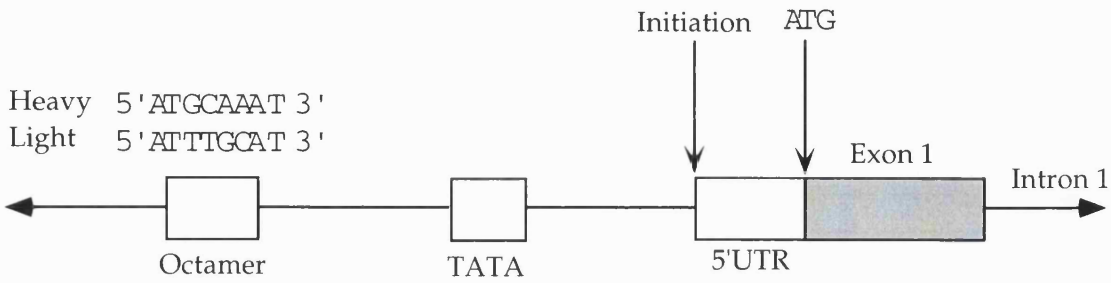


Figure 1.13 The immunoglobulin heavy- and light-chain gene proximal promoter region (adapted from Latchman, 1995). The promoter contains a TATA box and an octamer motif. The mRNA initiation start site and the ATG start codon is indicated. The 5' untranslated region (5'UTR) and exon one of the gene is also shown. Note that the octamer is in an opposite orientation in both genes.

can direct *lacZ* expression specifically to rod and cone cells (Yokoyama *et al.*, 1992). Transgenic mice containing a shorter human interphotoreceptor retinoid-binding protein gene promoter fragment of 141 bp has recently been shown to confer tissue specific expression to retinal photoreceptors and pinealocytes (Bobola *et al.*, 1995). Transgenic mice with bovine rod opsin promoter constructs have shown that fragments as small as 292 bp or as large as 2244 bp (both constructs include 70 bp of 5' untranslated region) are adequate enough to direct photoreceptor specific expression (Zack *et al.*, 1991).

Which sequences direct cone-cell-specific expression as opposed to rod-cell-specific expression? A 37 bp region termed the locus control region (LCR) within the human RCP gene promoter was shown to be essential for cone-specific expression in transgenic mice (Wang *et al.*, 1992). β-galactosidase reporter constructs containing the human RCP gene promoter with its LCR deleted, showed no expression of the *lacZ* transgene. However, the transgene of the human RCP gene promoter plus its LCR gave expression to both types of cones, the blue and the red/green cells, within the mouse retina (Wang *et al.*, 1992). This result was unexpected given the lack of sequence homology between the human BCP and RCP/GCP gene promoter. It is possible that the human RCP gene promoter construct, used in the above study, may lack certain regulatory elements which repress the expression of the RCP/GCP gene in mouse blue cones (Wang *et al.*, 1992).

A 6.4 kb DNA fragment derived from mouse BCP promoter, linked to a β-galactosidase (*lacZ*) reporter construct, was able to target expression exclusively to blue cones in transgenic mice (Chiu and Nathans, 1994a). A similar experiment using 3.8 or 1.1 kb of the human BCP gene 5′ flanking region gave reporter expression in blue cones, as well as in some cone bipolar cells, in transgenic mice (Chen *et al.*, 1994). A smaller reporter construct containing only 470 bp of the human BCP gene promoter was also found to give expression to both blue cones and a subset of cone bipolar cells (Chiu and Nathans, 1994b). This difference in expression patterns between the human and mouse BCP promoter constructs could be due to a variation of regulatory sequences between the mouse & human transgene fragments (Chiu and Nathans, 1994b). The above experiments imply that only 470 bp of 5′ flanking region from the human BCP gene is required for its blue cone cell-specific expression. Thus, the DNA sequence motifs responsible for this cell-specific expression should be contained within this short promoter region.

The LCR and the recent discovery of other photoreceptor-specific *cis*-acting elements are described in Section 5.1.2.

1.17 Aims of the project

- Which amino acid residues within the S group of pigments are responsible for their particular λ_{max}?
  To isolate and characterise a non-primate mammalian blue cone pigment (BCP) gene belonging to the S group of visual pigments, namely the porcine BCP gene.
  To compare the predicted porcine BCP protein with other members of the S group, to attain a better understanding of the visual pigments within this group.
  To predict potential spectral tuning sites, which may be responsible for causing shifts in λ_{max}, from a comparison of the S group of pigments with other vertebrate pigments.

- To test the feasibility of using a yeast (*Pichia pastoris*) expression system for the expression of human opsin genes.

- Which *cis*-acting regulatory elements direct cone-cell-specific expression as opposed to rod-cell-specific expression?
  To isolate and sequence the proximal 5′ flanking region from primate and porcine BCP genes.
  To characterise these isolated mammalian BCP gene 5′ flanking regions for evolutionary conserved *cis*-acting motifs.
  To characterise the WERI-Rb1 cell line. To establish whether or not this retinoblastoma cell line is expressing transcripts for the human BCP gene.
  To isolate nuclear proteins from Weri-Rb1 cells, cultured *in vitro*, when BCP gene expression is confirmed.
  To show that these Weri-Rb1 nuclear proteins could be used for electrophoretic mobility band shift assays on fragments of the human BCP gene promoter.

- To localise the human BCP gene to a specific region of human chromosome 7.
  To characterise a genomic cosmid clone containing the human BCP gene.

Chapter 2 Materials and Methods

2.1 DNA extraction

2.1.1 Liver tissue

The following protocol was carried out for each gram of tissue utilised. Frozen tissue samples (stored at -80°C) were ground up with liquid nitrogen, using a pestle and mortar. The resulting powder was resuspended in 5 ml of NE buffer in a Universal tube, before centrifugation at 4000 x g for 7 min. The supernatant was discarded before adding another 5 ml of NE buffer. The ground tissue was collected again by centrifugation at the same speed. This process was repeated until the supernatant was clear (approximately 4 times). After decanting the clear supernatant, the pellet was resuspended in 4 ml of NE buffer containing 1.2 mg of Proteinase K. The mixture was vortexed briefly before adding 2 ml of 10% SDS. After an overnight incubation step, at 50°C, 6 ml of equilibrated phenol was added to the sample and incubated for a further 20 min. After centrifugation at 200 x g for 10 min, the upper aqueous layer was transferred to a new tube. The aqueous DNA solution was extracted twice with a 1:1 mixture of phenol:chloroform and once with chloroform. The DNA was precipitated with one tenth volume NaCl (final concentration 0.4 M) and 2 volumes of absolute ethanol. The DNA was either spooled out using a hooked glass rod or collected by centrifugation at 3500 x g for 10 min, dried briefly under vacuum (~ 30 min) before resuspending in 500 µl of sterile TE pH 8.0. This resuspended solution was left overnight at room temperature to allow the genomic DNA to dissolve completely.

2.1.2 Yeast genomic DNA

Single yeast colonies were inoculated into sterile 10 ml of either MD or MM media, in a falcon tube or conical flask, and grown in a shaking 30°C incubator for two days. The cells were harvested by centrifugation at 1500 x g for 10 min and the supernatant discarded. The pellet of cells was washed with 10 ml sterile water by centrifugation at 1500 x g for 10 min. The cells were resuspended in 2 ml of SCE buffer, pH 7.5, containing DTT (1 M sorbitol, 10 mM sodium citrate, 10 mM EDTA, 10 mM DTT). Zymolase was added (0.3 mg) and this mixture was incubated at 37°C for 1 hour, for the formation of spheroplasts. 2 ml of 1% SDS was added to the

spheroplasts, before mixing gently, and then chilled on ice (0-4°C) for 5 min. 1.5 ml of 5 M potassium acetate, pH 8.9, was added and mixed gently, before centrifuging at 10000 x g for 5 min. The supernatant was transferred to a fresh tube. 2 volumes of ethanol was added to the supernatant and incubated at room temperature for 15 min. After centrifugation at 10000 x g for 20 min the pellet was gently resuspended in 0.7 ml of TE buffer, pH 7.4 (Tris-HCL, pH 7.4, 1 mM EDTA, pH 8.0) and transferred to a microcentrifuge tube. The yeast DNA was extracted with an equal volume of phenol:chloroform and then again with an equal volume of chloroform. The upper aqueous layer was split equally into two microfuge tubes. 0.5 of a volume of 7.5 M ammonium acetate, pH 7.5, and 2 volumes of ethanol was added to each tube. The tubes were placed either on dry ice for 10 min or at -20°C for 60 min. The precipitated DNA was collected by centrifugation at 10000 x g for 20 min. Each pellet was washed once with 1 ml of 70% ethanol before vacuum drying the DNA. Each DNA pellet was resuspended in 25 µl of sterile TE buffer, pH 7.5, and left overnight to completely dissolve.

2.1.3 Plasmid and cosmid mini-prep

All mini-preps of cloned DNA (plasmid and cosmid) were performed initially using a DNA binding resin supplied with the Wizard Miniprep kits (Promega) and later with self-prepared solutions and resin. Bacterial cultures, containing the cloned DNA, were grown overnight in 5 ml of LB broth in a 37°C incubator with shaking at 250 rpm. The cells were collected by centrifugation at 3000 x g for 5 min. The resulting pellets were resuspended in 200 µl of Cell Resuspension solution (50 mM Tris-HCL pH 7.5, 10 mM EDTA, 100 µg/ml RNAse A) and transferred to a microfuge tube (1.5 ml eppendorf). 200 µl of Cell Lysis solution (0.2 M NaOH, 1% SDS) was added and mixed by inverting the tube several times, until the mixture became clear/opaque. 200 µl of Neutralisation solution (1.32 M potassium acetate, pH 4.8) was added and mixed again by inverting the tube several times. The cell wall debris was collected by centrifugation at 13000 rpm in a Heraeus Biofuge A microcentrifuge. The supernatant was transferred to a new microfuge tube. 500 µl of either Wizard Miniprep or self-prepared DNA binding resin was added to the supernatant and mixed by gently inverting the tube several times. (The self-prepared DNA binding resin was made as follows: 5 g of 'Celite Analytical Filter Aid' (BDH Chemicals, cat. no. 33134) was suspended in 250 ml of deionised water by shaking vigorously. This mix was left to stand for 2-3 hours

before the supernatant was decanted off and discarded. 250 ml of 6 M guanidine hydrochloride, 50 mM Tris-HCL and 20 mM of EDTA was added to the prepared 'Celite'. This solution was shaken hard to mix and was now ready to use as a DNA binding resin.) The resin/cloned DNA mixture was vacuum pumped into a Wizard Mini-Column (Promega). The Wizard Mini-Column was washed through with 1.5 ml of Column Wash (80 mM NaCl, 8 mM Tris-HCL, 2 mM EDTA, 55% ethanol) using a vacuum pump. The Wizard Mini-Column with bound DNA was transferred to a microfuge tube and centrifuged in a microfuge for 10 seconds to flush out any remaining Column Wash. The bound DNA was stripped from the Wizard Mini-Column by centrifuging 90 µl of hot water (75-80°C) through it into a sterile microfuge tube. Usually 3 µl aliquots of this eluted DNA was either run on an agarose gel (see Section 2.4) to determine the approximate concentration or used for restriction enzyme digest to check for cloned inserts (as described in Section 2.3).

2.2 Polymerase chain reaction (PCR)

2.2.1 Standard PCR

PCR conditions were based on the parameters reviewed by Innis and Gelfand (1990) and Saiki *et al.* (1985). PCR was carried out using a $(NH_4)_2SO_4$ reaction buffer (Bioline UK) with *Taq* (*Thermus aquaticus*) polymerase in a 50 µl volume. A typical 50 µl volume PCR mix contained:

1 X concentration reaction buffer
dNTPs (200µM each of dATP, dCTP, dGTP and dTTP)
25 pmol of each oligonucleotide primer
1.5 mM of $MgCl_2$
0.02 units of *Taq* polymerase
DNA template to be amplified
(or no DNA, for negative control)
sdH_2O.

The reaction mixture was overlaid with mineral oil. PCR conditions varied according to the disassociation temperature of the primers and the expected length of the amplified product. Average conditions were as follows; an initial denaturisation step of 1 min at 94°C which was followed by 35 cycles of 3 steps consisting of denaturation at 94°C for 5 seconds, a primer annealing step for 10 seconds and an extension step at 72°C. A final extension step at 72°C for a minimum of 1 minute was always added. Roughly 30 seconds of extension time was given for

each kb of DNA to be generated. The temperature used for the annealing step was always at least 1°C lower than the estimated annealing temperature of the primers in the PCR. The optimal primer annealing temperature was estimated from its sequence, adding 4°C for each G or C and 2°C for each A or T nucleotide present. Occasionally extra $MgCl_2$ was added to the reaction buffer to optimise specificity of primer annealing. All PCRs were carried out on the same machine, a Hybaid OmniGene block.

Primer	Sequence
Ex1- 604/589R	5'-CTGCCGCAACTTTTTGTAGCGC-3'
Ex1- 529/514R	5'-GCCCATGAAAGCTGCCTGGAGG-3'
Ex1+ 410/431F	5'-ATGAGAAAAATGTCGGAGGAAG-3'
Ex1+ 469/484F	5'-GCCGTGGGATGGGCCTCAGTAC-3'
Ex2+ 784/799F	5'-TCACTGGCCTTCCTGGCCTTTGA-3'
Ex2- 929/908R	5'-GGCTCCAGCCAAAGAAGGGTGG-3'
Ex3+ 935/954F	5'-ATCCCTGAGGGCCTGCAGTG-3'
Ex3- 1069/1054R	5'-TGAGTAGGAGAAGCAGATGAGGG-3'
Ex4+ 1098/1117F	5'-TTGCAGCTCAGCAGCAGGAG-3'
Ex4- 1276/1254R	5'-TGCAGGAATGGTGACAAGCCGTA-3'
Ex5+ 1354F	5'-TCCAGTTCCGAGCTTGCATCATG-3'
Ex5- 1453/1433R	5'-GTTGGGGCCAACTTGGGTAGA-3'

Table 2.1 Primers specific to the coding region of the human BCP gene. The primer name refers to the one of the 5 exons (Ex2 = exon 2) of the BCP gene. A plus (+) or a F indicates the forward primer, and a minus (-) or a R refers to the reverse primer. The numbering system for each primer refers to the original notation given to the human BCP gene by Nathans and colleagues (1986a). (Note that the first 5 bp of the Ex5+ 1354F primer is intronic sequence).

2.2.2 PCR from phage

Heat denaturation of a phage lysate destroys the bacteriophage protein coat releasing DNA into the surrounding media, hence facilitating the use of PCR to amplify DNA inserts of interest. An aliquot (1µl) of the phage retinal cDNA library (Stratagene) was used as a template with the other standard components of PCR to a final volume of 50µl. The reaction mixture was overlaid with mineral oil (Sigma) to prevent evaporation. The PCR conditions were as follows; initial denaturation, 94°C for 3

minutes, then 35 cycles of 94°C for 5 seconds, 56°C for 5 seconds, 72°C for 30 seconds and then a final elongation step at 72°C for 2 minutes.

2.2.3 Unpredictably primed walking PCR (UPW-PCR)

UPW-PCR was adapted from a protocol developed by Dominguez and López-Larrea (1994). This method is used to amplify, by PCR, regions of unknown DNA which flank known DNA sequences. Undigested genomic DNA is used as template with four primers, 2 of which are sequence-specific (SS) and the other two are universal walking (WP). The SS primers are designed to the known sequence with the 3′ end facing towards the unknown region, with one nested ('Inner-SS') inside the other ('Outer-SS'; see Figure 5.5). One of the SS primer is the 'Outer-SS primer' while the nested one is the 'Inner-SS primer'. The two WP primers (WP-33 and WP-17) are of a non-specific nature with the sequence given below:

WP-33 = 5′-TTTTTTTTTTTTTTTGTTTGTTGTGGGGGGGTT-3′
WP-17 = 5′-TTTTTGTTTGTTGTGGG-3′

The WP-17 is completely within the sequence of the WP-33. The UPW-PCR consists of two successive rounds of amplification, the first utilising the Outer-SS primer with the WP-33 and the second using the Inner-SS primer with the WP-17.

A first round 50 µl PCR mix was set up as described in Section 2.2.1, but without the Outer-SS primer, and this was placed at 94°C for 1 min. This mixture was then incubated at 15°C for 2 min, then 25°C for 10 min, then 72°C for 1 min and finally at 90°C for 1 min before adding the Outer-SS primer to the mix. Then 35 cycles of a standard PCR was carried out with an extension step of 1 min and the annealing temperature (Ta) was at 1°C below the optimal Ta of the Outer-SS primer (see Section 2.2.1). Approximately 10 µl of the first round PCR was run on a 1.5% agarose gel (Section 2.4) to determine whether a smear of DNA was present.

Second round PCR was carried out using a 1/1000 dilution of the first round PCR as a template in conjunction with the Inner-SS primer and the WP-17. After an initial denaturisation step at 94°C for 1 min the PCR conditions were as follows; 5 seconds @ 94°C, 1 second @ Ta (Ta = 1°C below the optimal Ta of the Inner-SS primer), 30 seconds @ 72°C, for 35 cycles with a final extension step of 72°C for 1 min. 10 - 20 µl of the second round PCR product was run on a 1.5% agarose gel (see Section 2.4) to determine whether discrete bands were present.

Primer	Sequence
Alu 1	5'-GCCTCCCAAAGTGCTGGGATTACAG-3'
-747/-725F	5'-TTCAGTGAACAGACAGGAGGAAG-3'
-677/-707R	5'-GCCATGTTCATGCCTACCCCTACCCTTAGCT-3'
-651/-629F	5'-CCCTGGCTGGAGGTCTGAGTAGG-3'
-629/-651R	5'-CCTACTCAGACCTCCAGCCAGGG-3'
-572/-551F	5'-GGATGCCCTAGACCTCTAGGCA-3'
-551/-572R	5'-TGCCTAGAGGTCTAGGGCATCC-3'
-450/-473R	5'-GGAACCAGCCCTGGGCTGCAGGAA-3'
-424/-404F	5'-GCACCTTCCACAAAGGGGCAG-3'
-392/-409R	5'-TCCTCAACTCATCTGCCC-3'
-303/-281F	5'-TCTTCAGTGAATATGGGTCAGCA-3'
-275/-294R	5'-GGGAGGTGCTGACCCATATT-3'
-186/-208R	5'-GTTCTTCAGAGGATAAGAGATTG-3'
-137/-110F	5'-TCTAATCCCAAACTTTGTCCTTGGGAGG-3'
-99/-121R	5'-CTAATCCTAAACCTCCCAAGGAC-3'
-9/-29R	5'-CACACCCTCCTCTGAGTCCTC-3'

Table 2.2 Primers specific to the 5′ flanking region of human BCP gene. (Note that the Alu1 primer is an oligonucleotide designed from a consensus Alu repeat sequence (Nelson *et al.*, 1989)). These primers were used for the amplification of 5′ flanking regions from primate and porcine BCP genes, and for EBSA experiments (Sections 5.21, 5.2.2, 5.2.53 and 5.2.54). F and R after a primer name indicates the forward primer and reverse primer, respectively. The primer name indicates the position of this oligonucleotide upstream of the ATG start codon. Thus, the first nucleotide at the 5′ end of the -99/-121R primer sequence is the 99th bp, and the last nucleotide at the 3′ end is the 121th bp, upstream of the ATG start site (-1 being the first bp 5′ to the ATG codon).

Primer	Sequence
Rod Ex4+	5'-GCGGCTGCCCAGCAGCAGG-3'
Rod Ex5-	5'-GGCCGGGCCACCTGGCTCG-3'
RCP Ex4+	5'-GCTGCATCATCCCACTCGC-3'
RCP Ex5-	5'-GACGCAGTACGCAAAGATC-3'
GCP Ex4+	5'-GCTGCATCACCCCACTCAG-3'
GCP Ex5-	5'-GAAGCAGAATGCCAGGACC-3'

Table 2.3 Opsin primers used for the PCR amplification of human rod opsin (Rod), red cone pigment (RCP) and green cone pigment (GCP).

Primer	Sequence
WP-33mer	5'-TTTTTTTTTTTTTTTGTTTGTTGTGGGGGGGTT-3'
WP-17mer	5'-TTTTTGTTTGTTGTGGG-3'
Ex1-604/589R	5'-CTGCCGCAACTTTTTGTAGCGC-3'
Ex1-529/514R	5'-GCCCATGAAAGCTGCCTGGAGG-3'

Table 2.4 The primers which were used for unpredictably primed walking (UPW) PCR on primate BCP genes (see Section 5.2.2). First round UPW-PCR was carried out using the WP-33mer in conjunction with the Ex1-604/589R ('Outer Sequence-Specific') primer. Second round UPW-PCR was performed using the WP-17mer and the Ex1- 529/514R ('Inner Sequence-specific') primer.

Repeat Primer	Sequence
$(ACG)_6$	5' ACGACGACGACGACGACG 3'
$(CAG)_6$	5' CAGCAGCAGCAGCAGCAG 3'
$(GGC)_6$	5' GGCGGCGGCGGCGGCGGC 3'
$(GGA)_6$	5' GGAGGAGGAGGAGGAGGA 3'
$(GGT)_6$	5' GGTGGTGGTGGTGGTGGT 3'
$(CT)_{12}$	5' CTCTCTCTCTCTCTCTCTCTCTCT 3'
$(GT)_{12}$	5' GTGTGTGTGTGTGTGTGTGTGTGT 3'
BCP Repeat+	5' CCATCTCCCATATTTCTATCTAG 3'

Table 2.5 List of primers that were used in PCR with the vectorette protocol to test for the presence of triplet and dinucleotide repeat sequences within the human BCP cosmid clone, and the BCP Repeat+ primer which was used to sequence across the GGT repeat units found within the cosmid clone (see Section 6.2.2).

2.2.4 Vectorette PCR

The Vectorette PCR protocol is outlined schematically in Figure 2.1. Vectorette PCR is another method for generating DNA sequences from unknown regions. However unlike UPW, this protocol utilises restriction enzyme-digested DNA which is ligated to a double stranded DNA linker, which is known as the 'Vectorette' (ICI Biological Products). In addition this method only requires two primers, one of which is sequence-specific, with its 3' end towards the unknown region of DNA, and a 'Vectorette-specific' primer. The Vectorette linker contains a compatible 3' end which

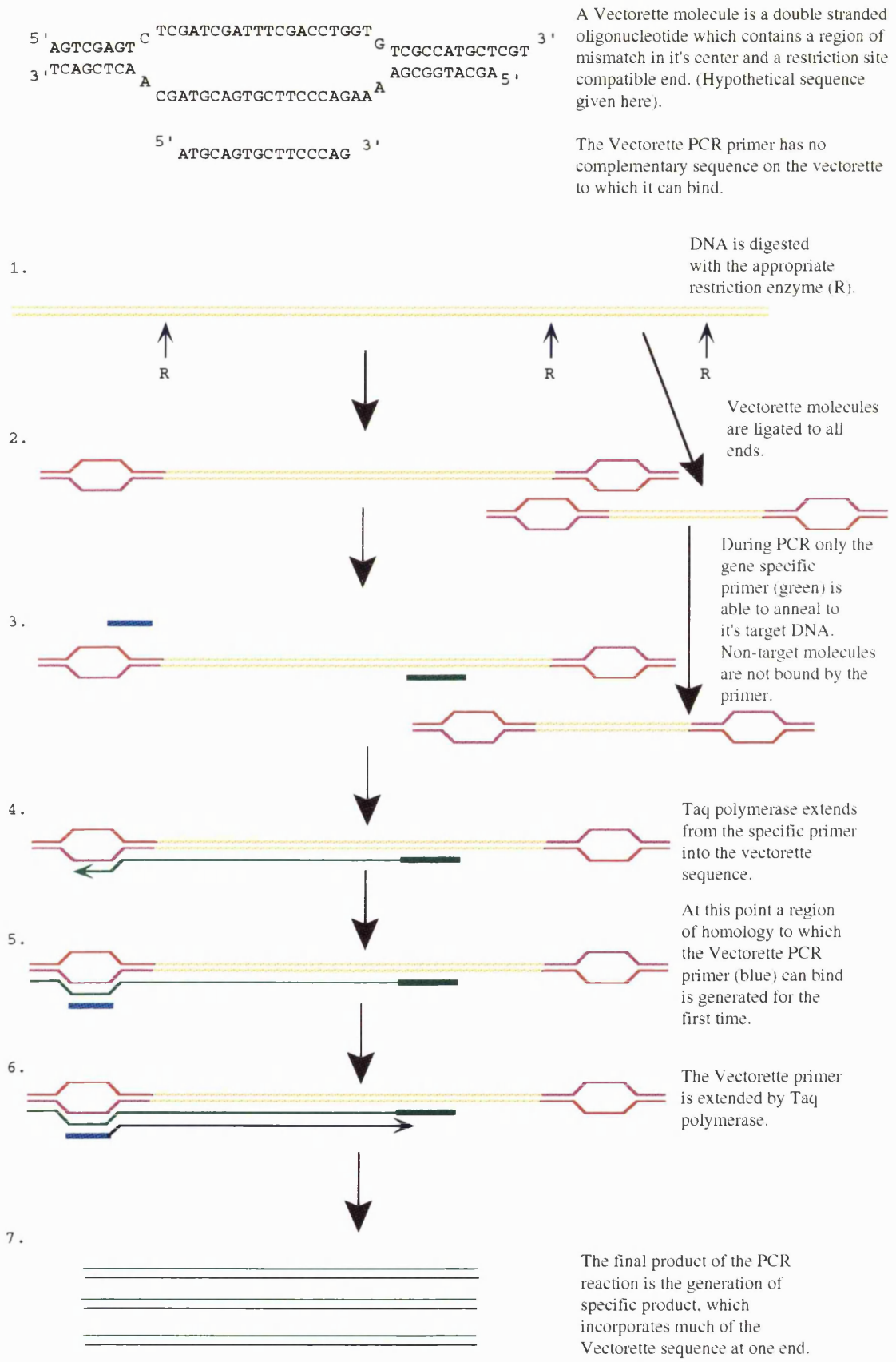


Figure 2.1 Schematic representation of the principle of Vectorette PCR (Dulai, 1996).

allows it to be ligated to an appropriately digested DNA fragment. The Vectorette is designed such that it contains a 22 bp stretch of mismatched nucleotides within it, which form a 'bubble' where the two non-complementary strands cannot anneal (Figure 2.1). The Vectorette-specific (VS) primer sequence is identical to the sequence of the anti-sense strand of the Vectorette linker which forms the 'bubble'. Thus, during a Vectorette PCR the VS primer does not have a complement strand to anneal to until this strand is synthesised by the sequence-specific primer (Figure 2.1). Although this method can be utilised for genomic DNA, in this study the Vectorette system was used on cloned cosmid DNA.

2.2.4.1 Construction and PCR of a Vectorette-cosmid clone

2 μg of cosmid DNA was digested for 3 hours with *Hae*III (see Section 2.3). 3 pmol of the Vectorette linker, with a *Hae*III compatible 3' end (blunt end linkers), was ligated to 1 μg of the digested cosmid DNA with 10 U of T4 DNA ligase (Pharmacia) in the presence of the restriction enzyme buffer, 2 mM ATP and 2 mM DTT at 20°C for 1 hour. This digestion/ligation mix was then cycled 10 times between 37°C for 30 min and 20°C for 1 hour. Ligation of the Vectorette linker to digested DNA abolishes the *Hae*III recognition sequence, thus incubation at 37°C will only digest religated cosmid DNA. The cycling of ligation/digestion promotes the eventual attachment of Vectorette linkers to all the digested fragments of cosmid DNA. 2 μl of this Vectorette/cosmid DNA was used as template for Vectorette PCR. A standard PCR mix was prepared with only one primer present, the sequence-specific primer. 30 cycles of a standard PCR was carried out before adding the Vectorette-specific primer and another 35 cycles was performed. An aliquot of this PCR (15-20 μl) was run on a 1.5% agarose gel (see Section 2.4) for visualisation of any discrete bands.

2.3 DNA digestion with restriction endonucleases

Plasmid, cloned DNA and PCR products were digested with a number of restriction endonucleases according to manufacturers instructions (BRL, NBL, and Cambio). Digestion reactions were generally prepared in 10 or 20 μl reaction mixes containing 1 x the appropriate buffer with 5 units of enzyme. Incubation times varied between a minimum of three hours for cloned DNA, such as mini-preped cosmids or plasmids, to overnight for genomic DNA. Once digestion was complete, reactions were

stopped by heating the samples at 65-85°C to denature the restriction endonucleases.

2.4 Electrophoresis and visualisation of DNA

Digested DNA and PCR generated products were separated for analysis by agarose gel electrophoresis (modified from Sambrook *et al.*, 1989). DNA fragments were size fractionated either on a 1% or 1.5% (w/v) agarose gel (20cm x 20cm) made in 1 x TAE buffer. ϕX174/*Hae*III, λ/*Hind* III or a 1 kb ladder marker were used molecular size standards (see below). DNA was loaded in 1 x loading buffer and electrophoresed in 1 x TAE buffer. DNA fragments were visualised by staining the gel in ethidium bromide solution (0.5μg/ml) for 15 minutes and then placing it on a UV transilluminator for photography.

Size of DNA Markers (kb)		
ϕX174/*Hae*III	λ/*Hind*III	1 kb ladder
1.358	23.130	12.216
1.078	9.416	11.198
0.872	6.557	10.180
0.602	4.361	9.162
0.310	2.322	8.144
0.281	2.027	7.126
0.271	0.564	6.108
0.234	0.125	5.090
0.194		4.072
0.118		3.054
		2.036
		1.636
		1.018
		0.517
		0.506
		0.396
		0.344

2.5 Ligation of DNA inserts into plasmid vectors

Ligase mediated cloning of PCR products was performed using reagents supplied by Invitrogen (TA cloning kit) and Promega (pGEM-T Vector System) according to the recommended protocols. Prior to ligation, PCR products were spun through a column containing Sephacryl S-400

(Pharmacia) to remove any residual primers and dNTPs. 50 ng of vector was ligated to an appropriate amount of purified PCR product such that the ratio of molar ends of vector to PCR product was from 1:1 to 1:3. The ligation reaction was carried out overnight at the recommended temperature of 12°C (Invitrogen) or 15°C (Promega). An aliquot of this ligation reaction was used for transformation of competent *E coli*, before storing the remainder at -20°C. Ligation of digested cosmid to dephosphorylated vector (see Section 2.14.1) was carried out in a 10 µl volume, (3 µl vector, 5 µl digested cosmid, 1 µl ligation buffer and 1 µl of ligase (Promega)).

2.6 Transformation of competent *E. coli*

DH5α competent cells were supplied by Invitrogen or Gibco BRL and the transformation procedure was adapted from the manufacturer's recommended protocol. Transformation was achieved by gently mixing 1-3 µl of a ligation reaction to 50 µl of competent cells. The cells were incubated on ice (4°C) for 30 min before a heat shock step at 42°C for 35 seconds. The cells were then immediately placed on ice for 2 min to quench the heat shock. 450 µl of SOC media was added to the cells before incubating them at 37°C to allow the expression of antibiotic resistance. Usually 150 µl of the recovered transformed *E coli* were plated out onto LB/agar plates containing the appropriate antibiotic (see next Section).

2.7 Growth media and selection of recombinants

LB (Luria-Bertaini) broth and agar (bacteriological, 15g/l) supplemented with the appropriate antibiotic were used for the growth of all bacterial strains (see below for preparation). Antibiotic selection of transformants was achieved with 50µg/ml of Ampicillin or Kanamycin for TA vector, and Ampicillin only for pGEM-T and Bluescript vectors. Blue/white colour selection of recombinant transformants by the process of α-complementation was achieved by spreading 40µl of X-gal (20 mg/ml in DMF) on to the surface of an agar plate and allowing this to evaporate for about an hour. After plating out, the recovered transformation reaction the plates were incubated overnight at 37°C.

1 litre of LB broth mix was prepared as follows:

10 g of tryptone,

5 g of yeast extract,

10 g sodium chloride and 1 ml of sodium hydroxide, made up to 1 litre with distilled water and then sterilised by autoclaving at 15 psi for

20 min. For 1 litre of LB/agar, for LB plates, 15 g of agar was added before autoclaving, and the plates were poured using aseptic techniques when the molten LB/agar was about 40-45°C. A stock solution of 50 mg/ml of Ampicillin was prepared in distilled water and filter sterilised by passing through a 0.22-micron filter and stored at -20°C. A 10 mg/ml stock solution of Kanamycin was prepared in the same way as the Ampicillin.

2.8 Screening recombinant *E. coli* for inserts

White colonies were screened for the presence of the correct size insert from the plates containing the transformed clones by either restriction endonuclease digestion or by colony PCR. In the former method, 7 µl of plasmid DNA isolated from each clone (see Section 2.1.3) was digested with one or two restriction endonucleases which were designed to cut out the insert from the multiple cloning site of the plasmid vector. The exact restriction endonuclease(s) utilised depended on the cloning vector (see Section 2.5) and the manufacturer's recommendations were followed for the digest conditions (see Section 2.3). Digests were then run on an agarose gel and visualised (Section 2.4). Recombinant clones containing inserts of the correct size were then sequenced (see Section 2.12).

Colony stab PCR was also utilised to screen white colonies for inserts. Single white colonies were sampled using the tip of a sterile toothpick and this was swirled in 10 µl of sterile distilled H_2O contained in a 0.5 ml Eppendorf. The same colony was either subcultured on to a gridded LB/agar plate or inoculated in to 5 ml of sterile LB broth with the appropriate antibiotic. The 10 µl of sdH_2O containing the inoculated colony was incubated at 95°C for 5 minutes to lyse the *E. coli* cells. The cell debris was pelleted by centrifugation 13000 rpm in a bench microcentrifuge. 1 µl of the cleared supernatant was used in a standard PCR. The plasmids from colonies which gave the correct size inserts by PCR were isolated and sequenced (see Section 2.12)

2.9 Screening human and porcine cosmid genomic libraries

The pig genomic library utilised in this study was constructed in the cosmid vector Lorist B, with an estimated titre of 2.9×10^8 cfu/ml (kindly donated by Dr M Vaiman, Laboratoire de Radiobiolgie Appliquée, 78352 Jouy-en-Josas, France). The human genomic library, consisting of approximately 5×10^5 independent recombinants, was also contained in the cosmid vector Lorist B (Cachon-Gonzalez, 1991).

2.9.1 Preparing genomic cosmid library master filters

The following protocol was carried out for both genomic libraries. A 20cm x 20cm charged nylon filter (Hybond-N^+, Amersham) was placed on the surface of a megaplate of LB/agar, containing 50 µg/ml of Kanamycin. A glycerol stock of the library, stored at -80°C, was thawed on ice and 2 µl was aliquoted into 2 ml of LB/Kanamycin media and mixed by gentle inversion. This homogenous mixture was evenly distributed over the surface of the filter in 200 µl aliquots and spread uniformly, using a sterile glass spreader, to cover the whole area of the filter. The master filter was incubated at 37°C overnight until single colonies were approximately 2 mm in diameter.

2.9.2 Preparing replica filters

Two replica filters were constructed, under aseptic conditions, from the master. Gloves were worn throughout and all manipulations of filters were carried out using ethanol soaked flamed forceps. The master filter was removed from its megaplate and placed colony side up on a large sheet of Whatmann 3MM paper. The replica filters were prewetted, by placing them on fresh LB/Kanamycin megaplates, and then carefully overlayed on top of the master so as not to introduce bubbles between the nylon membranes. Another sheet of Whatmann was placed over the filters. A glass plate, which was placed over the membranes, was used to apply equal pressure by hand over the whole surface area to transfer the colonies from the master to the replica filter. Asymmetric needle holes were introduced at the edges of the filters for orientation. Replica filters were carefully peeled off the master and placed on megaplates and incubated at 37°C for approximately 8-12 hours, whilst the master was recovered at 30°C overnight, then stored at 4°C.

2.9.3 Preparing colonies on replica filters for hybridisation

Whatmann 3MM filter paper was soaked in denaturing solution in a tray and the excess fluid was poured out. The replica filters were placed on top of the presoaked Whatmann paper, colony side up, and left to soak for 10 min. The replica filters were then transferred on to Whatmann paper which had been presoaked in neutraliser solution and left to soak for 10 min. the above two steps lyses the colonies and enables the DNA to bind to the Nylon membrane. The lysed cell debris was removed by gently wiping the surface of the filter by hand (with gloves) under 2 x SSC. The

filters were air dried. The DNA was fixed to the Nylon membrane by either baking at 80°C or by UV crosslinking (see Section 2.10.2).

2.9.4 Nucleic acid hybridisation to replica filters (primary screen)

Replica filters with bound cosmid DNA were prehybridised overnight at either 55°C or 42°C in 30 ml of Church's with 1% BSA (Church and Gilbert, 1984). The lower temperature was used if no primary signals were obtained after hybridisation at 55°C. 50-100 ng of purified α-P^{32}-dCTP labelled probe (see Section 2.10.3) was boiled for 10 minutes and quenched on ice. Approximately 10-15 ml of the prehybridisation solution was discarded and the labelled probe was added to the remaining mix. Hybridisation was carried out overnight at either 55°C or 42°C. Hybridisation was performed either in roller bottles or in sandwich boxes on a shaking platform.

Filters were washed twice in 2 x SSC, 0.1% SDS for 1 hour each wash. The filters were monitored with a Geiger counter. If necessary the filters were washed again, at a higher stringency, in 1 x SSC, 0.1% SDS. The replica filters were then autoradiographed.

2.9.5 Identification of positive colonies after a primary screen

The position of replicated positive signals on the autoradiographs were marked onto a clear acetate sheet, using the orientation needle marks for positional guidance. The acetate sheet was then placed on to a light box and a few drops of LB/antibiotic spread over it. The master filter was placed on to the acetate and aligned with the needle holes to facilitate the exact identification of the areas containing the positive signals. Colonies within a 5 mm diameter area of the positive signals were picked using a sterile plastic loop and inoculated in to 200 µl of LB/antibiotic. These colonies were incubated at 37°C, 225 rpm for 1 hour to recover the cells and then were utilised for secondary screening.

2.9.6 Secondary screening

A 1/1000 dilution of each pool of colonies recovered from primary positive screenings were prepared in LB broth/antibiotic. 5 µl, 20 µl and 100 µl of each dilution mix was plated out on to 8.5 cm diameter plates of LB/antibiotic agar and grown at 37°C for 10 hours. Glycerol stocks (2:8 ratio of glycerol:cell culture) were made with an aliquot of the remaining diluted colonies and the rest was stored for 10 hours at 4°C. Whichever of the 3 dilution plates, for each primary pool, gave approximately 250

colonies of even separation was noted and this dilution was used to prepare master filters for secondary screening. Replica filters were prepared from the master filters exactly as described in Section 2.9.2, and the procedures for orientation, lysing of cells, binding of DNA, hybridisation and washes for subsequent autoradiography were identical to that used for the primary screening.

Replicated positive signals were picked as single colonies from the secondary master filters and inoculated into 3 ml of LB/antibiotic for isolation of individual cosmids (Section 2.1.3). The isolated positive cosmids were digested, Southern blotted on to Hybond N filters (see Section 2.10) and probed for the correct insert.

2.10 Southern blotting

The following techniques were adapted from the original discovery that DNA can be transferred by capillary action from an agarose gel to a membrane described by Southern (1975).

Digested cosmid DNA (Section 2.3) was run on an 1% agarose gel and visualised (see Section 2.4). Electrophoresed DNA was denatured by washing the gel in 0.4 M NaOH for 15 min. The DNA was know ready for blotting. The equilibrated gel was placed upside down on a wick of 3MM Whatmann paper which was placed over a reservoir of 0.4 M NaOH (the transfer buffer). A Hybond N membrane was cut to the exact size of the gel and marked for orientation of the lanes. The membrane was placed on to the agarose gel and then two pieces of 3MM Whatmann, which were cut to the size of the gel and soaked in transfer buffer, were placed on top of this. When placing the membranes and Whatmann paper care was taken as not to introduce any bubbles between any layers. Several layers of paper towels were then added on top and a uniform weight was then placed on top of these. Capillary transfer can be achieved within 5 hours, but was normally left to proceed overnight. The DNA was fixed to the membrane either by baking at 80°C for 2 hours or by UV crosslinking according to the manufacturers instructions. UV crosslinking was done on an Amersham UV Crosslinker set at 70000 $\mu J/cm^2$, after the membrane was dried at 80°C for 10 min.

2.10.3 Radioisotope labelling of probes

This protocol is adapted from the 'random priming' method developed by Feinberg and Vogelstein (1983). This method involves the annealing of hexanucleotides, of random sequence, to the denatured

template of interest, and these act as primers for the incorporation of [α-^{32}P]dCTP into a newly sythesised strand by the Klenow fragment of DNA polymerase I. The radioactive probe encompassing the entire length of your fragment of DNA can then be used for hybridisation to denatured DNA bound to Nylon membranes.

The DNA probe of interest used in this study was generally a PCR product, which was purified from residual primers and dNTPs as described in Section 2.5. The purified template was heat denatured at 95°C for 3 min and then placed immediately on ice. 1 x reaction buffer (BRL), 5 units of Klenow (BRL) and 30 μCi of [α-^{32}P]dCTP was added to the denatured DNA, to a final volume of 30 μl, and incubated at 37°C for 1 hour. Excess unincorporated dCTP was removed by passing the reaction mixes through a Nick-Column (Pharmacia) using the manufacturers standard protocol. The radioactive probe was heat denatured at 95°C for 5 min before using it for hybridisation. The hybridisation procedure of the probe to the Southern blotted DNA was carried out exactly as described in Section 2.9.4.

2.10.4 Stripping radioactive probes off membranes

Hybond N and Hybond N$^+$ filters were stripped by washing them in sandwich boxes with boiling water. Washing of the filter with boiling water was repeated several times. Removal of bound probe was confirmed by autoradiography.

2.11 Autoradiography

For products which had been probed or labelled with α-^{32}P the autoradiograph was exposed at -80°C using Kodak X-Omat film within a cassette containing two intensifying screens. The time of exposure depended on the strength of the radioactive signal, but usually between 1 hour to overnight. For products labelled with ^{35}S the autoradiograph was exposed at room temperature overnight using Fuji XL film, within a cassette without intensifying screens.

2.12 DNA sequencing

All cloned plasmid inserts were sequenced on both strands either using a manual sequencing protocol or the automated ABI sequencer. It was found that cosmids could be sequenced directly on the automated ABI sequencer, thus dispensing with the need to subclone them. However, success with direct sequencing of intact cosmids on the ABI was found to depend on the particular primer utilised for sequencing. Some primers

would work fine for PCR but were no good in sequencing reactions. The exact reason for this discrepancy is undetermined.

2.12 .1 Manual sequencing of cloned DNA

Sequencing reactions were performed using a standard ^{35}S-dATP incorporation method utilising a commercial T7 polymerase kit (Pharmacia). This protocol is an adaptation of the original dideoxynucleotide sequencing method (Sanger *et al.*, 1977).

1-2 µg of double stranded plasmid DNA (about 5 µl of a mini-prep Section 2.1.3) was denatured in a 0.4 M NaOH solution, at room temperature for 10 min, in a total volume of 40 µl. The DNA was precipitated by adding 11 µl of 2 M sodium acetate (pH 4.8) and 110 µl of 100% ethanol and, after mixing by inverting, this was incubated at -80°C for 30 min. The DNA was recovered by centrifugation at 13000 rpm in a bench microfuge. The supernatant was discarded and the DNA pellet washed in 110 µl of cold 70% ethanol. The DNA was collected again by centrifugation and the supernatant discarded. After vacuum drying the DNA pellet was resuspended in 10 µl of sterile water, 2 µl of the sequencing primer (5 pmol/µl) and 2 µl of annealing buffer. The primer was annealed by incubating at 65°C for 5 min, then at 37°C for 10 min and finally at room temperature for 5 min before collecting the mix by a centrifugation step.

To the collected denatured DNA with annealed primer was added 3 µl of labelling mix A (containing dCTP, dGTP and dTTP), 1 µl of ^{35}S-dATP and 2 µl of diluted T7 polymerase. This was mixed and then incubated at room temperature for 5 min. While the ^{35}S-dATP mix was incubating, 2.5 µl of each of the four dideoxy termination mixes (ddATP, ddCTP, ddGTP and ddTTP) was heated at 37°C in separate tubes. 4.5 µl of the ^{35}S-dATP mix was added to each of the four termination mixes and incubated at 37°C for 5 min before stopping the sequencing reaction by adding 5 µl of Stop solution (95% formamide with loading dye) to each tube. These reactions were then run on a polyacrylamide gel (see Section 2.13).

2.12 .2 Automated fluorescent sequencing of cloned DNA

All automated sequencing was carried out on a ABI 373A DNA sequencer. This system utilised fluorescent dyes and lasers to read a particular DNA sequence which is analysed and converted to a graphic image on a computer. Full details of the mechanisms of this automated

system can be obtained from the suppliers (ABI cycle sequencer from Perkin Elmer, Foster City, California).

Mini-preped cosmids and plasmids were cycle sequenced using the ABI PRISM^{TM} AmpliTaq® DNA polymerase FS dideoxy kit and a sequence-specific primer for sequencing according to the manufacturers protocol. The sequencing reactions consisted of 8 µl reaction mix, 1 µl sequencing primer (3.2 µM) and 1 µg of template DNA in a total volume of 20 µl. The reaction mix was then incubated at 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 min for 25 cycles, performed either on a Perkin Elmer 9600 or a 2400 PCR machine. Excess fluorescent unincorporated dye was removed from the reaction mix by ethanol precipitation of the DNA. The samples were then run on the ABI 373A DNA sequencer according to the manufacturers protocol. Sequencing reactions were generally run overnight and the output data was converted to a text file and an analysis file on an Apple Macintosh computer which was linked to the ABI sequencer.

2.13 Denaturing polyacrylamide gel electrophoresis

Sequencing reaction samples were run on a 6%, 0.4 mm thick denaturing polyacrylamide gel, prepared on a 50 cm length Bio-Rad Sequencing apparatus. The sequencing gel was prepared as follows. The back sequencing plate of the apparatus was silanised before assembly (Sigmacote, Sigma). The 6% gel consisted of 42 g of urea (ICN), 15 ml of 40% (w/v) bis-acrylamide solution (BDH or NBL) and 10 ml of 10 x TBE which were mixed together and made up to 100 ml in water. For the polymerisation of a plug, 6 µl of 25% (w/v) ammonium persulphate (Bio-Rad)/ml of gel and 6 µl TEMED (BDH)/ml of gel was added, mixed and poured. For the polymerisation of the running gel, 2.5 µl of 25% (w/v) ammonium persulphate/ml of gel and 1 µl TEMED/ml of gel was added, mixed and quickly poured.

The ^{35}S-dATP sequencing reaction samples were denatured at 95°C for 3 min before loading on to a pre-warmed (50-55°C) 6% gel and electrophoresed at a constant power (50 W for 25 cm, and 90 W for 40 cm, wide gels) in 1 x TBE. After electrophoresis the gels were fixed in a 1:1 mixture of 10% acetic acid:10% methanol for 10 min. The gel was transferred to a sheet of 3MM Whatmann by placing the paper on the gel to which it preferentially stuck. The exposed surface of the gel was covered in cling film. The gel was dried on 3MM Whatmann paper for 1 to 2

hours under vacuum at 70°C. The cling film was removed from the dried sequencing gel and then autoradiographed (Section 2.11)

2.14 Subcloning of cosmid DNA into plasmids

Cosmids were subcloned into the plasmid vector Bluescript® (Stratagene) for ease of manipulation and sequencing of inserts. Both mini-preped cosmid (Section 2.1.3) and Bluescript were digested with the same restriction endonuclease in 10 µl reaction mixes (8 µl DNA, 1 µl One Phor All buffer (Pharmacia) and 1 µl enzyme). 3 µl of each digest was run on a 1% agarose gel to confirm that the digestion was complete.

2.14.1 Dephosphorylation of digested vector arms

The arms of the digested Bluescript vector were dephosphorylated, to prevent them from religating together, as follows. To the digested vector was added 1.5 µl of alkaline phosphotase buffer and 0.5 µl of calf intestinal alkaline phosphotase (1 unit/µl) made up to a total volume of 25 µl. This reaction incubated at 37°C for 15 min then at 56°C for 15 min. Another 0.5 units of the enzyme (alkaline phosphatase) was added to the reaction before incubating again at 37°C and 56°C. The reaction was stopped by heating at 80°C for 10 min. Then 3 µl of this dephosphorylated vector was used directly in a ligation reaction with 5 µl of the digested cosmid and ligated overnight at 15°C (see Section 2.5). This ligation mix was then used to transform *E. coli* (see Section 2.6 and 2.7) and screened for clones which contained the desired insert, as described below.

2.14.2 Identification of positive subclones

Transformed colonies were triplicate streaked on to gridded Hybond N^+ filters, which were on LB/antibiotic agar petri dishes. The colonies were grown on the filters at 37°C overnight. One of the three filters was stored at 4°C while the other two were processed as replica filters by lysing the colonies, binding the DNA, probing and autoradiography to find the subclone with the required insert (as described in Sections 2.9.2 to 2.9.4). Subclones containing the correct insert were picked from the master filter stored at 4°C and inoculated into LB/antibiotic media, subcultured and minipreped (see Section 2.1.3) for further analysis.

2.15 Routine cell culture

2.15.1 Growth and maintenance of the Weri-Rb1 cell line

The Weri-Rb1 cell line was obtained from the American Type Culture Collection (ATTCC). Manipulation of the retinoblastoma cells were carried out in a Microflow Laminar Flow Workstation. The cells were cultured in RPMI media containing 10% Glutamax™ (Gibco BRL), supplemented with 10% foetal calf serum (Sigma) and 1% penicillin/streptomycin (Gibco BRL). The Weri-Rb1 cells were grown in suspension in Greiner flasks in a 37°C humidified incubator with 5% CO_2. Cells were grown to a maximum density of 6×10^6 before sub-dividing these cells, into fresh media, to a density of 1×10^4. Cells were routinely sub-divided twice a week to maintain the appropriate density.

Frozen stocks were prepared by centrifuging approximately 3×10^6 cells at 125 x g for 10 min and the pellet was resuspended in 0.5 ml of normal culture medium. To this resuspension was added, a drop at a time with tapping to mix the cells/media, 0.5 ml of culture media containing 10% (v/v) dimethylsulphoxide (DMSO). The resuspended cells were first climatised at -20°C for two hours, then at -70°C for two hours before placing in liquid nitrogen for long term storage.

<u>RPMI Culture Medium</u>
89% (v/v) RPMI with Glutamax
10% (v/v) Foetal Calf Serum
1% (v/v) Streptomycin/Penicillin

<u>Freezing Medium</u>
9 ml of RPMI Culture Media
1 ml of DMSO

2.15.2 Human fibroblast cell line

A fibroblast cell line was obtained from Dr Nick Occleston (Dept of Pathology, Institute of Ophthalmology, London UK). This cell line was grown by Dr Occleston and was originally cultured from cells isolated from human Tenon's capsule. The growth and maintenance of these cells is described elsewhere (Occleston, 1996).

2.16 Isolation of poly-A + RNA

Plasticware and glassware which was utilised for the following procedures was DEPC treated, prior to use, to prevent RNAse contamination (as described in Sambrook *et al.*, 1989). Poly-A+ RNA was extracted using the equipment and solutions provided in the QuickPrep™ Micro mRNA Purification Kit (Pharmacia) as instructed by the

manufacturers protocol. This kit utilises salt buffer solutions and an Oligo(dT)-Cellulose mix to isolate the poly-A^+ RNA.

High-Salt Buffer	Low-Salt Buffer
10 mM Tris-HCL (pH 7.4)	10 mM Tris-HCL (pH 7.4)
1 mM EDTA	1 mM EDTA
0.5 M NaCl	0.1 M NaCl

1×10^7 cells were lysed in a solution containing guanidinium thiocyanate, N-lauroyl sarcosine, 10 mM Tris-HCL (pH 7.4) and 1 mM EDTA before adding this to the Oligo(dT)-Cellulose mix. The mix was pelleted by centrifugation at 16000 x g for 10 seconds. The Oligo(dT)-Cellulose pellet was washed four times with 1 ml of High-Salt Buffer before washing twice with 1 ml of Low-Salt Buffer. The Oligo(dT)-Cellulose mix with bound RNA was transferred to a MicroSpin Column (Pharmacia) in 0.3 ml of the Low-Salt Buffer. 0.5 ml of Low-Salt Buffer was passed through the MicroSpin Column, by centrifugation, a total of 3 times. Bound poly-A^+ RNA was eluted off the Column with two lots of 0.2 ml of warm Elution Buffer (10 mM Tris-HCL (pH 7.4), 1 mM EDTA at 65°C) into a sterile eppendorf. The RNA was precipitated by adding 10 µl of Glycogen solution (glycogen at 10 mg/ml in DEPC-treated water), 40 µl of 2.5 M potassium acetate (pH 5.0) and 1 ml of 95% ice cold ethanol and then incubating this mix at -80°C for 30 min. The precipitated mRNA was pelleted by centrifugation at 13000 rpm for 5 min at 4°C. The supernatant was decanted off and the tube allowed to dry by natural evaporation. The dried pellet of poly-A^+ RNA was resuspended in 26 µl of DEPC-treated water and an aliquot of this (50 ng) was used directly for first strand cDNA synthesis (as described in Section 2.17).

2.17 cDNA synthesis

First strand cDNA synthesis was carried out using the components supplied with a commercial kit (3′ RACE System for Rapid Amplification of cDNA Ends, Gibco BRL) as instructed by the manufacturers protocol. Poly-A^+ RNA is converted into cDNA using the enzyme reverse transcriptase (RT) and an oligo-dT adapter primer. Amplification of the cDNA of interest is achieved by PCR with either two gene-specific primers or one gene-specific primer in conjunction with an adapter primer (which is supplied in the kit). The gene-specific primers are designed to known exon sequences.

First strand cDNA synthesis was accomplished as follows: 1 µl of the oligo-dT adapter primer (10 µM stock) was added to 50 ng of poly-A^+ RNA

in 11 µl of DEPC treated water and mixed gently in a 0.5 ml sterile microfuge tube. The mixture was heated at 70°C for 10 min and then chilled on ice for 1 min before collecting the contents by brief centrifugation. The following components were added:

2 µl of 10 x PCR buffer
2 µl of 25 mM $MgCl_2$
1 µl of 10 mM dNTP mix
2 µl of 0.1 M DTT

and mixed gently before incubating at 42°C for 3 min. 1 µl of reverse transcriptase (SuperScript II) was added to the mixture and incubated for a further 50 min at 42°C. The reaction was stopped by incubating the mixture at 70°C for 15 min. After chilling on ice for 1 min the reaction mix was collected by centrifugation. 1 µl of RNAse H was added to the mix and incubated at 37°C for 20 min.

An aliquot (1 µl) of the newly synthesised first strand cDNA was used directly as template for amplification in a standard PCR with two gene-specific primers (using the protocol as described in Section 2.2.1). The remainder of the cDNA mixture was stored at -20°C.

2.18 Protein extractions and quantification

2.18.1 Nuclear protein extraction

The following protocol was adapted from that described by Dignam *et al.* (1983) and requires the following buffers:

Buffer A	Buffer C	Buffer D
10 mM Hepes pH 7.9	20 mM Hepes pH 7.9	20 mM Hepes pH 7.9
1.5 mM $MgCl_2$	0.42 M NaCl	0.5 mM KCl
10 mM KCl	1.5 mM $MgCl_2$	0.5 mM DTT
0.5 mM DTT	0.5 mM DTT	0.5 mM PMSF
0.5 mM PMSF	0.5 mM PMSF	20% v/v Glycerol
	25% v/v Glycerol	

Cells (1-2 x 10^7) were harvested by centrifugation at 125 x g, and the pellet was washed once in 5 ml ice cold PBS (phosphate buffered saline; Gibco BRL) and three times in 1 ml ice cold buffer A (see below). The remaining steps were carried out in the cold-room (4°C). Cells were resuspended in 2 x pellet-volume (≈50-150µl) with buffer B (buffer B = buffer A containing 0.1% Nonidet P-40 (Sigma) and incubated on ice for 5

minutes. The lysed cells were spun at 13000 rpm for 10 minutes in a microfuge and the supernatant was discarded. The nuclear pellet was resuspended in 2 x pellet-volume with buffer C (see below) and incubated for 15 minutes at 4°C on a rotating tumbler, followed by a gentle vortex of the mix. The lysed nuclei were then spun at 13000 rpm for 12 minutes in a microfuge. The supernatant, containing the nuclear proteins, was transferred to a fresh 0.5 ml eppendorf and diluted with buffer D (see below) keeping the ratio of buffer D to buffer C at 6:1. The extracts were then frozen quickly on dry ice, in aliquots (20 µl) to avoid repeated freeze-thawing of working stocks, and stored at -70°C.

2.18.2 Bradford assay for protein concentration

1 ml of Coomassie dye reagent was added to 10 µl of isolated protein extract, mixed well and left to stand for 15 min. The absorbance of the protein sample at 595 nm was measured using a spectrophotometer (Ultraspec® 2000; Pharmacia Biotech). An estimate of the protein concentration was determined from a standard curve produced using a stock BSA (bovine serum albumin; A_{595} v [BSA] mg/ml) sample as a standard.

<u>Coomassie Dye Reagent</u>
100 mg Coomassie Brilliant Blue G
30 mg SDS
50 mg 95% (v/v) Ethanol
100 ml 85% (v/v) Phosphoric Acid
Dilute this mixture to final volume of 1 litre, using distilled water.

2.18.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Isolated proteins were visualised by running denatured samples on a 12% SDS-PAGE and staining in Coomassie brilliant blue, prepared as described below (adapted from Sambrook *et al.*, 1989).

<u>12% running gel</u>
2.5 ml, 1.5 M Tris-HCL (pH 8.8)
4 ml, 30% Acrylamide-bis
100 µl, 10% SDS
3.35 ml, H_2O
43 µl, 10% APS
10 µl, TEMED

<u>5% stacking gel</u>
625 µl, 0.5 M Tris-HCL (pH 6.8)
385 µl, 30% Acrylamide-bis
25 µl, 10% SDS
1.45 ml, H_2O
12.5 µl, 10% APS
2.5 µl TEMED

SDS-polyacrylamide gels were prepared in an ATTO gel casting apparatus. The running gel was poured first and allowed to polymerise before pouring the stacking gel on top of the running gel. Protein samples were denatured by adding an equal volume of disruption buffer (20% SDS, 1 ml of 1M TrisHCl pH 8, 500 µl of glycerol, 600 µl of β-mercaptoethanol, 500 µl of 0.1% bromophenol blue and 7.4 ml of water) and then boiled for 5 min in a water bath and loaded on to the gel. Each gel was run at 35 mA current until the blue dye had reached the bottom of the gel. The proteins were stained, for visualisation, in Coomassie brilliant blue (0.2% in 10% acetic acid/40% methanol) and destained in 10% acetic acid/40% methanol for 5 hours to overnight. Molecular weight protein standards used are given below:

Protein standard	Molecular weight (D)
Bovine albumin	66000
Ovalbumin	45000
G3P dehydrogenase	36000
Carbonic anhydrase	29000
Trypsinogen	24000
Soybean trypsin inhibitor	20100
α-lactalbumin	14200

2.19 DNA mobility shift assay

The DNA mobility shift assay is based on the principal that the migration of a fragment of DNA on a non-denaturing polyacrylamide gel can be retarded by the action of proteins binding to the DNA in a sequence-specific manner. By utilising a radioactive labelled DNA the extent of retardation can be visualised by autoradiography. The following protocols were adapted from those described in Latchman (1993).

2.19.1 5′ end-labelling of oligonucleotides and generation of probe DNA fragments

Each PCR primer (10 pmol) was labelled using [γ-^{32}P]dATP in a final volume of 10µl by adding the following: 1µl of T4 polynucleotide kinase (5 units; Pharmacia), 3µl of [γ-^{32}P]dATP (3000Ci/mM) and 1µl of One-Phor-All buffer (Pharmacia). The reaction mix was incubated at 37°C for 1 hour. Unincorporated [γ-^{32}P]dATP was removed by running the mix through a NAP-5 (Pharmacia) column using the manufacturers standard protocol. The labelled primer (1 pmol) was utilised in a standard 50 µl PCR reaction

(as described in Section 2.2.1), in conjunction with 1 pmol of the appropriate unlabelled primer to generate the DNA fragment for use in the band shift assay.

2.19.2 Non-denaturing polyacrylamide gel electrophoresis

1 µl of radioactive labelled PCR products were loaded onto a 50 ml 6% non-denaturing polyacrylamide gel made up as follows:

5 ml of 1 x TBE,
10.1 ml of 30% acrylamide:bisacrylamide stock,
34.9 ml of H_2O,
350 µl of 10% Ammonium persulphate
and 17.5 µl of TEMED.

The gel was prepared in an ATTO apparatus (see next Section) and run in 1 x TBE for approximately 1-2 hours at 10 V/cm. The gels were dried on Whatmann filter paper (45 min at 80°C, with vacuum) and the radioactive signals were visualised by autoradiography (Section 2. 11).

2.19.3 Electrophoretic band shift assay (EBSA)

The components of the EBSA binding reaction mix are given below:

<u>Binding Reaction Mix</u>
20 mM Hepes (pH 7.9)
1 mM $MgCl_2$
10% glycerol
0.5 mM DTT
50 mM KCl
1 µg poly dI-dC (Pharmacia)
100 ng aprotinin
20 fmol of ^{32}P end-labelled PCR fragment/annealed oligonucleotides
4 µl nuclear protein extract (1-2 µg).

The binding reactions, containing 20 fmols of ^{32}P end-labelled probe, were carried out in a 20µl reaction volume at room temperature for 20 minutes, and then incubated on ice for 20 minutes. This mix was then loaded onto a 50 ml 4% non-denaturing polyacrylamide gel (12.5 ml of 1 x TBE, 6.8 ml of 30% acrylamide:bisacrylamide stock, 30.7 ml of H_2O polymerised with 350 µl of 10% Ammonium persulphate and 17.5 µl of TEMED). Electrophoresis was carried out in a cold room (at 4°C). Gels

were pre-run for 45 minutes prior to loading samples and were then run at 10 V/cm for 3 hours at 4°C in 0.25 x TBE (23 mM Tris-Borate, 0.5 mM EDTA pH 8.0). The gels were run on an ATTO apparatus (AE-6220 160x160 mm slab chamber from the ATTO corporation). The gels were dried on Whatmann filter paper (45 min at 80°C, with vacuum) and the radioactive signals were visualised by autoradiography (Section 2. 11).

2.19.4 Competitive EBSA

Competitive EBSA was carried out as described above for normal EBSA but with the addition of various concentrations of unlabelled probe, consisting of a forward oligonucleotide annealed to its antisense complementary sequence, which was included in the binding reaction. Probes of complementary unlabelled PCR primers were annealed together by heating equal concentrations of the forward primer and the reverse primer at 80°C for 2 min and then allowing this mix to cool to room temperature. Either 0.2, 2, 200 or 2000 fmols of unlabelled annealed complementary oligonucleotides was added to the binding reaction to compete for protein binding.

2.20 The *Pichia pastoris* yeast expression system

Expression studies using the yeast *Pichia pastoris* system was performed according to the manufacturers instruction manual (*Pichia* Expression Kit, Invitrogen®). The yeast pPIC9 vector was utilised in this study (Figure 2.2). A schematic diagram showing the steps involved in from cloning the gene of interest to the inducement of heterologous protein expression is given in Figure 2.3.

2.20.1 Growth media

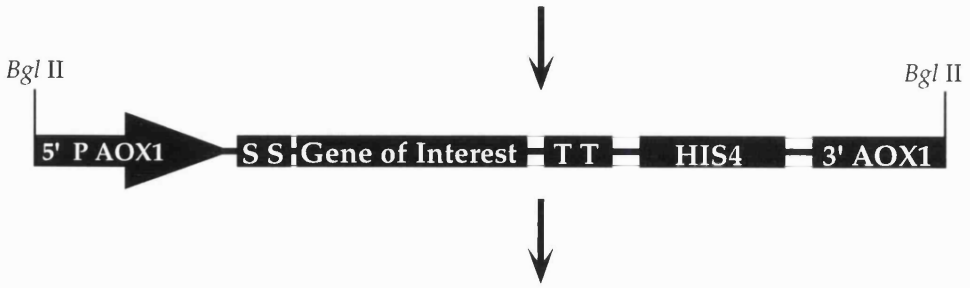
The following media and solutions were utilised for the growth, transformation and induction of protein expression of the *P pastoris* host strain GS115.

<u>SCED buffer</u>
1 M sorbitol
10 mM sodium citrate
10 mM EDTA
10 mM DTT

<u>CaS solution</u>
1M sorbitol
10 mM TrisCl (pH 7.5)
10 mM $CaCl_2$

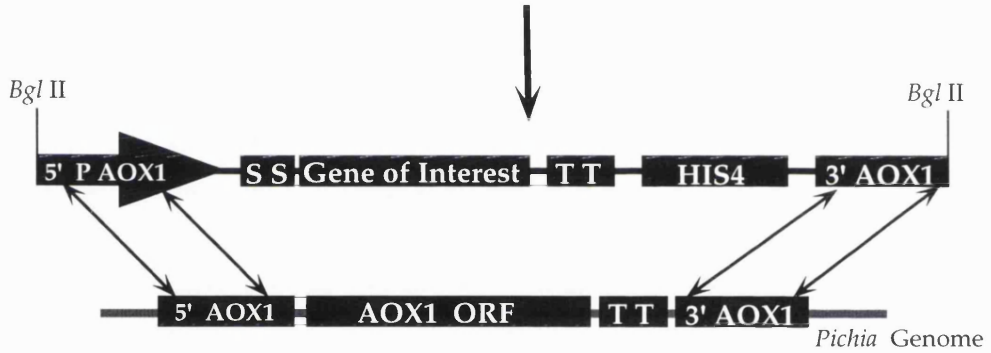
1. Clone the gene of interest into the yeast expression vector pPIC9, transform and mini-prep the construct.

2. Linearise the construct by digestion with *Bgl* II.

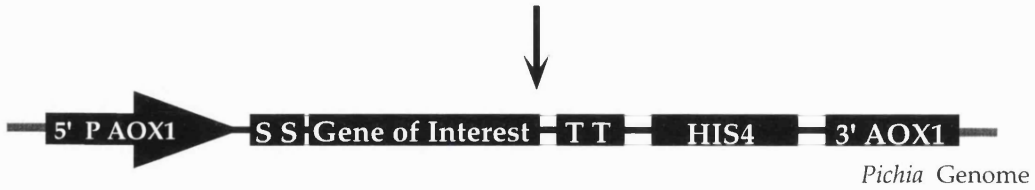


3. Prepare spheroplasts of the HIS4 deficient *Pichia pastoris* strain GS115. Transform spheroplasts with the linearised construct.

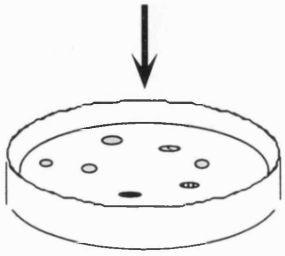
4. A recombination event occurs *in vivo* between the 5' and 3' AOX1 sequences in the pPIC9 vector and those in the *Pichia pastoris* genome. This results in the replacement of the AOX1 gene with the gene of interest.



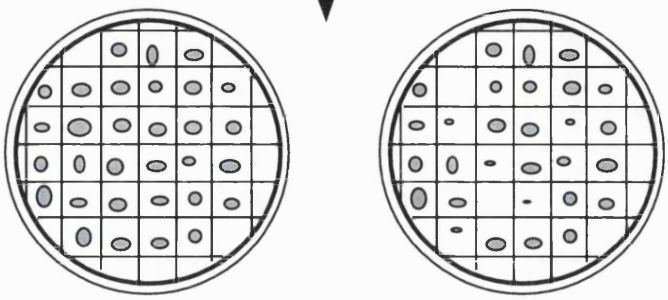
5. The *Pichia pastoris* genome now contains the gene of interest and the HIS4 gene.



6. Plate transformants on histidine-deficient media (-His plate). Cells in which recombination has occurred will grow (H^+ transformants), all others will not produce histidine and will die.



7. Screen H^+ transformants for integration at the correct loci. Select colonies from the -His plate and replica patch onto a -His/+Glucose plate (MD plate) and a -His/+Methanol plate (MM plate). Colonies which grow slowly on the MM plate no longer contain the AOX1 gene and have a H^+/M^- (methanol utilisation deficient) phenotype.



-His/+Glucose media (MD plate)

-His/+Methanol media (MM plate)

8. Select 10 H^+/M^- colonies and grow for 2 days in media containing glycerol as the carbon source (BMGY media). Pellet cells and remove the media.

9. To induce heterologous expression, resuspend pellet in media containing methanol as the carbon source (BMMY media). Grow cells for 2-6 days.

10. Analyse the expression level, which should have been secreted out into the media due to the attached secretion signal (SS) to the protein of interest, by running an aliquot of the media on a SDS-PAGE.

Figure 2.2 Outline of the procedures involved in obtaining a yeast clone which should be expressing the gene of interest.
Key: **pPIC9** is the yeast expression vector utilised in study (Figure 4.3).
5' P AOX1 is the 5' flanking promoter sequence for the *P pastoris* alcohol gene.
S S is the coding sequence for the α-factor secretion signal of *Saccharomyces cerevisiae* which is linked to the gene of interest.
T T is a transcription termination signal.
HIS4 is the gene which encodes for the enzyme histidinol dehydrogenase. Any clone containing the HIS4 gene can grow on histidine deficient media, whereas yeast lacking the gene cannot grow.
3' AOX1 is the 3' flanking region of the *P pastoris* alcohol oxidase gene.
AOX1 ORF indicates the coding region of the alcohol oxidase gene. Yeast which have the AOX1 gene are able to utilise methanol as a carbon source for growth.

PEG/CaT solution
5 mM Tris (pH 7.5)
5 mM $CaCl_2$
20% PEG (w/v)

SOS media
1 M sorbitol
0.3 x YPD
10 mM $CaCl_2$

YPD media and plates
10 g bacto yeast extract
20 g peptone
900 ml H_2O,
autoclave before adding 100 ml of 10 x Dextrose
(for plates add 20 g of agar before autoclaving)

10 x YNB (Yeast nitrogen base)
13.4 g YNB dissolved in 100 ml H_2O,
filter sterilised, by passing through a millipore filter

500 x Biotin
20 mg biotin dissolved in 100 ml H_2O and filter sterilised

10 x Dextrose
20 g D-glucose dissolved in 100 ml of H_2O and filter sterilised

10 x Methanol
Filter sterilised 5% methanol solution

10 x Glycerol
Filter sterilised 10% glycerol solution

100 x AA (Amino acids)
500 mg each of L-glutamic acid, L-methionine, L-lysine, L-leucine and L-isoleucine dissolved in 100 ml of H_2O and filter sterilised

RD (Regeneration dextrose-top agar)
Prepare the same as RDB, but with 10 g of agar
Maintain at 45°C prior to use
Should be prepared on day of transformation of yeast

RDB (Regeneration dextrose base)
186 g sorbitol
20 g agar
700 ml H_2O,
autoclave and cool to and maintain mix at 45°C,
then add prewarmed (45°C) stock solutions:
100 ml 10 x Dextrose
100 ml 10 x YNB
2 ml 500 x Biotin
10 ml 100 x AA
88 ml H_2O, then pour plates

2.20.2 Yeast cell growth

GS115 yeast were streaked on to a YPD plate and grown for 2 days at 30°C. A single colony from the plate was picked and inoculated into 100 ml of YPD media and grown overnight at 30°C with vigorous shaking (250 rpm). 5, 10 and 20 ml of the overnight culture was inoculated into three separate flasks each containing 200 ml of YPD and incubated overnight at 30°C with vigorous shaking. The cultures were harvested by centrifugation at 1500 x g for 10 min when the OD_{600} of the yeast cells were between 0.2 and 0.3. These cells were then utilised for spheroplast formation (see below).

2.20.3 Preparing yeast spheroplasts

A pellet of yeast cells, from an overnight culture with an OD_{600} between 0.2 and 0.3, were washed with 20 ml of sterile water by centrifugation at 1500 x g for 10 min. The cells were washed again in 20 ml of 1 M sorbitol. The cells were resuspended in 20 ml of SCED buffer, pH 7.5, and divided into two equal samples. Zymolase was added (0.3 mg), to one of the two samples, and this mixture was incubated at 30°C for 1 hour, for the formation of spheroplasts. The time of incubation to achieve approximately 70% spheroplasting was determined as follows:
200 µl of the incubating cells was added to 800 µl of 5% SDS at t = 0, 2, 4, 5, 6, 7, 8, 9, 10, 15, 20 and 30 min after adding Zymolase and the OD_{800} was measured. % spheroplasting = 100 - (OD_{800} at time t/ OD_{800} at t = 0) x 100.

Zymolase was added to the second sample of yeast cells and incubated at 30°C for the time required to establish 70% spheroplasting (about 20-25 min). Spheroplasts were harvested by gentle centrifugation at 750 x g for 10 min and the supernatant was discarded. The spheroplasts

were gently resuspended (by tapping side of tube softly) in 10 ml of 1M sorbitol. Spheroplasts were pelleted by gentle centrifugation again and the supernatant was discarded. Spheroplasts were washed in 10 ml of cold CaS solution and centrifuged gently as before. Spheroplasts were resuspended gently in 0.6 ml of cold CaS solution and placed on ice. The spheroplasts were used immediately for transformation.

2.20.4 Transformation of yeast

1-10 μg of linearised DNA (pPIC9 vector containing the gene of interest and the histidinol dehydrogenase gene) was added to 100 μl of spheroplasts and incubated at room temperature (RT°C) for 10 min. 0.5 ml of PEG/CaT solution was added to the spheroplasts/DNA and mixed gently and incubated for another 10 min at RT°C. This mix was collected by centrifugation at 750 x g for 10 min and the supernatant was discarded. The pellet of transformed cells were resuspended in 150 μl of SOS and incubated for 20 min at RT°C. 850 μl of 1 M sorbitol was then added before mixing 300 μl of this suspension with 10 ml of molten RD held at 45°C. This was quickly poured onto RDB plates and allowed to set (10-15 min). The plates were inverted and incubated at 30°C for 4-6 days, to allow histidine positive (H^+) transformant cells to form colonies.

2.20.5 Screening for histidine positive/alcohol oxidase deficient (H^+/M^-) transformants

A small percentage (5-35%) of the H^+ transformant colonies will have resulted from complete displacement of the *P pastoris* alcohol oxidase (AOX1) gene by the *Bgl*II linearised pPIC9 construct and are termed H^+/M^- transformants (see Figure 2.3). H^+/M^- transformants which have displaced the AOX1 gene cannot produce the enzyme alcohol oxidase. As a result of this displacement these transformants cannot efficiently metabolise methanol as a carbon source and therefore grow poorly on MM (minimal methanol) media (Invitrogen manual). However, these H^+/M^- transformants will grow normally on MD (minimal dextrose) plates, since they can still utilise glucose for growth. This slow growing attribute was used to distinguish the H^+/M^- colonies from the H^+ transformants. H^+ transformants with an intact AOX1 gene will grow at the same rate on both MD and MM plates, since they can still utilise methanol as a carbon source. By comparing the growth rates of duplicate H^+ colonies on both MD and MM media it was possible to select for H^+/M^- clones, which grew slower on MM plates

Using a sterile toothpick individual H^+ transformants were picked and duplicate streaked on to gridded MM and MD plates (making sure to streak the MM plate first). A new toothpick was used for each transformant which was picked. The plates were then incubated at 30°C for 2-4 days. On the 2nd, 3rd and 4th day of incubation the rate of growth of each transformant on each plate was compared and noted. Colonies which grew normally on the MD plate but showed little or no growth on MM plates (H^+/M^- clones) were analysed further for heterologous protein expression (described below).

2.20.6 Heterologous protein expression studies on H^+/M^- transformants

The heterologous gene of interest which was used to transform the H^+/M^- yeast, is linked to the secretion signal of the *Saccharomyces cerevisiae* α-mating factor gene (see Figure 4.3). Alcohol induced expression from the AOX1 gene promoter results in high levels of a fusion protein product which consists of the protein of interest linked to the α-mating factor signal sequence. This signal sequence targets the fusion protein to be secreted out of the yeast cell into the surrounding media. Thus the protein of interest can be isolated directly from the methanol containing media (MM) in which the H^+/M^- clones were grown in.

To induce the heterologous protein expression, each H^+/M^- transformant was first inoculated into 10 ml of BMGY media in 50 ml Falcon tubes (Falcon) and incubated at 30°C with vigorous shaking (>200 rpm). The cells were grown to saturation (2 days) such that the OD_{600} of the cells was between 10-20. The caps of the tubes were left completely loose during growth. The cells were harvested by centrifugation at 4000 x g for 10 min. The supernatant was discarded and the pellet resuspended in 2 ml of BMMY media. The top of the tube was covered with a piece of sterile gauze (cheese cloth) instead of the cap. The tubes were then incubated again at 30°C with vigorous shaking for 4 days. The cells were pelleted as before and aliquot of the supernatant was run on an SDS-PAGE for visualisation of any secreted heterologous protein products (as described in Section 2.18.3).

2.21 Computing and software

This thesis was written using Microsoft Word version 5.1a ran on a Apple Macintosh Performa 6200. General analysis of DNA sequences was

performed using GeneWorks version 2.1 (Intelligenetics, Inc., CA 94040) or Clustal V (Higgins and Sharp, 1988). 5′ flanking sequences were analysed for protein binding sites using the program TESS (Transcription Element Sequence Software) which was accessed from the World Wide Web via Netscape version 1.1. Other analysis was carried out using the computing programs at the UK Human Genome Mapping (HGMP) resource centre (Clinical Research Centre, Watford Road, Harrow, Middlesex, HA1 3UJ) via networked facilities.

2.22 Reagents and suppliers

2.22.1 General reagents

- TE buffer: 10 mM Tris-HCL pH 7.5, 0.1 mM sodium EDTA pH 8, autoclaved
- Ficoll-orange G loading buffer (10 X stock): 25 g Ficoll, 250 mg Orange G and 9.3 g EDTA in 1 X TAE made up to a volume of 100 ml
- TAE buffer (10 X stock): 0.4 M Tris-acetate, 10 mM sodium EDTA, pH 8
- 20 X SSC: 175.3 g NaCl and 88.2 g sodium citrate, pH 7.0 in one litre of distilled water (3 M NaCl, 0.3 M sodium citrate)

2.22.2 Solutions used for DNA preparation and Southern blotting

- 10 X lysis buffer: 50 mM Tris-HCl pH 7.5, 25 mM $MgCl_2.6H_2O$, 0.6 M sucrose, 5%(v/v) Triton X 100
- 10 X suspension buffer: 0.1 M Tris-HCl pH 7.5, 0.1 M NaCl, 0.1 M sodium EDTA pH 8
- Denaturer: 1.5 M NaCl, 0.5 M sodium hydroxide
- Neutraliser: 3 M NaCl, 0.3 M sodium citrate, pH 5.5
- Colony blot neutraliser: 1.5 M NaCl, 0.5 M Tris-HCl pH 7.2, 1 mM EDTA

2.22.3 Media for microbiological manipulations

- LB broth: 1%(w/v) bactotryptone, 0.5%(w/v) yeast extract, 1%(w/v) NaCl, autoclaved
- LB agar: LB broth containing 1.5%(w/v) agar, autoclaved
- Cell resuspension solution: 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 μg/ml RNAse A
- Cell lysis solution: 0.2 M NaOH, 1%(w/v) SDS
- Neutralisation solution: 1.32 M potassium acetate pH 4.8

2.22.4 Manual sequencing solutions

- Stop solution: 0.3% Bromophenol blue, 0.3% Xylene cyanol, 10 mM EDTA pH 7.5, 97.5% deionised formamide
- 10 X TBE: 90 mM Tris borate, 10 mM sodium EDTA pH 8.3
- Sequagel: Sequencing system concentrate (#EC-830), National Diagnostics

2.22.5 Automated sequencing reagents

- 10 X TBE: 90 mM Tris borate, 10 mM sodium EDTA pH 8.3
- Loading solution: Deionised formamide: 0.05 M EDTA (5:1 v/v)
- SequaGel™ ready-to-use 6% sequencing gel solution (#EC-836), National Diagnostics
- 10% Ammonium persulphate (IBI), freshly prepared
- 2 M Sodium acetate
- Phenol/water/chloroform (Applied Biosystems: #400765)

2.22.6 Suppliers

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

Chemicals not specified were purchased from:

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

- Radiochemicals were purchased from Amersham.

2.22.7 Oligonucleotides

Oligonucleotides were either synthesised by Neil Ebenezer on a MilliGen DNA synthesiser at the Institute of Ophthalmology or purchased from British Biotechnology, Cruachem and Pharmacia.

2.22.8 Glassware and plasticware

All glassware was washed with Teepol detergent and then rinsed thoroughly with double distilled water before use. Microcentrifuge tubes and tips were used direct from the supply bags and were only sterilised, by autoclaving, for use with bacterial and yeast culture and for the setting up of PCR reactions. 1.5 ml and 0.5 ml microcentrifuge tubes were purchased from Elkay. 1.0 ml (blue) and 200 µl (yellow) were purchased from Algen. Tubes and tips were sterilised by autoclaving at 15 psi for at least 20 min. Sterile 10 ml and 50 ml disposable, screw cap tubes were obtained from Falcon.

Chapter 3

Characterisation of the Porcine Blue Cone Pigment Gene and Identification of Spectral Tuning Sites

3.1 Introduction

How each visual pigment attains its particular λ_{max} has yet to be completely determined. The following study attempts to increase our knowledge of the S group of visual pigments. The S group are vertebrate visual pigments which possess a λ_{max} of less than 440nm (see Section 1.13 and Table 1.1).

However, of the S group members, only two non-primate mammalian blue cone pigment (BCP) genes have been characterised to date, namely the murine and bovine (Chiu *et al.*, 1994). In this study, the genomic structure and sequence of the porcine BCP gene was determined. The porcine BCP amino acid sequence was compared with that of other vertebrate visual pigments and analysed for any residues which could affect spectral tuning. Thus, as well as increasing our knowledge of BCP genes in mammals, it was possible to investigate which amino acids may be responsible for determining the wavelength of maximum absorption (λ_{max}) within the S group of pigments.

3.1.1 Porcine visual pigments

The domestic pig (*Sus scrofa* , Order: Artiodactyla; Family: Suidae) is an ungulate. Ungulate refers to the group of mammals with hooves, which includes horses, rhinoceroses, cattle, sheep and goats (Clutton-Brock, 1987). Ungulates possess arrhythmic eyes, that is their eyes are large and show other characteristics suitable for both diurnal and nocturnal life (Neitz and Jacobs, 1989). Approximately 7% of all the photoreceptors in the porcine retina are cone cells (Miller and Snyder, 1977). Results from electroretinogram (ERG) flicker photometry indicate that the porcine retina contains two classes of cones with λ_{max} of approximately 439nm and 556nm (Neitz and Jacobs, 1989). Thus, the pig like most other mammals, possesses dichromatic colour vision.

The cone opsins from porcine have yet to be sequenced and categorised into specific visual pigment groups. The short-wavelength opsins of two other non-primate mammals, bovine and murine, are members of the S group of visual pigments (Chiu *et al.*, 1994). Since

bovine species are also of the Order Artiodactyla it is probable that the 439nm cone class detected in porcine contains an S group opsin. Characterisation of the gene coding for the porcine BCP will clarify which of the visual pigment groups this opsin belongs to and also contribute to our knowledge of mammalian opsins.

3.1.2 Terminology

The three letter and one letter codes given for specific amino acid residues are used throughout this chapter. For reference, a list of these codes are given in the appendix. The biochemical properties of the side chains of each amino acid and residues which are considered to possess conserved properties and which are non-conserved, in spectral tuning terms, are grouped accordingly in the appendix. The opsin numbering system referred to throughout this chapter is in accordance with the bovine rod opsin sequence (Wang *et al.*, 1980) unless stated otherwise.

3.1.3 Spectral tuning

If sequence differences between opsins mediate λ_{max}, the next questions to arise are what are these differences and where do they occur within the opsin? Comparison of the amino acid sequences of different classes of opsins may highlight the residues responsible for these variations in λ_{max}.

As mentioned previously (Section 1.14), all the members of the L group of pigments contain two conserved positively charged residues, His-181 and Lys-184. These two residues bind chloride ions (Cl^-) resulting in a shift of λ_{max} to longer wavelengths (Wang *et al.*, 1993). The λ_{max} of human red and green cone pigments are shifted from 535nm and 515nm in a Cl^- depleted environment to 560nm and 530nm upon addition of Cl^-, respectively. 560nm and 530nm are the approximate values of λ_{max} obtained by microspectrophotometry of single human red and green cone cells, respectively (Dartnall *et al.*, 1983). Thus, chloride ions within the L group of cone cells play a role in spectral tuning and this in turn is dependent on the presence of the residues His-181 and Lys-184 within the opsin. All other vertebrate opsins outside of the L group (S/M1/M2/Rh) contain Glu-181 and Gln-184 (except the goldfish ultraviolet pigment which contains Gly-184). Although this explains how the L group of opsins attain red shifts in their λ_{max} in relation to the other visual groups, this does not explain the differences in λ_{max} between individual members

of the L group. For example, why is there a 30nm difference in λ_{max} between the human red and green cone pigments?

The nucleotide sequences of the human red cone pigment (RCP) and green cone pigment (GCP) genes are 98% identical, and their respective proteins only differ at 15 out of 364 amino acid residues (Nathans *et al.*, 1986a). The 30nm red shift from the GCP to the RCP was initially thought to be attributed to the substitution of non-polar residues with hydroxyl-bearing amino acids at only three of these 15 sites, site 164, 261 and 269, (site numbering using the bovine rod opsin system) (Neitz *et al.*, 1991). An additional position, site 217, which also involves substitution of a hydroxyl-bearing amino acid for a nonpolar residue between the human RCP and GCP, was implicated from studies on New and Old World monkey visual pigments belonging to the L group (Ibbotson *et al.*, 1992; Williams *et al.*, 1992). Site-directed mutagenesis studies finally established that substitution at 7 of these 15 residues resulted in complete conversion of the λ_{max} observed in one pigment to the other (Asenjo *et al.*, 1994). These seven spectral tuning sites and the residues within the human RCP and GCP at these positions are given in Table 3.1. Substitution at any of one these seven sites results in a shift of λ_{max}. Every extra replacement on top of the initial substitution produces an additive effect until the 30nm difference between the two pigments is compensated for when all seven residues are substituted. Substitutions at sites 261 and 269 are attributed to causing two thirds of the 30nm shift (Asenjo *et al.*, 1994).

The hydroxyl groups of an amino acid are thought to interact with the chromophore either in a direct or indirect manner to produce shifts in λ_{max} (Chan *et al.*, 1992). In a similar fashion, it was hypothesised that an electrostatic interaction between the Schiff base and charged amino acid residues within the opsin could cause shifts in λ_{max} (Kropf and Hubbard, 1958). This is true for the charged residue Glu-113, the counter-ion to the protonated Schiff base, which influences spectral tuning by electrostatic interaction with the chromophore (Sakmar *et al.*, 1989; Zhukovsky and Oprian, 1989). In the search for the counter-ion another spectral tuning site was discovered, namely site 122. Substitution of Glu-122 to Gln-122 or Ala-122, in bovine rod opsin, caused a blue shift in λ_{max} down from 498nm to 482nm and 476nm , respectively (Nakayama and Khorana, 1991). Both Ala and Gln are uncharged residues. Thus, these non-conserved substitutions for the charged Glu residue at site 122 result in shifts of λ_{max}.

Amino acid position	RCP	GCP
100	Ser	Tyr
164	Ser	Ala
214	Ile	Thr
217	Ala	Ser
261	Tyr	Phe
269	Thr	Ala
293	Tyr	Phe

Table 3.1 Seven amino acid sites with their respective residues within the human red cone pigment (RCP) and green cone pigment (GCP). Exchange of residues at these sites between the RCP and GCP result in shifts of λ_{max} (Asenjo *et al.*, 1994). Replacement of all these 7 residues found at the same site in the GCP for the residues in the RCP results in a complete conversion of the GCP λ_{max} to the λ_{max} of the RCP (Asenjo *et al.*, 1994).

Both Glu and Asp residues are negatively charged amino acids. Substitution of Glu-122 for Asp-122 is considered a conservative change, in spectral tuning terms. However, substitution of Glu-122 for Asp-122, in bovine rod opsin, also gives a λ_{max} of 476nm (Nakayama and Khorana, 1991). Thus, the spectral tuning effect at site 122 is thought to be a steric one rather than an interaction of electrostatic charge.

Thus, it would seem that conservative amino acid substitutions can affect the λ_{max} of a visual pigment, as well as non-conservative replacements, depending on the position of the amino acid within the opsin.

3.1.4 Predicting spectral tuning sites

Which amino acid sites are responsible for the λ_{max} of the S group of pigments? Is there a way of predicting which amino acid positions are likely to influence λ_{max}, without substituting every residue within each and every visual pigment?

Spectral tuning sites can be predicted by comparing closely related visual pigments that differ in λ_{max} by a few nm from each other. The above method was applicable to the human RCP and GCP, where only a few amino acids differed between them. This method was ideal for identifying spectral tuning sites within the rod opsins of 11 species of teleost fish, from the same sub order and endemic to Lake Baikal (Hunt *et al.*, 1996). The rod opsins of these 11 species of fish fall into 5 spectral classes ranging in λ_{max} from 516nm down to 484nm and share over 90% amino acid identity. From a comparison of differing amino acids within the fish rod opsin sequences, 3 potential spectral tuning sites were identified. These sites are given in Table 3.2, (reference D).

To identify sites that were more likely to influence λ_{max} within these fish rod opsins, the following assumptions were made by the authors: (i) that the amino acid should sit within one of the transmembrane helices, (ii) this residue should face into the lumen of the chromophore binding-pocket, and (iii) the substitution at this site relative to the equivalent site in the other rod opsins should be of a non-conservative nature. How does one predict which residues within the opsin are likely to face into the chromophore binding-pocket? In this case the authors used the aid of a model predicted for the three-dimensional structure of the transmembrane helices in G protein-coupled receptors (Baldwin, 1993).

Helix	Amino acid position	Reference
II	83	A and D
	86	A
III	113	B
	114	B
	115	B
	117	B
	118	B
	121	B
	122	A and B
	124	C
	125	B
	126	B
	127	B
IV	160	B
	164	B
V	207	B
	211	A and B
	214	B
VI	261	B and D
	265	B
	268	B
	269	B
	272	B
	276	B
VII	289	C
	292	C and D
	307	C

Table 3.2 List of amino acid sites which are likely to cause spectral shifts. These sites were identified from models of opsin structure and comparisons of amino acid sequences of vertebrate opsins.
Reference A: Kosower (1988); B: Alkorta and Du (1994); C: Chang *et al.* (1995); D: Hunt *et al.* (1996). Note, site 113 is the counterion for the Schiff base of retinal, and sites 164, 214, 261 and 269 are spectral tuning sites for the human red and green cone pigments (see Section 3.1.3).

The 3D structure of the 7 α-helices was predicted from an alignment of 204 G protein-coupled receptors, which included 32 visual pigments (Baldwin, 1993 and Figure 3.1). This model proposed that each transmembrane region consists of 26 amino acids (1-26), of which only residues 5 to 22 sit within the cell membrane. It is possible to identify residues that face into the binding-pocket by aligning any opsin sequence to this model.

Other models have been generated in an attempt to unravel the 3D structure of opsins. These models also identify specific amino acid sites which will have close contact with the chromophore itself. A structural model based on charge-charge interactions was developed from an alignment of human and bovine rod opsin and human cone opsins (Kosower, 1988). This charge-charge interaction model correctly predicted that the λ_{max} difference between the human RCP and GCP was due to the net effect of seven hydroxyl groups around the chromophore (described in Section 3.1.3). This model also proposed that the side chains of residues at four particular positions are situated very close to the chromophore, thereby causing a shift in λ_{max}. These four sites are given in Table 3.2, (reference A).

A method known as sequence divergence analysis, which calculates the degree of amino acid variability between proteins, the secondary structure, and orientation of the transmembrane regions, was utilised to predict a 3D model of human rod opsin (Alkorta and Du, 1994). This 3D model proposes that residues at five sites, site 125, 126, 164, 207 and 265, within the opsin form a hydrophobic binding pocket for the β-ionone ring of retinal. These five sites plus 16 other positions are in direct contact with the chromophore, according to this 3D model. Hence, the position of these 21 residues are candidates for the regulation of spectral tuning. (These 21 sites are also listed in Table 3.2,(reference B)). A subset of the above sites have been shown to cause spectral shifts by site-directed mutagenesis analysis. For example, amino acid substitutions at sites 115, 117, 122, 126, 127 and 265, within bovine rod opsin, result in spectral shifts (Nakayama and Khorana, 1991). In addition, the amino acids at these 7 positions, 115, 117, 122, 126, 127 and 265, are responsible for cross-linking with the β-ionone ring of 11-*cis*-retinal (Nakayama and Khorana, 1990).

By utilising the data predicted by these models of opsin structure and the data from site-directed mutagenesis studies it should be possible to identify the amino acids within any opsin that are likely to affect λ_{max}.

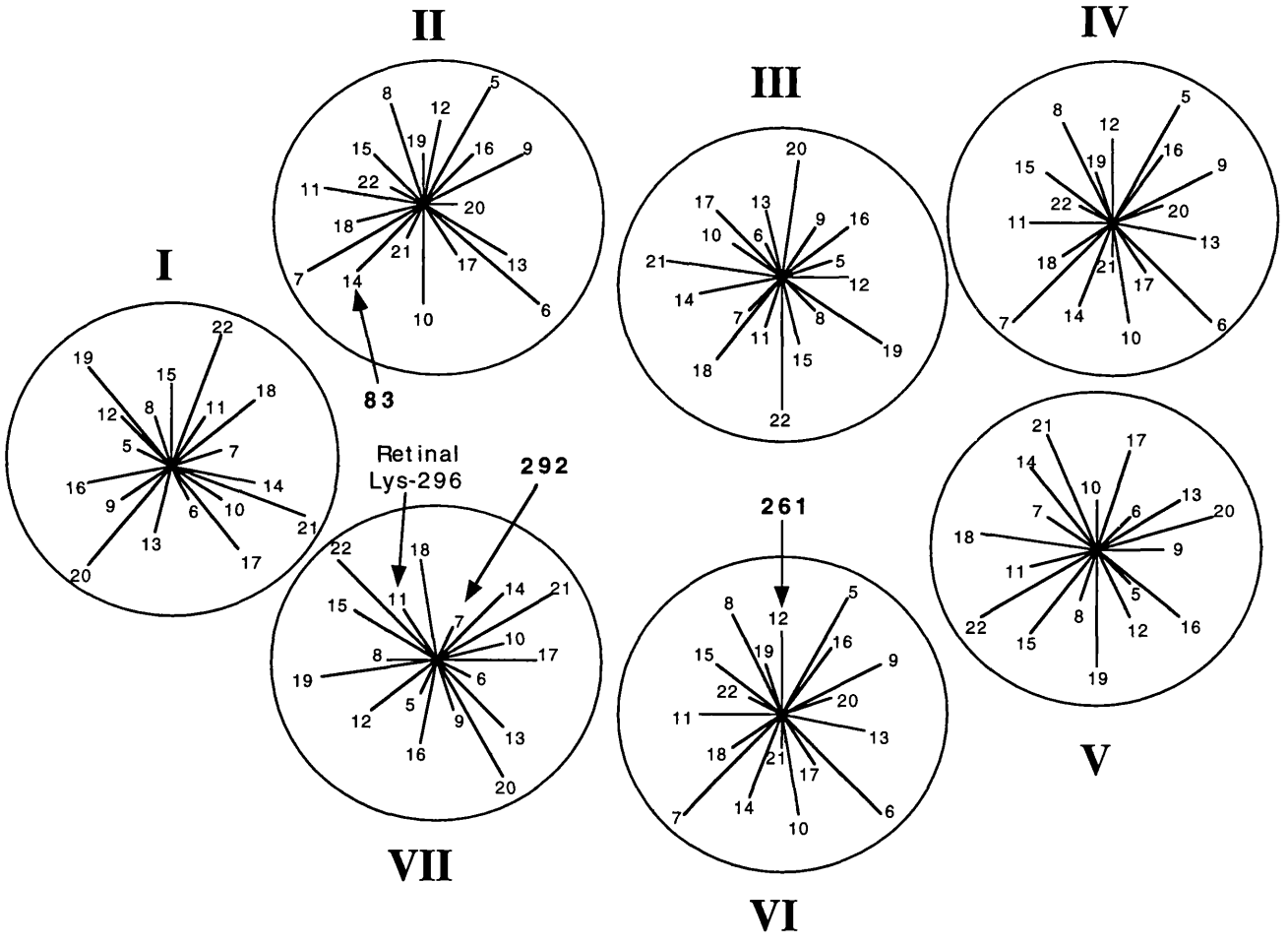


Figure 3.1 Baldwin's model of the seven helical transmembrane regions of G protein-coupled receptors (Baldwin, 1993). Each helix consists of 26 amino acids, only residues 5 to 22 are actually situated within the cell membrane. The relative position and orientation of each amino acid from 5 to 22 within each helix is indicated. The position of the chromophore binding site Lys-296 within helix VII is indicated. Also shown are sites 83, 261 and 292 identified as spectral tuning sites utilising this model (Hunt *et al.*, 1996; Section 3.1.4).

3.2 Results

3.2.1 Isolation and subcloning of a genomic cosmid clone containing the porcine BCP gene

An amplified porcine genomic cosmid library was plated out and screened with the human full length BCP cDNA of 1044 bp (as described in Section 2.9). (Isolation of the human BCP cDNA is described in Section 4.2.1.) Primary screening of the library after hybridising at 55°C, with a [α^{32}P] dCTP-radioactive labelled PCR product of the human BCP cDNA, gave no positive signals (as described in Section 2.9.4). Primary screening of the porcine library again after hybridisation at 42°C gave two positive signals. Secondary screening resulted in the identification of a single cosmid clone, which was isolated and digested with various restriction endonucleases for Southern blot analysis (Figure 3.2A; as described in Sections 2.3 and 2.10). Hybridisation of this Southern blot with the human BCP cDNA probe, revealed two bands of approximately 3 and 2 kb in size for the *Bam*HI digest (Figure 3.2B). This suggested that a single *Bam*HI site was present within the porcine BCP genomic sequence. The porcine cosmid was digested with *Bam*HI and subcloned into Bluescript plasmid vector (as described in Section 2.14). Bluescript subclones were streaked in duplicate onto gridded filters (see Section 2.14.2). Probing of these replica filters with the human BCP cDNA identified six subclones as positives for containing *Bam*HI digested porcine BCP gene (Figure 3.3). The positive subclones were sequenced, using BCP specific primers (as described in Section 2.12). PCR products generated from the porcine cosmid itself, using BCP-specific primers (Table 2.1 and Section 2.2.1) were also cloned and sequenced (Sections 2.5-2.8 and 2.12) .

3.2.2 The porcine BCP is a member of the S group of opsins

The genomic structure of the human BCP gene has been characterised previously (Nathans *et al.*, 1986a). The DNA sequencing data obtained for the porcine BCP gene was compared with the human BCP gene. The protein coding regions of the porcine and human BCP genes share a 87.7% identity, at the nucleotide level.

The porcine BCP amino acid sequence was deduced from the nucleotide sequence of the coding region of the porcine BCP gene (Figure 3.4). Figure 3.5 shows the deduced amino acid sequence for the porcine BCP aligned to the human BCP. The porcine BCP is composed of 351 amino acids, an extra 3 residues compared to the human BCP. Essential

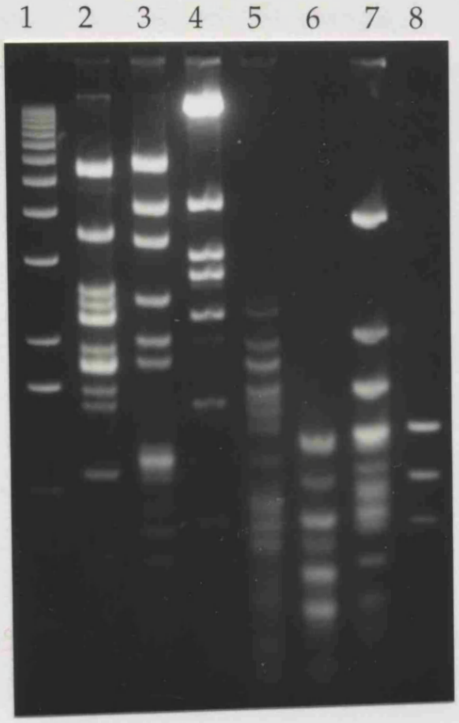


Figure 3.2 A Restriction endonuclease digestion of a genomic cosmid clone containing the porcine BCP gene.

Lane 1: 1kb DNA ladder
Lane 2: *Bam*HI digest
Lane 3: *Pst*I digest
Lane 4: *Apa*I digest
Lane 5: *Alu*I digest
Lane 6: *Hae*III digest
Lane 7: *Rsa*I digest
Lane 8: ϕX174/*Hae*III DNA ladder

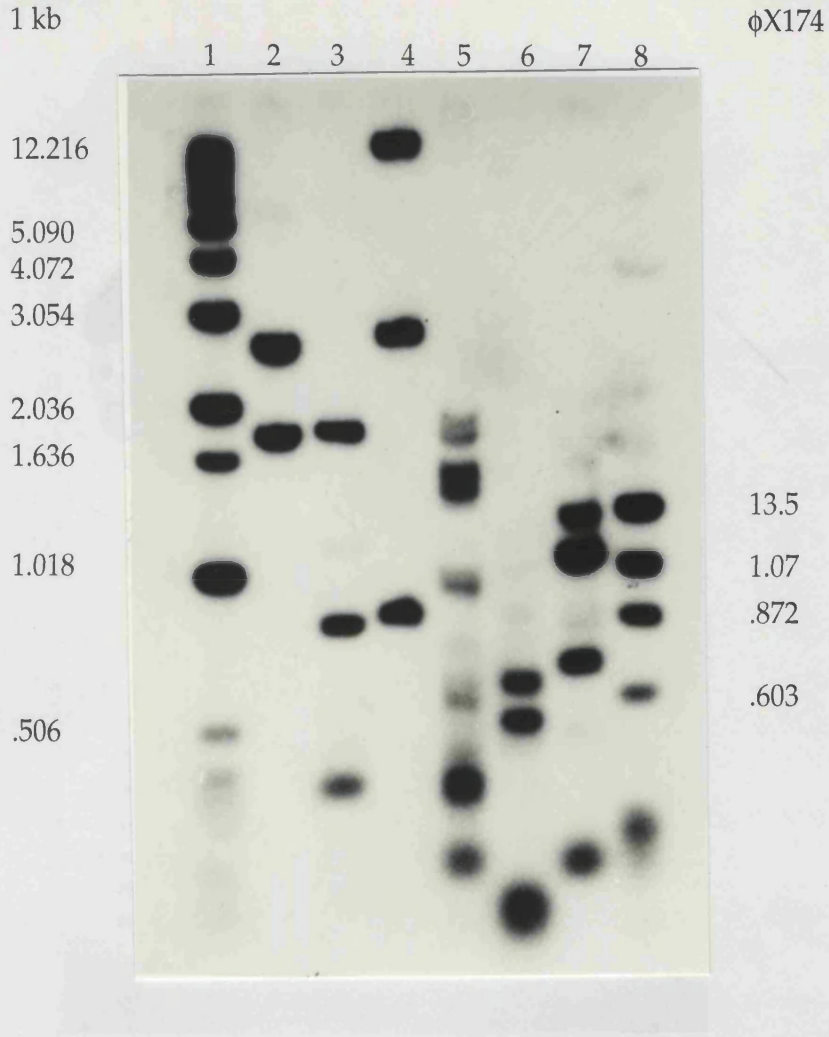


Figure 3.2 B Southern blot of the genomic porcine BCP gene cosmid clone. The blot was first probed with the full length human BCP cDNA and after autoradiography, the blot was probed again with the 1 kb and ϕX174/*Hae*III DNA ladders.

Lane 1: 1 kb DNA ladder
Lane 2: *Bam*HI digest
Lane 3: *Pst*I digest
Lane 4: *Apa*I digest
Lane 5: *Alu*I digest
Lane 6: *Hae*III digest
Lane 7: *Rsa*I digest
Lane 8: ϕX174/*Hae*III DNA ladder

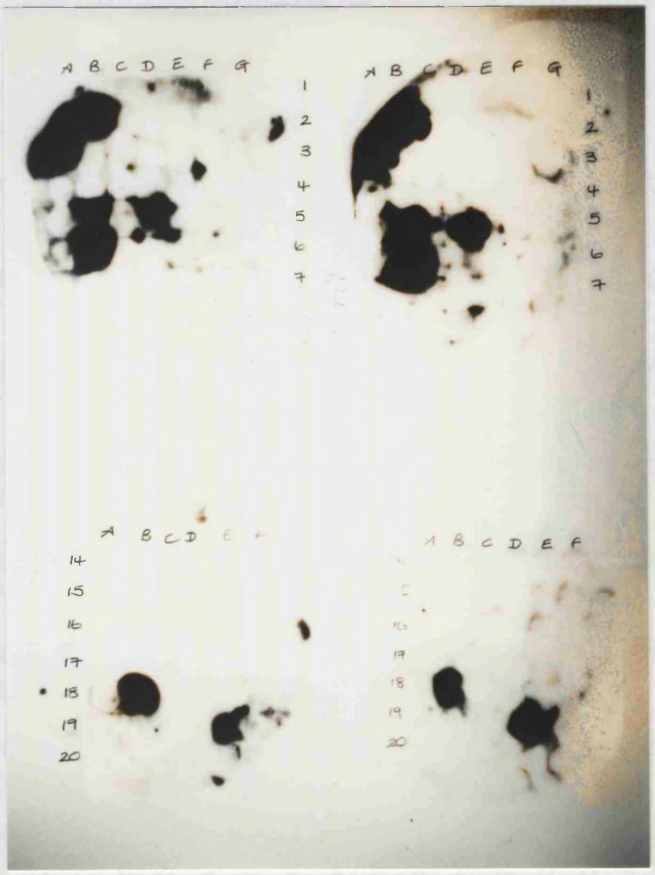


Figure 3.3 Probing of replica filters for *Bam*HI subclones of the porcine BCP gene. The filters were gridded for reference. Subclones within grids A3, B2, B5, B6, D5 (top filters), B18 and E19 (bottom filters) gave positive replicated signals when probed with the full length human BCP cDNA.

```
ATGAGCAAGATGCCAGAGGAGGAGGAGGAGGAGTTTCTTCTGTTCAAGAACATCTCCTTG   60
 M  S  K  M  P  E  E  E  E  E  E  F  L  L  F  K  N  I  S  L    20

GTGGGGCCGTGGGATGGGCCTCAGTACCACCTCGCCCCTGTCTGGGTCTTCCACCTCCAG  120
 V  G  P  W  D  G  P  Q  Y  H  L  A  P  V  W  V  F  H  L  Q    40

GCAGCCTTCATGGGCTTTGTCTTCCTTGTAGGGACACCACTCAATGCCACGGTTCTGGTG  180
 A  A  F  M  G  F  V  F  L  V  G  T  P  L  N  A  T  V  L  V    60

GCCACCCTGCGCTACAGAAAGTTGCGGCAGCCCCTCAACTATATTCTGGTCAACGTGTCC  240
 A  T  L  R  Y  R  K  L  R  Q  P  L  N  Y  I  L  V  N  V  S    80

CTGGGGGGCTTCATCTACTGCATCTTCTCTGTCTTCAGCGTCTTCATTGCCAGCTGTCAT  300
 L  G  G  F  I  Y  C  I  F  S  V  F  S  V  F  I  A  S  C  H   100

GGGTACTTCGTCTTTGGCCGCCGTGTTTGTGCTATGGAGGCCTTCCTAGGCTCTGCAGCA  360
 G  Y  F  V  F  G  R  R  V  C  A  M  E  A  F  L  G  S  A  A   120

GGTCTGGTGACAGGCTGGTCACTGGCCTTCTTGGCCTTTGAGCGCTACATCATCATCTGT  420
 G  L  V  T  G  W  S  L  A  F  L  A  F  E  R  Y  I  I  I  C   140

AAGCCCTTCGGCAACTTCCGCTTCAGCTCCAAGCATGCACTGATAGCGGTCCTGGCCACC  480
 K  P  F  G  N  F  R  F  S  S  K  H  A  L  I  A  V  L  A  T   160

TGGGCAATTGGGATCGGCGTCTCCATCCCACCCTTCTTTGGCTGGAGCCGGTTCCTGCCT  540
 W  A  I  G  I  G  V  S  I  P  P  F  F  G  W  S  R  F  L  P   180

GAGGGCCTGCAGTGTTCCTGTGGCCCCGACTGGTACACCGTGGGCACCAAGTACTACAGC  600
 E  G  L  Q  C  S  C  G  P  D  W  Y  T  V  G  T  K  Y  Y  S   200

GAGTATTATACCTGGTTCCTCTTCATCTTCTGCTACATCGTACCTCTCGCCCTCATCTGC  660
 E  Y  Y  T  W  F  L  F  I  F  C  Y  I  V  P  L  A  L  I  C   220

TTCTCCTACTCTCAGCTGCTGGGGGCCCTCAGAGCTGTTGCAGCTCAGCAGCAGGAGTCA  720
 F  S  Y  S  Q  L  L  G  A  L  R  A  V  A  A  Q  Q  Q  E  S   240

GCTTCAACCCAGAAGGCTGAGCGGGAGGTGAGCCACATGGTGGTTGTGATGGTGGGGTCC  780
 A  S  T  Q  K  A  E  R  E  V  S  H  M  V  V  V  M  V  G  S   260

TTTTGTGTCTGTTACACGCCCTACGCTGCCCTGGCCATGTATATAGTCAACAACCGGAAC  840
 F  C  V  C  Y  T  P  Y  A  A  L  A  M  Y  I  V  N  N  R  N   280

CACGGGGTGGACCTACGGCTTGTCACCATTCCTGCCTTCTTCTCCAAGAGCGCTTGCATC  900
 H  G  V  D  L  R  L  V  T  I  P  A  F  F  S  K  S  A  C  I   300

TACAATCCGATCATCTACTGCTTCATGAATAAGCAGTTCCGAGCTTGCATCATGGAGATG  960
 Y  N  P  I  I  Y  C  F  M  N  K  Q  F  R  A  C  I  M  E  M   320

GTGTGCGGAAAGCCCATGACAGATGAGTCTGACATGTCTAGCTCCCAGAAAACCGAAGTT 1020
 V  C  G  K  P  M  T  D  E  S  D  M  S  S  S  Q  K  T  E  V   340

TCTACTGTCTCTTCTAGCCAAGTTGGCCCCAACTAA  1056
 S  T  V  S  S  S  Q  V  G  P  N stop  351
```

Figure 3.4 Nucleotide sequence of the coding region of the porcine BCP gene, with the deduced amino acid sequence given under each codon.

conserved amino acids at important sites within the porcine BCP are present, confirming its identity as an opsin (these residues are indicated in green in Figure 3.5). For example, lysine-296, the binding site of the chromophore 11-*cis* retinal via a protonated Schiff base (Wang *et al.*, 1980), glutamate-113, which serves as the counterion for the retinylidene Schiff base (Sakmar *et al.*, 1989), and cysteine-110 and -187, which form a disulfide bond necessary for the formation of a functional opsin (Karnik and Khorana, 1990) are all present. Conserved serine and threonine residues are present in the C-terminal end of the protein. These residues are sites for multiple phosphorylation by protein kinase, and for the binding of arrestin (Wilden *et al.*, 1986). The charged pair of residues Glu-134 and Arg-135 at the C-terminal end of cytoplasmic loop III, which play a role in the binding and activation of transducin (Franke *et al.*, 1992), are also present in the porcine BCP. Asn-Xaa-Ser are recognition sites for glycosylation reactions of the Asn residue, where Xaa can be any amino acid (Creighton, 1993). Thus, the asparagine residue at the N-terminal end of the protein, Asn-17, may be a site of glycosylation.

Table 3.3 shows the percentage identity between the predicted amino acid sequence for the porcine BCP with a variety of other opsins. The porcine BCP shares over 60% amino acid identity with all members of the S group of visual pigments, with the highest identity of 94% to the bovine BCP. Based on this high degree of amino acid identity, and the above sequence analysis, the porcine BCP predicted from the BCP gene isolated in this study is clearly another member of the S group of visual pigments.

3.2.3 Genomic structure of the porcine BCP gene

Comparison of the nucleotide sequence of the porcine BCP gene with human gene sequence resulted in the identification of the exon/intron boundaries. These two gene sequences are shown aligned to each other in Figure 3.6. The exon/intron boundaries in the porcine BCP gene were found to be at identical positions to those within the human BCP gene (Figure 3.6). The GT-AG rule, which refers to the consensus splice-site at the junction of exon and introns found in many nuclear genes in eukaryotes (Lewin, 1990), was identified for the exon/intron boundaries in the porcine gene (Table 3.4).

The porcine BCP gene contains 5 exons and 4 introns. The protein coding region of the porcine BCP gene, from the ATG start codon to the stop codon, is 1056 bp (Figure 3.4). Exons 2 to 5 are the same size in both human and porcine. Exon 1 in the porcine contains an extra 9 bp

```
Porcine  MSKMPEEEEE EFLLFKNISL VGPWDGPQYH LAPVWVFHLQ AAFMGFVFLV  50
Human    MRKMS---EE EFYLFKNISS VGPWDGPQYH IAPVWAFYLQ AAFMGIVFLI  47
         * **    ** ** ******  **********  **** * ** ***** ***

Porcine  GTPLNATVLV ATLRYRKLRQ PLNYILVNVS LGGFIYCIFS VFSVFIASCH  100
Human    GFPLNAMVLV ATLRYKKLRQ PLNYILVNVS FGGFLLCIFS VFPVFVASCN   97
         * **** *** ***** **** **********  ***  **** ** ** ***

Porcine  GYFVFGRRVC AMEAFLGSAA GLVTGWSLAF LAFERYIIIC KPFGNFRFSS  150
Human    GYFVFGRHVC ALEGFLGTVA GLVTGWSLAF LAFERYIVIC KPFGNFRFSS  147
         ******* ** * * ***  * ********** ********** **********

Porcine  KHALIAVLAT WAIGIGVSIP PFFGWSRFLP EGLQCSCGPD WYTVGTKYYS  200
Human    KHALTVVLAT WTIGIGVSIP PFFGWSRFIP EGLQCSCGPD WYTVGTKYRS  197
         ****  **** * ******** ******** * ********** ******** *

Porcine  EYYTWFLFIF CYIVPLALIC FSYSQLLGAL RAVAAQQQES ASTQKAEREV  250
Human    ESYTWFLFIF CFIVPLSLIC FSYTQLLRAL KAVAAQQQES ATTQKAEREV  297
         * ******** * **** *** *** *** **  ********* * ********

Porcine  SHMVVVMVGS FCVCYTPYAA LAMYIVNNRN HGVDLRLVTI PAFFSKSACI  300
Human    SRMVVVMVGS FCVCYVPYAA FAMYMVNNRN HGLDLRLVTI PSFFSKSACI  337
         * ******** ***** ****  *** ***** ** ******* * ********

Porcine  YNPIIYCFMN KQFRACIMEM VCGKPMTDES DMSSSQKTEV STVSSTQVGP N  351
Human    YNPIIYCFMN KQFQACIMKM VCGKAMTDES DTCSSQKTEV STVSSTQVGP N  348
         ********** *** **** * **** ***** *  ******* ********** *
```

Figure 3.5 Alignment of the predicted porcine BCP amino acid sequence to the human BCP. Conserved amino acids are undermarked with an *. Residues indicated in green are mentioned in Section 3.2.2.

	Phylogenetic Group	Human Blue	Porcine Blue
Human red	L	42.2	41.9
Human green	L	43.4	42.5
Human rod	Rh	44.5	41.1
Chicken blue	M1	49.4	47.9
Goldfish blue	M1	50.0	47.3
Goldfish Ultraviolet	S	62.5	64.0
Murine blue	S	85.6	86.5
Bovine blue	S	86.5	94.0
Chicken violet	S	81.6	81.0
Talapoin blue	S	95.6	85.7
Marmoset blue	S	92.5	86.6
Human blue	S	100	86.8
Porcine blue	?	86.8	100

Table 3.3 The percentage identity between the human blue cone pigment and the deduced porcine blue cone pigment amino acid sequences, compared with a variety of other opsins. The phylogenetic group to which each visual pigment belongs is also given (see Section 1.14 for description of phylogenetic groups).

```
Porcine        GAGGATGATCTATAAGAGGAATCAGAGG-GGGTG-GGGACATCC ATG AGC
Human          GAGGATCACCTATAAGAGGACTCAGAGGAGGGTGTGGGGCATCC ATG AGA

Porcine   AAG ATG TCA GAG GAG GAG GAG GAG GAG TTT CTT CTG TTC AAG AAC
Human     AAA ATG TCG GAG GAA --- --- --- GAG TTT TAT CTG TTC AAA AAT

Porcine   ATC TCC TTG GTG GGG CCG TGG GAT GGG CCT CAG TAC CAC CTC GCC
Human     ATC TCT TCA GTG GGG CCG TGG GAT GGG CCT CAG TAC CAC ATT GCC

Porcine   CCT GTC TGG GTC TTC CAC CTC CAG GCA GCC TTC ATG GGC TTT GTC
Human     CCT GTC TGG GCC TTC TAC CTC CAG GCA GCT TTC ATG GGC ACT GTC

Porcine   TTC CTT GTA GGG ACA CCA CTC AAT GCC ACG GTT CTG GTG GCC ACC
Human     TTC CTT ATA GGG TTC CCA CTC AAT GCC ATG GTG CTG GTG GCC ACA

Porcine   CTG CGC TAC AGA AAG TTG CGG CAG CCC CTC AAC TAT ATT CTG GTC
Human     CTG CGC TAC AAA AAG TTG CGG CAG CCC CTC AAC TAC ATT CTG GTC

Porcine   AAC GTG TCC CTG GGG GGC TTC ATC TAC TGC ATC TTC TCT GTC TTC
Human     AAC GTG TCC TTC GGA GGC TTC CTC CTC TGC ATC TTC TCT GTC TTC

Porcine   AGC GTC TTC ATT GCC AGC TGT CAT GGG TAC TTC GTC TTT GGC CGC
Human     CCT GTC TTC GTC GCC AGC TGT AAC GGA TAC TTC GTC TTC GGT CGC

                                                        Intron 1
Porcine   CGT GTT TGT GCT ATG GAG GCC TTC CTA GGC TCT GCA GCA GGT CTG
Human     CAT GTT TGT GCT TTG GAG GGC TTC CTG GGC ACT GTA GCA GGT CTG

Porcine   GTG ACA GGC TGG TCA CTG GCC TTC TTG GCC TTT GAG CGC TAC ATC
Human     GTT ACA GGA TGG TCA CTG GCC TTC CTG GCC TTT GAG CGC TAC ATT

Porcine   ATC ATC TGT AAG CCC TTC GGC AAC TTC CGC TTC AGC TCC AAG CAT
Human     GTC ATC TGT AAG CCC TTC GGC AAC TTC CGC TTC AGC TCC AAG CAT

Porcine   GCA CTG ATA GCG GTC CTG GCC ACC TGG GCA ATT GGG ATC GGC GTC
Human     GCA CTG ACG GTG GTC CTG GCT ACC TGG ACC ATT GGT ATT GGC GTC

                                        Intron 2
Porcine   TCC ATC CCA CCC TTC TTT GGC TGG AGC CGG TTC CTG CCT GAG GGC
Human     TCC ATC CCA CCC TTC TTT GGC TGG AGC CGG TTC ATC CCT GAG GGC

Porcine   CTG CAG TGT TCC TGT GGC CCC GAC TGG TAC ACC GTG GGC ACC AAG
Human     CTG CAG TGT TCC TGT GGC CCT GAC TGG TAC ACC GTG GGC ACC AAA

Porcine   TAC TAC AGC GAG TAT TAT ACC TGG TTC CTC TTC ATC TTC TGC TAC
Human     TAC CGC AGC GAG TCC TAT ACG TGG TTC CTC TTC ATC TTC TGC TTC
```

```
Porcine   ATC GTA CCT CTC GCC CTC ATC TGC TTC TCC TAC TCT CAG CTG CTG
Human     ATT GTG CCT CTC TCC CTC ATC TGC TTC TCC TAC ACT CAG CTG CTG

                        Intron 3
Porcine   AGG GCT CTC AGA GCT GTT GCA GCT CAG CAG CAG GAG TCA GCT TCA
Human     AGG GCC CTG AAA GCT GTT GCA GCT CAG CAG CAG GAG TCA GCT ACG

Porcine   ACC CAG AAG GCT GAG CGG GAG GTG AGC CAC ATG GTG GTT GTG ATG
Human     ACC CAG AAG GCT GAA CGG GAG GTG AGC CGC ATG GTG GTT GTG ATG

Porcine   GTG GGG TCC TTT TGT GTC TGT TAC ACG CCC TAC GCT GCC CTG GCC
Human     GTA GGA TCC TTC TGT GTC TGC TAC GTG CCC TAC GCG GCC TTC GCC

Porcine   ATG TAT ATA GTC AAC AAC CGG AAC CAC GGG GTG ACC TTA CGG CTT
Human     ATG TAC ATG GTC AAC AAC CGT AAC CAT GGG CTG GAC TTA CGG CTT

Porcine   GTC ACC ATT CCT GCC TTC TTC TCC AAG AGC GCT TGC ATC TAC AAT
Human     GTC ACC ATT CCT TCA TTC TTC TCC AAG AGT GCT TGC ATC TAC AAT

                                        Intron 4
Porcine   CCG ATC ATC TAC TGC TTC ATG AAT AAG CAG TTC CGA GCT TGC ATC
Human     CCC ATC ATC TAC TGC TTC ATG AAT AAG CAG TTC CAA GCT TGC ATC

Porcine   ATG GAG ATG GTG TGC GGA AAG CCC ATG ACA GAT GAG TCT GAC ATG
Human     ATG AAG ATG GTG TGT GGG AAG GCC ATG ACA GAT GAA TCC GAC ACA

Porcine   TCT AGC TCC CAG AAA ACC GAA GTT TCT ACT GTC TCT TCT ACG CAA
Human     TGC AGC TCC CAG AAA ACA GAA GTT TCT ACT GTC TCG TCT ACC CAA

Porcine   GTT GGC CCC AAC TAA GGAGCC-ACACTGGCATATCTG---CAGCGAG-----GA
Human     GTT GGC CCC AAC TGA GGACCCAATATTGGCCTGTTTGCAACAGCTAGAATTAAA

Porcine  TTTCACATTTAAGGAAATTTCTACTTTCTCTATTAAAATCCCAACCACTAACAGCACAG
Human    TTTTACTTTTAAGTAAGTTTCTATTGTCTCCGTCAGAAACCAAACTACTAAAAACACAA

Porcine  TAAA-ACAGGAAGGGGAGTGACAGCAATTTGGGGAGTCAATTTTCC----TACCAGTTC
Human    AAAAGATGGTAAAAGGAGTGATGGCAGTTTGGGGAGTCAATTTTTCATTTTCTTACTAT

Porcine  CTCCCTACTGAC-ACAAAGCTACTGTTCCAGCTGGGTCTACTTCAGATCAACCCAAAGG
Human    TGCCTTCTTGCCTACAAAGCTACTGTTTCCACT-GGTCTATTTCAGACC-ACCCAAAGG

Porcine  CCATCTCAACAGTCATCAGTTTCCACTCTTTTTTTTTTTTTTTTTGCT
Human    CCATTTCAACAATCATCAGTTTCTACTC-------------------CT
```

Figure 3.6 An alignment of the coding region of the porcine BCP gene to the human BCP gene. The TATAA box is shown in bold. The ATG start codon is underlined. The polyadenylation signal sequence (ATTAAA) is given in bold and underlined. The location of the four introns is indicated above the contiguous coding sequence.

Junction	Species	Sequence
EXON 1-Intron 1	Porcine	...GCA GCA G **gt**acctcagg..
	Human	...GTA GCA G **gt**actgcagg..
Intron 1-EXON 2	Porcine	...ttcatcgc**ag** GT CTG GTG..
	Human	...ttcaccac**ag** GT CTG GTG..
EXON 2-Intron 2	Porcine	...TGG AGC CG **gt**gagagtgt..
	Human	...TGG AGC CG **gt**gagagtgc..
Intron 2-EXON 3	Porcine	...tcccttgc**ag** G TTC CTG..
	Human	...tcctttgc**ag** G TTC ATC..
EXON 3-Intron 3	Porcine	... AGA GCT **gt**gagtggc..
	Human	... AAA GCT **gt**gagtgga..
Intron 3-EXON 4	Porcine	...tgcttccc**ag** GTT GCA...
	Human	...tccaccccc**ag** GTT GCA...
EXON 4-Intron 4	Porcine	... AAG CAG **gt**aaagctct..
	Human	... AAG CAG **gt**aaagctct..
Intron 4-EXON 5	Porcine	...tttccctc**ag** TTC CGA...
	Human	...ttctctcc**ag** TCC CAA...

Table 3.4 Comparison of the exon and intron junction splice sites of the porcine BCP gene (this study) and the human BCP gene (Nathans *et al.*, 1986a). Coding sequence is shown in uppercase and intron sequence is shown in lowercase lettering. The BCP gene splice sites in both species obey the gt-ag rule, and these nucleotides are indicated in bold.

compared to the equivalent exon in the human gene (Figure 3.6). The Goldberg-Hogness (TATA) box in the porcine sequence is found in the same position upstream of the BCP gene when compared with the human gene. Based on the position of the TATA box, the mRNA transcription start-site for the porcine gene is presumed to be at the same position as the human (Figure 3.6). The predicted polyadenylation signal (ATTAAA) for the porcine BCP mRNA is 30 bp further downstream relative to the position of the human polyadenylation signal (Nathans *et al.*, 1986a and Figure 3.6).

Amplification of individual introns from the porcine BCP gene and the human BCP gene, using exon-specific primers (see Section 2.2.1 and Table 2.1), indicated that there are differences in intron sizes, notably for intron 3 (see Figure 3.7). Intron 3 in the human BCP gene has been reported to be 606 bp long (Nathans *et al.*, 1986a). The same intron in the porcine BCP gene appeared to be approximately 300 bp shorter (Figure 3.7, lanes 7 and 8). (The extra sequence in intron 3 of the human BCP gene was found to be as a result of an insertion of an *Alu* element (see Section 6.2.3).) The size differences, if any, between the other introns are small in comparison. Intron 2 appears to be of the same size in both species (Figure 3.7, lanes 4 and 6). Introns 1 and 4 are approximately 50 bp larger and 70 bp smaller in the porcine relative to the equivalent introns in human, respectively (Figure 3.7). Thus, the genomic structure of the porcine and human BCP genes are very similar.

3.2.4 Comparison of the S group proteins

Figure 3.8 shows the deduced amino acid sequence for the porcine BCP aligned to the vertebrate visual pigments that have been classified into the S group (Nathans *et al.*, 1986a; Okano *et al.*, 1992; Chiu *et al.*, 1994; Hunt *et al.*, 1995). The goldfish ultraviolet pigment was recently placed within the S group on the basis of its higher degree of amino acid identity to the S opsins than to any of the other group of visual pigments (Hisatomi *et al.*, 1996). Amino acid sequences were initially aligned using the computer package Clustal V (Section 2.21), and then refined by eye. Note that the numbering given for the porcine BCP amino acid sequence (shown in bold in Figure 3.8) is identical to the bovine rod opsin system, where the Schiff base lysine is at site 296 (Wang *et al.*, 1980). Figure 3.8 indicates those residues which are identical between all the S group members (undermarked with an asterix (*)).

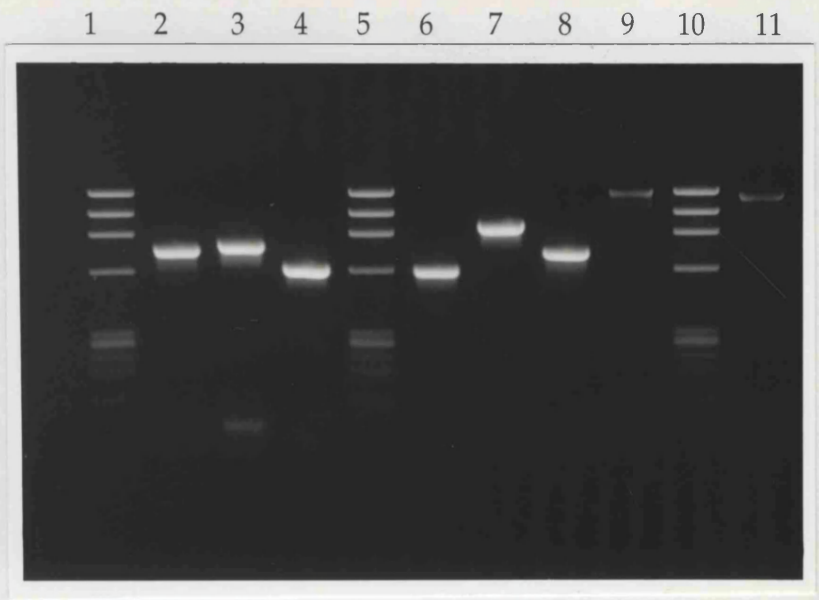


Figure 3.7 Amplification of the 4 introns within the porcine and human BCP genes. PCR was performed using primers specific to BCP exonic sequences flanking the introns (see Table 2.1 for sequence of primers).

		Primers		
Lane	Template	Forward	Reverse	Size (bp)
2	Human	Ex1+ 469/484F	Ex2- 929/908R	747
3	Porcine	Ex1+ 469/484F	Ex2- 929/908R	~ 797
4	Human	Ex2+ 784/799F	Ex3- 1069/1054R	616
6	Porcine	Ex2+ 784/799F	Ex3- 1069/1054R	~ 616
7	Human	Ex3+ 935/954F	Ex4- 1276/1254R	949
8	Porcine	Ex3+ 935/954F	Ex4- 1276/1254R	~ 670
9	Human	Ex4+ 1098/1117F	Ex5- 1453/1433R	1343
11	Porcine	Ex4+ 1098/1117F	Ex5- 1453/1433R	~ 1270

Lanes 1, 5 & 10: ϕX174/*Hae*III DNA ladder, sizes (bp):		1353
		1078
		872
		603
		310
	(appears as one band)	281/271
		194
		118
		72

```
                                                                  Helix I
                                                                 ======================
Porcine     MSKMPEEEEE EFLLFKNISL VGPWDGPQY-H LAPVWVFHLQ AAFMGFVFLV GTPLNATVLV   60
Human       MRKMS---EE EFYLFKNISS VGPWDGPQY-H IAPVWAFYLQ AAFMGTVFLI GFPLNAMVLV   57
Talapoin    MRKMS--EEE EFYLFKNISS VKPWDGPQY-H IPPVWAYYLQ AAFMGTVFLA GFPLNAMVLV   58
Marmoset    MSKMS--EEE EFYLFKNISS VGPWDGPQYPH IAPSWAYYLQ AAFMGIVFLA GLPLNTMVLV   59
Mouse       MS-----GED DFYLFQNISS VGPWDGPQY-H LAPVWAFRLQ AAFMGFVFFV GTPLNAIVLV   55
Bovine      MSKMS--EEE EFYLFKNISL VGPWDGPQY-H LAPVWAFHLQ AVFMGFVFFV GTPLNATVLV   58
Chick Vi    MS-----SDD DFYLFTNGSV PGPWDGPQY-H IAPPWAFYLQ TAFMGIVFAV GTPLNAVVLW   55
Gf UltraVi  MDA-W--TYQ -FG---NLSK ISPFEGPQY-H LAPKWAFYLQ AAFMGFVFFV GTPLNAIVLF   53
            *          *    * *    *  **** *   * *   **   *** **   * ***  **

                                  Helix II
            =    CyI      ======================      ExI        ========
Porcine     ATLRYRKLRQ PLNYILVNVS LGGFIYCIFS VFSVFIASCH GYFVFGRRVC AMEAFLGSAA   120
Human       ATLRYKKLRQ PLNYILVNVS FGGFLLCIFS VFPVFVASCN GYFVFGRHVC ALEGFLGTVA   117
Talapoin    ATVRYKKLRQ PLNYILVNVS FGGFLLCIFS VFPVFVNSCK GYFVFGRHVC GFEAFLGTVA   118
Marmoset    ATVRYKKLRH PLNYVLVNVS VGGFLLCIFS VFPVFVASCH GYFVFGRHVC ALEGFLGTVA   119
Mouse       ATLHYKKLRQ PLNYILVNVS LGGFLFCIFS VFTVFIASCH GYFLFGRHVC ALEAFLGSVA   115
Bovine      ATLRYRKLRQ PLNYILVNVS LGGFIYCIFS VFIVFITSCY GYFVFGRHVC ALEAFLGCTA   118
Chick Vi    VTVRYKRLRQ PLNYILVNIS ASGFVSCVLS VFVVFVASAR GYFVFGKRVC ELEAFVGTHG   115
Gf UltraVi  VTMKYKKLRQ PLNYILVNIS LGGFIFDTFS VSQVFFSALR GYYFFGYTLC AMEAAMGSIA   113
             *  *  **  **** *** *   **     * *  **      **  **   *   *   *

            Helix III                                   Helix IV
            ==========          CyII          ======================
Porcine     GLVTGWSLAF LAFERYIIIC KPFGNFRFSS KHALIAVLAT WAIGIGVSIP PFFGWSRFLP   180
Human       GLVTGWSLAF LAFERYIVIC KPFGNFRFSS KHALTVVLAT WTIGIGVSIP PFFGWSRFIP   177
Talapoin    GLVTGWSLAF LAFERYIVIC KPFGNFRFSS KHALTVVLAT WIIGIGVSIP PFFGWSRFIP   178
Marmoset    GLVTGWSLAF LAFERYIVIC KPFGNFRFSS KHALMVVLTT WTIGIGVSIP PFFGWSRYIA   179
Mouse       GLVTGWSLAF LAFERYVVIC KPFGSIRFNS KHALMVVLAT WIIGIGVSIP PFFGWSRFIP   175
Bovine      GLVTGWSLAF LAFERYIIIC KPFGNFRFSS KHALMVVVAT WTIGIGVSIP PFFGWSRFVP   178
Chick Vi    GLVTGWSLAF LAFERYIVIC KPFGNFRFSS RHALLVVVAT WLIGVGVGLP PFFGWSRYMP   175
Gf UltraVi  GLVTGWSLAV LAFERYVVIC KPFGSFKFGQ SQALGAVALT WIIGIGCATP PFWGWSRYIP   173
            *********  ******  ** ****   *     **  *  * * ** *   * ** ****

                                                Helix V
            ExII                      ========================      CyIII
Porcine     EGLQCSCGPD WYTVGTKYYS EYYTWFLFIF CYIVPLALIC FSYSQLLGAL RAVAAQQQES   240
Human       EGLQCSCGPD WYTVGTKYRS ESYTWFLFIF CFIVPLSLIC FSYTQLLRAL KAVAAQQQES   237
Talapoin    EGLQCSCGPD WYTVGTKYRS ESYTWFLFIF CFIVPLSLIC FSYTQLLRAL KAVAAQQQES   238
Marmoset    EGLQCSCGPD WYTVGTKYRS EYYTWFLFIF CFIVPLALIC FSYAQLLRAL KAVAAQQQES   239
Mouse       EGLQCSCGPD WYTVGTKYRS EYYTWFLFIF CFIIPLSLIC FSYSQLLRTL RAVAAQQQES   235
Bovine      EGLQCSCGPD WYTVGTKYYS EYYTWFLFIF CYIVPLSLIC FSYSQLLGAL RAVAAQQQES   238
Chicken Vi  EGLQCSCGPD WYTVGTKYRS EYYTWFLFIF CFIVPLSLII FSYSQLLSAL RAVAAQQQES   235
Gf UltraVi  EGIGTACGPD WYTKNEEYNT KSYTYFLLVS CFMMPIMIIT FSYSQLLGAL RAVAAQQAES   233
            **    **** ***    *     ** **    *   *   *  *** ***  *  ****** **

                                  Helix VI                          Helix VII
                                ======================  ExIII   ==========
Porcine     ASTQKAEREV SHMVVVMVGS FCVCYTPYAA LAMYIVNNRN HGVDLRLVTI PAFFSKSACI   300
Human       ATTQKAEREV SRMVVVMVGS FCVCYVPYAA FAMYMVNNRN HGLDLRLVTI PSFFSKSACI   297
Talapoin    ATTQKAEREV SRMVVVMVGS FCVCYVPYAA FAMYMVNNRN HGLDLRLVTI PAFFSKSACI   298
Marmoset    ATTQKAEREV SRMVVVMVGS FCVCYVPYAA LAMYMVNNRN HGLDLRLVTI PAFFSKSSCI   299
Mouse       ATTQKAEREV SHMVVVMVGS FCLCYVPYAA LAMYMVNNRN HGLDLRLVTI PAFFSKSACV   295
Bovine      ASTQKAEREV SHMVVVMVGS FCLCYTPYAA LAMYIVNNRN HGVDLRLVTI PAFFSKSSCV   298
Chicken Vi  ATTQKAEREV SRMVVVMVGS FCLCYVPYAA LAMYMVNNRD HGLDLRLVTI PAFFSKSACV   295
Gf UltraVi  ASTQKAEKEV SRMVVVMVGS FVVCYGPYAI TALYFSYAED SNKDYRLVAI PSLFSKSSCV   293
            * ***** ** * ******** *  ** ***   * *           * *** * *  **** *

            ========
Porcine     YNPIIYCFMN KQFRACIMEM V-CGKPMTDES DMSSS-QKTEV STVSSSQVGP N          351
Human       YNPIIYCFMN KQFQACIMKM V-CGKAMTDES DTCSS-QKTEV STVSSTQVGP N          348
Talapoin    YNPIIYCFMN KQFQAHIMKM V-CGKAMTDES DISSS-QKTEV STVSSTQVGP N          349
Marmoset    YNPIIYCFMN KQFRACIMEM V-CGKAMTDES DISSS-QKTEV STVSSTQVGP N          350
Mouse       YNPIIYCFMN KQFRACILEM V-CRKPMADES DVSGS-QKTEV STVSSSKVGP N          346
Bovine      YNPIIYCFMN KQFRACIMEM V-CGKPMTDES ELSSS-QKTEV STVSSSQVGP N          349
Chicken Vi  YNPIIYCFMN KQFRACIMET V-CGKPLTDDS DASTSAQRTEV SSVSSSQVGP T          346
Gf UltraVi  YNPLIYAFMN KQFNACIMET V-FGKKI-DES SEVSS--KTET SSVS------ A          336
            *** ** *** *** * *    *   *   * *     *   **  * **
```

Figure 3.8 Deduced amino acid sequence of the porcine BCP compared with the other S group visual pigments. Conserved amino acids are undermarked with an *. The seven alpha-helical transmembrane regions are overlined, and indicated as Helix I-VII. Cytoplasmic loops I, II and III are labelled CyI, CyII and CyIII. Extracellular loops I, II and III are labelled ExI, ExII and ExIII. Amino acid insertions are indicated by a dash. Lysine-296, glutamate-113, cysteine -110 and -187 and the charged pair of residues Glu-134 & Arg-135 which are conserved in all opsins are indicated in green. Serine and threonine residues, which are potential sites of phosphorylation, at the C-terminal end of the protein are also indicated in green. Residues in red are amino acids which are found conserved in all of 24 other vertebrate opsins (listed in Table 3.5) and are also present within members of the S group of pigments at that particular site. Residues in blue are amino acids which are found only within members of the S group of opsins when compared with 24 other vertebrate pigments at that particular site. The other 24 vertebrate opsins which were used for comparison are listed in Table 3.5. Chicken Vi = Chicken violet cone pigment, Gf UltraVi = Goldfish ultraviolet pigment.

The above alignment of the eight S group pigments was also compared with 24 other vertebrate opsin amino acid sequences aligned by Yokoyama (1994). These 24 non-S group opsins are listed in Table 3.5. Residues which are conserved in all of these 24 opsins and are also found at the same position within members of the S group of pigments were noted (indicated in red in Figure 3.8). For example, Ser-240 is conserved in all opsins (Figure 3.8). Whereas, with the exception of the marmoset BCP, which contains Ala-180, Pro-180 is conserved in all the vertebrate opsins (Figure 3.8). Thus, one can identify residues which are conserved within the S group of visual pigments, and whether they are also conserved in all the other 24 vertebrate visual pigments, utilised in this study, from Figure 3.8. For example, a continuous stretch of conserved amino acids, between the S group members, was identified within cytoplasmic loop III (from Ala-232 to Ser-251). A subset of these residues, within cytoplasmic loop III, are also conserved in the other 24 vertebrate opsins (Ala-234, Gln-237, Ser-240, Thr-243, Gln-244, Ala-246, Glu-247, Glu-249 and Val-250). This high degree of conservation at this particular region of the protein is not surprising, since residues within cytoplasmic loop III are essential for the binding and activation of transducin (Konig *et al.*, 1989).

3.2.41 Identification of amino acids which are exclusive to individual S opsins

The comparison of the S group of opsins with 24 other vertebrate visual pigments (Table 3.5) has permitted the identification of those residues which are exclusive to members of the S group. For example, Arg-67 is a chicken violet pigment-specific residue and is indicated in blue in Figure 3.8. Other residues found only within members of the S group and not in any of the other visual pigments, at a particular position are indicated in blue in Figure 3.8. Examples of residues which are exclusive to members of the S group are Trp-60 and Ala-120 (Figure 3.8). Trp-60 is found only in the chicken violet opsin, whereas Val-60 and Phe-60 are present in the other members of the S group. Val-60 and Phe-60 are also found within the other 24 opsins, and thus these residues are shown in plain text in Figure 3.8. Ala-120 is found only within 7 members of the S group, whereas Cys-120 is conserved in all members of the L group and Gly-120 is found in all the other visual pigments including the chicken violet opsin (Figure 3.8).

Residues which are identified as being exclusive to one or two members of the S group can be used to follow the evolution of individual

Rh group

Lamprey	(*Lamptera japonica*)
Sand Goby	(*Pomatoschistus minutus*)
Goldfish	(*Carassius auratus*)
Frog	(*Xenopus laevis*)
Chicken	(*Gallus gallus*)
Human	(*Homo sapiens*)
Cow	(*Bos tauras*)
Mouse	(*Mus musculus*)
Hamster	(*Cricetulus griseus*)

M2 group

Goldfish (G1)
Goldfish (G2)
Chicken

M1 group

Goldfish	
Gecko	(*Gekko gekko*)
Chicken	
Fish	(*Astyanax fasciatus*, ($_{Af}$))

L group

Fish (G101 $_{Af}$)
Fish (G103 $_{Af}$)
Gecko
Human (GCP)
Fish (ROO7$_{Af}$)
Goldfish (R)
Chicken (R)
Human (RCP)

Table 3.5 List of the 24 vertebrate visual pigments which were used to produce an amino acid sequence alignment by Yokoyama (*Mol Biol Evol* (1994) 11(1):32-39). The amino acid sequences of these 24 opsins were compared with the residues in the S group of visual pigments, for the production of Figure 3.8.

species. For example, Thr-46, Thr-155 and Gln-314 are found only in the two Old World primate BCPs (human and talapoin monkey; Figure 3.8). This implies that these substitutions occurred originally in a primitive ancestor of the Old World primates after the divergence from the New World primates (marmoset) about 43 million years ago, indicated as time 1 in Figure 3.9 (Hunt *et al.*, 1995). The human BCP-specific residue Ile-50 (Figure 3.8) would probably have originated about 14 million years ago after the split of the Old World primates into the Hominoid (humans and great apes) and the Cercopithecoid (talapoin monkey) branches (Time 2 in Figure 3.9). Evolutionary analysis of other residues which are exclusive to members of the S group of opsins (identified in blue in Figure 3.8) can be carried out in a similar fashion.

Residues which involve non-conservative amino acid substitutions and that are exclusive to individual members of the S group of opsins may be involved in altering the function of the protein, such as its spectral sensitivity. An alternative interpretation is that these substitutions do not affect the normal action of the protein. The residue Lys-22 is specific to the talapoin monkey BCP and found in the N-terminal region of the protein (Figure 3.8). Lys-22 is the only residue carrying a positive charge at this position. The non-polar Gly-22 is present in all other members of the S group and in the human red and green pigments. The hydroxyl bearing Ser-22 is found in all vertebrate rod opsins. Arg-324 is another amino acid which carries a positive charge and is found to be specific to the murine BCP, whereas every other vertebrate opsin, including the rest of the S group of pigments, contain Gly-324 (Figure 3.8). Whether or not Lys-22 or Arg-324 have an effect on the normal function of these visual pigments is difficult to predict without experimental evidence. Ala-180 is a marmoset BCP-specific residue, whereas Pro-180 is present in all the other vertebrate opsins including the other S group pigments (Figure 3.8). Proline residues possess a unique structure, which is not seen in any other amino acid (Stryer, 1988). Substitution of proline for any other residue is considered a non-conserved change. Thus, Ala-180, in the marmoset, may well alter the properties of the opsin (see Section 3.3.2).

3.2.42 Identification of S group-specific amino acids

Any residues indicated in blue in Figure 3.8 which are conserved, at the same position, in at least 7 of the 8 members of the S group are listed in Table 3.6, and are termed 'S group-specific'. For example, the residues Ala-120 and Cys-299 were found to be exclusive to members of the S group and

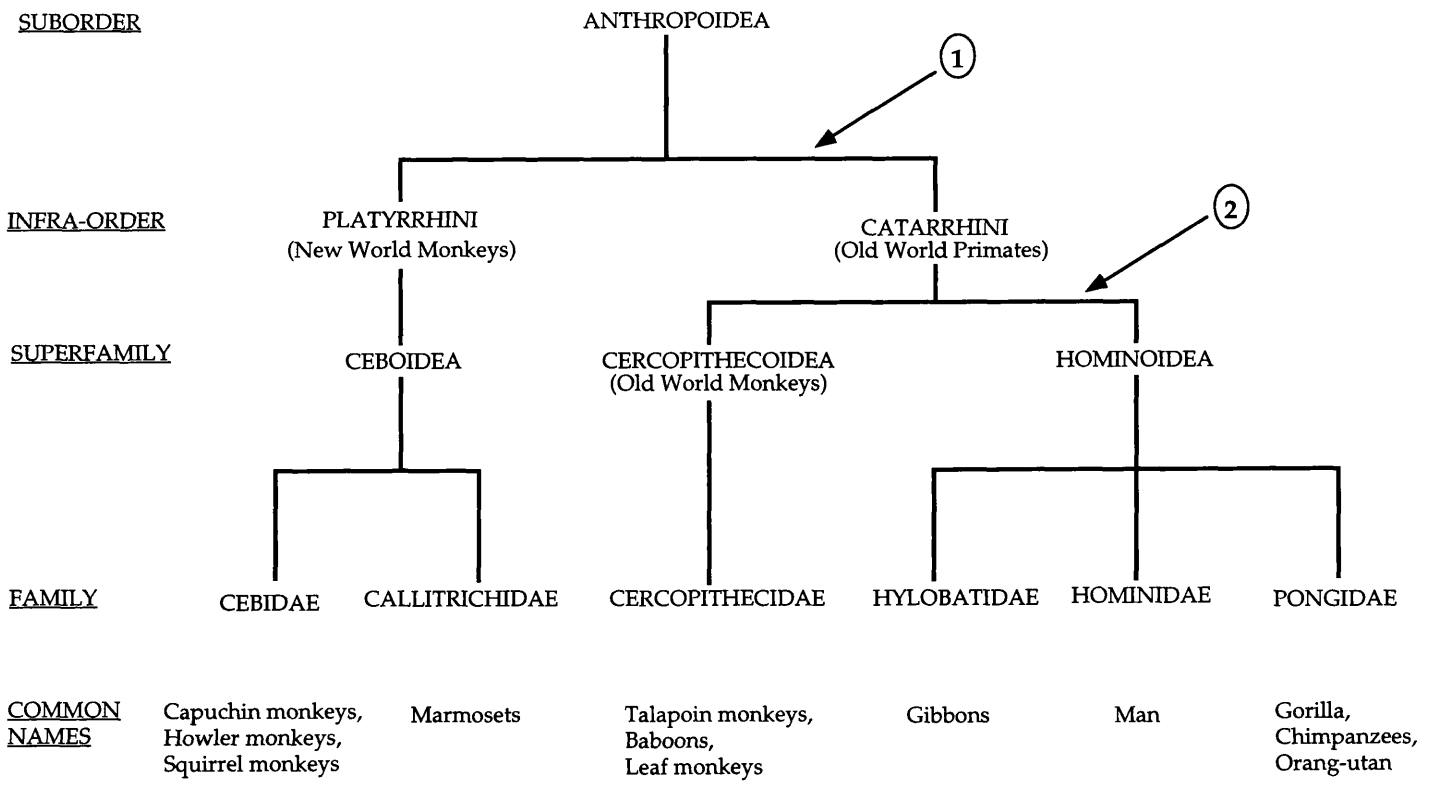


Figure 3.9 Classification of the primate suborder Anthropoidea (adapted from Napier, 1985).

Positions 1 and 2 refer to times within the evolution of the primates where certain species-specific residues found in their respective BCPs have arisen for a particular group (see Section 3.2.41).

Region	Position
N-terminal	Trp-24#
Helix I	Gln-40 Gly-45
Cytoplasmic Loop I	None
Helix II	Gly-82♦ Phe-84 Cys/**Asp**-87 Ser-90 Val-91 Val-94
Extracellular Loop I	Val-109#
Helix III	Gly-117 Ala-120♦ Leu-122 Thr-124 Gly-125 Phe-130#
Cytoplasmic Loop II	None
Helix IV	Leu-154 Ile-163 Val-167# Phe-172
Extracellular Loop II	Val/**Lys**-194 Gly-195# Thr-196# Ser/**Thr**-200
Helix V	Thr-204 Trp/**Tyr**-205 Leu-218#
Cytoplasmic Loop II	None
Helix VI	Gly-259 Ser-260 Tyr-265 Met-273#
Extracellular Loop III	Asn/**Tyr**-277 His-281# Gly-282#
Helix VII	Val-288 Cys-299 Cys/**Ala**-307
C-terminal	Ala-315

Table 3.6 Amino acid residues which are exclusive to and conserved within the S group of pigments (S group-specific).
Any residues indicated in blue in Figure 3.8 which are conserved at the same position within at least 7 members of the S group are termed 'S group-specific' and are listed in the above table. In all cases where only 7 of the 8 S group members contained the conserved S group-specific residue at a particular position, it was found that the species which lacked the conserved residue was either the chicken violet opsin or the goldfish ultraviolet.

♦ Gly-82 and Ala-120 are found in all the S pigments except the chicken violet, which contains residues that are also found in 24 other non-S group opsins at these positions (Figure 3.8).

Trp-24, Val-109, Phe-130, Val-167, Gly-195, Thr-196, Leu-218, Met-273, His-281 and Gly-282 are found in all the S pigments except the goldfish ultraviolet, which contains residues that are also found in 24 other non-S group opsins at these positions (Figure 3.8).

Residues indicated in bold are goldfish ultraviolet opsin-specific amino acids, which are also at positions containing S group-specific amino acids identified in the Table above. For example, Ser/**Thr**-200 indicates that Ser-200 is conserved in all members of the S group except the goldfish ultraviolet, which contains (**Thr**-200) an amino acid which is exclusive to a member of the S group at this position (indicated in blue in Figure 3.8).

The other 24 vertebrate opsins which were used for comparison are listed in Table 3.5. The first 20 and last 23 amino acids from the N-terminal and C-terminal ends, respectively were not compared due to the high degree of heterogeneity between these regions.

conserved in 7 and all members of the S group, respectively. Thus, Ala-120 and Cys-299 are S group specific residues, based on the above criteria (Figure 3.8 and Table 3.6).

There are no S group-specific residues situated within any of the cytoplasmic loops (Table 3.6). In contrast, there are S group-specific amino acids in the extracellular loops (Table 3.6). Some of the S group-specific amino acids within the extracellular loops involve non-conservative substitutions compared to the residues at the same site in other opsins. For example, the S group-specific residue Ser-200, found in extracellular loop II, contains a hydroxyl group side chain, whereas non-charged residues Asn-200 and Val-200 are found in the Rh/M1/M2 and the L group of visual pigments, respectively (see Section 1.14 for definition of Rh, M1, M2 and L groups). Amino acid changes within the extracellular loops between groups of pigments resulting in loss/gain of a hydroxyl or a charge group may be involved in the spectral tuning of the visual pigment (see Section 3.2.53).

3.2.5 Identification of spectral tuning sites

Some of the S group-specific amino acids identified in Table 3.6, and other residues which are indicated in blue lettering in Figure 3.8, may be involved in the regulation of wavelength of absorption. Of particular interest are residues which may cause a shift from the 530-540nm range, corresponding to the λ_{max} for green cone pigments, down to the 360-440nm range corresponding to the λ_{max} of the S group of pigments.

A three step approach was used to identify amino acids which were more likely to be involved in producing spectral shifts within the S group of pigments. First, S group-specific amino acids situated within the α-helical transmembrane regions were identified (Figure 3.8 and Table 3.6). Second, these α-helical transmembrane residues were compared with the amino acids in other vertebrate visual pigments (listed in Table 3.5). From this comparison, non-conservative amino acid substitutions which involve a gain or a loss of either a hydroxyl or a charged side chain were noted. Third, the relative position of each substitution within the transmembrane regions was determined using a model of the opsin protein derived by Baldwin (1993; see Figure 3.1). Table 3.7 includes a list of the amino acids which were compared (identified using the above criteria) and indicates the relative position of these residues on Baldwin's model. This three step approach allowed for the identification of the non-

Helix	*Amino acid site	Human Opsins RCP	GCP	Rod	S Group Opsins Hum	Tal	Mar	Por	Bov	Mur	Ch	Gf	∞Position in helix
I	40	T	T	L	Q	Q	Q	Q	Q	Q	Q	Q	3 †
	45	I	I	F	G	G	G	G	G	G	G	G	8
II	82	A	A	A	G	G	G	G	G	G	S	G	13
	83	D	D	D	G	G	G	G	G	G	G	G	14
	84	L	L	L	F	F	F	F	F	F	F	F	15
	86	E	E	M	L	L	L	Y	Y	F	S	F	17
	87	T	T	V	C	C	C	C	C	C	C	D	18
	90	A	A	G	S	S	S	S	S	S	S	S	21
	91	S	S	F	V	V	V	V	V	V	V	V	22
	94	S	S	T	V	V	V	V	V	V	V	V	25 †
III	117	V	V	A	G	G	G	G	G	G	G	G	7
	120	C	C	G	A	A	A	A	A	A	G	A	10
	122	I	I	E	L	L	L	L	L	L	L	L	12
	124	G	G	A	T	T	T	T	T	T	T	T	14
	125	L	L	L	G	G	G	G	G	G	G	G	15
	130	I	I	V	F	F	F	F	F	F	F	V	20
IV	163	W	W	M	I	I	I	I	I	I	I	I	13
	164	S	A	A	G	G	G	G	G	G	G	G	14
	167	W	W	C	V	V	V	V	V	V	V	C	17
	172	I	I	L	F	F	F	F	F	F	F	F	22
V	204	M	M	V	T	T	T	T	T	T	T	T	3 †
	205	I	I	I	W	W	W	W	W	W	W	Y	4 †
	214	I	T	I	V	V	I	V	V	V	V	M	13
	217	A	S	I	S	S	A	A	S	S	S	M	16
	218	I	I	I	L	L	L	L	L	L	L	I	17
	220	M	V	F	C	C	C	C	C	C	I	T	19
	224	L	L	G	T	T	A	S	S	S	S	S	23 †

Helix	*Amino acid site	Human Opsins RCP	GCP	Rod	S Group Opsins Hum	Tal	Mar	Por	Bov	Mur	Ch	Gf	∞Position in helix
VI	259	F	L	I	G	G	G	G	G	G	G	G	10
	260	A	A	A	S	S	S	S	S	S	S	S	11
	261	Y	F	F	F	F	F	F	F	F	F	F	12
	265	W	W	W	Y	Y	Y	Y	Y	Y	Y	Y	16
	269	T	A	A	A	A	A	A	A	A	A	A	20
	273	C	C	F	M	M	M	M	M	M	M	L	24 †
VII	288	A	A	M	V	V	V	V	V	V	V	V	3 †
	289	A	A	T	T	T	T	T	T	T	T	A	4 †
	292	A	A	A	S	A	A	A	A	A	A	S	7
	293	Y	F	F	F	F	F	F	F	F	F	L	8
	299	T	T	A	C	C	C	C	C	C	C	C	14
	307	V	V	I	C	C	C	C	C	C	C	A	22

Table 3.7 A comparison of S group-specific residues situated within the seven α-helical transmembrane regions, (indicated in blue) with amino acids found at the equivalent sites within human rod opsin and human red and green cone pigments.

Key: Hum = human; Tal = talapoin monkey; Mar = marmoset; Por = porcine; Bov = bovine; Mur = murine; Ch = chicken; Gf = Goldfish; RCP = Red cone pigment; GCP = green cone pigment.

*Underlined amino acid sites are positions which are known or predicted to cause spectral shifts from previous studies (see Table 3.1 and 3.2). Amino acids at these positions are also compared between the above opsins.

∞Position in helix indicates the position of that residue within the α-helical transmembrane region based on the model proposed by Baldwin (1993; see Figure 3.1). Underlined positions of residues in helix indicate which amino acids clearly face into the interior of the chromophore binding-pocket as proposed by Baldwin.

† beside a particular position in helix indicates that although this residue is one of the 26 amino acids situated within the helix, it is not embedded in the cell membrane (see Figure 3.1).

Residues shown in red are conserved in all the other vertebrate opsins listed in Table 3.5.

conservative substitutions which face into the interior of the chromophore binding pocket (see Section 3.2.51).

Other models of opsin structure have implicated a number of sites which are presumed to have direct contact with the chromophore (Kosower, 1989; Alkorta and Du, 1994; see Section 3.1.4 and Table 3.2). Any residues identified as being exclusive to members of the S group and that are also situated at these predicted sites are included in Table 3.7. Residues identified as being exclusive to members of the S group of pigments which also matched to sites known to be responsible for the spectral difference between the human red and green cone pigments (see Section 3.1.3 and Table 3.1 for these sites) are also included in Table 3.7. The potential spectral tuning sites identified using the above criteria are described below.

3.2.51 S group-specific spectral tuning sites:

(i) Non-conserved amino acid substitutions

Four potential S group-specific spectral tuning sites involving non-conservative amino acid substitutions were identified, which should shift the λ_{max} of all the S group members. These sites are Ser-90, Thr-124, Tyr-265 and Cys-299 (Table 3.7). These residues were identified using the criteria given in Section 3.25. These four S group-specific amino acids are proposed as the residues which contribute to the large shift of λ_{max} from the 500nm range down to the λ_{max} observed in the S group members.

The four amino acids at these potential spectral tuning sites, Ser-90, Thr-124, Tyr-265 and Cys-299, are conserved in all the members of the S group analysed in this study (Table 3.7). Three of these S group-specific residues involve a gain of a hydroxyl group in the S group of pigments when compared with other vertebrate opsins, namely Ser-90, Thr-124 and Tyr-265. Ser-90, situated within transmembrane helix II, is replaced by Ala-90 in the L group pigments whilst Gly-90 is found in every other pigment. Tyr-265 is replaced by Trp-265 in all other vertebrate pigments and the sulphur containing S group-specific residue Cys-299 is replaced by either a hydroxyl bearing or a non-polar residue in all the other pigments (Table 3.7). Amino acids at site 124 and site 265 are predicted to have close contact with the chromophore (Alkorta and Du, 1994; Chang *et al.*, 1995). Ser-90 and Cys-299 were recently, in an independent study, predicted to cause blue shifts in λ_{max} (Hisatomi *et al.*, 1996). Thus, the four S group-specific residues presented as putative spectral tuning sites are supported by other data.

(ii) Conservative amino acid substitutions

The substitution of amino acids with conserved chemical properties at sites which interact with the chromophore, may still alter the λ_{max} of a pigment (see Section 3.1.3). Three residues identified as S group-specific, Gly-117, Leu-122 and Gly-125, represent substitutions of a conservative nature when compared with the amino acids at the same sites within other visual pigments (Table 3.7). Gly, Ala, Leu and Val are considered to be non-polar amino acid residues (see appendix). Val-117 is found in the L group of pigments, whereas Ala-117 is present in all the other opsins except the S group. Leu-122 in the S group is considered a conservative substitution for Ile-122, which is found in all the L group pigments. The S group-specific Gly-125 is replaced by Leu-125 in all other vertebrate opsins.

The above three positions, 117, 122 and 125, were found to match to sites that are predicted to have close contact with the chromophore (Table 3.2). In addition, amino acid substitutions at sites 117 (Ala to Phe) and 122 (Glu to Ala, Gln or Asp) have previously been shown to cause spectral shifts (Ridge *et al.*, 1991; Zhukovsky and Oprian, 1989). Amino acid substitutions of a conservative nature at site 122 have been shown to cause spectral shifts (Nakayama and Khorana, 1991; see Section 3.1.3). Thus, the S group-specific residues Gly-117 and Leu-122 are candidates for causing spectral shifts. Although, Gly-125 has not been shown to cause a shift in λ_{max} by experimental evidence, site 125 is predicted to interact directly with the chromophore (Alkorta and Du, 1994). In addition, the model predicted by Baldwin places site 125 facing directly into the chromophore binding-pocket (Table 3.7 and Figure 3.10, position 15 in helix III). On this basis, Gly-125 is presented here as another putative S group-specific spectral tuning site.

3.2.52 Other spectral tuning sites

Are there any other sites within the transmembrane helices which may shift the λ_{max} of members of the S group? As described in Section 3.1.3, substitutions at sites 164, 214, 217, 261, 269 and 293 involving a loss or gain of a hydroxyl group differentiate the human red and green cone pigments (Table 3.1). A comparison of the amino acids at these sites between human RCP, GCP and members of the S group of pigments is included in Table 3.7. Gly-164, Phe-261 and Ala-269, which are conserved in all the S opsins, are not S group-specific residues. At sites 261 and 269 all the S group pigments contain the same residue as that found in the

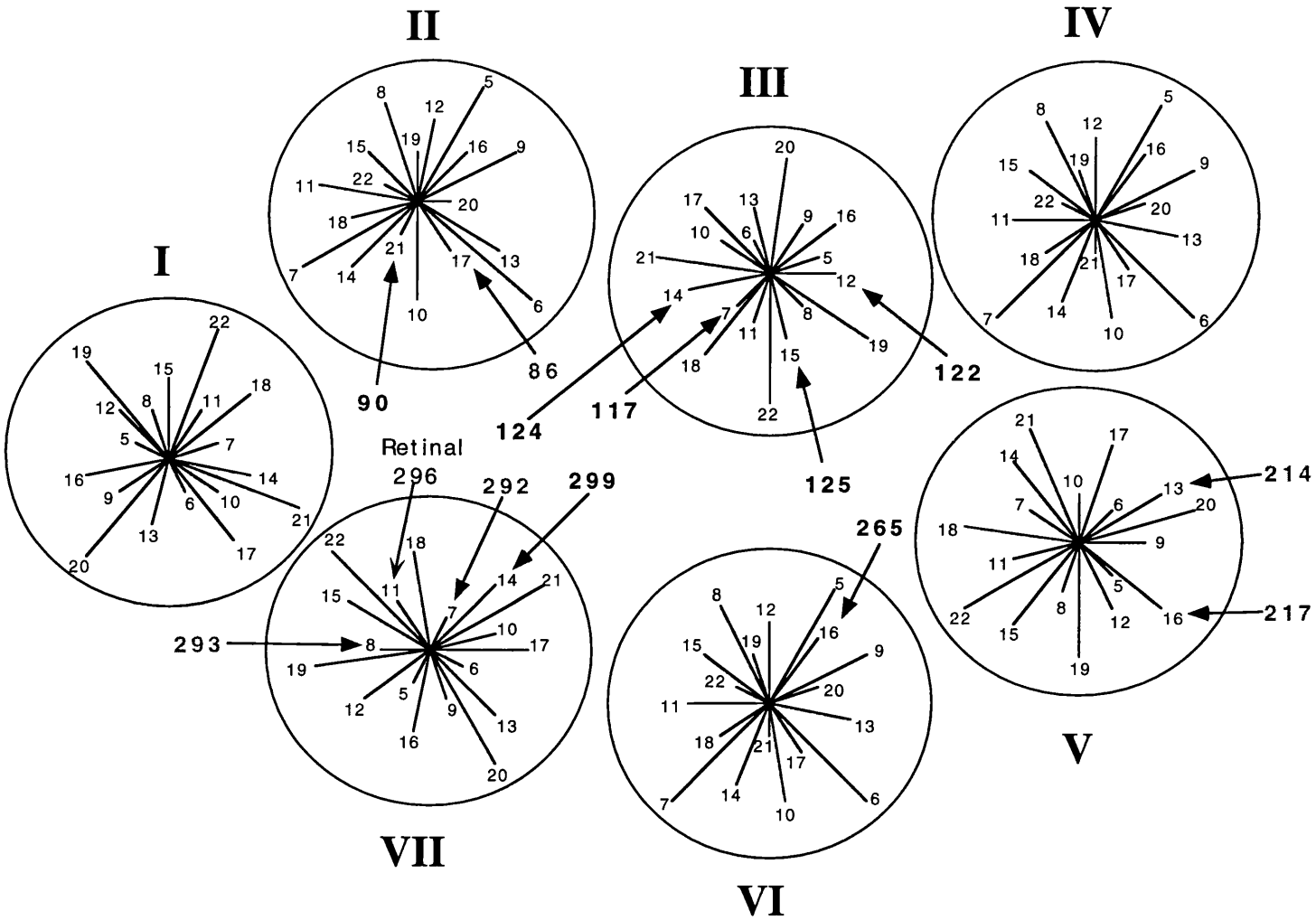


Figure 3.10 Baldwin's model of the helical transmembrane regions. Potential spectral tuning sites identified in this study, situated within the cell membrane bound regions of the transmembrane helices, which may shift the λ_{max} of the S group of pigments are indicated in bold. Potential spectral tuning sites which may fine tune the peak wavelength of absorption between members of the S group are shown in grey.

green cone pigment. Thus, there is no variation of λ_{max} between members of the S group of pigments caused by substitutions at site 164, 261 and 269. The other three sites, 214, 217 and 293, are postulated to cause fine tuning of the λ_{max} between individual members of the S group (described below).

With the exception of the goldfish ultraviolet (UV) pigment, all the other S group pigments contain Phe-293, which is identical to the residue found in the human GCP (Table 3.7). Again with the exception of the goldfish UV pigment, all the other S group of pigments contain a conserved (Ile-214) or a conservative residue (Val-214) to the one found in the human RCP at site-214. The goldfish ultraviolet pigment contains Leu-293 and Met-214, and these residues may fine tune the λ_{max} of the goldfish UV pigment within the S group. The marmoset and porcine BCP and the human RCP contain Ala-217, whereas the goldfish UV pigment contains Met-217 and the other S group members and the human GCP contain Ser-217 (Table 3.7). Hence, substitutions at site 217 may also act to fine tune the λ_{max} between individual members of the S group.

Site 292 has been postulated to cause spectral shifts (Chang *et al.*, 1995, Hunt *et al.*, 1996). Within the S group, the hydroxyl bearing Ser-292 is only present in the human BCP and the goldfish UV pigment (Figure 3.8 and Table 3.7). Outside of the S group Ser-292 is found only in chicken, goldfish and blind cave fish BCPs (M1 group pigments, see Table 3.5; Yokoyama, 1994). All other vertebrate pigments analysed in this study contain Ala-292, which results in the loss of a hydroxyl group in these opsins. Site 292, situated within helix VII, faces directly into the chromophore binding pocket according to Baldwin's model (Figure 3.10, position 7 in helix VII). It seems that Ser-292 may be responsible for a spectral shift in two members of the S group and all members of the M1 group.

3.2.53 Spectral tuning sites in extracellular loops

Are there sites which are not embedded within the transmembrane helices that may influence the λ_{max} and thus contribute to the blue shift? Amino acids at site 100 are situated within extracellular loop I. A 4nm blue shift is caused when Ser-100, the amino acid found in the human red cone pigment (RCP), is substituted for Tyr-100, the residue found in the human green cone pigment (GCP) (Asenjo *et al.*, 1994). Thus, there is evidence to suggest that residues within the extracellular loops may influence λ_{max}. The S group pigments contain either uncharged (Asn),

charged (His, Lys or Arg) or hydroxyl bearing (Tyr) residues at site-100 (Figure 3.8). These non-conservative substitutions at site 100 may result in fine tuning of the λ_{max} between individual members of the S group.

The S group-specific residues Ser/Thr-200, situated within extracellular loop II, are the only hydroxyl bearing amino acids at this position (Table 3.6 and Table 3.7). All the L group of pigments contain the non-polar Val-200 (except the gecko GCP which contains the sulphur carrying Cys-200). All the other vertebrate visual pigment groups (Rh/M1/M2) contain the uncharged Asn-200. Thus, Ser/Thr-200, within extracellular loop II, may blue shift the λ_{max} of all members of the S group.

Thr-204 and Trp/Tyr-205 are S group-specific residues (Table 3.6). Thr-204 is the only hydroxyl bearing residue at this site. Met-204 is found in all the L group of pigments and Val-204 is present in all the Rh/M1/M2 opsins. Tyr- and Trp-205 are non-conservative substitutes for Val-, Leu-, Ile-, and Met-205 found in all the other visual pigments. However, site 204 and site 205 on Baldwin's model are situated at positions 3 and 4 within helix V, respectively (Table 3.7). According to Baldwin's model residues at positions 3 and 4 of a helix are not situated within the cell membrane (Section 3.1.4). Since the S group-specific residues Thr-204 and Trp-205 are found at the junction of extracellular loop II and helix V it is difficult to predict the orientation of these residues (Figure 3.1). However, the conserved nature of these residues at these sites suggests they play a role in maintaining the function of the group of pigments with a low λ_{max}. Thus, Thr-204 and Trp/Tyr-205 are potential residues which may contribute to the short-wavelength λ_{max} of the S group of pigments.

3.2.54 Ultraviolet spectral tuning sites

The λ_{max} of the goldfish UV pigment is between 355-360nm (Bowmaker *et al.*, 1991b). ERG and behavioural studies on mice suggest that the mouse BCP absorbs maximally within the 360-370nm range (Jacobs *et al.*, 1991). Thus, residues within these two UV sensitive proteins which maintain their very short-wavelength λ_{max} may be conserved, between the two pigments.

A comparison of mouse BCP-specific, goldfish UV-specific and other S opsin-specific residues (identified from Figure 3.8) is given in Table 3.8. Ile-57 and Phe-86 are the only residues which are present in both the mouse BCP and the goldfish UV but are absent from the other S group members and all the other vertebrate pigments. Site 57 is at position 20 within helix one on Baldwin's model, which is situated facing away from

Amino Acid Position	Residues which are exclusive to the S group of opsins Primate, porcine, bovine & chicken	Goldfish Ultraviolet	Mouse BCP
21	Val/Pro	Ile	Val
57	Thr/Val	**Ile**	**Ile**
86	Tyr/Leu/Ser	**Phe**	**Phe**
87	Cys	Asp	Cys
194	Val	Lys	Val
200	Ser	Thr	Ser
220	Cys	Thr	Cys
205	Trp	Tyr	Trp
224	Ser/Thr/Ala	**Ser**	**Ser**
277	Asn	Tyr	Asn
280	Asn/Asp	Asp	Asn
283	Leu/Val	Lys	Leu
307	Cys	Ala	Cys

Table 3.8 A comparison of S opsin-specific residues found at the equivalent position between the mouse BCP and the goldfish ultraviolet pigment with the other members of the S group. Residues conserved in both the mouse and goldfish pigments are shown in bold. Sites identified as potential spectral tuning sites in this study are underlined (Sections 3.2.51-3.2.53.

the chromophore binding pocket, thus is discounted as a potential spectral tuning site (Figure 3.10). However, amino acids at site 86 are predicted to have close contact with the chromophore, thus are likely to affect the λ_{max} (Kosower, 1989). In addition, site 86, situated in helix II, is also predicted to face directly into the chromophore binding pocket by Baldwin's model (Figure 3.10; position 17, helix II). The other S group members contain either the non-polar Leu-86 or a hydroxyl bearing Ser/Tyr-86 (Table 3.8 and Figure 3.8). The sulphur containing Met-86 is found in all the Rh and M2 pigments, while the charged residue Glu-86 is present in all the L group of opsins. The M1 group of pigments contain either Thr-86 or Val-86. Thus, all the other visual pigments with a λ_{max} higher than the UV range possess non-conservative substitutions for Phe-86. Hence, Phe-86 is presented as a potential spectral tuning site which may contribute to shift the $\lambda_{max,}$ of members of the S group, into the ultraviolet region of the spectrum. The amino acid positions identified as putative spectral tuning sites in this study are listed in Table 3.9a and b.

3.3 Discussion

3.3.1 The porcine BCP: a new S group opsin

Sequencing and PCR analysis of the genomic cosmid isolated from a porcine library confirmed this product to be a gene encoding for a BCP. This is the first porcine cone pigment gene to be sequenced. Analysis of the porcine BCP amino acid sequence, predicted from its gene sequence, determined that this opsin is a new member of the S group of visual pigments. This is based on the high degree of amino acid identity of the porcine BCP protein sequence to other S group members (Table 3.3). The higher degree of amino acid identity with the bovine sequence (94%) is not too surprising since both the porcine and bovine species belong to the same Order, Artiodactyla (see Section 3.1.6). Prior to this study the murine and bovine BCP genes were the only non-primate mammalian members of the S group characterised (Chiu *et al.*, 1994).

The genomic structure of the porcine BCP gene was found to be very similar to that of the human BCP gene (Nathans *et al.*, 1986a). Both genes contained 5 exons and 4 introns and were about 3.2 kb in length, from their ATG start codon in exon 1 to the stop codon in exon 5. The mRNA transcription initiation site for the porcine BCP gene was predicted to be situated at the same position as that in the human gene (Figure 3.6). Interestingly, this makes the 5′ untranslated region (5′UTR) of the mRNA

a

Amino acid position	Predicted before?	Site-directed Mutagenesis?	
Ser-90	Yes	NO	
Gly-117	Yes	Yes	[1](Ala to Phe => 498 to 490nm)
Leu-122	Yes	Yes	[1](Glu to Asp => 498 to 476nm)
Thr-124	Yes	NO	
Gly-125	Yes	NO	
Ser/Thr-200	NO	NO	
Thr-204	NO	NO	
Trp/Tyr-205	NO	NO	
Tyr-265	Yes	Yes	[1](Trp to Tyr => 498 to 483 nm)
Cys-299	Yes	NO	

b

Amino acid position	Predicted before?	Site-directed Mutagenesis?	
86	Yes	NO	
100	Yes	Yes	[2](Ser to Tyr => 563 to 559nm)
214	Yes	Yes	[2](Ile to Thr => 559 to 555nm)
217	Yes	Yes	[2](Ala to Ser => 563 to 561nm)
292	Yes	Yes	[1](Ala to Asp => 498 to 488nm)
293	Yes	Yes	[2](Tyr to Phe => 563 to 561nm)

Table 3.9 The amino acid sites within the S group of opsins predicted to cause a shift in wavelength of maximum absorption (λ_{max}) according to the criteria presented in this thesis. Sites involving non-conservative substitutions, in spectral tuning terms, are given in plain text. Sites involving conservative substitutions are underlined. Sites situated within extracellular loops are given in bold lettering. Sites situated within the border of a helix are given in italics. Whether these sites have been previously predicted (see Table 3.2) or previously shown to cause shifts by site-directed mutagenesis on bovine rod opsin ([1]Nakayama and Khorana, 1991) or the human RCP and GCP ([2]Asenjo *et al.*, 1994) is indicated.

a: Spectral tuning sites that shift λ_{max} of the S group of pigments as a whole Sections 3.2.51 and 3.2.53);

b: Spectral tuning sites that shift λ_{max} of individual members of the S group (see Sections 3.2.52 and 3.2.54).

for both species of BCP gene only 7 bp long. The length of the 5'UTR is very short considering that the average size of a 5'UTR in a human gene is about 100 bp (Strachan and Read, 1996). However, the predicted polyadenylation signal in the porcine gene was found to be 30 bp further downstream than the polyadenylation signal in the human gene (Figure 3.6). This variation in the position of the polyadenylation site presumably arises in a longer 3' untranslated region of the mRNA transcript for the porcine BCP compared to the human BCP gene. The coding region of the porcine BCP gene was calculated to share a 87.7% identity with the human BCP gene, at the nucleotide level.

3.3.2 Analysis of S group proteins

The alignment presented in Figure 3.8 is the first comprehensive comparison of all members of the S group of visual pigments, for which sequences are now available. A comparison of this alignment with other vertebrate opsins has allowed the identification of amino acids that are unique to the group of pigments which have a λ_{max} lower than 440 nm. Other comparisons of vertebrate opsins have been made to evaluate the evolution of the S group of pigments and to identify spectral tuning sites (Yokoyama, 1994; Chang *et al.*, 1995). Yokoyama (1994) compared 32 visual pigment amino acid sequences, 26 from vertebrates and 6 from invertebrates, to evaluate the evolution of the S group of pigments. A similar alignment of vertebrate and invertebrate opsins plus other G-proteins was produced to predict residues which may be involved in blue shifts of λ_{max} (Chang *et al.*, 1995). However, both of these alignments only utilised two members of the S group, namely the human BCP and the chicken violet.

The following observations arise from the alignment of the amino acid sequences of the S group members (Figure 3.8). Within the S group, the porcine BCP possesses the longest protein. The extra 3 residues in the porcine protein, relative to the human BCP, are accounted for by the insertion of 3 glutamic acid residues within the N-terminal region of the protein. This insertion results in a continuous stretch of 6 Glu in the porcine BCP compared to only 3 Glu in the human BCP. The bovine and non-human primate BCP proteins contain a continuous stretch of 4 Glu residues at the same site. Thus, there seems to be a variation in the number of continuous Glu residues at the N-terminal region of the BCP protein, within and between artiodactyl and primate species. This difference in the number of Glu residues between the human and the

other primates probably originated after the split of the Old World primates into the Hominoid and the Cercopithecoid branches (Time 2 in Figure 3.9), about 14 million years ago. It would be interesting to determine the number of Glu residues at this position within the BCP proteins of other Hominoid primates and other mammalian species. This would show whether 3 Glu residues is a human-specific feature and the extent of variation of the number of Glu residues between species. Alternatively, the number of Glu residues may be polymorphic within species.

The high conservation of amino acids observed within extracellular loop II and the cytoplasmic loops II and III, between the S group pigments, is not surprising since mutation of residues within these loops are deleterious to the function of the opsin (Figure 3.8). These cytoplasmic loops are required for the binding and activation of transducin (Konig *et al.*, 1989). Mutations within the cytoplasmic loops, particularly loop III, are known to affect the binding and activation of transducin severely (Franke *et al.*, 1988 and 1992). The extracellular loops are involved in opsin transport from the endoplasmic reticulum to the cell membrane, as well as chromophore binding (Khorana, 1992). The importance of these loops is reflected in the high number of residues which are conserved between all vertebrate opsins within these loops (indicated in red in Figure 3.8).

It was interesting to note the disparity of S group-specific residues between the extracellular and cytoplasmic loops (see Section 3.2.42). The absence of S group-specific residues from all the cytoplasmic loops contrasts with the relatively high number found within extracellular loops II and III (Table 3.6). This, suggests that the individuality of a visual pigment may be partially determined by the presence of certain residues within the extracellular loops.

Residues found to be exclusive to individual members of the S group of pigments (shown in blue in Figure 3.8) which result in non-conservative substitutions when compared with amino acids at the same site in other opsins were noted (Section 3.2.41). Whether or not the non-conservative residues Lys-22 (talapoin BCP-specific) or Arg-324 (murine BCP-specific) have an effect on the normal function of the visual pigment is difficult to predict without experimental evidence. It is prudent to assume that a non-conservative substitution at a site that is conserved in every other vertebrate opsin is likely to cause an alteration of function, as is predicted for the marmoset BCP-specific residue Ala-180 (Figure 3.8). Pro-180, situated in extracellular loop II, is present in all the other

vertebrate opsins compared in this study. Pro-180 is also found in the four opsins (Rh1-4) of the fruit fly (*Drosophila melanogaster*) and in the opsin of octopus (*Paroctopus defleini*) (Yokoyama, 1994). This degree of conservation suggests that Pro-180 has an essential role within the opsin molecule. A mutation within the human rod opsin gene that changes Pro-180 to Ala-180 results in dominant retinitis pigmentosa, a form of retinal degeneration leading to blindness (Daiger *et al.*, 1995). A Pro-180-Ala mutation has not been recorded for the human BCP protein, thus the actual physiological effect of this substitution is not certain. Such an alteration may not lead to complete blindness but may disrupt the functioning of the blue cone cell, similar to the affects seen in tritanopia. Human tritanopia is an autosomal dominant colour-vision disorder which is characterised by a defect in the BCP. One alteration causing this defect is a Pro-267 to Ser-267 substitution (Weitz *et al.*, 1992). Pro-267, situated in helix VI, is conserved in all vertebrate opsins and invertebrate opsins sequenced to date. However, an amino acid substitution which causes retinal degeneration in human rod opsin does not necessarily result in the same phenotype within a S group pigment. For example, a Phe-220 to Cys-220 mutation in human rod opsin results in dominant retinitis pigmentosa (Bunge *et al.*, 1993), but Cys-220 is found in all the mammalian members of the S group (see Figure 3.8). Similarly, with the exception of the goldfish UV pigment, which contains Ala-186, the amino acid Ser-186 is conserved in all the vertebrate pigments (see Figure 3.8). Mutation of this conserved Ser-186 in human rod opsin is known to result in dominant retinitis pigmentosa (Dryja *et al.*, 1991). Thus, Ala-180 in the marmoset BCP may give this opsin distinct properties which is not seen in other vertebrate pigments.

3.3.3 Evaluation of the putative spectral tuning sites

The shift from the λ_{max} range of the L group (about 530-560nm) to the S group λ_{max} range is more than 100nm. Mutagenesis studies and analysis of opsins from natural populations indicate that non-conservative amino acid substitutions result in small shifts of wavelength of absorbance, between 2-15nm, to fine tune λ_{max} (Asenjo *et al.*, 1994; Hunt *et al.*, 1996). There are only a limited number of spectral tuning sites that have been identified using the criteria that a non-conserved substitution within the transmembrane region which faces into the binding-pocket results in a spectral shift. In this study only 4 sites have been identified that may shift the λ_{max} of all members of the S-group, using the above

criteria (Table 3.9a). The number of potential spectral tuning sites identified seem too few to account for the large shift from the L group λ_{max} range to the S group λ_{max} range. These 4 sites, Ser-90, Thr-124, Tyr-265 and Cys-299, together would need to blue shift the absorbance by at least a 100 nm to reach the λ_{max} range of the S group from the 530-560nm range.

Chemically, substitutions involving the non-polar residues Gly, Ala, Val, Leu and Ile are considered to be of a conservative nature in spectral tuning terms. However, substitution of conservative residues at site 122, which faces towards the chromophore binding pocket, has previously been shown to cause shifts in λ_{max} (Nakayama and Khorana, 1991; Ridge *et al.*, 1992; Section 3.1.3). Using the above criteria the S group-specific residues Gly-117, Leu-122 and Gly-125 were identified, which may also effect spectral tuning (see Section 3.2.51 and Table 3.9a). Without further experimental evidence, such as site-directed mutagenesis analysis, it is not possible to state for certain whether these predicted spectral tuning sites cause shifts. Thus, our knowledge of which residues alter opsin function, in spectral tuning terms, is far from complete.

It was very interesting to note the presence of Phe-86, a murine BCP-specific and goldfish UV pigment-specific residue (Figure 3.8). The amino acid phenylalanine possesses an aromatic ring (see appendix for properties of amino acids). An aromatic ring is associated with π-electron clouds, which gives the amino acid unique properties (Stryer, 1988). This π-electron cloud influences the transfer of electrons by interacting with other π-systems. Thus, this π-electron cloud, associated with phenylalanine, can influence the π-electron system which is associated with the chromophore of visual pigments thereby affecting the λ_{max} (Mathies and Stryer, 1976). It is possible that Phe-86 which is found only within the mouse BCP and goldfish UV pigment may contribute to the shift of the λ_{max} to the ultraviolet region of the spectrum.

Experimental evidence is required to show whether the S group-specific amino acids at sites 200, 204 and 205, identified as potential spectral tuning sites in this study, are involved in causing shifts in λ_{max}. Ser/Thr-200, situated within extracellular loop II, was proposed as a spectral tuning site on the basis that amino acid substitutions at site 100, within extracellular loop I, have been shown to cause spectral shifts between the human RCP and GCP (Asenjo *et al.*, 1994; see Section 3.2.53). Baldwin's model cannot predict whether any of the amino acids within the extracellular loops or at the junction of the transmembrane helices, such

as sites 204 and 205, interact with the chromophore. It should be noted that residues at sites 214, 217, 269 and 293, which contribute to the spectral difference observed between the human RCP and GCP, do not clearly face into the chromophore binding pocket according to Baldwin's model (see Table 3.7 and Figure 3.10). There are other S group-specific residues that do not face into the chromophore binding pocket (such as Phe-84, Val-91, Phe-130, Phe-172 and Ser-260, see Table 3.7), which are of a non-conservative nature compared to the amino acids situated at the same position within other opsins. Thus, these S group-specific amino acids may also play a role in maintaining the short λ_{max} observed in the S pigments. However, without experimental evidence, or an established 3-dimensional crystallised structure of opsin, to determine which of the other S group-specific amino acids within the transmembrane regions, or the extracellular loops, interact with the chromophore it is not possible to confidently choose other potential spectral tuning sites.

3.4 Summary

The porcine blue cone pigment (BCP) gene has been sequenced and its genomic structure characterised. Comparison of the deduced amino acid sequence for the porcine BCP with other opsins indicated that this opsin is a new member of the S group of visual pigments. Analysis of the S group of pigments has identified amino acids that are specific to this family. 10 potential spectral tuning sites, containing S group-specific residues, are presented as the ones which blue shift the λ_{max} of the S group of pigments. Of these 10 sites, 3 are newly identified sites, 4 support previous predictions and the remaining 3 have been shown previously to cause shifts by site-directed mutagenesis on bovine rod opsin (summarised in Table 3.9a). Six other sites are also postulated to fine tune spectral shifts between individual members of the S group (Table 3.9b). Five of these sites are supported by previous site-directed mutagenesis studies, either on the human RCP and GCP or bovine rod opsin. The validation of the potential spectral tuning sites identified in this study requires site-directed mutagenesis with subsequent expression and measurement of absorption spectra of the mutant opsins.

Without a proper 3D crystallised structure of opsin that shows exactly the spatial arrangement of each amino acid and its side chain, it is difficult to truly state with confidence which residues interact directly or indirectly with the chromophore. It is unwise to rely on only one model

of opsin to propose which residues are involved in spectral tuning or to assume that only non-conserved amino acid substitutions affect the λ_{max}.

INSTITUTE OF OPHTHALMOLOGY
LIBRARY

Chapter 4

Expression of the Human Blue Cone Pigment Gene in the *Pichia pastoris* Yeast Expression System

4.1 Introduction

As an alternative to cultured mammalian cells, the use of the yeast (*Pichia pastoris*) as an *in vitro* expression system for opsins, specifically for the human BCP, was tested in this study. Successful expression of the human BCP gene would allow quick and easy isolation of high concentrations of opsin. This yeast system could then be utilised for the expression of site-directed mutated BCP genes, to study which specific amino acids within the opsin are involved in wavelength modulation.

4.1.1 *In vitro* expression of opsin genes

Why do we need to artificially express opsin genes? Most of the data obtained about the structural and functional properties of visual pigments has been obtained from analysing artificially expressed wild-type and mutant forms of opsin. For instance, the identification of the counterion Glu-113 (Sakmar *et al.*, 1989), Cys-110 and Cys-187 which form an essential disulphide bridge (Karnik and Khorana, 1990), and amino acids which interact with the chromophore (Nakayama and Khorana, 1991) all utilised rod opsin which had been manufactured and isolated from an expression system. Mammalian tissue culture cells, transfected with cloned cone opsin cDNA, has allowed the direct determination of the absorption spectra of individual cone pigments which were produced and purified from these artificial expression systems (Oprian *et al.*, 1991; Merbs and Nathans, 1992; Johnson *et al.*, 1993). Analysis of hybrid forms of the human red and green cone pigments, expressed *in vitro*, permitted the identification of the exact residues which result in the spectral difference observed between these two opsins (Asenjo *et al.*, 1994). Thus, the availability of an expression system which allows the production of mutant forms of opsin is an essential tool for visual pigment scientists.

All the experiments mentioned above utilised embryonic monkey or human kidney cell (293S) cultures, the typical mammalian system, for the expression of transfected rod or cone opsin genes. Although the use of mammalian cell lines, to express opsin, is an established protocol which works, this system is a time consuming procedure which results in the

production of only a relatively low concentrations of opsin protein (100-200 μg of bovine rod opsin per litre of saturated growth media; Nathans *et al.*, 1989). Mammalian cell lines can be grown from a single transfectant, containing the appropriate gene, which involves a 1 month delay before the cells are ready for isolation of the expressed opsin (Nathans *et al.*, 1989). A transient transfection system for expressing synthetic bovine rod opsin from mammalian cultures has also been developed, where only a brief (72 hours) delay separates transfection and the extraction of the opsin protein (Oprian *et al.*, 1987). An alternative expression system which produces high concentrations of protein, and involves a simple and quick procedure for the isolation of the opsin, would be an ideal improvement of the current mammalian system (mentioned above).

4.1.2 The *Pichia pastoris* yeast expression system

The yeast *Pichia pastoris* has been developed for use as an expression system by Invitrogen® (*Pichia* expression kit manual, 1994). The *P pastoris* expression system is faster, easier and cheaper to use than the mammalian tissue culture method. The yeast can be grown in a similar fashion to *E coli* thus, unlike mammalian cell lines, they do not require manipulation within a safety cabinet or growth in a humidified incubator supplied with a 5% CO_2 atmosphere.

The use of *P pastoris* as an expression system arises from its ability to metabolise methanol as a sole carbon source. The utilisation of methanol, by *P pastoris*, requires the initial conversion of this compound to formaldehyde by the enzyme alcohol oxidase. The expression of the gene that codes for alcohol oxidase (AOX1) is induced by methanol to very high levels, such that when *P pastoris* is grown in a methanol containing environment this enzyme represents greater than 30% of the total soluble protein within the cell. Thus, a foreign gene linked to the AOX1 gene promoter within the *P pastoris* genome can be induced with methanol to produce high levels of heterologous protein.

A list of some of the heterologous proteins which have been expressed using the *P pastoris* system, with the relative concentrations obtained for each product, is given in Table 4.1. An added advantage of the *P pastoris* system is that the protein of interest can be secreted out of the cell if a signal sequence is attached to the expressed peptide. The attached signal sequence acts to target the protein to be transported out of the cell, into the growth media. *P pastoris* secretes very low levels of native protein, thus methanol induced expression of a heterologous

Protein	Expression (g/L)	Intracellular (I) or Secreted (S)
Human serum albumin	4.0	S
Human tumor necrosis factor	10.0	I
Invertase	2.3	S
Bovine lysozyme c2	0.55	S
Tetanus toxin fragment C	12.0	I
Human epidermal growth factor	0.5	S
Mouse epidermal growth factor	0.45	S
Hepatitis B surface antigen	0.4	I

Table 4.1 Examples of heterologous protein products obtained with the *Pichia pastoris* expression system (adapted from *Pichia* expression kit manual, Invitrogen® 1994). The concentration of the protein and whether the secretion method was utilised is indicated.

protein with an attached secretion signal will result in this being the major product found within the surrounding growth media. Hence, secretion of high quantities of the protein of interest provides the initial purification step without the need for further manipulation to isolate the expressed product, such as lysing the cells. The process of cloning the gene of interest, identification of a positive transformed yeast clone and inducement of heterologous protein expression, using the *P pastoris* system is described in Section 2.20.

4.2 Results

4.2.1 Isolation and cloning of a human BCP complementary DNA into a yeast expression vector

A human retinal cDNA library (Stratagene) was used as template for the amplification of a full length BCP cDNA, utilising the PCR primers Ex1+ 410/431F and Ex5- 1453/1433R (as described in Section 2.2.2; see Table 2.1 for sequence of primers). First round PCR, using the Ex1+ and Ex5- primers, resulted in amplification of a faint product of the correct size (Figure 4.1A) which was utilised as a template for second round PCR. The second round PCR product (Figure 4.1B) was cloned into the TA cloning vector (as described in Sections 2.5-2.8). Sequencing of the cloned product verified it as a human BCP cDNA clone and that there were no PCR errors (Figure 4.2; as described in Section 2.12).

Figure 4.3 shows a diagrammatic representation of the yeast expression vector pPIC9 which was utilised in this study. PCR primers, with linkers containing appropriate restriction enzyme cutting sites, were designed for the amplification and subsequent cloning of the full length BCP cDNA into the cloning site of pPIC9 (Table 4.2). An Ex1+ primer with an *EcoR*I linker and an Ex5- primer with a *Not*I linker was used to amplify the human BCP cDNA by PCR (as described in Section 2.2.1). This PCR product was then digested with *EcoR*I and *Not*I before cloning the product into the pPIC9 expression vector, which had been linearised with the same two restriction enzymes (as described in Sections 2.3 and 2.5-2.8). 10 individual pPIC9 clones were confirmed to be containing the BCP cDNA by digestion with *Bam*HI (Figure 4.4). A single *Bam*HI site is present within the human BCP cDNA (Figure 4.2) and a single *Bam*HI is also contained within the pPIC9 vector sequence (Figure 4.5). Digestion of pPIC9 with a BCP cDNA insert with *Bam*HI results in a single fragment of 1045 bp, which is cut out of the pPIC9/BCP clone (Figure 4.4). This 1045 bp

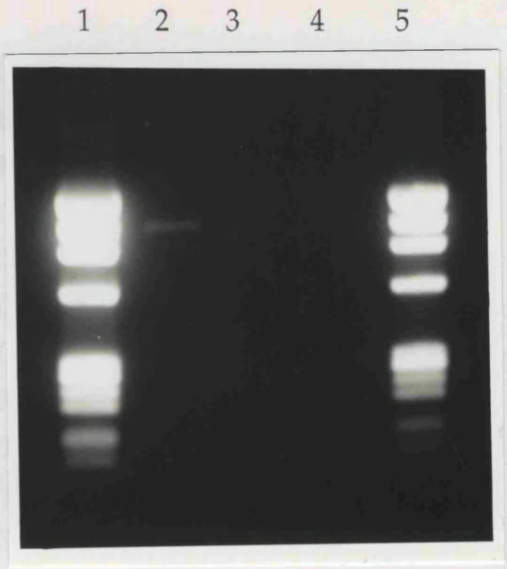


Figure 4.1A First round PCR amplification of a human retinal cDNA library using the BCP-specific primers Ex1+ 410/431F and Ex5- 1453/1433R.
Lanes 1 & 5: ϕX174/*Hae*III DNA ladder
Lane 2: Retinal cDNA template
Lane 3: Retinal cDNA template
Lane 4: No DNA negative control

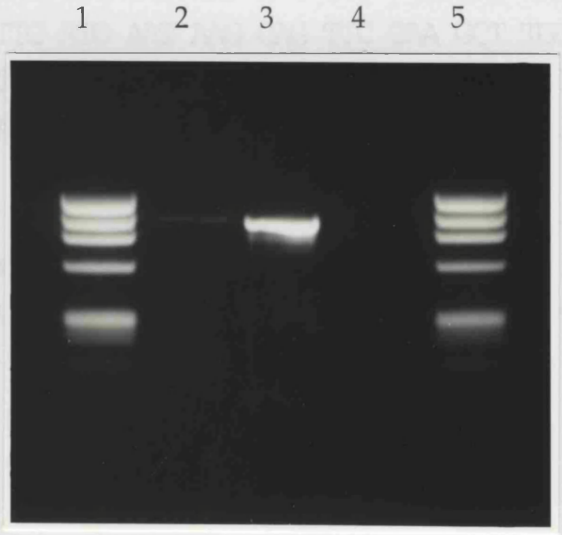


Figure 4.1B Second round PCR amplification of a human retinal cDNA library using the BCP-specific primers Ex1+ 410/431F and Ex5- 1453/1433R.
Lanes 1 & 5: ϕX174/*Hae*III DNA ladder
Lane 2: 1/1000 of the first round PCR product from Lane 3 **4.1A** as template
Lane 3: 1/1000 of the first round PCR product from Lane 2 **4.1A** as template
Lane 4: No DNA negative control

ATG AGA AAA ATG TCG GAG GAA GAG TTT TAT CTG TTC AAA AAT ATC 45
TCT TCA GTG GGG CCG TGG GAT GGG CCT CAG TAC CAC ATT GCC CCT 90
GTC TGG GCC TTC TAC CTC CAG GCA GCT TTC ATG GGC ACT GTC TTC 135
CTT ATA GGG TTC CCA CTC AAT GCC ATG GTG CTG GTG GCC ACA CTG 180
CGC TAC AAA AAG TTG CGG CAG CCC CTC AAC TAC ATT CTG GTC AAC 225
GTG TCC TTC GGA GGC TTC CTC CTC TGC ATC TTC TCT GTC TTC CCT 270
GTC TTC GTC GCC AGC TGT AAC GGA TAC TTC GTC TTC GGT CGC CAT 315
GTT TGT GCT TTG GAG GGC TTC CTG GGC ACT GTA GCA GGT CTG GTT 360
ACA GGA TGG TCA CTG GCC TTC CTG GCC TTT GAG CGC TAC ATT GTC 405
ATC TGT AAG CCC TTC GGC AAC TTC CGC TTC AGC TCC AAG CAT GCA 450
CTG ACG GTG GTC CTG GCT ACC TGG ACC ATT GGT ATT GGC GTC TCC 495
ATC CCA CCC TTC TTT GGC TGG AGC CGG TTC ATC CCT GAG GGC CTG 540
CAG TGT TCC TGT GGC CCT GAC TGG TAC ACC GTG GGC ACC AAA TAC 585
CGC AGC GAG TCC TAT ACG TGG TTC CTC TTC ATC TTC TGC TTC ATT 630
GTG CCT CTC TCC CTC ATC TGC TTC TCC TAC ACT CAG CTG CTG AGG 675
GCC CTG AAA GCT GTT GCA GCT CAG CAG CAG GAG TCA GCT ACG ACC 720
CAG AAG GCT GAA CGG GAG GTG AGC CGC ATG GTG GTT GTG ATG GTA 765
<u>GGA TCC</u> TTC TGT GTC TGC TAC GTG CCC TAC GCG GCC TTC GCC ATG 810
TAC ATG GTC AAC AAC CGT AAC CAT GGG CTG GAC TTA CGG CTT GTC 855
ACC ATT CCT TCA TTC TTC TCC AAG AGT GCT TGC ATC TAC AAT CCC 900
ATC ATC TAC TGC TTC ATG AAT AAG CAG TTC CAA GCT TGC ATC ATG 945
AAG ATG GTG TGT GGG AAG GCC ATG ACA GAT GAA TCC GAC ACA TGC 990
AGC TCC CAG AAA ACA GAA GTT TCT ACT GTC TCG TCT ACC CAA GTT 1035
GGC CCC AAC 1044

Figure 4.2 Sequence of the full length human BCP cDNA amplified from a retinal cDNA library. A single *Bam*HI site which is present within the cDNA is underlined.

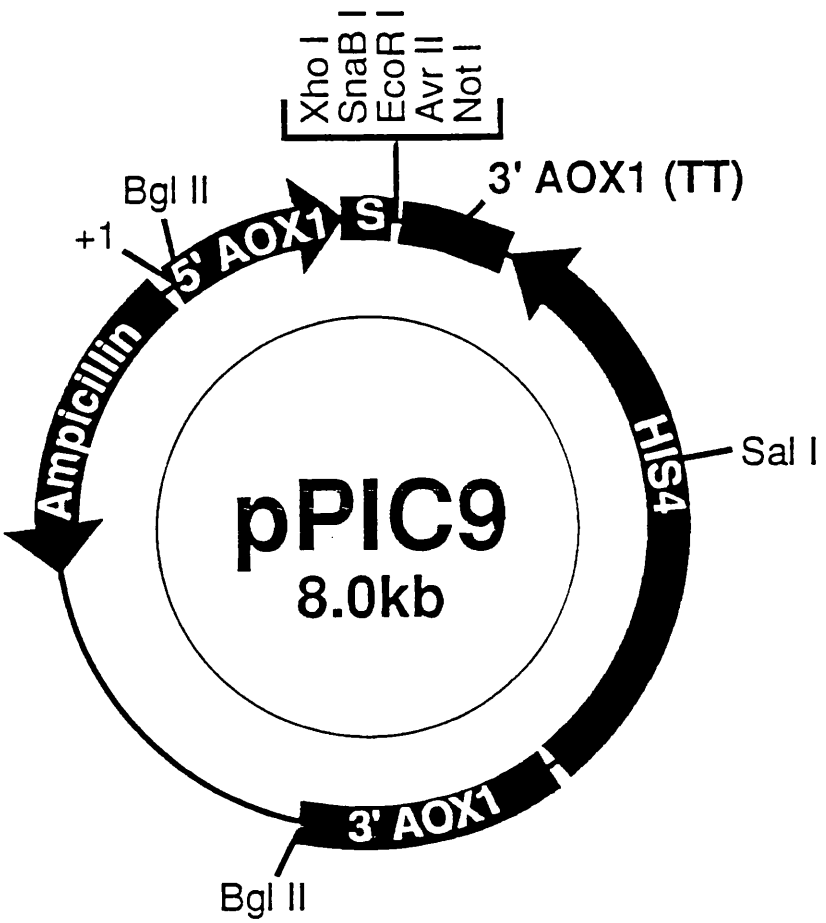


Figure 4.3 pPIC9 the *Pichia pastoris* expression vector.
Vector size: 8034 bp.
5' AOX1, approximately 1000 bp region of the alcohol oxidase promoter fused to the *Saccharomyces cerevisiae* α-factor secretion signal sequence with *Xho* I, *Hind*III, *Sna*BI, *Eco* RI, *Avr*II and *Not*I cloning sites.
3' AOX1, approximately 256 bp of the alcohol oxidase terminating sequence.
P pastoris histidinol dehydrogenase gene, HIS4, contained on a 2.4 kb fragment.
Region of 3' AOX1 DNA about 650 bp in size, which together with the 5' AOX1 region is necessary for site-directed integration.

Primer	Sequence
Ex1+EcoRI	5′ CGCGATCGTAGAATTC**ATGAGAAAAATGTCGGAGGAA** 3′
Ex5-NotI	5′ CGCGCTCTATGCGGCCGC**TCAGTTGGGGCCAACTTGCGT** 3′

Table 4.2 Sequences of the Ex1+ and Ex5- primers with linkers containing the restriction sites for *Eco*RI and *Not*I, respectively. The primer region corresponding to coding sequence of the human BCP cDNA is given in bold and the restriction enzyme recognition sequence is underlined. A non-specific sequence of nucleotides were added to the 5′ end of each primer to act as anchor sites for the restriction enzyme.

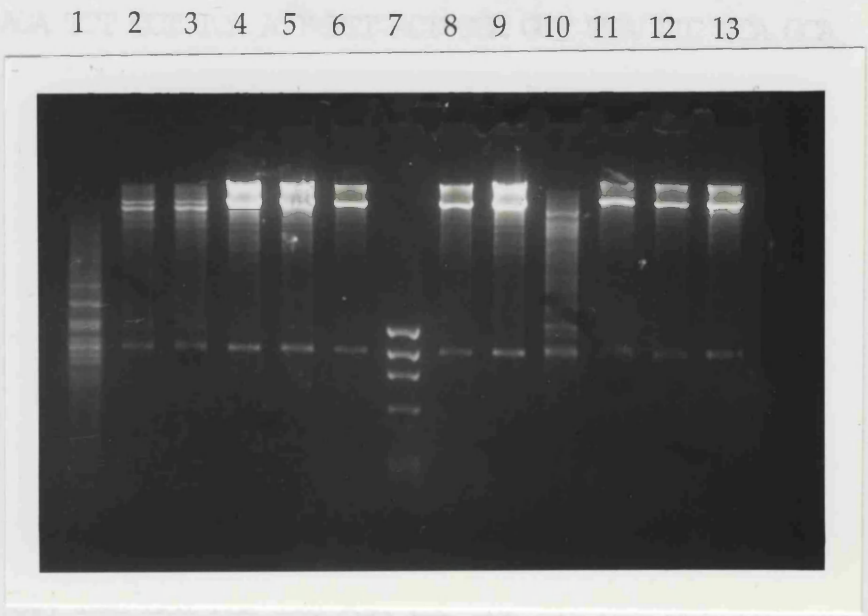


Figure 4.4 *Bam*HI digestion of pPIC9 clones containing the human BCP cDNA.

Lanes 1-6: pPIC9/BCP cDNA clones digested with *Bam*HI

Lane 7: ϕX174/*Hae* III DNA ladder

Lanes 8-13: pPIC9/BCP cDNA clones digested with *Bam*HI

mRNA AOX1 5′ end

5′ACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCATC
ATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAAC

*Bam*HI

GACTTTTAACGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAGGA*TCC* AAA

αF start

CG ATG AGA TTT CCT TCA ATT TTT ACT GCA GTT TTA TTC GCA GCA
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala

TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA
Ser Ser Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu

ACG GCA CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA
Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu

GAA GGG GAT TTC GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA
Glu Gly Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr

AAT AAC GGG TTA TTG TTT ATA AAT ACT ACT ATT GCC AGC ATT GCT
Asn Asn Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala

^

GCT AAA GAA GAA GGG GTA TCT CTC GAG AAA AGA GAG GCT GAA GCT
Ala Lys Glu Glu Gly Val Ser Leu Glu Lys Arg Glu Ala Glu Ala

*Eco*RI *Not*I

TAC GTA GAA TTC CCT AGG GCG GCC GCG AAT TAA
Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn Stop

TTCGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCT
TGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAGGCTTCA
TTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTGTCA 3′

Figure 4.5 The pPIC9 expression vector multi-cloning site with attached αF secretion signal sequence (adapted from *Pichia* expression manual Invitrogen ®, 1994). The *Eco*RI and *Not*I sites utilised as cloning sites in this study are underlined. The *Bam*HI site that was used to check for the presence of the human BCP cDNA insert is shown in italics (see Section 4.2.1).

fragment contains 766 bp of the 5′ end of the BCP cDNA and 279 bp of the expression vector, consisting of the yeast αF secretion signal sequence and 10 bp of the 5′ untranslated end of the AOX1 gene (Figure 4.5). Multiple fragments were obtained for two pPIC9/BCP clones after digestion with *Bam*HI, which were thought to be due to degradation of the vector (Figure 4.4, Lanes 1 and 10). pPIC9/BCP clones which were used for subsequent transformation and expression studies were sequenced to confirm that were no nucleotide errors within the BCP cDNA, which may have been incorporated by the *Taq* polymerase during PCR.

4.2.2 Transformation and selection of yeast recombinants containing the human BCP cDNA

A schematic diagram and an explanation of the procedures involved in transformation and selection of recombinant *P pastoris* colonies is given in Figure 2.2 and Section 2.20. Positive expression vector clones containing the human cDNA (pPIC9/BCP) were digested with the restriction endonuclease *Bgl*II, and confirmed by visualisation on an agarose gel (Figure 4.6). The *Bgl*II-digested 6 kb pPIC9/BCP fragment (Figure 4.6) contains the *P pastoris* histidinol dehydrogenase gene (HIS4) and the BCP cDNA with the 5′ AOX1 gene promoter fused to the *Saccharomyces cerevisiae* α-factor secretion signal sequence (see Figure 2.2).

Spheroplasts of the GS115 *P pastoris*, histidinol dehydrogenase deficient, host strain were prepared (as described in Section 2.20.3). The GS115 spheroplasts were transformed with the *Bgl*II linearised pPIC9/BCP DNA and plated out on to histidine-deficient regeneration-dextrose (RD) agar plates (as described in Section 2.20.4). The histidine positive (H^+) transformants were then screened for clones in which a recombination event had occurred between the 6 kb linearised pPIC9/BCP fragment and the native *P pastoris* AOX1 locus (which were termed H^+/M^- clones, see Figure 2.2). Identification of H^+/M^- clones was achieved by duplicate streaking of individual H^+ transformants on to gridded minimal dextrose (MD) and minimal methanol (MM) plates, and then monitoring the growth of these colonies at 30°C over a period of 2-4 days (as described in section 2.20.5). H^+/M^- transformants, which have displaced the AOX1 gene, cannot produce alcohol oxidase and as a result cannot efficiently metabolise methanol as a carbon source, thus grow poorly on MM media. However, H^+/M^- transformants will grow normally on MD plates. H^+ transformants with an intact AOX1 gene will grow at the same rate on

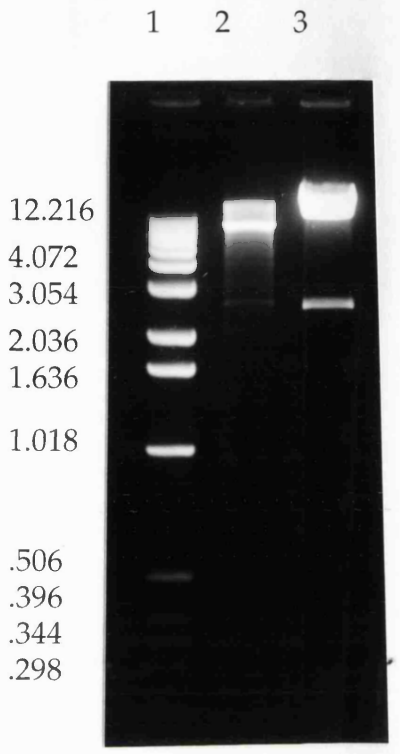


Figure 4.6 Digestion of pPIC9 clones containing the human BCP cDNA with *Bgl*II.

Lane 1: 1 kb DNA ladder

Lane 2-3: pPIC9/BCP cDNA clones digested with *Bgl*II

both MD and MM plates, since they can still utilise methanol as a carbon source. By comparing the growth rates of duplicate H^+ colonies on both MD and MM media it was possible to select for H^+/M^- clones, which grew slower on MD and MM plates (see Figure 2.2, step 7). A total of 513 H^+ transformants were duplicate screened on MD and MM plates, of which 46 clones were found to be H^+/M^-, according to the criteria given above.

4.2.3 Pilot expression studies of H^+/M^- yeast clones

Pilot expression studies, utilising MM media, for 23 of the 46 yeast clones identified as H^+/M^- transformants were carried out (as described in Section 2.20.6). After inducing expression from the AOX1 gene promoter, an aliquot of MM media from each H^+/M^- clone, containing the secreted proteins, was run on an SDS-PAGE for visualisation (as described in Section 2.18.3). The positive control yeast strain (H^+/M^- AlbS) which secretes albumin when grown in MM media was shown to have secreted a 67 kD protein that was clearly visible on a Coomassie Blue stained SDS protein gel (Figures 4.7 A, B and C). A high concentration of a 40 kD secreted protein product was expected for the human BCP, but this was not observed for any of the H^+/M^- clones (Figures 4.7 A, B and C). Similar expression studies on the other 23 remaining H^+/M^- transformants also indicated that these clones were not secreting large quantities of the human BCP (data not shown).

4.2.4 Further analysis of H^+/M^- yeast clones

To test for the presence of the BCP cDNA within the yeast genome, total genomic DNA was isolated from each of the 46 yeast colonies identified as H^+/M^- clones (as described in Section 2.1.2). Each sample of extracted DNA was used as template in a PCR in conjunction with human BCP gene-specific primers (Ex1+ and Ex5-, as described in Section 2.2.1). Only 2 of the 46 H^+/M^- clones (clones 19 and 43) gave a PCR product, suggesting that only two of the transformants contained a human BCP cDNA which was integrated into the yeast genome (Figure 4.8, Lanes 20 and 45). Further studies to optimise expression of the human BCP was concentrated on the H^+/M^- transformants 19 and 43, which gave positive signals with the above PCR.

To optimise heterologous protein expression, the yeast cells should be grown in conical flasks with baffled bottoms, which increases the level of aeration to the growing cells (as recommended by Invitrogen®, 1994). The level of secreted proteins were compared by SDS-PAGE for the H^+/M^-

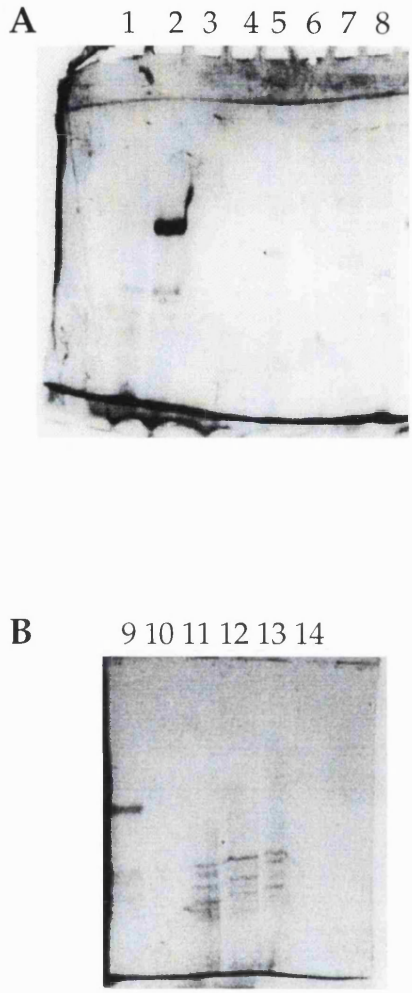


Figure 4.7 A & B SDS-PAGE of secreted proteins from identified H+/M- yeast clones. Lanes 2 and 9 contain the secreted Albumin protein (66 kDa) from the positive control (see Section 4.2.3).

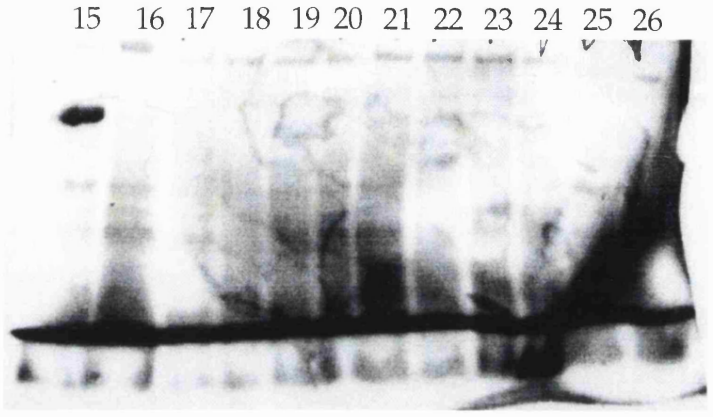


Figure 4.7 C SDS-PAGE of secreted proteins from identified H^+/M^- yeast clones. The high concentration protein in lane 15 is the secreted Albumin positive control (size 66 kDa).

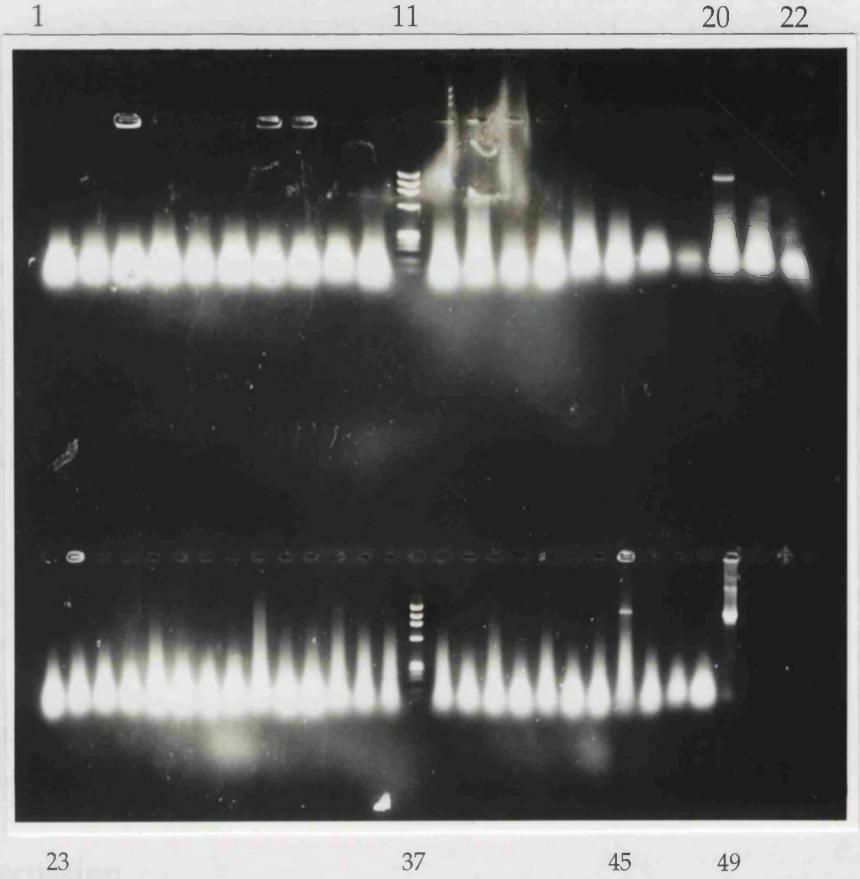


Figure 4.8 PCR of yeast genomic DNA, isolated from *P pastoris* colonies identified as containing the human BCP cDNA. Amplification was performed with the Ex1+ 410/431F and Ex5- 1453/1433R primers to test for the presence of the full-length BCP cDNA within the genome. Lanes 11 and 37 contain the ϕX174/*Hae*III DNA ladder. Lane 49 contains the product for the postive control (Ex1+ 410/431F and Ex5- 1453/1433R PCR on the isolated human BCP gene cloned in Bluescript vector).

clones 19 and 43 after growth in falcon tubes, conical flasks and baffled bottomed conical flasks (Figure 4.9, Lanes 2-7). No large concentration of secreted protein product corresponding in size to the human BCP (40 kD) was observed for any of the growth methods utilised (Figure 4.9). It is possible that because the opsin is a membrane bound protein, the expressed BCP is targeted to the yeast plasma membrane, and thus is not secreted. Crude cell membrane protein preps of the H^+/M^- clones 19 and 43 and the H^+/M^- AlbS yeast were prepared by preparing spheroplasts (as described in Section 2.20.3) and lysing the cell membrane with SDS (as described in Section 2.1.2) and then collecting the pellet. The crude cell membrane extract was visualised by SDS-PAGE (Figure 4.9, Lanes 10-12, and Figure 4.10; as described in 2.18.3). Although it is not clear from Figures 4.9 and 4.10, there was no difference in the number or concentration of the proteins found within the cell membranes of the H^+/M^- clones 19 and 43 and the H^+/M^- AlbS yeast.

To check for the presence of human BCP transcripts, cDNA was synthesised from mRNA isolated from the H^+/M^- clones 19 and 43 and the AlbS yeast (as described in Sections 2.16 and 2.17). PCR using the isolated cDNA as template with human BCP gene-specific primers suggests that both the H^+/M^- clones, 19 and 43, are expressing the BCP cDNA (Figure 4.11).

4.3 Discussion

A full length human BCP cDNA has been isolated and cloned into the yeast expression vector pPIC9. Two clones (H^+/M^- 19 and 43) were identified as containing the human BCP cDNA, after transformation of the yeast *Pichia pastoris* with the pPIC9/BCP clone. Heterologous mRNA expression of the human BCP was confirmed by amplification of cDNA prepared from the H^+/M^- 19 and 43 clones. There was no evidence for the production of the BCP protein. However, the method utilised in this study (SDS-PAGE) is not the most sensitive to detect the protein of interest. An antibody based technique (Western blotting) will detect whether any BCP protein is being produced. If the opsin is translated then steps can be taken to try and induce maximum expression of the BCP from the H^+/M^- 19 and 43 clones.

The main advantages of the *P pastoris* expression system, if successful, over the mammalian protocol is the abundance of protein produced (>0.5 g/l as opposed to μg/l) and the ease of opsin isolation when it is secreted out of the cell (see Section 4.1). It was found that the length of

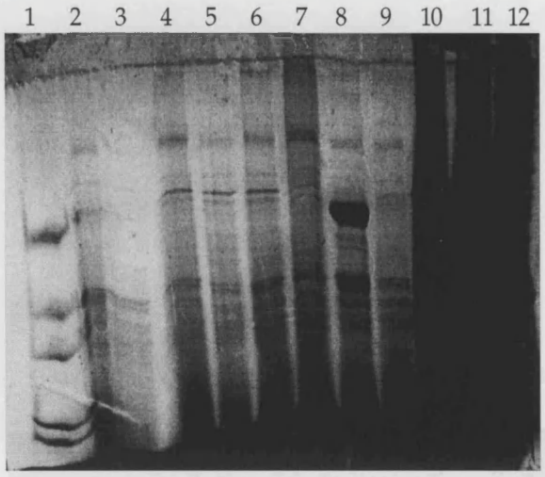


Figure 4.9 Optimisation of secretion of proteins for H^+/M^- clones 19 and 43.

Lane 1: Protein standard: 66 kD; 45 kD; 36 kD; 29 kD and 24 kD.

Lane 2: Clone 19 grown in falcon tube; Lane 3: Clone 43 grown in falcon tube;

Lane 4: Clone 19 grown in conical flasks; Lane 5: Clone 43 grown in conical flasks;

Lane 6: Clone 19 grown in baffled flasks;

Lane 7: Clone 43 grown in baffled flasks;

Lane 8: AlbS (albumin secreting) yeast clone grown in falcon tube;

Lane 9: *Pichia pastoris* wild type H^+/M^+ grown in falcon tube.

Lanes 10, 11 and 12: cell membrane proteins of clones 19, 43 and AlbS, respectively.

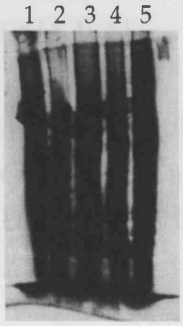


Figure 4.10 Cell membrane proteins of the H^+/M^- clones 19 (lanes 1 and 2) and 43 (lanes 3 and 4) and the AlbS yeast.

Figure 4.11 PCR of yeast cDNA isolated from H^+/M^- clones 19 and 43, to test for the expression of the human BCP. Amplification was performed with the Ex1+ 410/431F and Ex5- 1453/1433R primers. Lanes 5 and 10 contain the ϕX174/*Hae*III DNA ladder.

Lanes 1-3: cDNA from clone 19 as template

Lanes 4, 6 and 7: cDNA from clone as template

Lane 8: Positive control, cDNA cloned in Bluescript plasmid as template

Lane 9: Negative control, no DNA.

time from transformation to the isolation and detection of secreted proteins from H^+/M^- clones was a minimum of 15 days, which is half the total time it takes to transform, select for positive clones and isolate opsin from a mammalian system (Nathans *et al.*, 1989). However, 95% (44 out of 46) of the clones which were identified as being H^+/M^- did not contain the BCP gene (Figure 4.8). Thus, the identification H^+/M^- clones by observing the difference in rates of growth on MM and MD plates is not a precise procedure. It is advisable to test for the presence of the gene of interest within the yeast genome before proceeding with the expression analysis. It should be noted that there is clonal variation in the levels of expression of the heterologous protein of interest, that is each H^+/M^-/BCP clone will show differences in the level of opsin which is secreted (personal communication, Invitrogen®). In addition, there would appear to be a correlation between the number of copies of the integrated vector and the level of expression (more copies = greater protein expressed; Clare *et al.*, 1991a and b). It could be that the two positive transformants identified in this study (H^+/M^- 19 and 43) are low level expressing clones. Transformation and analysis of more yeast may identify other positive H^+/M^-/BCP clones which express and secrete the opsin at higher levels. Further work is required, to establish a higher transformation efficiency of yeast and to show the actual level of variation of expression between positive clones.

Although in this study it was found that a very small percentage of the yeast transformants did contain an integrated pPIC9/BCP vector, it was decided not to proceed with the further development of this system for opsin gene expression. However, there is still the potential for developing the *P pastoris* expression system for the production of high concentrations of opsins because of the low maintenance and the ease with which yeast cultures can be grown relative to mammalian cells.

Chapter 5

Characterisation of the 5′ Proximal Promoter Region of the Mammalian Blue Cone Pigment Gene

5.1 Introduction

Which DNA elements found within the 5′ proximal promoter region of the BCP gene bind proteins to influence transcription? The minimum 5′ flanking region from the human BCP gene which results in blue cone-specific expression in transgenic mice has been determined (see Section 1.16.4). However, the sequences within the 5′ flanking BCP gene promoter (*cis*-acting elements) that are likely to bind transcription factors have yet to be defined.

One approach to identify functionally important DNA motifs involved in the regulation of transcription is to compare sequences between different species (Nie *et al.*, 1996; Di Polo *et al.*, 1996). Sequences which have important biological functions will be highly conserved through evolution, whereas unimportant sequences will tend to diverge. Thus, a comparison of analogous sequences between species should highlight the conserved regions.

5.1.1 Aims

The initial aim of this study was to isolate and sequence the 5′ flanking region of the BCP gene from New and Old World primate species, including human, and from one non-primate mammalian species. A comparison of these sequences should identify conserved elements between these species. A comparison of these conserved elements with other known transcription factor binding sites will identify which are BCP-specific, photoreceptor-specific, retinal-specific and which are general ubiquitous elements.

The next step of this study is to determine whether or not the human retinoblastoma cell line Weri-Rb1 are expressing cone cell transcripts, and more specifically transcripts of the BCP gene. Assuming the BCP gene is expressed within the Weri-Rb1 cells, these cells can be used as a source of retinal nuclear proteins. These Weri-Rb1 nuclear proteins can then be utilised in electrophoretic band shift assays (EBSA) on fragments of the human BCP gene 5′ flanking DNA. This should provide the preliminary data for further investigation of the sequences responsible

for transcription factor binding within the proximal promoter of the mammalian BCP gene .

5.1.2 Photoreceptor gene-specific enhancer elements

The minimal 5′ flanking promoter region of a gene required to regulate tissue/cell-specific expression can be determined from *Lac* Z or other reporter constructs in mice (see Section 1.16.4). Analysis of the 5′ flanking region of various photoreceptor genes, such as BCP, RCP, arrestin and IRBP, has identified the minimal region required to confer photoreceptor-specific expression in transgenic mice (Chen *et al.*, 1994; Wang *et al.*, 1992; Kikuchi *et al.*, 1993; Bobola *et al.*, 1995). Subsequent analysis of these promoter regions has identified conserved sequences which may act as *cis*-acting elements.

A 37 bp sequence, termed the locus control region (LCR), located about 3.5 kb upstream of the human RCP gene was found to confer cone-specific expression in mice (Wang *et al.*, 1992). This LCR was found to share homology to a conserved region located upstream of mammalian rod opsin genes, termed the rod opsin enhancer region (RER) (Nie *et al.*, 1996). The RER, a consensus sequence of about 100 bp, was discovered from a comparison of the 5′ flanking region of rod opsin genes from bovine, human, mouse and rat (Nie *et al.*, 1996). The RER is situated approximately between 1.5-2.4 kb upstream of the mRNA start site, although the actual position varies depending on the species. The RER acts to enhance the expression of the rod opsin gene. A hexamer sequence, 5′-CGATGG-3′, present in all the rod opsin gene RERs was found to be absent from the LCR (Nie *et al.*, 1996). This sequence is thought to be involved in producing rod-specific enhancement of expression.

Directly adjacent to the hexamer 5′-CGATGG-3′ was a 8 bp conserved sequence present in all the rod opsin genes and the LCR, 5′-CC/TTAATCA-3′. This sequence with a core TAAT motif was originally noted in the 5′ proximal promoter of the rod opsin genes from *Drosophila melanogaster* (Mismer *et al.*, 1988; Fortini and Rubin, 1990). In *D melanogaster* the motif could be either 5′-CTAATC-3′ or 5′-CTAATT-3′ depending on the particular rod opsin promoter (Fortini and Rubin, 1990). A 5′-CTAATT-3′ motif within the first 150 bp of the rat rod opsin proximal promoter was shown to interact with retinal nuclear proteins (Morabito *et al.*, 1991). Thus, this suggests that this motif is a *cis*-acting element. This motif has since been discovered in the 5′ flanking regions of the human, mouse and bovine interphotoreceptor retinoid-binding protein (IRBP)

gene, in the chicken rod opsin gene and in the human and murine β-subunit of PDE gene (Liou *et al.*, 1994; Sheshberadaran and Takahashi, 1994; Di Polo *et al.*, 1996). The 5′-CTAATC/T-3′ motif has been termed the photoreceptor conserved element one (PCE1).

Sequence comparison and DNAse I footprinting of the 5′ flanking regions of human and murine β-PDE genes identified several *cis*-acting sequences, including PCE1 (Di Polo *et al.*, 1996). Both β-PDE promoters contained an AP-1 (Activator protein) binding site. The AP-1 sequence, 5′-TGAGTCAG-3′, binds the Jun and Fos transcription factors. The Jun and Fos proteins form dimers, via leucine zipper domains, and bind to the AP-1 site to greatly enhance transcription of genes (Latchman, 1995). Sp1 and CAAT box elements were also found upstream of the β-PDE promoters, which implies that these elements also play a role in contributing to the efficient expression of photoreceptor genes (see Section 1.16.1 for description of Sp1 and CAAT sites).

Although various photoreceptor gene promoters have been analysed, the actual sequences which may be involved in cone-specific expression have yet to be defined. The characterisation of the 5′ flanking region of the BCP gene will further our knowledge of *cis*-acting sequences which contribute to cone-specific expression. Identification of these *cis*-acting elements will be useful for the application of gene therapy for inherited cone dystrophies, which have recently been mapped (Evans *et al.*, 1994; Hong *et al.*, 1994; Balciuniene *et al.*, 1995; Small *et al.*, 1996). Inherited cone dystrophies with preferential or early blue cone involvement have also been noted (Bresnick *et al.*, 1989; Went *et al.*, 1992). The availability of the 5′ flanking region of the human BCP gene will allow the identification of mutations within this promoter region that may cause early blue cone dystrophies.

5.1.3 Retinoblastoma

Retinoblastoma is a cancer of the retina (reviewed by Chen and Mukai, 1993). Retinoblastoma is the most common form of childhood cancer. The gene, when inactivated, which causes retinoblastoma has been identified and characterised (Lee*et al.*, 1987a; Hong *et al.*, 1989). This gene, known as the retinoblastoma (RB) gene, codes for a tumour supressor protein (Latchman, 1995). The RB gene product is expressed in the nucleus of all normal adult tissues. The protein product of the RB gene possesses DNA-binding properties and is found within the nucleus of the cell (Lee *et al.*, 1987b). The RB protein acts as a tumour supressor by regulating the

expression of specific genes involved in the cell-cycle (Sanford and Mukai, 1993).

Retinoblastoma tumours are thought to arise in the retina from a primitive photoreceptor cell (McFall *et al.*, 1977). In support of this theory, retinoblastoma tumour cells do take the form of rossettes and fleurettes possessing cell processes resembling photoreceptors (Ts'o *et al.*, 1970). Northern blot analysis from a number of different retinoblastoma cell lines, established from individual patients, could not detect any expression of the rod cell-specific α subunit of transducin or rod opsin in any of these cell lineages (Bogenmann *et al.*, 1988). However, transcripts for the cone cell-specific α subunit of transducin and for the red and green cone pigment genes were detected in these RB cell lines, suggesting that retinoblastoma may be a neoplasm originating from a primitive cone cell. It should be noted that the blue cone pigment transcript was undetectable in any of the above RB cell lines by Northern blot analysis (Bogenmann *et al.*, 1988). Y79 was the first human retinoblastoma cell line to be established as a stable continuous tumour cell culture (Reid *et al.*, 1974). Farber and colleagues showed that the Y79 cells express rod photoreceptor-specific gene products by RT-PCR (Di Polo and Farber, 1995). The Y79 retinoblastoma cell line may have originated from a rod photoreceptor.

Weri-Rb1 is a human retinoblastoma cell line which was established in 1974 (McFall *et al.*, 1977). It was derived from a tumour removed from a 1-year-old female with no family history of this disease. Weri-Rb1 cells possess similar morphology to the Y79 cells. Both grow as a suspension of small round cells and form grape-like clusters (McFall *et al.*, 1977). However, differences in growth rates, cloning abilities in soft agar and morphology were found between the two cell lines (McFall *et al.*, 1977; McFall *et al.*, 1978). These differences between the Weri-Rb1 and Y79 cells indicate a difference in cell type of origin of tumourogenesis. The Weri-Rb1 cell line was chosen in this study for further analysis.

5.2 Results

5.2.1 Isolation and sequencing of the human and porcine blue cone pigment (BCP) gene 5′ flanking region

A genomic cosmid clone containing the human BCP gene was isolated for the characterisation of its upstream region. An amplified human genomic cosmid library was plated out and screened with the human full length BCP cDNA (as described in Section 2.9). Primary and

secondary screening of this genomic library resulted in the identification of a cosmid clone containing the human BCP gene (as described in Section 2.9.1 to 2.9.6). The cosmid clone was verified to be containing the BCP gene by PCR using the human gene-specific primers to amplify exons 1 to 2 and exons 4 to 5 (as described in Section 2.2.1; see Figure 5.1).

The *Hae*III vectorette library prepared from the human BCP cosmid allowed the amplification of 919 bp of 5′ flanking region using the vectorette protocol (as described in Section 2.2.3). 962 bp of human BCP upstream region, including the above 919 bp, was obtained by PCR using the primers Ex1- 529/514R and Alu1 (Figure 5.2; as described in Section 2.2.1). The Ex1- 529/514R primer was designed from an alignment of exon one of primate BCP genes (described in the next Section). Alu1 is a consensus primer designed to the human repeat elements known as *Alu* (see Table 2.2 for sequence of primer and Section 6.1.3 for a description of *Alu* elements). The 5′ flanking sequence obtained for the human BCP gene is given in Figure 5.3a.

The isolation and subcloning of a genomic cosmid clone containing the porcine BCP gene has been described in Section 3.2.1. Sequencing of the porcine BCP gene from these subclones determined which contained exon 1 and the 5′ flanking region (as described in Section 3.2.1). 949 bp of the 5′ flanking region of the porcine BCP was sequenced (given in Figure 5.3b). An alignment of the human and porcine sequences obtained in this study is given in Figure 5.4. The alignment was produced using the software package Clustal V (see Section 2.21 for list of software), and refined by eye where necessary. This alignment was utilised for the design of consensus upstream sequence specific primers (-572/-551F and -424/-404F see Table 2.2 for sequence of primers). These primers were then used for the amplification of 5′ flanking regions of the BCP genes from other primate species (as described in the next Section). The upstream numbering system refers to the human sequence. Minus one (-1) is the first nucleotide 5′ of the ATG start codon. Thus, the -572/-551F primer sequence is 22 bp long, starts at 572 bp (5′ end) and ends at 551 bp (3′ end) upstream of the ATG codon.

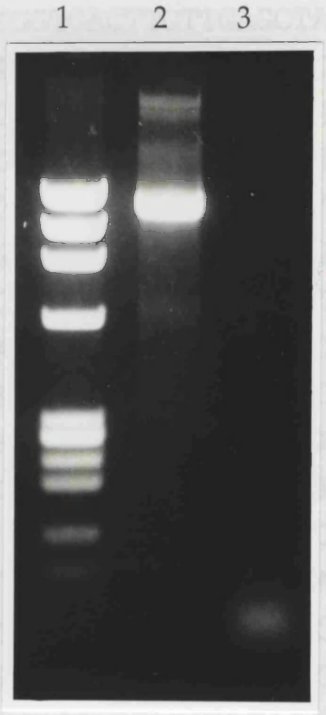


Figure 5.2 PCR of a genomic cosmid clone containing the human BCP gene with the Ex1- 604/589R and Alu1 primers (see Table 2.1 and 2.2 for sequence of primers. Amplification resulted in a 1163 bp product containing 201 bp of exon 1 and 962 bp of 5' flanking region of the BCP gene (Lane 2).

Lane 1: ϕX174/*Hae*III DNA ladder

Lane 2: Ex1- 604/589R and Alu1 primers on BCP cosmid

Lane 3: No DNA negative control

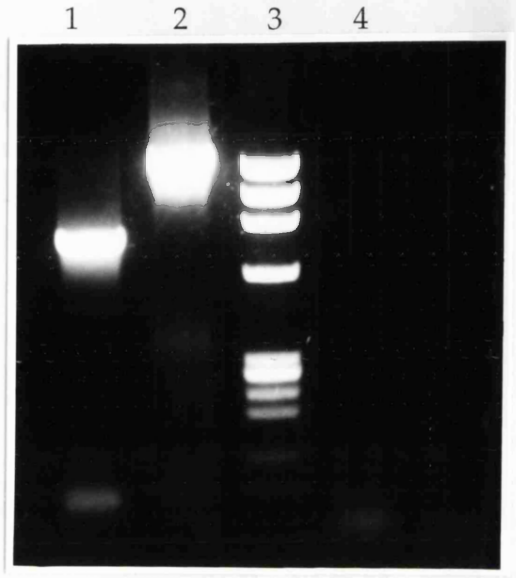


Figure 5.1 PCR of a genomic cosmid clone containing the human BCP gene. Amplification of exons 1 to 2 (Lane 1), utilising the primer Ex1+ 469/482F with the Ex2- 929/908R primer (747 bp product), and exons 4 to 5 (Lane 2), using the primer Ex4+ 1098/1117F with the Ex5- 1453/1433R primer (1343 bp product). Lane 4: No DNA negative control.

Lane 3 : ϕX174/*Hae*III DNA ladder, sizes (bp):	1353
	1078
	872
	603
	310
(appears as one band)	281/271
	194
	118
	72

GCCTCCCAAAGTGCTGGGATTACAGGCTTGAGCCACCACGCCC -920
GGCCAACTCATCTATATATTTATGATCTGCTGAGGTGAAAAGATT - 875
CATTCATAAAGCCTAAAGTGTAGTACTGAGTACAAGTGCAGGTGG - 830
AACTTAAAGCAGAGTTGGTGTCACTGGGGTGGGCTGCACAGAAGA - 785
GGTGTTTGTGAGGTAGGCTGTGAGGGCTAGATAAGACTTCAGTGA - 740
ACAGACAGGAGGAAGACTGTGGTGACTTGTTGAGCTAAGGGTAGG - 695
GGTAGGCATGAACATGGCGTAAGGACAGTGAACAGACCCACCTCC - 650
CTGGCTGGAGGTCTGAGTAGGAAAGGCACACTCTTTGAGTAGATT - 605
CAACACAAAGAAATGGAGGCAGCCTGGCTCAGGGATGCCCTAGAC - 560
CTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCATCAAGGCTG - 515
TTTACTGATGTGCTTCCAGGGCACTCCCCACTCCCAGCCCTTTCC - 470
TGCAGCCCAGGGCTGGTTCCTAGCCTCTCAGCAGACTTAAGATGG - 425
GCACCTTCCACAAAGGGGCAGATGAGTTGAGGAAAACTTAACTGA - 380
TACAGTTGTGCCAGAAGCCAAAATAAGAGGCGTGCCCTTTCTATA - 335
GCCCCATTAAAAGAACAAAAAAGTGGAAGCATCTTCAGTGAATAT - 290
GGGTCAGCACCTCCCAGACCTCAGGGAGTCCACTTCTGTTCATCC - 245
CAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATCTCTT - 200
ATCCTCTGAAGAACACAATCGGCTTTGGGGCCACAGAAGGTTTAG - 155
GTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTCCTTGGGAGG - 110
TTTAGGATTAGTATTGATCATTCACAGAGCCCAAGTGTTTTTAGA - 65
GGAGGGGTTTTGTGGGGTGGGAGGATCACC**TATAA**GAGGACTCAG - 20
AGGAGGGTGTGGGGCATCC**ATG** - 1

Figure 5.3a 962 bp of the human blue cone pigment gene 5' flanking region. The beginning of the *Alu* repeat sequence is underlined, which includes the 25 bp Alu1 primer sequence given in red which was used to generate this upstream region (Section 5.2.1). The **GGCC** sequence, the last 4 bp of the *Alu* repeat, represents the *Hae*III site that was utilised for the vectorette protocol to obtain the initial 919 bp of upstream sequence (see Section 5.2.1). The ATG start codon and the TATA box are also given and shown in bold lettering. The numbering system refers to the nucleotide position upstream of the start site, -1 being the first bp 5′ to the ATG start codon.

CCTGTATCTAAAGGGTGGCACTGTTATCTATGTATCTATAGGCAA -905
CTGAGGTGGGAGAGAAGAGACACTCGTAAGGCCCAAAGTGTAACG -860
TTGAATATAAGCGTGAGCAGAACTTAGAGCCATTCTGGGGTCCCT -815
GGGGTGGGCTTCTCATCAGAGGTCTTTGTGAGTCATTCTTAGAGC -770
TGAGCAGGTCTTGGATGAGCAGACAGGAGGAAGCCTGGCAGCTTG -725
GGCACTGCTCAGCTAAGGGCAGGGGCAGGTGTGAACATGGCATGA -680
GACGGACAGAGAAAAAACCAACTTCAACTGGCCTGGGGACATCTG -635
AATAGGAAAGATACCTCTCTGGGGAGACACACCATAGAAGAAACT -590
GGCAGCCTGGATGCCTTAGATCTCTAGGCATTGGCGAATCACCCT -545
TGAACTCTTCCATGAAGGCTCTTGCTAAGGTGCTCCAGACACTCT -500
CCATTCCCTGTCTCTTCCTGCAGCCCGAGGCTGGTTCCCAGGCTG -455
CTGTGAGCCCAAGATCCCTGGAGAGGGAGGCTTAGGGGGTACCTT -410
CCACAAAGGGACAGATGGGGTACAGACACCTATAGGGATTTAGTG -365
GTGCCTGAAGCCAAATGAAGAGGGGTGCCCTTGAGAGTTCTACTC -320
AAAGGGGAAAAAAGTGGTGGAAGGATCTTTATTAAATATGGGTCA -275
GTATCTCCCTCCCAGGCCCTAGATCCTTTGAGCCCCATCTTTTAT -230
CCTCTGAAGAACACACAATCAGCTCTGGGGGACCAACAAAAGGAG -185
GTTTAGGTGGGGGCTTGAGGTCTCTAATCCCAAACTTTGTGGCTT -140
AGGAGGCTTAGGATCAGTACTGAATCCTTTCAGGGGAGCCCGAAA -95
GTGGGCTTTGGAGGAGAGGCTCTTTGGCAGGGTGGGGGGAAGTGG -50
CGGTGGGAGGATGATC**TATAA**GAGGAATCAGAGGGGGTGGGGACA -5
TCCC**ATG**

Figure 5.3b 949 bp of the porcine blue cone pigment gene 5′ flanking region. The **ATG** start codon and the **TATAA** box for the BCP gene is shown in bold

```
Human                          5' GCCT---CCCAAAGTGCTGGGA--   -943
Porcine                        5' -CCTGTATCTAAAG-GGTGGCACT
                                   ***    * **** * *** *

Human     -TTACAGCTTGAGCCACCACGCCCGGCCAACTCATCTATATATTTA-TG   -896
Porcine   GTTA--TCT------------------ATGT-ATCTATA-GGCAACTG
           ***   **                  *  * *******     * **

Human     ATCTGCTGAGGTGAAAAGATTCATTCATAAAGCCTAAAGTGTAGTACTG   -847
Porcine   AGGTGG-GAGA-GAAGAGA--CACTCGTAAGGCCCAAAGTGTAACGTTG
          *  **  ***  *** ***  ** ** *** *** ********    **

Human     AGTACAAG--TGCAGGTGGAACTTAAAGC-AGAGTTGGTGTCACTGGGG   -801
Porcine   AATATAAGCGTG-AGC-AGAACTTAGAGCCAT-TCTGGGGTCCCTGGGG
          * ** ***  ** **   ******* *** *    *** *** ******

Human     TGGGCTGCACAGAAGAGGTGTTTGTGAGGTAGGCTGTGAGGGCTAGATA   -752
Porcine   TGGGCTTCTCATCAGAGGTCTTTGTGAGTCATTCT-T-AGAGCT-GAGC
          ****** * **  ****** ********  *  ** * ** *** **

Human     AGA-CTTCAGTGAACAGACAGGAGGAAGACT----G--TGGTGACTTGT   -710
Porcine   AGGTCTTGGATGAGCAGACAGGAGGAAGCCTGGCAGCTTGGGCACT-GC
          **  ***   *** ************** **    *  ***  *** *

Human     TGAGCTAAGGGTAGGGGTAGGCATGAACATGGCGT-A-A-GGACAGTGA   -664
Porcine   TCAGCTAAGGGCAGGGGCAGGTGTGAACATGGCATGAGACGGACAGAGA
          * ********* ***** ***  ********** * * * ****** **

Human     ACAGACCCACCTCC-CTGG-CTGG--AGGTCTGAGTAGGAAAGGCA-CA   -620
Porcine   AAAAACCAACTTCAACTGGCCTGGGGACATCTGAATAGGAAAGATACCT
          * * *** ** **  **** ****  *  ***** ********  * *

Human     CTCTTT-GAGT-AGATTCAACACAAAGAAATGGAGGCAGCCTGGCTCAG   -573
Porcine   CTCTGGGGAGACACAC-C-ATAG-AAGAAACT--GGCAGCCT-------
          ****   ***  * *  * * *  ******    ********

Human     GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCA   -524
Porcine   GGATGCCTTAGATCTCTAGGCATTGGCGAATCACCCTTGAACTCTTCCA
          ******* **** ************ * * *  **   *  ** *****

Human     TCAAGGCTGTTTACTGATGTGCTTCCAGGGCACTCCCCACTCCCAGCCC   -475
Porcine   TGAAGGCTCTT-GCTAAGGTGCT-CCA-GACACTCTCCATTCCCTGTCT
          * ****** **  ** *  **** ***   ***** *** ****   *

Human     TTTCCTGCAGCCCAGGGCTGGTTCCTAGCCT-CTC--AGC---AGA--C   -434
Porcine   CTTCCTGCAGCCCGAGGCTGGTTCCCAGGCTGCTGTGAGCCCAAGATCC
           ************  ********** ** ** **   ***   ***  *
```

```
Human     TT----A---AG-AT----GGGCACCTTCCACAAAGGGGCAGATGAGTT  -397
Porcine   CTGGAGAGGGAGGCTTAGGGGGTACCTTCCACAAAGGGACAGATGGGGT
           *    *   **       *** ************** ****** * *

Human     GAGGAAAACTTAACTGATACAGTTGTGCCAGAAGCCAAAATAAGAGGCG  -348
Porcine   ACAGACACCTATAGGGATTTAGTGGTGCCTGAAGCCAAATGAAGAGGGG
             ** * **  *  ***  *** ***** *********  ****** *

Human     TGCCCTTTCTATAGCCCCATTAAAAGAACAAAAAAGTGG---AAGCATC  -302
Porcine   TGCCCTT--GAGAGTTCTACTCAAAGGGGAAAAAAGTGGTGGAAGGATC
          *******   * **  * * * ****   **********   *** ***

Human     TTCAGTGAATATGGGTCAGCACCTCCCAGACCTCAGGGAGTCCACTTCT  -253
Porcine   TTTATTAAATATGGGTCAGTATCTCCC--TCC-CAGG------------
          ** * * ************ * *****   ** ****

Human     GTTCATCCCAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATC  -204
Porcine   -------------CCC-------------TAGATCCTTTGAGCCCCATC
                       ***              ****  ********* ***

Human     TCTTATCCTCTGAAGAACA--CAATCGGCTTTGGGG--CCA-CAGAAGG  -160
Porcine   TTTTATCCTCTGAAGAACACACAATCAGCTCTGGGGGACCAACAAAAGG
          * *****************  ***** *** *****  *** ** ****

Human     ---TTTAGGTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTC-CTTG  -115
Porcine   AGGTTTAGGTGGGGGCTT-GAGGTCTCTAATCCCAAACTTTGTGGCTTA
             ******* * ** ** * * * *****************  ***

Human     GGAGGTTTAGGATTAGTATTGA-TC-ATTCACAG-AGCCC--AAGTGT-  -72
Porcine   GGAGGCTTAGGATCAGTACTGAATCCTTTCAGGGGAGCCCGAAAGTGGG
          ***** ******* **** *** **  ****  * *****  *****

Human     TTTTAGAGGAGGGG-T-TTTGTG-GGGTGGG----------------AG  -42
Porcine   CTTTGGAGGAGAGGCTCTTTGGCAGGGTGGGGGGAAGTGGCGGTGGGAG
           *** ****** ** * ****   *******                **

Human     GATCACCTATAAGAGGACTCAGAGGAGGGTGTGGGGCATCC-ATG
Porcine   GATGATCTATAAGAGGAATCAGAGG-GGGTG-GGGACATCCCATG
          *** * *********** ******* ***** *** *****  ***
```

Figure 5.4 Sequence alignment of the human and porcine BCP gene 5′ flanking promoter region. The ATG start-codon and the TATA box is shown in bold. Primer sequences designed from this alignment are listed below (see Section 5.2.2).

5′-GGATGCCCTAGACCTCTAGGCA-3′ -572/-551F primer sequence
5′-GCACCTTCCACAAAGGGGCAG-3′ -424/-404F primer sequence

5.2.2 Isolation and sequencing of the 5′ flanking regions of the blue cone pigment (BCP) gene from an Old World and a New World primate

Initially, 55 bp and 49 bp of 5′ flanking region of the BCP gene from an Old World primate (talapoin) and two New World primates (capuchin and marmoset), respectively, was isolated using the Unpredictably Primed Walking PCR (UPW-PCR) protocol (described in Section 2.2.3). Genomic DNA from talapoin, capuchin and marmoset, isolated from liver tissue samples from these primates (Section 2.1.1), was used as template for UPW-PCR. The usage of UPW-PCR required the design of the consensus nested primers Ex1- 604/589R and Ex1- 529/514R, which are situated within exon one of primate BCP genes. These consensus nested primers were designed from a comparison of the first 201 bp of exon one of the human BCP gene with the equivalent region of the marmoset and talapoin monkey gene sequences (see Figure 5.5). UPW-PCR using the 33-mer walking primer (see Table 2.4) with the gene-specific Ex1- primers for the first round of extension steps, on talapoin, marmoset and capuchin genomic DNA resulted in a 184 bp, 181 bp and 178 bp product, respectively. Figure 5.6 shows an alignment of sequences obtained for the talapoin, marmoset and capuchin monkey using the UPW-PCR method, to the equivalent human region. Attempts to amplify larger upstream fragments, using the 33-mer walking primer (WP-33mer), by altering the PCR conditions had no effect (Figure 5.6). The reason for this is presumably due to the WP-33mer clamping down on to a conserved sequence just upstream of the TATA box in the primates.

UPW-PCR generated DNA sequence corresponding to approximately half of exon 1 of the BCP gene for the non-human primates (Figure 5.7). The deduced amino acid sequence obtained for part of exon one for these non-human primates is shown aligned to the equivalent human sequence in Figure 5.8. The capuchin sequence presented here is original new data. Although the talapoin and marmoset BCP gene sequences for exon one have been published previously, the nucleotides coding for the first seven amino acids at the N-terminus of the BCP protein was not determined (Hunt *et al.*, 1995; Figure 5.5). Due to the high degree of homology and close evolutionary relationship between New and Old World primates it was presumed these seven amino acids in the talapoin and marmoset would be similar to the human. The data presented here, generated using the UPW-PCR method, completes the

```
Human       ATG AGA AAA ATG TCG GAG GAA --- GAG TTT TAT CTG TTC AAA AAT ATC TCT TCA GTG GGG CCG TGG GAT
Talapoin    ::: ::: ::: ::: ::: ::: ::: GAG GAG TTT TAT CTG TTC AAA AAT ATC TCT TCA GTG AAG CCG TGG GAT
Marmoset    ::: ::: ::: ::: ::: ::: ::: GAG GAG TTT TAT CTG TTC AAG AAC ATC TCC TCG GTG GGG CCG TGG GAT
                                        *** *** *** *** *** **  **  *** **  **  ***   * *** *** ***

Human       GGG CCT CAG TAC --- CAC ATT GCC CCT GTC TGG GCC TTC TAC CTC CAG GCA GCT TTC ATG GGC ACT GTC
Talapoin    GGG CCT CAG TAC --- CAC ATT CCC CCT GTC TGG GCC TAC TAC CTC CAG GCA GCT TTC ATG GGC ACT GTC
Marmoset    GGG CCT CAG TAC CCC CAC ATT GCC CCT TCC TGG GCC TAC TAC CTC CAG GCA GCT TTC ATG GGC ATT GTC
            *** *** *** ***     *** ***  ** ***   * *** *** * * *** *** *** *** *** *** *** *** * * ***

Human       TTC CTT ATA GGG TTC CCA CTC AAT GCC ATG GTG CTG GTG GCC ACA CTG CGC TAC AAA AAG TTG CGG CAG
Talapoin    TTC CTT GCA GGG TTT CCA CTC AAT GCC ATG GTG CTG GTG GCC ACA GTG CGC TAC AAA AAG TTG CGG CAG
Marmoset    TTC CTT ATA GGG TTA CCA CTG AAT ACC ATG GTG CTG GTG GCC ACA GTG CGC TAC AAA AAG TTG CGG CAT
            *** ***   * *** **  *** *** ***  ** *** *** *** *** *** ***  ** *** *** *** *** *** *** **
```

Figure 5.5 Alignment of sequences for part of exon one of the blue cone pigment gene from talapoin monkey and marmoset, to the equivalent human region, for the design of primers (sequence from Hunt *et al.*, 1995). The consensus gene-specific primers Ex1-604/589R ('Outer-SS primer') and Ex1- 529/514R ('Inner-SS primer') were designed from this comparison, for use with Unpredictably Primed Walking PCR and upstream sequence-specific primers (see Sections 2.2.3 and 5.2.2). The complement of the Ex1- 604/589-primer sequence is shown in **red**. The complement of the Ex1- 529/514- primer sequence is underlined. The actual sequence of these primers are given in Table 2.4. The first 21 bp of sequence, from the ATG start codon onwards, was not determined for the talapoin monkey and marmoset (Hunt *et al.*, 1995) and are indicated by dots (:).

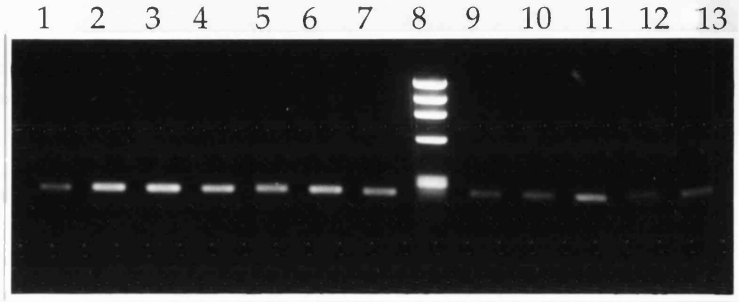


Figure 5.6 Amplification products from marmoset, capuchin and talapoin monkey genomic DNA using the UPW-PCR protocol, with various PCR conditions.

		PCR Conditions	
Lane	DNA Template	$MgCl_2$ Concentration	Annealing Temperature
1	Marmoset	1.5 mM	62 °C
2	Capuchin	1.5 mM	62 °C
3	Talapoin	1.5 mM	62 °C
4	Marmoset	3.0 mM	62 °C
5	Capuchin	3.0 mM	62 °C
6	Talapoin	3.0 mM	62 °C
7	Marmoset	1.5 mM	68 °C
8		DNA ladder	ϕX174/*Hae*III
9	Capuchin	1.5 mM	68 °C
10	Talapoin	1.5 mM	68 °C
11	Marmoset	3.0 mM	68 °C
12	Capuchin	3.0 mM	68 °C
13	Talapoin	3.0 mM	68 °C

```
          -55                                                  -7      -1
Human     TTGTGGGGTGGGAGGATCACCTATAAGAGGACTCAGAGGAGGGTGTGGGGCATCC ATG AGA AAA ATG TCG GAG GAA --- GAG TTT
Talapoin  TTCTGGGGTGGGAGGATCACCTATAAGAGGACTCAGAGGGGGGTGTGGGGCATCC ATG AGA AAG ATG TCA GAG GAA GAG GAG TTT
Capuchin  TTGTGGGGTGG----ATCATCTATAAGAGGACTCAGAGG--GGTGTGGGGTATCC ATG AGC AAG ATG TCA GAG GAA GAG GAG TTT
Marmoset  TTGTGGGGTGG----ATCATCTATAAGAGGACTCAGAGG--GGTGTGGGGTATCC ATG AGC AAG ATG TCA GAG GAA GAG GAG TTT
          ** ********    **** ******************  ********* **** *** **  **  *** **  *** ***     *** ***
                                                                  ^

Human     TAT CTG TTC AAA AAT ATC TCT TCA GTG GGG CCG TGG GAT GGG CCT CAG TAC --- CAC ATT GCC CCT GTC TGG
Talapoin  TAT CTG TTC AAA AAT ATC TCT TCA GTG AAG CCG TGG GAT GGG CCT CAG TAC --- CAC ATT GCC CCT GTC TGG
Capuchin  TAT CTG TTC AAG AAC ATC TCT TCG GTG GGG CCG TGG GAT GGG CCT CAG TAC --- CAC ATT GCC CCT GTC TGG
Marmoset  TAT CTG TTC AAG AAC ATC TCC TCG GTG GGG CCG TGG GAT GGG CCT CAG TAC CCC CAC ATT GCC CCT TCC TGG
          *** *** *** **  **  *** **  **  ***   * *** *** *** *** *** *** ***     *** *** *** ***   * ***

Human      GCC TTC TAC CTC CAG GCA GCT TTC ATG GGC     55 bp of 5′ flanking promoter sequence and 126 bp of coding region
Talapoin   GCC TAC TAC CTC CAG GCA GCT TTC ATG GGC     55 bp of 5′ flanking promoter sequence and 129 bp of coding region
Capuchin   GCC TTC TAC CTC CAG GCA GCT TTC ATG GGC     49 bp of 5′ flanking promoter sequence and 129 bp of coding region
Marmoset   GCC TAC TAC CTC CAG GCA GCT TTC ATG GGC     49 bp of 5′ flanking promoter sequence and 132 bp of coding region
          *** * * *** *** *** *** *** *** *** ***
```

Figure 5.7 Alignment of New and Old World primate sequences, to the equivalent human region, obtained using the unpredictably primed walking PCR (UPW-PCR) method. The Old World monkey, talapoin, and the two New World monkey, the marmoset and capuchin monkey sequences were obtained using the consensus gene-specific primers Ex1- 604/589-, Ex1- 529/514- with the UPW-PCR primers, WP 33-mer and WP 17-mer (as described in Section 5.2.2). The complement of the Ex1- 529/514- primer sequence is shown underlined. Nucleotides which are identical between all four species are undermarked with an asterix (*). The TATAA box and the ATG start codon is given in bold. Sequences representing coding regions are shown as triplet codons, beginning with the initial ATG start-site. The mRNA start site for the non-human primates, is presumed be to at the same position as in the human (at -7), and is undermarked with a ^ .

```
Human      M R K M S E E E - F Y L F K N I S S V  20
Talapoin   M R K M S E E E E F Y L F K N I S S V  21
Capuchin   M S K M S E E E E F Y L F K N I S S V  21
Marmoset   M S K M S E E E E F Y L F K N I S S V  21
           *   * * * * * *   * * * * * * * * * *

Human      G P W D G P Q Y - H I A P V W A F Y L  38
Talapoin   K P W D G P Q Y - H I A P V W A Y Y L  39
Capuchin   G P W D G P Q Y - H I A P V W A F Y L  39
Marmoset   G P W D G P Q Y P H I A P S W A Y Y L  40
             * * * * * * *   * * * *   * *   * *

Human      Q A A F M G                            44
Talapoin   Q A A F M G                            45
Capuchin   Q A A F M G                            45
Marmoset   Q A A F M G                            46
           * * * * * *
```

Figure 5.8 Alignment of the first 44-46 amino acids of the primate BCP protein. The amino acid sequence for the non-human primates were predicted from the BCP gene sequence obtained for part of exon one by UPW-PCR (Section 5.2.2; Figure 5.7). The human protein sequence is from Nathans *et al.*, 1986a. The amino acids are given in there single letter codes. The first M in the alignment being the initial methionine of exon one. The asterix (*) shown below a specific column of amino acids indicates conservation between all the four primate species.

nucleotide and the amino acid sequence for these genes. The second amino acid in the BCP protein is an arginine residue in the two Old World primates (Figure 5.8). However, the second amino acid in the New World primate BCP protein is a serine residue (Figure 5.8). The most parsimonious explanation for this amino acid difference is that this change must have occurred soon after the divergence of the New and Old World primate lineages (Time 1 in Figure 3.9).

5′ flanking regions from talapoin and capuchin monkey BCP genes were extended by PCR of genomic DNA, using consensus sequence-specific primers designed to the promoter region (-572/-551F and -424/-404F, see previous Section) and exon 1 (Ex1- 529/514; PCR conditions are described in Section 2.2.1). Amplified PCR products from talapoin and capuchin monkey were ligated into pGEM cloning vector, before transforming and selecting for positive recombinants (as described in Sections 2.5-2.8). The BCP upstream regions were sequenced from these plasmids (as described in Section 2.12). Figure 5.9a and 5.9b shows the sequence of the 5′ flanking regions obtained for these two primates.

5.2.3 Analysis of the promoter regions of the porcine and primate BCP genes

The 573 bp and 570 bp of 5′ flanking region obtained for the talapoin and capuchin monkey were aligned to the equivalent region of upstream sequences obtained for the human and porcine BCP genes, and the percentage nucleotide identity calculated (Table 5.1). The published bovine and murine 5′ flanking BCP gene sequences are compatible in length to the amplified products obtained for talapoin and capuchin monkey, for the region from the ATG start codon to the consensus upstream primer -572/-551F (Chiu and Nathans, 1994b). Thus, sequences corresponding to the 5′ flanking region of the BCP gene downstream of the -572/-551F primer were compared for the porcine, human, talapoin, capuchin, bovine and murine, and the percentage nucleotide identity calculated (Table 5.1).

The highest degree of nucleotide identity is observed from a comparison between the 5′ flanking region of the two Old World primate species (95.5%; Table 5.1). The least degree of identity between 5′ flanking regions is observed between a comparison of the porcine and murine species (64.6%). The most informative alignments, that clearly delineate conserved motifs which may be involved in transcriptional regulation, will come from the comparison of sequences that share the least overall identity with each other. Thus, an alignment of the porcine and bovine

a

GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGT -529
TCCATCAAGGCTGTTTACTGATGTGCTTCCAGGGCACTCCCCACT -484
CCCAGCCCTTTCCTCCAGCCCGAGGCTGTTCCCAGCTCTCAGGAG -439
AATTAAAATGGGCACCTTCCACAAAGGGGCAGATGAGGTGAGGAA -394
AACTTAACCGATACAGTTGTACCAGAAGCCAAAACAAAAGGCATG -349
CCCTTTCTATAGCCCCATTAAAAGAACAGCAAAAAGGTGGAAGCA -304
TCTTCAGTGAATATGGGTCAGCACCTCCCAGACCTCAGGGAGTTC -259
ACTTCTATTCATTCCAGCACCCAGCATTGCATATCCAGATTATTT -214
GAGCCCAATCTTTTATCCTCTGAAGAACACAATCAGCTTTGGGGC -169
CACAAAAGGTTTAGGTAGTGGTTTAGCGATTTCTAATCCCAAACT -124
TTGTCCTTGGGAGGTTTAGGATTAGTATTGATCATTCACAGAGCC -79
CAAGTGTTTTCAGAGGAGGGGTTTTCTGGGGTGGGAGGATCACC**T** -35
ATAAGAGGACTCAGAGGGGGGTGTGGGGCATCC**ATG**

b

GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTATGTCCTAT -526
TCCATCAAGGCTGTTTACTGACCTGCTTCCAGGGCACTCCCCACT -481
CCCAGCCTCTTCCTGCAGACCGAGGCTGCTTCCCAGCCTCTCAGG -436
AGATTTAAGATGGGCACCTTCCACAAACGGGCAGATGAGGTGAGG -391
AAAACTTTACAGACACAGTTGTGCCAGAAGCCAAAATCAGAGGGG -346
TGCCCTTTCTGTAGCCCCATTATAAAAGAACTGCAAAAAGGTGGA -301
AGCATCTTCAACGAACATGGGTCAGCACCTCCCAGACCTCAGGGA -256
GTTCACTTCTGTTCATCCCAGCACCCAACATTGCATATCCAGATT -211
ATTTGAGCCCAATATTTTATCCTCTGAAGAACACAATCGGCTTTG -166
GGAACACAAAAGGTTTAGGTAGTGGTTTAGGGATTTCTAATCCAA -121
ACTTTGTCCTTAGGAGGTTTAGGATTAGTGTTCATCAGTCACAGT -76
GCCCAAGTGTTTTTAAAGGAGGGGTTTTGTGGGGTGGATCATC**TA** -31
TAAGAGGACTCAGAGGGGTGTGGGGTATCC**ATG**

Figure 5.9a 573 bp of the blue cone pigment gene 5′ flanking region from talapoin monkey; (**b**) 570 bp of the blue cone pigment gene 5′ flanking region from capuchin monkey, plus the ATG start codon. The -572/-551+ consensus upstream primer, which was used to generate this upstream fragment, is shown in blue lettering (Section 5.2.2). The TATA box and the ATG start codon is given in bold.

	Hum	Tal	Cap	Por	Bov	Mur
Human	100	95.5	91.1	74.0	74.5	70.2
Talpoin		100	90.5	73.3	73.1	68.3
Capuchin			100	70.9	70.7	70.2
Porcine				100	81.3	64.6
Bovine					100	64.8
Murine						100

Table 5.1 Percentage nucleotide identity of the proximal 5′ flanking region for the BCP genes between human (572 bp), talapoin monkey (573 bp), capuchin monkey (570 bp), porcine (582 bp), bovine (566 bp) and murine (526 bp). (Bovine and murine sequences were obtained from Chiu and Nathans, 1994b).

sequences, which gives a percentage nucleotide identity of 81.3%, is less informative than an alignment of the murine sequence to the equivalent porcine sequence, which gives only a 64.6% identity (Table 5.1). In addition, the greater the number of sequences compared, the easier it is to ascertain conserved sequences. An alignment of the human, talapoin, capuchin and the murine sequences together results in a 53.9% nucleotide identity between all four species. This 53.9% nucleotide identity is reduced to 40.6% when the 5′ flanking regions of all six species are combined together in one alignment. Hence, this latter alignment is the most informative in delineating evolutionary conserved regions (see Figure 5.10).

5.2.4 Evolutionary conservation of upstream sequences in mammalian BCP genes

In order to identify potential transcription factor binding sites that may be involved in the expression of the BCP gene, the human, talapoin, capuchin, porcine, bovine and murine 5′ flanking sequences were compared (Figure 5.10). In addition the 5′ flanking sequences of these six mammalian BCP genes were analysed for potential transcription factor binding sites, utilising the computer program TESS (Transcriptional Element Search Software; see Section 2.21 for list of computer software). The TESS analysis tested both the sense and anti-sense strand of the DNA for putative protein binding sites. The numbering system refers to the human sequence, where -1 is the first nucleotide 5′ to the ATG start codon.

The TESS program identified only one potential transcription factor binding site which is situated in a conserved region at the same position within all six species (see Figure 5.10). The sequence of this conserved motif (5′-AGAACA-3′, at position -190 to -185) is the reverse complement of the potential binding site for the GR (glucocorticoid receptor) and PR (progesterone receptor) transcription factors (Latchman, 1995). Genes containing the GR/PR motif (5′ GGCTACANNNTGTTCT 3′) bind the GR and PR proteins in response to the action of glucocorticoid and progesterone hormones, respectively. Thus, genes which contain the GR/PR motif are inducible by hormone treatment.

Interestingly, the six bp motif (5′-AGAACA-3′), within the proximal promoter region of the mammalian BCP genes, is situated at the 3′ end of a 17 bp stretch of highly conserved nucleotides. This 17 bp region is conserved through all six species and is the longest continuous region of conserved nucleotides identified by the alignment (Figure 5.10, from

```
Human       GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCAT  -523
Talapoin    GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCAT
Capuchin    GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTATGTCCTATTCCAT
Porcine     GGATGCCTTAGATCTCTAGGCATTGGCGAATCACCCTTGAACTCTTCCAT
Bovine      ---------------CTAGGCATTGTCGAGTTGCCCCTGGACTGTTCCAT
Murine      GGATGC-----------AG---TTGTT---TCTGCACTGTCCCG---CA-
                             **   ***     *   *   *  *     **

                                                            _
Human       CAAGGCTGTTTACTGATGTGCTTCCAGGGCACTCCCCAC-TCCCAG-CCC  -475
Talapoin    CAAGGCTGTTTACTGATGTGCTTCCAGGGCACTCCCCAC-TCCCAG-CCC
Capuchin    CAAGGCTGTTTACTGACCTGCTTCCAGGGCACTCCCCAC-TCCCAG-CCT
Porcine     GAAGGCTCTT-GCTAAGGTGC-TCCA-GACACTCTCCAT-TCCCTG-TCT
Bovine      GAAGGCTGTT--------------CAAGA-ACTCTCTCTATCCCTTACCC
Murine      GT-GGCTATTTA--AA-ATGCTCCCAAGGGGGCCTCCCCACTCAT--CCT
               **** **     *        ** *     * *           *

            ________
Human       TTTCCTGCAGCCCAGGGCTGGTTCCTAGC---CT-----CTCAGCAGACT -433
Talapoin    TTTCCTCCAGCCCGAGGCTG-TTCCCAG----CT-----CTCAGGAGAAT
Capuchin    CTTCCTGCAGACCGAGGCTGCTTCCCAGC---CT-----CTCAGGAGATT
Porcine     CTTCCTGCAGCCCGAGGCTGGTTCCCAGGCTGCTGTGAGCCCAAGATCCC
Bovine      TTTCCTGCAGCCTGAGGCTGG-----------CT--GAGCCCAAGACCCC
Murine      CTTCCTGTTTCCCCAAGCGGATTCCCAGG-T-CT---------------C
             *****     *    ** *            **

                                                   ______
Human       T----A---AGA-T-----GGGCACCTTCCACAAAGGGGCAGATGAGTTG  -396
Talapoin    T----A---AAA-T-----GGGCACCTTCCACAAAGGGGCAGATGAGGTG
Capuchin    T----A---AGA-T-----GGGCACCTTCCACAAACGGGCAGATGAGGTG
Porcine     TGGAGAGGGAGGCTTAGG-GGGTACCTTCCACAAAGGGACAGATGGGGTA
Bovine      TGGAGAGGGAGACTTAGG-GGGCACCTACCACAATGGGGCAGATGGTGTG
Murine      TGGATAGGGACACTCAGGTGAGCATCTTTCACAAAGGGCCAGGTGAGATG
            *    *   *   *     * * * **  *****  ** *** **   *

Human       AGGAAAACTTAACTGATACAGTTGTGCCAGAAGCCAAAATAAGAGGCGTG  -346
Talapoin    AGGAAAACTTAACCGATACAGTTGTACCAGAAGCCAAAACAAAAGGCATG
Capuchin    AGGAAAACTTTACAGACACAGTTGTGCCAGAAGCCAAAATCAGAGGGGTG
Porcine     CAGACACCTATAGGGATTTAGTGGTGCCTGAAGCCAAATGAAGAGGGGTG
Bovine      TAAACACCTACATGGATGTAGTGGTGCCTGAAGTGAAACAAAGAGGCTTG
Murine      CAGA---------GGA---------GCCTGAAGCCAAAGGCAGAGG--TG
               *         **         ** ****  ***   * ***  **

Human       CCCTTTCTATAGCCCCATTA--AAAGAACA--AAAAA-GTGGAAGCATCT  -301
Talapoin    CCCTTTCTATAGCCCCATTA--AAAGAACAGCAAAAAGGTGGAAGCATCT
Capuchin    CCCTTTCTGTAGCCCCATTATAAAAGAACTGCAAAAAGGTGGAAGCATCT
Porcine     CCCTT--GAGAGTTCTACT-CAAAGGGG-AAAAAAGTGGTGGAAGGATCT
Bovine      CCTTC--AACAGTTCCACT-CAAAGGGA-AAAAGAGTGGTGGAAGGGTCT
Murine      ------------------------GG----AATGAGGTATGGCAGCATCT
                                     *        *    *** **  ***

Human       TCAGTGAATATGGGTCAGCACC-TCCCAGACCTCAGGGAGTCCACTTCTG  -252
Talapoin    TCAGTGAATATGGGTCAGCACC-TCCCAGACCTCAGGGAGTTCACTTCTA
Capuchin    TCAACGAACATGGGTCAGCACC-TCCCAGACCTCAGGGAGTTCACTTCTG
Porcine     TTATTAAATATGGGTCAGTATC-TCCC---TCCCAGG-------CCCT--
Bovine      TCATTAAATATGGGTCAGCATC-TCCC---TACTGAG---TCCACCCTTA
Murine      CCAAAGAATATAGGTCAGCATCCTCCAGGCCACAGAAA--CCAATACTGA
              *   ** ** ****** * * ***

Human       TTCATCCCAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATCTC  -202
Talapoin    TTCATTCCAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATCTT
Capuchin    TTCATCCCAGCACCCAACATTGCATATCCAGATTATTTGAGCCCAATATT
Porcine     ----------------------------CAGATCCTTTGAGCCCCATCTT
Bovine      TTTATTCCAGCCTGCAGCATCGCATATGTGGATTCTGTGAGCCCAATCTT
Murine      --CCC-AAAGCACCCAGCCTTGT-TATCCAGGTT-TCTGACCTCAATCTC
                                         * *  * *** * * ** *
```

```
                ________________
Human           TTATCCTCTGAAGAACA----CAATCGGCTTTGGGG--CCA-CAGA---    -163
Talapoin        TTATCCTCTGAAGAACA----CAATCAGCTTTGGGG--CCA-CAAA---
Capuchin        TTATCCTCTGAAGAACA----CAATCGGCTTTGGGA--ACA-CAAA---
Porcine         TTATCCTCTGAAGAACACA--CAATCAGCTCTGGGGGACCAACAAAAGG
Bovine          TTATCCTCTGAAGAACACA--CAATCGGCTTTGGGG--CCA-CAAA---
Murine          TTATCCTCTGAAGAACACAATCAATCAGCTTTGGG---CTCATGAA---
                *****************    ***** *** ****          *

                __________                          _______
Human           AGGTTTAGGTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTC-CTTGG   -114
Talapoin        AGGTTTAGGTAGTGGTTTAGCGATTTCTAATCCCAAACTTTGTC-CTTGG
Capuchin        AGGTTTAGGTAGTGGTTTAGGGATTTCTAATCC-AAACTTTGTC-CTTAG
Porcine         AGGTTTAGGTGGGGGCTT-GAGGTCTCTAATCCCAAACTTTGTGGCTTAG
Bovine          AGGTTTAGGTGGGGGTTT-GGGGTCTCTAATCTCAAACTTTGTC-CTTAG
Murine          AGGTTTAGGTGAGAACTAAGAGATCTCTAATCTGGGACTTTGTC-TTTGG
                **********      *  * * * *******    *******   ** *

Human           GAGGTTTAGGATTAGTATTGA--TCATTCACAG-AGCCC--AAGTGT--T  -71
Talapoin        GAGGTTTAGGATTAGTATTGA--TCATTCACAG-AGCCC--AAGTGT--T
Capuchin        GAGGTTTAGGATTAGTGTTCA--TCAGTCACAG-TGCCC--AAGTGT--T
Porcine         GAGGCTTAGGATCAGTACTGAATCCTTTCAGGGGAGCCCGAAAGTGGG-C
Bovine          GAGGCTTAGGATCAGTACTGA--TCTTTCAGGG-AGCCCCAAAGTGG--C
Murine          GAGGCTTAGGAT-AG-----------TTCAG---AGTCTC-AACTGTATC
                **** ******* **            ***     * *   ** **

Human           TTTAGAGGAGGGGTTTTGTGG-------------------GGTGGGAGGA  -40
Talapoin        TTCAGAGGAGGGGTTTTCTGG-------------------GGTGGGAGGA
Capuchin        TTTAAAGGAGGGGTTTTGTGG-------------------GGTGG----A
Porcine         TTTGGAGGAGAGGCTCTTTGGCAGGGTGGGGGGAAGTGGCGGTGGGAGGA
Bovine          TTTGGAGGAGAGGGTCTTTGGCAGGTGGGGGAGAAGTGATGG-GGGAGGA
Murine          TTTAAAGGAG-GGTGCTGT---AG-----GGAAGGGTGATAGCAGAAGAA
                **   ***** **   *                        *  *   *

Human           TCACCTATAAGAGGACTCAGAGGAGGGT-GTG-GGGCATCC---------  -1
Talapoin        TCACCTATAAGAGGACTCAGAGGGGGGT-GTG-GGGCATCC---------
Capuchin        TCATCTATAAGAGGACTCAGAGG--GGT-GTG-GGGTATCC---------
Porcine         TGATCTATAAGAGGAATCAGAGG-GGGT-G-G-GGACATCCC--------
Bovine          TGATCTATAAGAGGAATCAGAGG-GGGT-GGG-GGACATCC---------
Murine          TCGTCTATAAGAGGGACCGGAGGT-GGTAGTGAGGGCATCCCAAGCGGGA
                    **********   * ****  *** * * **  ****
```

Figure 5.10 Alignment of primate and porcine BCP upstream sequences to the published bovine and murine sequences. The alignment extends over the region from the -572/-551F primer to -1 (the first nucleotide 5' prime to the ATG start site). Overlined and underlined motifs are mentioned in Section 5.2.4.

GGATGCCCTAGACCTCTAGGCA -572/-551F primer sequence

5'-AGCCAA-3' Binding site for NF-1 (Nuclear Factor 1)

5'-TGGGTCA -3' (AP-1) Binding site for products of *c-Jun* and *c-Fos*

5'-TGAGCC-3' Binding site for Zeste

5'-AGAACA-3' The reverse complement of this sequence is the binding site for GR, PR and AR transcription factors

position -201 to -185). 11 bp of this 17 bp conserved region is not part of the GR and PR binding site and did not match to any previously identified transcription factor binding sites. The fact that this 17 bp stretch has been selectively maintained between all these species, indicates that this sequence is likely to be involved in an essential biological function. Thus, this region represents a potential binding site for a new as yet unidentified transcription factor or factors, that may be exclusive to the retina or even specific to the photoreceptors.

The TESS program identified three other potential transcription factor binding sites, NF-1 (nuclear factor-1), AP-1 (activator protein-1) and Zeste, which are conserved at equivalent positions, within all but one of the 6 species compared (Figure 5.10). The Zeste (-214 to -209; 5′-TGAGCC-3′) and AP-1 (-290 to -284; 5′-TGGGTCA-3′) binding sites were present in all species except murine at these particular positions (Figure 5.10). The NF-1 binding site (-364 to -359; 5′-AGCCAA-3′) was conserved in all species except bovine (Figure 5.10). The AP-1 motif bind the Jun and Fos protein factors to propagate the enhancement of transcription of genes which contain the AP-1 binding site (as described in Section 5.1.2). NF-1 is a ubiquitous protein factor, like the TATA binding protein, which can bind to the promoter region of a gene containing the appropriate binding site to facilitate the initiation of transcription (Latchman, 1995). Zeste is a protein factor found within the fruit fly *Drosophila* which acts as a homeotic transcriptional regulator (Latchman, 1995).

Other sequence motifs of interest that were not identified by the TESS program, were noted from this alignment by eye. The photoreceptor conserved element (PCE1) was found to be present in the proximal 5′ flanking regions of all the species compared in this study. The PCE1 sequence 5′-CTAATC-3′ is shown in red lettering in Figure 5.10 (position -136 to -131). This PCE1 motif (described in Section 5.1.2) may confer photoreceptor-specific expression of the BCP gene for the six mammalian species analysed in this study.

A number of conserved motifs were not recognised by the TESS program and these may be responsible for enhancing, or even conferring cone-specific, expression. The following conserved elements mentioned below are overlined in Figure 5.10. A 7 bp conserved sequence 5′-ACTTTGT-3′ situated 4 bp downstream of the PCE1, resembles an element known as an E-box. A feature of the E-box is that the outer two conserved nucleotides are in reverse complementation to each other. This kind of dyad symmetry is a characteristic seen in other transcription factor binding

motifs (Latchman, 1995). This 5′-ACTTTGT-3′ motif is similar to an E-box sequence 5′-CATTTG-3 (consensus 5′-CANNTG-3′) which was found within the promoter region of the human rod cGMP-phosphodiesterase β-subunit gene (Di Polo *et al.*, 1996). On the basis of this similarity, and that the 7 bp conserved motif in the BCP gene promoter displays the reverse complement of its two outer nucleotides, this motif is termed 'E-box-like' for ease of reference. The 'E-box-like' motif is conserved in the 5′ flanking region of all the species compared in this study (see Figure 5.10), making it a good candidate for transcription factor binding.

The consensus sequence for an E-box (5′-CANNTG-3′) is conserved in all the species compared in this study, at position -406 to -401 (in green and overlined Figure 5.10). The E-box motif is a binding site for transcription factors with basic-helix-loop-helix (bHLH) DNA binding features, such as the MyoD and E2A proteins (Blackwell and Weintraub, 1990). The sequence of the E-box motif present within the 5′ flanking region of the mammalian BCP genes is 5′-CAGA/GTG-3′ (Figure 5.10).

Most of the conserved sites, identified by eye, exhibit some kind of palindromic or symmetrical feature (overlined Figure 5.10), which is usually indicative of transcription factor binding sites. For example, the conserved sequences 5′-AGGTTTAGG-3′ or 5′-TAGGAT-3′ lie between -162 to -154 and -107 to -102 bp upstream of the ATG start codon, respectively. Both sequences are conserved in all six species (Figure 5.10). Interestingly, the former motif of 9 bp is repeated in the primates and situated within a 13 bp stretch of conserved sequence (5′-GG<u>AGGTTTAGG</u>AT-3′; position -114 to -102; Figure 5.10). In all of the non-primate species, the fourth nucleotide T is substituted for a C within the 9 bp motif at position -112 to -104 (5′-AGGCTTAGG-3′; Figure 5.10). This 9 bp consensus sequence motif 5′-AGG(T/C)TTAGG-3′ may be a core recognition domain for a particular, as yet unidentified, transcription factor. The 5′-TAGGAT-3′ sequence is also present within the 5′ flanking region of the bovine cyclophilin gene (Ferreira *et al.*, 1995). The bovine cyclophilin protein is thought to play a role in the proper folding or the transport of opsin and is found to be predominantly present in cone cells (Ferreira *et al.*, 1995). Interestingly, another 9 bp sequence 5′-T/CT/CTTCCTG/CC/T-3′ lying between -475 and -467 bp upstream of the BCP genes (Figure 5.10) is also found in the 5′ flanking region of the bovine cyclophilin gene (5′-CTTTCCTGC-3′; Ferreira *et al.*, 1995). A similar sequence with a core 5′-TTCCT-3′ motif is also present within the 5′ proximal promoter of the human cone α-PDE subunit gene (5′-TCTTCCTTGC-3′; Piriev *et al.*, 1995).

5.2.5 Identification of putative protein binding sites within the extended porcine and human BCP gene 5′ flanking regions

Additional upstream sequences were obtained for the porcine and human BCP genes (see Section 5.2.1). Motifs identified in this extra region of 5′ flanking sequence may act as enhancer elements which act to increase the level of BCP gene expression.

The computer program TESS was used to analyse 949 bp and 962 bp of 5′ flanking regions obtained for the porcine and the human BCP gene, respectively (see Section 2.2.1 for list of computer software). Figure 5.11 and Figure 5.12 lists the numerous potential transcription factor binding sites that were identified for the human and porcine sequences, respectively. This analysis tested both the sense and anti-sense strand of the DNA for putative protein binding sites. These transcription factor binding sites that are common to both species and situated within conserved regions were noted (see Figure 5.13).

Analysis of the TESS results indicates that there are several potential transcription factor binding sites which are common to both the porcine and human 5′ flanking sequences (such as LyF-1, ER, GR, Sp1 or AP-2, see Figure 5.11 and 5.12). Only 6 of these potential binding sites are situated within conserved regions at equivalent positions within both the human and porcine sequences, namely AP-1, GR/PR, NF-1, Zeste, PuF and PEA3 (Figure 5.13). The former 4 potential binding sites are situated within the promoter regions aligned for all six species analysed in this study (Figure 5.10), and have been mentioned in the previous Section (Section 5.2.4). The PEA3 binding site (5′-AGGAAA-3′) is found at position -631 to -626 within the extra 5′ flanking sequence obtained for the porcine and human BCP genes (Figure 5.13). The PEA3 site is a recognition binding sequence for the Ets protein factor, which is encoded by a proto-oncogene (Latchman, 1995). PEA3 motifs are often found near AP-1 sequences, which also bind oncogene products (Jun and Fos). The 5 bp at the 3′ end of the PEA3 binding site (5′-A<u>GGAAA</u>-3′) is part of another NF-1 binding motif, 5′-GGAAAG-3′ at position -630 to - 625 (NF-1 actually binds to the reverse complement of 5′-GGAAAG-3′; see Figure 5.12). Thus, the PEA3 and NF-1 binding sites overlap via this 5 bp common region in the porcine/human alignment (see Figure 5.13). There are two PuF motifs (5′-

```
GCCTCCCAAAGTGCTGGGATTACAGGCTTGAGCCACCACGCCCGGCCAAC     -913
                                <----- Zeste
                                        <----- Sp1
TCATCTATATATTTATGATCTGCTGAGGTGAAAAGATTCATTCATAAAGC     -863
        ------> ETF
                                  ---------> Pit-1
                                        <-------- TBP
                                        <-------- TFIID
CTAAAGTGTAGTACTGAGTACAAGTGCAGGTGGAACTTAAAGCAGAGTTG     -813
GTGTCACTGGGGTGGGCTGCACAGAAGAGGTGTTTGTGAGGTAGGCTGTG     -763
        <------ PuF
                     -----> GR
AGGGCTAGATAAGACTTCAGTGAACAGACAGGAGGAAGACTGTGGTGACT     -713
     -----> NIT2
      <----- GATA-1
      <----- GATA-2
                 -------> GR
                 -------> PR
                    <----- LVa
                                              -----> GCN4
TGTTGAGCTAAGGGTAGGGGTAGGCATGAACATGGCGTAAGGACAGTGAA     -663
                              -----> UCRF-L
                                     -----> GR
                                          -------> PR
                                             <----- LVa
CAGACCCACCTCCCTGGCTGGAGGTCTGAGTAGGAAAGGCACACTCTTTG     -613
                              <----- PEA3
                               -----> NF-1
                                     -----> GR
AGTAGATTCAACACAAAGAAATGGAGGCAGCCTGGCTCAGGGATGCCCTA     -563
                <------ HiNF-A
                                -----> Zeste
GACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCATCAAGGCTGTT     -513
                                -----> LVa
TACTGATGTGCTTCCAGGGCACTCCCCACTCCCAGCCCTTTCCTGCAGCC     -463
                    ------> p300
                       -----> AP-2
                           ------> p300
                                  <----- NF-1
                                   -----> PEA3
CAGGGCTGGTTCCTAGCCTCTCAGCAGACTTAAGATGGGCACCTTCCACA     -413
AAGGGGCAGATGAGTTGAGGAAAACTTAACTGATACAGTTGTGCCAGAAG     -363
  ------> CAC-binding protein
  ------> NF-E2
   <----- Sp1
                <----- PEA3
                                <----- c-Myc
                                     <----- GR
                                                <----- NF-1
CCAAAATAAGAGGCGTGCCCTTTCTATAGCCCCATTAAAAGAACAAAAAA     -313
                    -------> TBP
                                     -----> AR
                                     -----> GR
                                     -----> PR
GTGGAAGCATCTTCAGTGAATATGGGTCAGCACCTCCCAGACCTCAGGGA     -263
                      <------ c-Fos
                      <------ c-Jun
                      <------ ER
                                <-------- LyF-1
GTCCACTTCTGTTCATCCCAGCACCCAGCATTGCATATCCAGATTATTTG     -213
        -----> LVa
                              <------- Pit-1
                               -----> EFII
                                                <----- Zeste
AGCCCAATCTCTTATCCTCTGAAGAACACAATCGGCTTTGGGGCCACAGA     -163
   -----> CCAAT-binding factor
   -----> CP1
   -----> CTF
       <------ c-Myc
                      -----> AR
                      -----> GR
                      -----> PR
                                     <----- Sp1
AGGTTTAGGTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTCCTTGGG     -113
                <----- TBF1
                     <----- H4TF-1
                                        <----- GR
AGGTTTAGGATTAGTATTGATCATTCACAGAGCCCAAGTGTTTTTAGAGG     -63
              ------> Pit-1
                                            <------- C/EBP
                                              ------> Sp1
AGGGGTTTTGTGGGGTGGGAGGATCACCTATAAGAGGACTCAGAGGAGGG     -13
            <------ PuF
                                          <------- C/EBP
                                              <------- TEF-2
TGTGGGGCATCC                                            -1
```

Pos	M	Tf AccNo	Sequence	Factor Name
-15	N	R02251	GGGTGTGG	TEF-2
-20	N	R03657	GAGGAGGG	C/EBP
-50	N	R02239	GGGTGGG	PuF
-64	C	R02029	GGAGGGG	Sp1
-66	N	R03657	GAGGAGGG	C/EBP
-98	C	R02791	TATTGAT	Pit-1
-122	N	R01030	TGTCCT	GR
-141	N	R00680	GATTTC	H4TF-1
-146	N	R03755	TTAGGG	TBF1
-173	N	R04195	GGGGCC	Sp1
-190	C	R01121	AGAACA	GR
-190	C	R01117	AGAAC	PR
-190	C	R01101	AGAACA	AR
-205	N	R02207	TCTCTTA	c-Myc
-209	C	R00768	CCAATC	CTF
-209	C	R00768	CCAATC	CP1
-209	C	R00765	CCAATC	CCAAT-binding
-214	N	R01599	TGAGCC	Zeste
-231	C	R01342	TGCATA	EFII
-232	N	R02783	TTGCATA	Pit-1
-254	C	R01135	CTGTTC	LVa
-280	N	R03510	CCTCCCAGA	LyF-1
-290	N	R01195	TGGGTCA	c-Jun
-290	N	R01195	TGGGTCA	c-Fos
-290	N	R01195	TGGGTCA	ER
-323	C	R01121	AGAACA	GR
-323	C	R01117	AGAACA	PR
-323	C	R01101	AGAACA	AR
-342	C	R03165	TTTCTATA	TBP
-367	N	R01681	AGCCAA	NF-1
-373	N	R03535	TGTGCC	GR
-377	N	R02893	CAGTTG	c-Myc
-395	N	R02095	AGGAAA	PEA3
-409	N	R01540	GGGCAG	Sp1
-410	C	R01251	GGGGCAG	NF-E2
-410	C	R01251	GGGGCAG	CAC-binding
-474	C	R02095	TTTCCT	PEA3
-475	N	R00802	CTTTCC	NF-1
-486	C	R04098	CACTCCC	p300
-490	C	R01054	TCCCCA	AP-2
-493	C	R04098	CACTCCC	p300
-531	C	R01135	CTGTTC	LVa
-579	C	R01599	GGCTCA	Zeste
-596	N	R00686	AGAAATG	HiNF-A
-625	C	R03535	GGCACA	GR
-630	C	R00802	GGAAAG	NF-1
-631	N	R02095	AGGAAA	PEA3
-665	N	R01135	GAACAG	LVa
-668	C	R01553	AGTGAACA	PR
-673	C	R01030	AGGACA	GR
-681	C	R01148	ATGGCG	UCRF-L
-717	C	R00831	TGACTT	GCN4
-741	N	R01135	GAACAG	LVa
-744	C	R01554	AGTGAACA	GR
-744	C	R01553	AGTGAACA	PR
-756	N	R03884	AGATAA	GATA-2
-756	N	R03884	AGATAA	GATA-1
-757	C	R02728	TAGATA	NIT2
-790	C	R01313	AGAAGA	GR
-803	N	R02239	GGGTGGG	PuF
-872	N	R03167	TTCATAAA	TBP
-872	N	R03166	TTCATAAA	TFIID
-877	C	R01263	ATTCATTCAT	Pit-1
-903	C	R01371	TATTTAT	ETF
-925	N	R01498	ACGCCC	Sp1
-934	N	R01599	TGAGCC	Zeste

Pos: indicates the postion of the transcription factor binding upstream of the ATG start site within human BCP gene (-1 being the first bp 5′ to the A of the ATG).

M: indicates whether the binding site is either N (normal) or C (reverse-complement).

Transfac Site Accession Number: This is the Transfac binding site accession number. You can click on accession numbers to browse their Transfac entry on the WWW.

Sequence: This is the binding site sequence taken from Transfac.

Name: This is the name of the factor, taken from Transfac. You can click on the names to browse the Transfac entries on the WWW.

Figure 5.11 TESS results for 962 bp of 5′ flanking region of the human BCP gene. The data presented is adapted from the output generated by the TESS program. The TESS program has compared the human sequence (given above) and matched it with transcription factor binding sites (given under the sequence) which have been placed on the Transfac database (listed above and to the right of the human sequence).

```
CCTGTATCTAAAGGGTGGCACTGTTATCTATGTATCTATAGGCAACTGAG     -900
   <----- NIT2
                        -----> GATA-1
                        -----> GATA-2
                         <----- NIT2
                                 <----- NIT2
                                                 -----> c-Myc
GTGGGAGAGAAGAGACACTCGTAAGGCCCAAAGTGTAACGTTGAATATAA     -850
       -----> GR
                 -----> Zeste
                                              <----- GAL4
GCGTGAGCAGAACTTAGAGCCATTCTGGGGTCCCTGGGGTGGGCTTCTCA     -800
        -----> GR
                                        <------ PuF
TCAGAGGTCTTTGTGAGTCATTCTTAGAGCTGAGCAGGTCTTGGATGAGC     -750
             ------> AP-1
             ------> c-Fos
             ------> c-Jun
             ------> FraI
             ------> NF-E2
             ------> v-Jun
              -----> GCN4
                                     <----- TGT3
AGACAGGAGGAAGCCTGGCAGCTTGGGCACTGCTCAGCTAAGGGCAGGGG     -700
      <------- c-Ets-1 54
      <------- c-Ets-1 68
      <------- c-Ets-1
      <------- PEA3
                                           <----- Sp1
                                                 ------> CAC
                                                 ------> NF-E2
                                                  <----- Sp1
CAGGTGTGAACATGGCATGAGACGGACAGAGAAAAAACCAACTTCAACTG     -650
                                              -----> c-Myc
GCCTGGGGACATCTGAATAGGAAAGATACCTCTCTGGGGAGACACACCAT     -600
   <----- AP-2
     -----> GR
      ------> SRF
                  <----- PEA3
                   -----> NF-1
                          <------- UBP-1
                                <----- AP-2
                                   <----- GR
                                                  -----> GR
AGAAGAAACTGGCAGCCTGGATGCCTTAGATCTCTAGGCATTGGCGAATC     -550
                                         ------> CTF
                                           ------> NF-1
                                            <----- F-ACT1
                                                -----> GCN4
ACCCTTGAACTCTTCCATGAAGGCTCTTGCTAAGGTGCTCCAGACACTCT     -500
    <----- GR
CCATTCCCTGTCTCTTCCTGCAGCCCGAGGCTGGTTCCCAGGCTGCTGTG     -450
                                                 <----- Zeste
AGCCCAAGATCCCTGGAGAGGGAGGCTTAGGGGGTACCTTCCACAAAGGG     -400
                            <----- TBF1
                                                 -----> GR
ACAGATGGGGTACAGACACCTATAGGGATTTAGTGGTGCCTGAAGCCAAA     -350
                                            <----- NF-1

TGAAGAGGGGTGCCCTTGAGAGTTCTACTCAAAGGGGAAAAAAGTGGTGG     -300
          <------- myogenin
                    <----- GR
AAGGATCTTTATTAAATATGGGTCAGTATCTCCCTCCCAGGCCCTAGATC     -250
        ------> TBP
                 <------ c-Fos
                 <------ c-Jun
                 <------ ER
                         <----- NIT2
                              ------> MAZ
CTTTGAGCCCCATCTTTTATCCTCTGAAGAACACACAATCAGCTCTGGGG     -200
   <----- Zeste
              -----> TFIID
                           -----> AR
                           -----> GR
                           -----> PR
GACCAACAAAAGGAGGTTTAGGTGGGGGCTTGAGGTCTCTAATCCCAAAC     -150
TTTGTGGCTTAGGAGGCTTAGGATCAGTACTGAATCCTTTCAGGGGAGCC     -100
CGAAAGTGGGCTTTGGAGGAGAGGCTCTTTGGCAGGGTGGGGGGAAGTGG      -50
                            <----- NF-1/L
                                ---------> CACCC-binding-f
                                 ---------> Yi
                                  <------ PuF
CGGTGGGAGGATGATCTATAAGAGGAATCAGAGGGGGTGGGGACATCCC        -1
                          -----> GCN4
                                   <------ PuF
                                     <----- AP-2
                                       -----> GR
                                        ------> SRF
```

Pos	M	Tf AccNo	Sequence	Factor Name
-9	C	R00465	GGACATC	SRF
-10	C	R03536	GGGACA	GR
-12	N	R01054	TGGGGA	AP-2
-15	N	R02239	GGGTGGG	PuF
-25	C	R00832	GAATCA	GCN4
-65	N	R02239	GGGTGGG	PuF
-67	C	R00531	CAGGGTGGGG	CACCC-binding factor
-71	N	R00078	TTGGCA	NF-1/L
-222	C	R01121	AGAACA	GR
-222	C	R01117	AGAACA	PR
-222	C	R01101	AGAACA	AR
-234	C	R01944	TTTATC	TFIID
-246	N	R01599	TGAGCC	Zeste
-268	C	R02306	CCCTCCC	MAZ
-273	N	R02729	TATCTC	NIT2
-281	N	R01195	TGGGTCA	c-Jun
-281	N	R01195	TGGGTCA	c-Fos
-281	N	R01195	TGGGTCA	ER
-288	C	R00061	TTAAATA	TBP
-329	N	R01313	AGTTCT	GR
-339	N	R00018	TGCCCTTG	myogenin
-356	N	R01681	AGCCAA	NF-1
-402	C	R03536	GGGACA	GR
-423	N	R03755	TTAGGG	TBF1
-451	N	R01599	TGAGCC	Zeste
-544	N	R01815	TGAACT	GR
-554	C	R00832	GAATCA	GCN4
-558	N	R02154	TGGCGA	F-ACT1
-559	C	R03434	TTGGCGA	NF-1
-561	C	R00956	CATTGGC	CTF
-599	C	R01313	AGAAGA	GR
-609	N	R03537	GACACA	GR
-615	N	R01054	TGGGGA	AP-2
-620	N	R02258	CTCTCTGG	UBP-1
-630	C	R00802	GGAAAG	NF-1
-631	N	R02095	AGGAAA	PEA3
-643	C	R00465	GGACATC	SRF
-644	C	R03536	GGGACA	GR
-646	N	R01054	TGGGGA	AP-2
-655	C	R02893	CAACTG	c-Myc
-702	N	R01540	GGGCAG	Sp1
-703	C	R01251	GGGGCAG	NF-E2
-703	C	R01251	GGGGCAG	CAC-binding protein
-708	N	R01540	GGGCAG	Sp1
-765	N	R02680	CAGGTC	TGT3
-785	C	R00655	GAGTCA	GCN4
-786	C	R01946	TGAGTCA	NF-E2
-786	C	R01037	TGAGTCA	v-Jun
-786	C	R01037	TGAGTCA	c-Jun
-786	C	R01037	TGAGTCA	c-Fos
-786	C	R01037	TGAGTCA	FraI
-786	C	R01037	TGAGTCA	AP-1
-813	N	R02239	GGGTGGG	PuF
-841	C	R01313	AGAACT	GR
-855	N	R00497	ATATAA	GAL4
-884	C	R01599	CACTCG	Zeste
-892	C	R01313	AGAAGA	GR
-907	C	R02893	CAACTG	c-Myc
-917	N	R02728	TATCTA	NIT2
-925	N	R02728	TATCTA	NIT2
-926	C	R03884	TTATCT	GATA-2
-926	C	R03884	TTATCT	GATA-1
-945	N	R02728	TATCTA	NIT2

Pos: indicates the postion of the transcription factor binding upstream of the ATG start site within porcine BCP gene (-1 being the first bp 5′ to the A of the ATG).

M: indicates whether the binding site is either N (normal) or C (reverse-complement).

Transfac Site Accession Number: This is the Transfac binding site accession number. You can click on accession numbers to browse their Transfac entry on the WWW.

Sequence: This is the binding site sequence taken from Transfac.

Name: This is the name of the factor, taken from Transfac. You can click on the names to browse the Transfac entries.

Figure 5.12 TESS results for 949 bp of 5′ flanking region of the porcine BCP gene. The data presented is adapted from the output generated by the TESS program. The TESS program has compared the porcine sequence (given above) and matched it with transcription factor binding sites (given under the sequence) which have been placed on the Transfac database (listed above and to the right of the porcine sequence).

```
Human                                5' GCCT---CCCAAAGTGCTGGGA--   -943
Porcine                              5' -CCTGTATCTAAAG-GGTGGCACT
                                        ***    * **** * *** *

Human     -TTACAGCTTGAGCCACCACGCCCGGCCAACTCATCTATATATTTA-TG   -896
Porcine   GTTA--TCT------------------ATGT-ATCTATA-GGCAACTG
           ***   **                  *  * *******     * **

Human     ATCTGCTGAGGTGAAAAGATTCATTCATAAAGCCTAAAGTGTAGTACTG   -847
Porcine   AGGTGG-GAGA-GAAGAGA--CACTCGTAAGGCCCAAAGTGTAACGTTG
          *  **  ***  *** ***  ** ** *** *** ********    **

Human     AGTACAAG--TGCAGGTGGAACTTAAAGC-AGAGTTGGTGTCACTGGGG   -801
Porcine   AATATAAGCGTG-AGC-AGAACTTAGAGCCAT-TCTGGGGTCCCTGGGG
          * ** ***  ** **   ******* *** *    *** *** ******

Human     TGGGCTGCACAGAAGAGGTGTTTGTGAGGTAGGCTGTGAGGGCTAGATA   -752
Porcine   TGGGCTTCTCATCAGAGGTCTTTGTGAGTCATTCT-T-AGAGCT-GAGC
          ****** * **  ****** ********  *  ** * ** *** **

Human     AGA-CTTCAGTGAACAGACAGGAGGAAGACT----G--TGGTGACTTGT   -710
Porcine   AGGTCTTGGATGAGCAGACAGGAGGAAGCCTGGCAGCTTGGGCACT-GC
          **  ***   *** ************** **    *  ***  *** *

Human     TGAGCTAAGGGTAGGGGTAGGCATGAACATGGCGT-A-A-GGACAGTGA   -664
Porcine   TCAGCTAAGGGCAGGGGCAGGTGTGAACATGGCATGAGACGGACAGAGA
          * ********* ***** ***  ********** * * * ****** **

Human     ACAGACCCACCTCC-CTGG-CTGG--AGGTCTGAGTAGGAAAGGCA-CA   -620
Porcine   AAAAACCAACTTCAACTGGCCTGGGGACATCTGAATAGGAAAGATACCT
          * * *** ** **  **** ****  *  ***** ********  * *

Human     CTCTTT-GAGT-AGATTCAACACAAAGAAATGGAGGCAGCCTGGCTCAG   -573
Porcine   CTCTGGGGAGACACAC-C-ATAG-AAGAAACT--GGCAGCCT-------
          ****   ***  * *  * * *  ******    ********

Human     GGATGCCCTAGACCTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCA   -524
Porcine   GGATGCCTTAGATCTCTAGGCATTGGCGAATCACCCTTGAACTCTTCCA
          ******* **** ************ * * *  **   *  ** *****

Human     TCAAGGCTGTTTACTGATGTGCTTCCAGGGCACTCCCCACTCCCAGCCC   -475
Porcine   TGAAGGCTCTT-GCTAAGGTGCT-CCA-GACACTCTCCATTCCCTGTCT
          * ****** **  ** *  **** ***   ***** *** ****   *

Human     TTTCCTGCAGCCCAGGGCTGGTTCCTAGCCT-CTC--AGC---AGA--C   -434
Porcine   CTTCCTGCAGCCCGAGGCTGGTTCCCAGGCTGCTGTGAGCCCAAGATCC
           ************  ********** ** ** **   ***   ***  *

Human     TT----A---AG-AT----GGGCACCTTCCACAAAGGGGCAGATGAGTT   -397
Porcine   CTGGAGAGGGAGGCTTAGGGGGTACCTTCCACAAAGGGACAGATGGGGT
           *    *   **       *** *************** ****** * *

Human     GAGGAAAACTTAACTGATACAGTTGTGCCAGAAGCCAAAATAAGAGGCG   -348
Porcine   ACAGACACCTATAGGGATTTAGTGGTGCCTGAAGCCAAATGAAGAGGGG
             ** * **  *  ***  *** ***** *********  ****** *
```

```
Human     TGCCCTTTCTATAGCCCCATTAAAAGAACAAAAAAGTGG---AAGCATC  -302
Porcine   TGCCCTT--GAGAGTTCTACTCAAAGGGGAAAAAAGTGGTGGAAGGATC
          *******   * **  * * * ****   **********   *** ***

Human     TTCAGTGAATATGGGTCAGCACCTCCCAGACCTCAGGGAGTCCACTTCT  -253
Porcine   TTTATTAAATATGGGTCAGTATCTCCC--TCC-CAGG------------
          ** * * ************ * *****   ** ****

Human     GTTCATCCCAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATC  -204
Porcine   -------------CCC-------------TAGATCCTTTGAGCCCCATC
                      ***              ****  ********* ***

Human     TCTTATCCTCTGAAGAACA--CAATCGGCTTTGGGG--CCA-CAGAAGG  -160
Porcine   TTTTATCCTCTGAAGAACACACAATCAGCTCTGGGGGACCAACAAAAGG
          * *****************  ***** *** *****  *** ** ****

Human     ---TTTAGGTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTC-CTTG  -115
Porcine   AGGTTTAGGTGGGGGCTT-GAGGTCTCTAATCCCAAACTTTGTGGCTTA
             ******* * ** ** * * * *****************  ***

Human     GGAGGTTTAGGATTAGTATTGA-TC-ATTCACAG-AGCCC--AAGTGT-  -72
Porcine   GGAGGCTTAGGATCAGTACTGAATCCTTTCAGGGGAGCCCGAAAGTGGG
          ***** ******* **** *** **  ****  * *****  *****

Human     TTTTAGAGGAGGGG-T-TTTGTG-GGGTGGG----------------AG  -42
Porcine   CTTTGGAGGAGAGGCTCTTTGGCAGGGTGGGGGGAAGTGGCGGTGGGAG
           *** ****** ** * ****   *******                **

Human     GATCACCTATAAGAGGACTCAGAGGAGGGTGTGGGGCATCC-ATG
Porcine   GATGATCTATAAGAGGAATCAGAGG-GGGTG-GGGACATCCCATG
          *** * *********** ******* ***** *** *****  ***
```

Figure 5.13 Sequence alignment of the human and porcine blue cone pigment gene 5′ flanking promoter region. The ATG start-codon and the TATA box is shown in bold. Potential transcription factor binding sites which are situated within conserved regions, are listed below (see Section 5.2.42).

GGGTGGG — Binding site for PuF

AGGAAA — Binding site for PEA3

GGAAAG — The reverse complement of this sequence is a binding site for NF-1

GGATGCCCTAGACCTCTAGGCA — -572/551F primer sequence
GCACCTTCCACAAAGGGGCAG — -424/-404F primer sequence

AGCCAA — Binding site for NF-1 (Nuclear Factor 1)

TGGGTCA — (AP-1) Binding site for products of *c-Jun* and *c-Fos*

TGAGCC — Binding site for Zeste

AGAACA — The reverse-complement of this sequence is a binding site for GR, PR and AR

GGGTGGG — Binding site for PuF

GGGTGGG-3′) within the human/porcine sequences, at positions -803 to -797 and -50 to -44. The PuF motif situated at -50 to -44 is also present at this position within the promoter region of the talapoin monkey, but is not found within in any of the other species analysed in this study (Figure 5.10). The PuF transcription factor is known to bind to the promoter and regulate the expression of the human *c-myc* gene *in vitro* (Postel *et al.*, 1993). As yet there is no evidence to suggest that PuF plays a role in the expression of the human BCP gene.

5.2.6 Characterisation of the Weri-Rb1 cell line

The Weri-Rb1 cells were grown successfully in RPMI media supplemented with Glutamax, foetal calf serum and antibiotics as described in Section 2.15.1. Initially the Weri-Rb1 cells, when grown in RPMI media supplemented with L-glutamine instead of Glutamax, grew very poorly or stopped proliferating altogether. This poor growth of the cells was thought to be the result of unused L-glutamine in the media being broken down to components that inhibited the growth of the cells (Gibco BRL, 1995). Glutamax is a peptide (L-Alanyl-L-Glutamine) and is not broken down if it is not completely utilised by the growing cells. Cells cultured in media containing Glutamax doubled in number once every two to three days, and required splitting into fresh media twice a week (as opposed to fresh media four times a week when grown with L-glutamine).

Weri-Rb1 cells were cultured *in vitro*, as described in Section 2.15.1 Cultured Weri-Rb1 cells were harvested and poly-A^+ mRNA was isolated as described in Section 2.16. Single stranded complementary DNA (cDNA) was prepared from this poly-A^+ mRNA by RT-PCR (as described in Section 2.17). The synthesised cDNA was used as template in PCR analysis of Weri-Rb1 transcripts. PCR amplification reactions with the human blue, green and red cone opsin specific primers (Section 2.2.1 and see primers table) gave products of the expected size whereas the rod opsin primers gave no detectable product (Figure 5.14). Sequencing analysis of these PCR products verified that they were opsin transcripts from cone pigment genes.

Thus, it was confirmed that Weri-Rb-1 cells cultured *in vitro* do express cone opsin genes and in particular that they do express the BCP gene. This implies that the Weri-Rb1 cells can be utilised as a source of the regulatory proteins which are involved in the control of expression of the BCP gene. Hence, proteins isolated from the nuclei of Weri-Rb1 cells

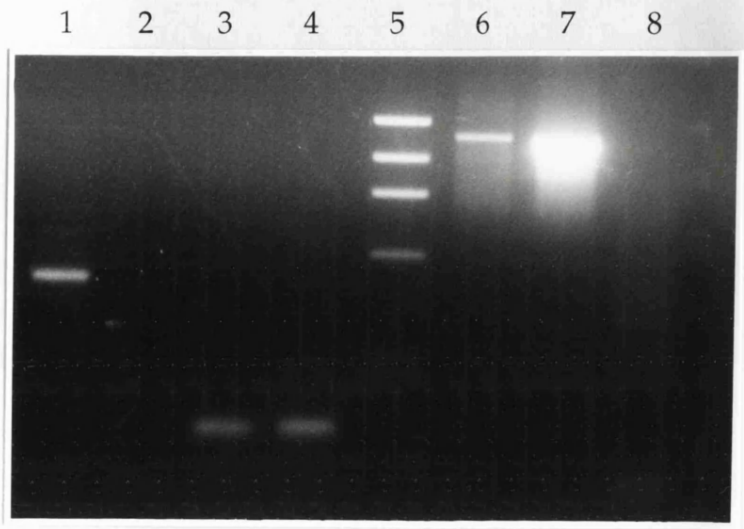


Figure 5.14 Amplification products for the human blue, red and green cone pigment genes from cDNA prepared from Weri-Rb1 cells.

		Opsin Specific Primers	
Lane	DNA Template	Forward	Reverse
1	Weri-Rb1 cDNA	BCP Ex2+784/799F	BCP Ex4-1276/1254R
2	Weri-Rb1 cDNA	Rod Ex4+	Rod Ex5-
3	Weri-Rb1 cDNA	RCP Ex4+	RCP Ex5-
4	Weri-Rb1 cDNA	GCP Ex4+	GCP Ex5-
5		ϕX174 DNA Ladder	
6	Human genomic	Rod Ex4+	Rod Ex5-
7	Porcine BCP cosmid	BCP Ex2+784/799F	BCP Ex4-1276/1254R
8	None (water)	BCP Ex2+784/799F	BCP Ex4-1276/1254R

Rod = Rod opsin; BCP = Blue cone pigment; RCP = Red cone pigment;
GCP = Green cone pigment;
See Table 2.1 and 2.4 for the sequence of the primers given above.

The Rod Ex4+ and Rod Ex5- primers are designed to amplify exons 4 and 5 of the human rod opsin gene which represents 240 bp (exon 4) and 108 bp (exon 5) which should result in a 348 bp product after amplification of cDNA or a 1181 bp genomic fragment (Lane 6; which includes 833 bp of intron 4). The RCP and GCP primers give a 159 bp cDNA product (Lanes 3 & 4). The Ex2+784/799F and Ex4-1276/1254R primers amplify from exons 2 to 4 for the human or porcine BCP gene resulting in a 498 bp cDNA product (Lane 1) or a 1120 bp genomic product for the porcine BCP (Lane 7).

should contain transcription factors involved in the expression of the BCP gene.

5.2.61 Isolation of nuclear proteins from Weri-Rb1 and fibroblast cell lines

Nuclear proteins were isolated from cultured Weri-Rb1 cells as described in section 2.18.1. Nuclear proteins were also obtained from a fibroblast cell line, originally isolated from Tenon's capsule (Section 2.15.2) for use as a control in electrophoretic band shift assays (EBSA). The concentration of isolated proteins was quantified using the Bradford assay system (section 2.18.2 and Table 5.2a). The concentrations for both the Weri-Rb1 and fibroblast nuclear extracts were found to be approximately 0.5 μg/μl (Table 5.2b). An aliquot of these extracts was electrophoresed on a denaturing polyacrylamide gel (as described in Section 2.18.3) for visualisation of the size range of proteins (see Figure 5.15). The Weri-Rb1 nuclear proteins ranged in size from about 14 kD to greater than 64 kD (Figure 5.15).

5.2.7 Electrophoretic band shift assays (EBSA) on the human BCP gene promoter

EBSA required the amplification of small contiguous fragments of DNA which spanned the length of the isolated human BCP gene 5′ flanking region (Section 5.2.1). Radioactive labelled DNA fragments, consisting of twelve overlapping regions spanning a total of 953 bp of the human BCP 5′ proximal promoter region, were generated by PCR (as described in Section 2.2.1). The sequence of the forward and reverse primers used to generate these overlapping fragments are given in Table 2.2. For ease of reference, these overlapping DNA fragments were termed fragments 1 to 12 (F1 to F12). F1 to F12 are all between 71 to 285 bp in length. Figure 5.16 shows the size and relative position of the generated fragments within the human 5′ flanking promoter region. The forward primers were end labelled with [γ ^{32}P] ATP, as described in Section 2.19.1, prior to amplification of these upstream fragments. 1 μl of each PCR reaction (each fragment) was run on an 6% non-denaturing polyacrylamide gel and the gels were dried on to filter paper before autoradiography (as described in Sections 2.19.2 and 2.11). The relative radioactivity of each amplified fragment was estimated from the autoradiograph, by eye, to ensure that the concentration of each fragment was approximately the same (Figure 5.17). Equal concentrations of each of

a

Bovine Serum Albumin Protein Stock (μg)	Absorption Reading at 595nm
0	0
0.1	0.070
1	0.135
5	0.392
10	0.572
15	0.822
20	0.939

b

10μl of Isolated Nuclear Protein Extract	Absorption Reading at 595nm
Weri-Rb1 cells	0.390
Fibroblast cells	0.377

Table 5.2 a Absorption readings at 595nm for stock bovine serum albumin protein concentrations, (**b**) readings obtained for the Weri-Rb1 and fibroblast nuclear protein extracts, using the Bradford protein concentration assay (as described in Section 2.18.2). These readings were used to estimate the protein concentrations of the isolated nuclear extracts (see Section 5.2.61).

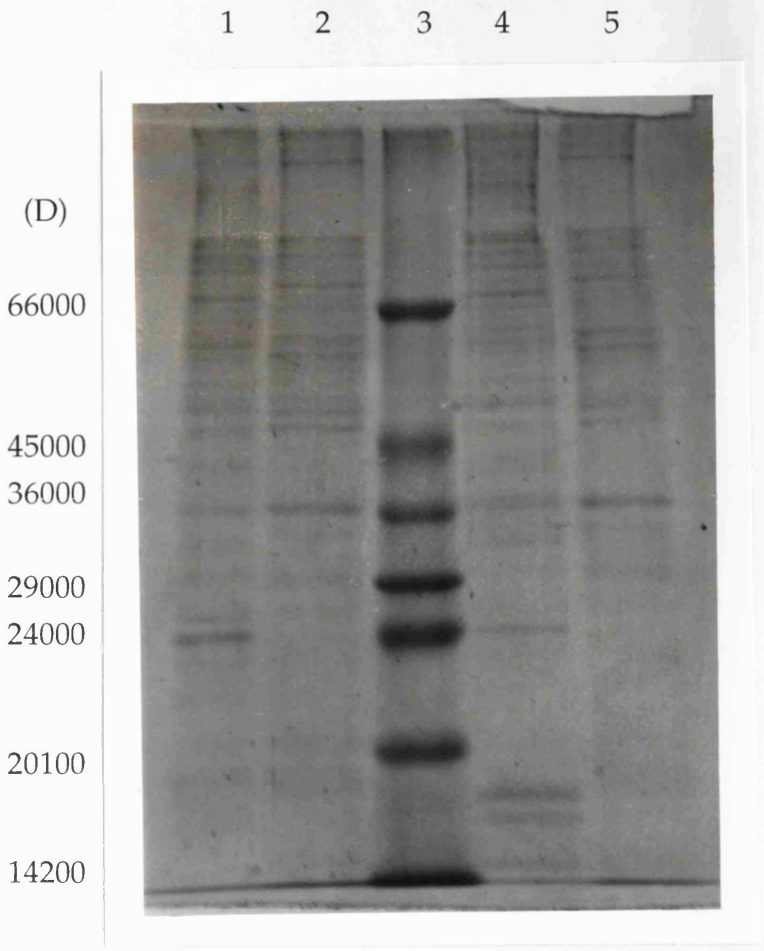


Figure 5.15 SDS-PAGE showing isolated nuclear protein extracts for the Weri-Rb1 and a fibroblast cell line.

Lane 1: Weri-Rb1 extract
Lane 2: Fibroblast extract
Lane 3: Protein standards (D)
Lane 4: Weri-Rb1 extract
Lane 5: Fibroblast extract

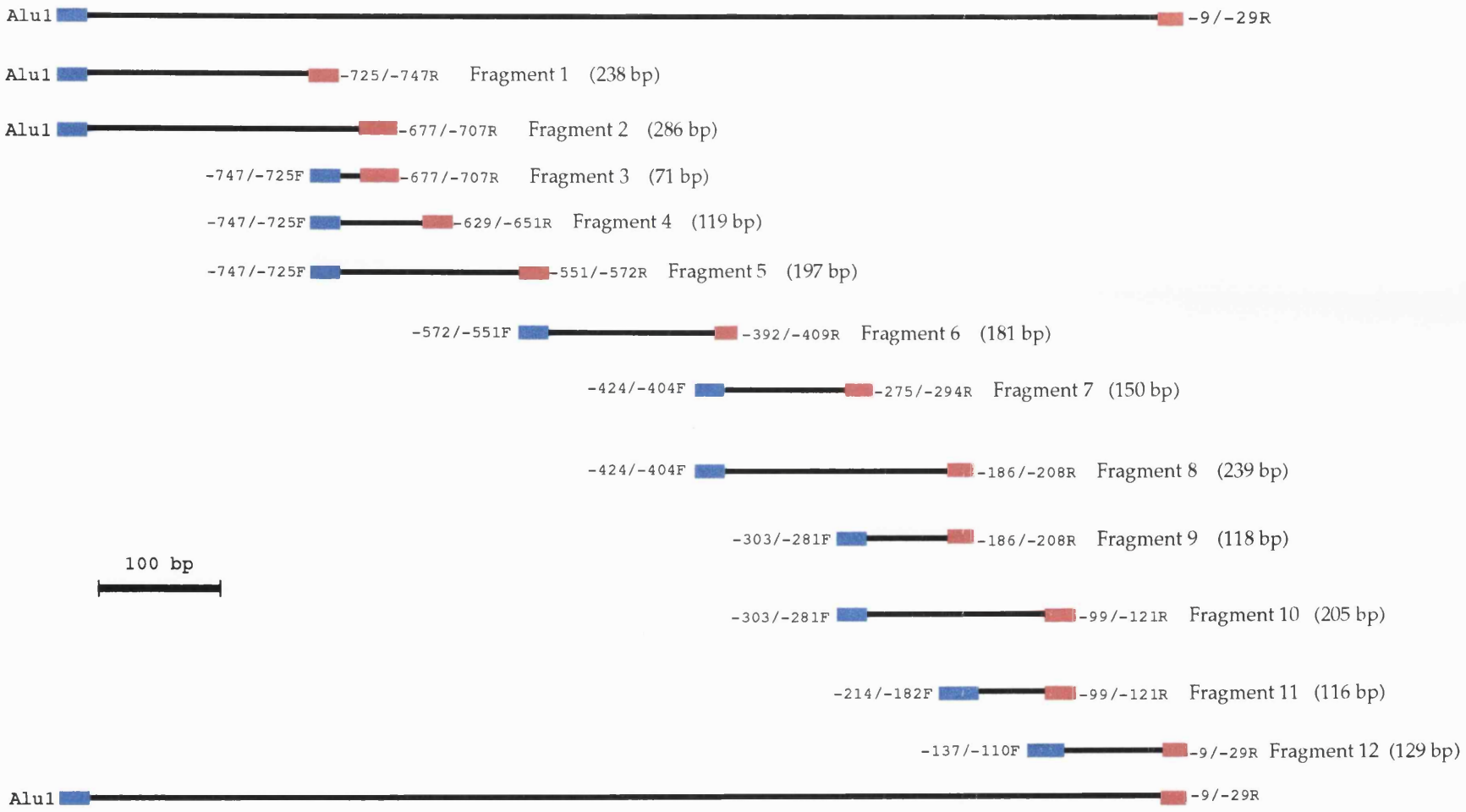


Figure 5.16 Twelve overlapping fragments generated by PCR, representing 954 bp of the 5' flanking region of the human blue cone pigment gene. ▬ indicates the forward (F) primer; ▬ indicates the reverse (R) primer. These amplified 5' flanking pieces of DNA were termed Fragments 1 to 12 (F1 to F12), and were utilised in electrophoretic band shift assays (see Section 5.2.7). Alu1 is a consensus forward primer designed to an *Alu* repeat element.

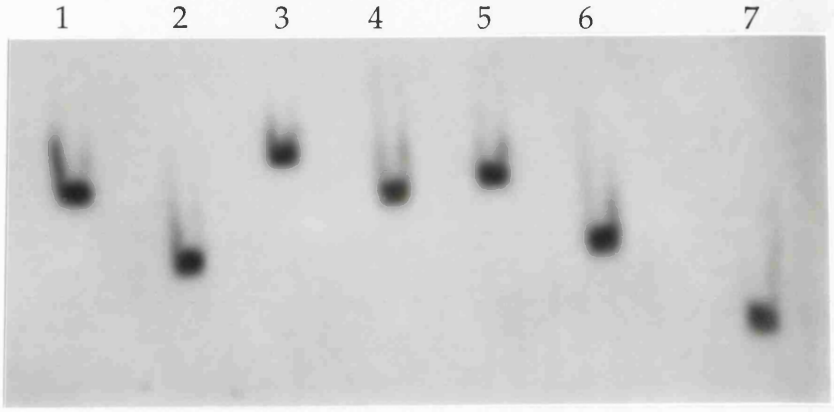


Figure 5.17 Autoradiograph showing equal amounts of radioactive labelled 'Fragments' of BCP 5' flanking sequence. These fragments were run on a polyacrylamide gel to check that the concentration of each Fragment (F) was approximately the same before proceeding with electrophoretic band shift assay experiments.

Lane 1: F10 (205 bp)
Lane 2: F9 (118 bp)
Lane 3: F8 (239 bp)
Lane 4: F6 (181 bp)
Lane 5: F5 (197 bp)
Lane 6: F4 (119 bp)
Lane 7: F3 (71 bp)

the twelve radioactive labelled PCR fragments were then used in separate protein to DNA binding reactions, for EBSA (as described Section 2.19.3).

Problems were encountered in optimising the conditions for binding of protein to the DNA. Initially, the binding reactions (see Section 2.9.3 for contents of binding reaction) were incubated on ice for 45 min prior to electrophoresis (as recommended by Latchman, 1993). Incubation at room temperature for 45 min was also tried, but both these binding conditions failed to give consistent results. Eventually, it was found that the fairly consistent results was obtained when the binding reaction was carried out for 20 min at room temperature then for 20 min on ice before electrophoresis on a pre-run 4% polyacrylamide gel at 4°C (see Section 2.9.3 for full description of binding reaction and EBSA).

All the fragments F1 to F12 were found to give some kind of shift, when mixed with the Weri-Rb1 nuclear protein extract (Figure 5.18, 5.19 and 5.20). This suggests that all these fragments contain binding sites for proteins that are found within the nucleus of Weri-Rb1 cells. The band shifts seen in Figure 5.19 are clearer and more pronounced then those observed in Figure 5.18. The binding reaction for the EBSA seen in Figure 5.19 was incubated at room temperature (RT) for 20 min prior to an incubation of 20 min on ice (as described in Section 2.9.3). Whereas, for the EBSA seen in Figure 5.18, the binding reaction incubation step was done completely on ice for 45 min. Thus, it appears the band shifts which were obtained were dependent on the type of binding conditions applied. Also a greater number of DNA/protein binding complexes were observed, for a particular fragment, when an incubation at room temperature was included in the binding reaction. Note the difference in DNA binding patterns observed in Figure 5.19 for F3, F4, F6, F8, F9, and F10, when bound to the Weri-Rb1 nuclear proteins, compared to the band shifts seen in Figure 5.18. Thus, it seems that a binding reaction with an incubation step at room temperature and on ice somehow produces a more favourable environment for the action of DNA binding proteins, which resulted in cleaner and more visible band shifts.

Two strong-shifted DNA-protein complexes can be seen for F3 when mixed with the Weri-Rb1 nuclear extract (Figure 5.19, Lane 1; Figure 5.20, Lane 4). Two shifted bands from one single fragment suggests that this piece of DNA contains two protein binding sites, and that each shifted band is generated from a protein or proteins interacting with one or the other site. Another interpretation is that F3 contains only one binding

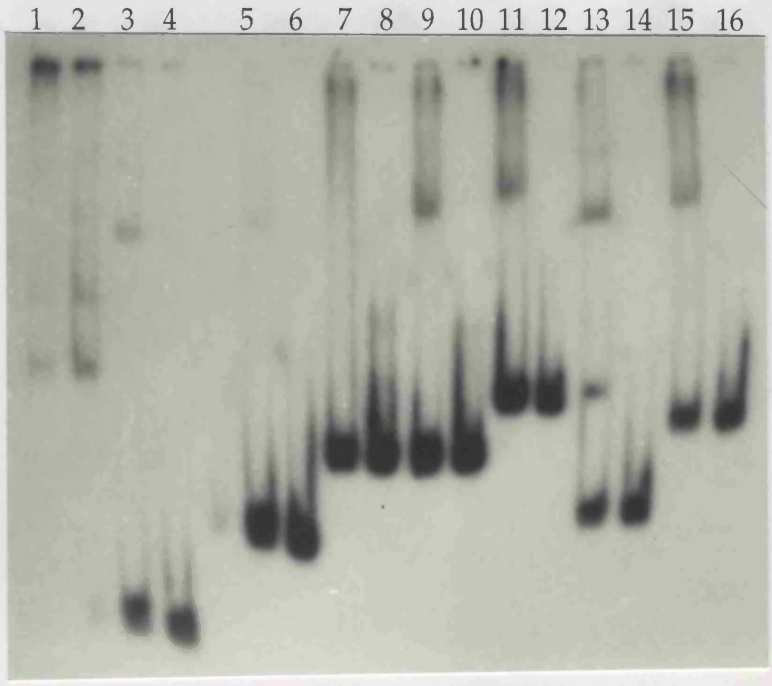


Figure 5.18 Electrophoretic band shift assay (EBSA) on F2, F3, F4, F5, F6, F8, F9 and F10 with Weri-Rb1 nuclear extracts.

Lane 1: F2 plus Weri-Rb1
Lane 2: F2 only (285 bp)
Lane 3: F3 plus Weri-Rb1
Lane 4: F3 only (71 bp)
Miss a lane on gel
Lane 5: F4 plus Weri-Rb1
Lane 6: F4 only (119 bp)
Lane 7: F5 plus Weri-Rb1
Lane 8: F5 only (197 bp)
Lane 9: F6 plus Weri-Rb1
Lane 10: F6 only (181 bp)
Lane 11: F8 plus Weri-Rb1
Lane 12: F8 only (239 bp)
Lane 13: F9 plus Weri-Rb1
Lane 14: F9 only (118 bp)
Lane 15: F10 plus Weri-Rb1
Lane 16: F10 only (205 bp)

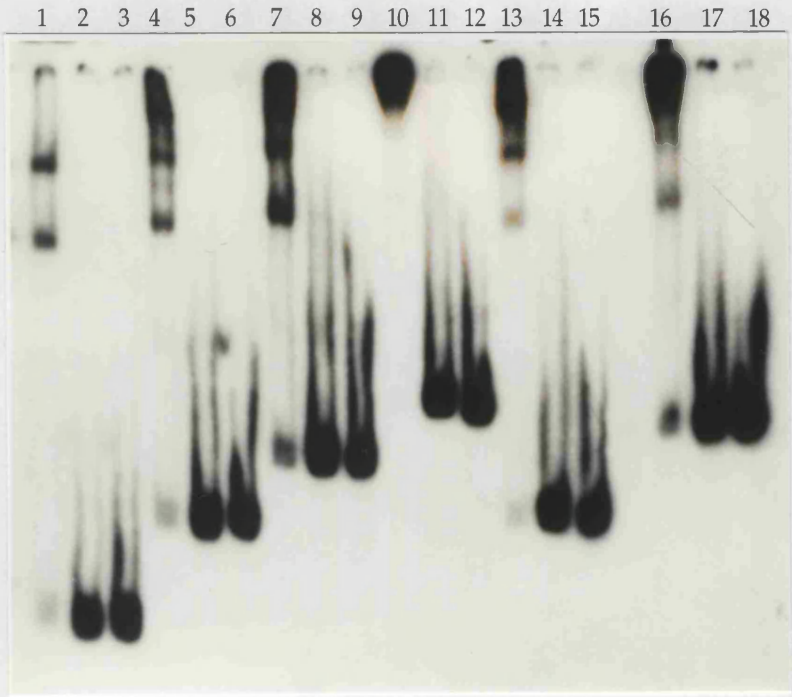


Figure 5.19 Electrophoretic band shift assay (EBSA) on F3, F4, F6, F8, F9 and F10 with Weri-Rb1 and fibroblast nuclear extracts.

Lane 1: F3 plus Weri-Rb1 extract
Lane 2: F3 plus fibroblast extract
Lane 3: F3 only (71 bp)
Lane 4: F4 plus Weri-Rb1 extract
Lane 5: F4 plus fibroblast extract
Lane 6: F4 only (119 bp)
Lane 7: F6 plus Weri-Rb1 extract
Lane 8: F6 plus fibroblast extract
Lane 9: F6 only (181 bp)
Lane 10: F8 plus Weri-Rb1 extract
Lane 11: F8 plus fibroblast extract
Lane 12: F8 only (239 bp)
Lane 13: F9 plus Weri-Rb1 extract
Lane 14: F9 plus fibroblast extract
Lane 15: F9 only (118 bp)
Lane 16: F10 plus Weri-Rb1 extract
Lane 17: F10 plus fibroblast extract
Lane 18: F10 only (205 bp)

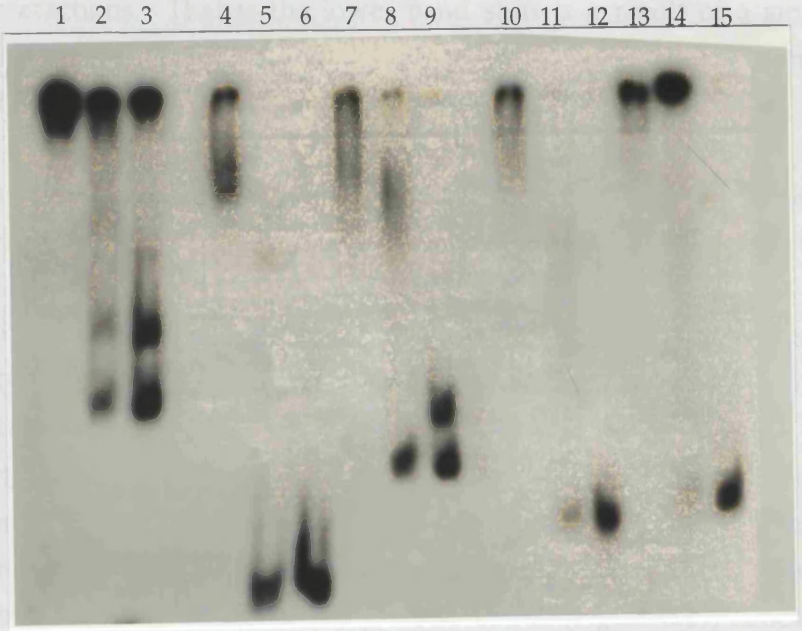


Figure 5.20 Electrophoretic band shift assay (EBSA) on F1, F3, F7, F11 and F12 with Weri-Rb1 and fibroblast nuclear extracts.

Lane 1: F1 plus Weri-Rb1 extract
Lane 2: F1 plus fibroblast extract
Lane 3: F1 only (237 bp)
Lane 4: F3 plus Weri-Rb1 extract
Lane 5: F3 plus fibroblast extract
Lane 6: F3 only (71 bp)
Lane 7: F7 plus Weri-Rb1 extract
Lane 8: F7 plus fibroblast extract
Lane 9: F7 only (150 bp)
Lane 10: F11 plus Weri-Rb1 extract
Lane 11: F11 plus fibroblast extract
Lane 12: F11 only (116 bp)
Lane 13: F12 plus Weri-Rb1 extract
Lane 14: F12 plus fibroblast extract
Lane 15: F12 only (129 bp)

site which binds more than one protein, either directly or via protein-protein interactions. That is the lower band shift is a result of a single protein binding to this specific site, and the second higher band shift is caused by another protein binding at the same position.

F3 is 71 bp in length and is also the first 71 bp at the 5′ end of F4, which is 119 bp in length (see Figure 5.16). If the extra 48 bp of sequence in F4 does not contain a motif for protein binding, then an EBSA with the Weri-Rb1 nuclear proteins should give similar two band shifts as seen for F3. F4 gave 3 strong band shifts with the Weri-Rb1 nuclear extract, the two lower DNA-protein complexes being similar to the shifts produced with F3 (Figure 5.19). The unique upper DNA-protein complex seen for F4 indicates that the extra 48 bp of sequence in F4, relative to F3, contains a protein binding site. The band shifts produced for the other fragments can be interpreted in a similar fashion (see Figure 5.18, 5.19 and 5.20). Faint band shifts, as seen for the lower bands for F9 or F10 (Figure 5.19, Lanes 13 and 16), may represent weak binding of a protein to a particular site such that a few fragments of DNA are retarded in the gel.

Fibroblast nuclear extracts were used as a source of non-retinal proteins. EBSA with fibroblast nuclear proteins on F4, F6, F8, F9 and F10 did not show any band shifts (Figure 5.19; Lanes 2, 5, 8, 11, 14, and 17). This suggests that the band shifts seen with the Weri-Rb1 nuclear proteins for these particular fragments were due to the binding of factors which are specific to the Weri-Rb1 cells. The fibroblast nuclear extract gave a faint DNA-protein shift which was similar to one of the two complexes observed for F3 with the Weri-Rb1 nuclear proteins (Figure 5.20; Lane 5 and Figure 5.21; Lane 5). This suggests that only one of the DNA-protein complexes observed for F3 is due to a Weri-Rb1-specific factor. Other fragments were also retarded by the fibroblast proteins, such as F11 and F12 (Figure 5.20; Lanes 11 and 14). This, suggests that general nuclear proteins which are present within other tissues also bind to the BCP gene promoter.

5.2.71 Competitive electrophoretic band shift assays

To show that proteins binding to the DNA were actually the cause of a particular fragment to shift, unlabelled double-stranded primers were incorporated into the binding reactions to test whether they could compete for binding sites (as described in Section 2.19.4). F3 and F4 were chosen initially for competitive EBSA, due to their relative small size and the fact that they also gave clear visible band shifts with the Weri-Rb1 nuclear proteins (Figure 5.19; Lanes 2 and 4). In addition, the forward primer

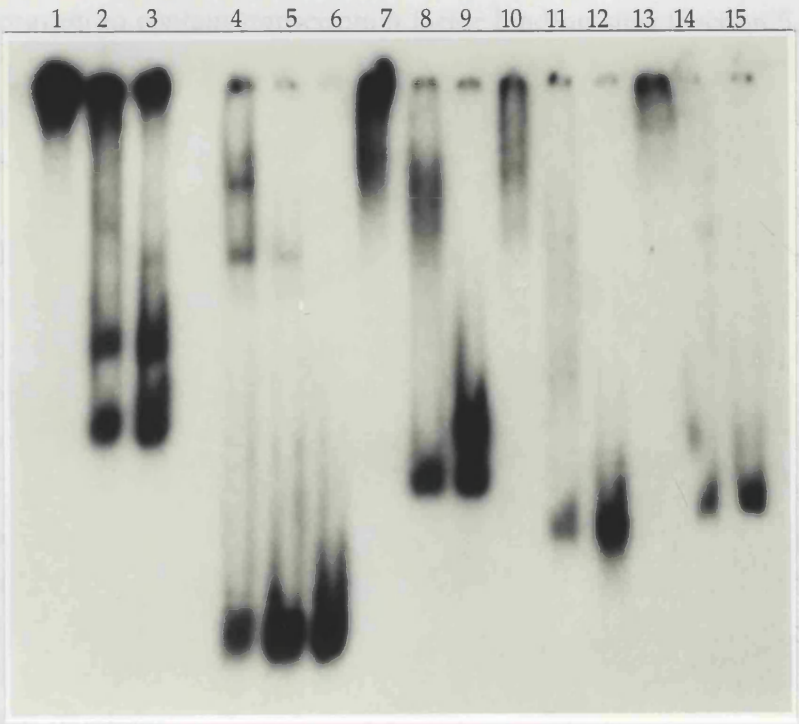


Figure 5.21 Electrophoretic band shift assay (EBSA) on F1, F3, F7, F11 and F12 with Weri-Rb1 and fibroblast nuclear extracts.

Lane 1: F1 plus Weri-Rb1 extract
Lane 2: F1 plus fibroblast extract
Lane 3: F1 only (237 bp)
Lane 4: F3 plus Weri-Rb1 extract
Lane 5: F3 plus fibroblast extract
Lane 6: F3 only (71 bp)
Lane 7: F7 plus Weri-Rb1 extract
Lane 8: F7 plus fibroblast extract
Lane 9: F7 only (150 bp)
Lane 10: F11 plus Weri-Rb1 extract
Lane 11: F11 plus fibroblast extract
Lane 12: F11 only (116 bp)
Lane 13: F12 plus Weri-Rb1 extract
Lane 14: F12 plus fibroblast extract
Lane 15: F12 only (129 bp)

-747/-725F, which is common to both F3 and F4, was shown by the TESS program to contain transcription factor binding sites (Section 5.2.5). Thus, a complementary reverse primer (-725/-747R) can be annealed to this forward primer and used as a competitor for protein binding sites in EBSA for both F3 and F4.

Radioactive probes F3 and F4 were generated for use in competitive EBSA (as described previously in Section 5.2.53). Equal concentrations of the unlabelled primer pair -747/-725F and -725/-747R were annealed together (see Section 2.19.4). The annealed primers were diluted such that they would be either 1:100, 1:10, 10:1 or 100:1 the concentration of F3 or F4 when the primers were added to the binding reactions (see Section 2.19.4). Competitive EBSA was performed using F3 and F4 with the Weri-Rb1 extracts, and the above annealed unlabelled primers as competitors for protein binding sites (as described in Section 2.19.4). Competitive EBSA showed that the band shifts observed in earlier EBSA experiments were due to Weri-Rb1 nuclear proteins binding to specific regions of the human BCP promoter.

Figure 5.22 shows the result for this competitive EBSA. Competition for binding sites has occurred. F3 is shifted by the Weri-Rb1 extract (Figure 5.22; Lane 1) and as the concentration of annealed competitor increases, these DNA-protein complexes are competed out (Figure 5.22; Lanes 4, 5, 6 and 7). The lower DNA-protein complex observed for F3 remains visible at the 1:10 level of competition (Figure 5.22; Lane 6), but is competed out at 10:1 and 100:1 (Figure 5.22; Lanes 5 and 7). The fact that all the DNA-protein complexes observed for F3 are competed out suggests that all the protein binding sites within F3 are contained within the competing double-stranded sequence (-747/-725F/R).

The competing probe (-747/-725F/R) also competes out some of the band shifts seen for F4 (Figure 5.22). This was to be expected since F3 is 71 bp long and is situated within the 5′ end of F4 (Figure 5.16). However, not all the DNA-protein complexes are competed out for F4. A strong band shift is still seen at a 10:1 (Figure 5.22; Lane 13) and even at a 100:1 (Figure 5.22; Lane 14) level of competition. This unique shifted complex is probably due to proteins binding to the extra 48 bp of sequence within F4 which is not present in F3 (Figure 5.16).

The above competitive EBSA result substantiates the potential transcription factor binding sites which were identified within the human BCP gene 5′ flanking region by the computer program TESS (see Section 5.2.42 and Figure 5.10). Figure 5.23 shows the sequence of the -747/-725F

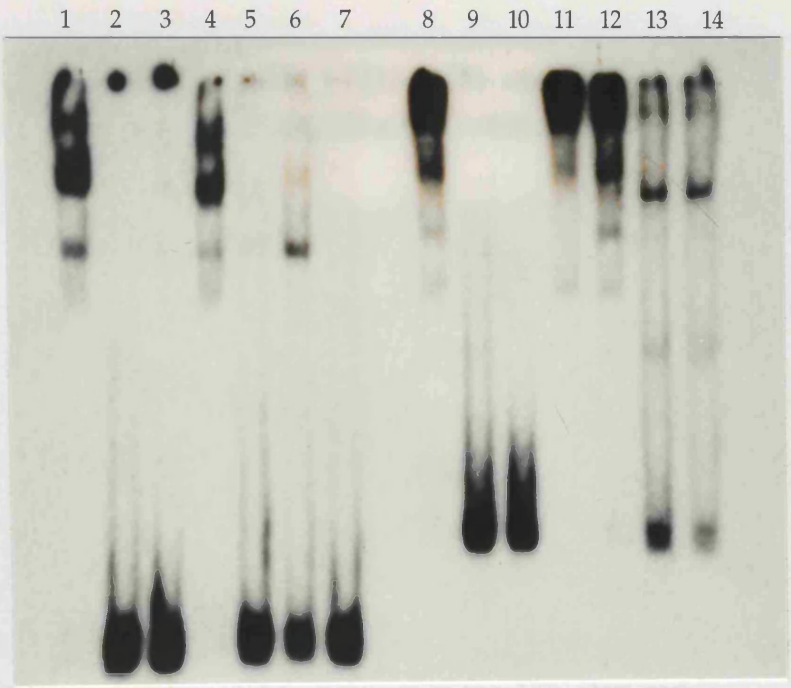


Figure 5.22 Autoradiograph of competitive EBSA on F3 and F4 using the Weri-Rb1 and fibroblast nuclear protein extracts. The annealed unlabelled complementary oligonucleotides -747/-725F and -725/-747R were used as competitors for this assay (see Table 2.2 for sequence of oligonucleotides).

Lane 1: F3 plus Weri-Rb1 extract
Lane 2: F3 plus fibroblast extract
Lane 3: F3 only (71 bp)
Lane 4: F3 plus Weri-Rb1 extract plus 100:1 of F3:competitor
Lane 5: F3 plus Weri-Rb1 extract plus 1: 10 of F3:competitor
Lane 6: F3 plus Weri-Rb1 extract plus 10:1 of F3:competitor
Lane 7: F3 plus Weri-Rb1 extract plus 1:100 of F3: competitor

Lane 8: F4 plus Weri-Rb1 extract
Lane 9: F4 plus fibroblast extract
Lane 10: F4 only (119 bp)
Lane 11: F4 plus Weri-Rb1 extract plus 100:1 of F4:competitor
Lane 12: F4 plus Weri-Rb1 extract plus 10:1 of F4:competitor
Lane 13: F4 plus Weri-Rb1 extract plus 1:10 of F4:competitor
Lane 14: F4 plus Weri-Rb1 extract plus 1:100 of F4:competitor

- 747 TTCAGTGAACAGACAGGAGGAAGACTGT -720

- 719 GGTGACTTGTTGAGCTAAGGGTAGGGGT -692

- 691 AGGCATGAACATGGCGTAAGGACAGTGA -664

- 663 ACAGACCCACCTCCCTGGCTGGAGGTCT -636

- 635 GAGTAGG -629

Figure 5.23 119 bp of human BCP gene 5′ flanking region corresponding to the region from -747 to -629 bp upstream of the ATG start codon.
This sequence represents fragments F4 (-747 to -629) and F3 (-747 to -677) which were used in competitive EBSA (see Section 5.2.71 and Figure 5.16a). The first 71 bp of F4 at the 5′ end represents the complete sequence of F3. The -747/-725F primer is given in blue, the complement of the -677/-707R primer is given in red and the complement of the -629/-651R primer is given in green. The underlined 5′-AGTGAACA-3′ indicates the binding site for GR/PR transcription factors. The underlined 5′-TGACTT-3′ indicates the binding site for the yeast protein GCN4. The above two sites are completely within F3. Flanking F3 and F4 is the binding site for UCRF-L (5′-ATGGCG-3′). Within F4 the underlined 5′-AGTGAACAG-3′ represents another binding site for GR/PR proteins as identified by the computer program TESS.

INSTITUTE OF OPHTHALMOLOGY LIBRARY BATH STREET EC1V 9EL

primer, situated within fragments 3 and 4, indicating the potential protein binding sites identified by TESS. According to the TESS program, the binding site for the GR/PR transcription factors is present within the -747/-725F primer sequence (see Figure 5.11). The GR/PR transcription factors are activated by hormones and are known to interact with human promoter fragments (Latchman, 1995). TESS indicates that F3 also contains a binding site for GCN4, which lies 5′ to the -747/-725F primer sequence (Figure 5.23). GCN4 is the binding site for the yeast transcription factor GCN4 (Latchman, 1995). Hence, GCN4 can be discounted as a potential binding site within the human BCP gene 5′ flanking region on this basis as well as on the competitive EBSA results. Thus, competitive EBSA can be used in a similar fashion on the other fragments to ascertain which regions of the human BCP promoter are binding nuclear proteins.

5.3 Discussion

The molecular events which determine the tissue/cell-specific expression of certain genes within the retina are not completely understood. *Cis*-acting regulatory elements are known to be important in the cell-specific regulation of genes. Which *cis*-acting elements confer specificity to a specific class of photoreceptor, rod or cone, have yet to be determined. The characterisation of the BCP gene 5′ flanking regions from primates and porcine and the comparison of these with other mammalian BCP gene upstream regions in this study is a step towards understanding cone-specific expression.

5.3.1 Evaluation of the primate products obtained by UPW-PCR

The UPW-PCR method for the isolation of 5′ flanking regions resulted in the amplification of the first half of exon one of the BCP genes from the non-human primates utilised in this study (Section 5.2.2). The capuchin, a New World monkey, is a member of the family Cebidae from the Order primates (Figure 3.9). To date, a complete BCP gene has not been sequenced for any member of the Cebidae. The capuchin BCP gene sequence for the first half of exon one, characterised in this study, is the first obtained for any member of the Cebidae. Thus, an opportunity arose to compare, from an evolutionary point of view, the predicted amino acid sequence of this short region of the capuchin BCP protein with the other primate BCP proteins (Figure 5.8). The separation of the New and Old World primate lineages is estimated to have occurred 43 million years ago (Hunt *et al.*, 1995). Thus, the substitution arginine-2, in the Old World

primates, to serine-2, in the New World primates, has been maintained since the split of the two lineages (Time 1 in Figure 3.9). The insertion of proline-30 in the marmoset BCP (Figure 5.8) also occurred after the split of the Old and New World primates, since the Old World primates as well as the bovine and murine BCP proteins lack this extra residue (Hunt *et al.*, 1995). However, the capuchin monkey does not contain this extra proline residue in its BCP protein (Figure 5.8). This suggests that the insertion event that gave rise to the extra proline residue originated much more recently after the sub-division of the New World primates into the Cebidae (Cebid monkeys) and the Callitrichidae (marmosets) (see Figure 3.9).

The UPW-PCR method initially used to generate BCP gene upstream regions for the Old and New world primates was found to result in the same size fragment for various PCR conditions (see Section 5.2.2). Attempts to allow the 33mer walking primer (WP-33mer) to anneal at a different position further upstream of the BCP gene by altering PCR conditions had no effect (Figure 5.6). The reason for this was almost certainly due to the 3′ end of the WP-33mer clamping down onto a conserved sequence situated just upstream of the TATAA box (Figure 5.24a and b). Figure 5.24a gives the sequence of the WP-33mer, with the 9 bp at the 3′ end underlined, which was annealing to a highly homologous region within the promoter region of the primate BCP genes (underlined in Figure 5.24b). The use of consensus upstream primers, designed from a comparison of the human and porcine upstream sequences (Figure 5.4), overcame this problem and generated extra upstream sequences for the primate species. These primers may work on other mammalian species to generate new BCP gene 5′ flanking sequences for analysis.

5.3.2 The photoreceptor conserved element (PCE1)

The high degree of nucleotide identity (over 80%) between the 5′ flanking regions of human, talapoin and capuchin monkey (over 90%) and between porcine and bovine (81.3%; Table 5.1) was not surprising since the former 3 species belong to the Suborder Anthropoidea and the latter two are members of the Order Artiodactyla. The alignment of the four species of BCP gene 5′ flanking regions isolated in this study (porcine, human, talapoin and capuchin) to the previously published bovine and murine sequences proved to be the most useful in identifying conserved elements (Figure 5.10).

a

```
WP-33mer    5'-TTTTTTTTTTTTTTTGTTTGTTGTGGGGGGGTT-3'
primer
```

b

```
Human       5'GCCCAAGTGTTTTTAGAGGAGGGGTTTTGTGGGGTGGGAGG -41
Talapoin    5'GCCCAAGTGTTTTCAGAGGAGGGGTTTTCTGGGGTGGGAGG
Capuchin    5'GCCCAAGTGTTTTTAAAGGAGGGGTTTTGTGGGGTGG----
              ************* * ************ ********

Human       ATCACCTATAAGAGGACTCAGAGGAGGGTGTGGGGCATCC 3' -1
Talapoin    ATCACCTATAAGAGGACTCAGAGGGGGGTGTGGGGCATCC 3'
Capuchin    ATCATCTATAAGAGGACTCAGAGG--GGTGTGGGGTATCC 3'
            **** ******************  ********* ****
```

Figure 5.24a Sequence of the 33mer UPW-PCR primer;

(**b**) Alignment of 81 bp of human BCP gene 5′ flanking region with the equivalent region from talapoin and capuchin monkey. The complement strand of the underlined GGAGGGGTT conserved sequence between these primates is the site at which the 3′ end of the WP-33mer primer is strongly annealing to (see Section 5.3.1). The TATA box is in bold.

PCE1 was found to be conserved in all six species analysed (Figure 5.10, red lettering). PCE1 has been suggested to confer rod-specific expression to a particular gene (Di Polo *et al.*, 1996). PCE1 is present within the promoter regions of various genes which show rod cell-specific expression, such as in the mouse and bovine arrestin genes (Kikuchi *et al.*, 1993; Kikuchi and Shinohara, 1994) and the mouse and human PDE β-subunit gene (Di Polo *et al.*, 1996). Mutation of the PCE1 found within the human interphotoreceptor retinoid-binding protein (IRBP) gene promoter abolishes reporter expression in transgenic mice, implying that this motif is essential for cell-specific expression (Bobola *et al.*, 1995). However, the PCE1 has now been identified in the promoter regions of genes known to be exclusively expressed in cone cells, such as the human cone PDE α subunit gene (Pittler *et al.*, 1995) and within the locus control region (LCR) found 3.2 kb upstream of the human red cone pigment gene (Nie *et al.*, 1996). Since the CTAATC motif is also present and highly conserved within the proximal promoter region of the BCP gene of the species analysed in this study, this suggests that the PCE1 can confer both rod- and cone-specific gene expression. These data imply that PCE1, as the name of the motif suggests, is a photoreceptor-specific element and is not exclusive to the rods or to a particular class of cone. Other *cis*-acting elements associated with PCE1 may confer rod or cone cell specificity.

An exact complement of the PCE1 sequence (5′-GATTAG-3′, position -104 to -99) was found to be conserved only in the primate species (Figure 5.10). Interestingly, PCE1 is also found in its complement form, 5′-GATTAA-3′, within the proximal promoter region of the human, mouse and bovine IRBP gene, and was shown to be protected in footprinting assays (Liou *et al.*, 1994). This suggests that the 5′-GATTAG-3′ at -104 to -99 within the primate BCP promoter may be an additional PCE1 binding site for these species (Figure 5.10).

5.3.3 Putative protein binding sites

A conserved AP-1 motif (5′-TGAGTCA-3′) was recently identified in both the murine and human β-PDE gene 5′ flanking regions (Di Polo *et al.*, 1996). Thus, it was interesting to note that the AP-1 binding site which was identified by the TESS program in both porcine and human upstream sequences, was also present within talapoin, capuchin, and bovine upstream sequences (Figure 5.13 and 5.10; 5′-TGGGTCA-3′ at position -290 to-284). The murine sequence, at the AP-1 conserved region, differs by only a single bp from the conserved 7 bp binding site. The murine

sequence at this position is 5′-TAGGTCA-3′ (Figure 5.10). The consensus AP-1 binding sites that have been identified so far are 5′-TGANTCA-3′ (where N is usually a G or a C) and 5′-TGGGTCA-3′ (Latchman, 1995). It is possible that the 7 bp sequence found in mouse is another AP-1 binding site. To test this theory, the ability of Jun and Fos proteins to bind to this site and regulate expression needs to be examined.

The presence of an AP-1 binding site in the promoter of a gene indicates that the expression of this gene is inducible by treatment with phorbol esters, growth factors, cytokines and neurotransmitters (Latchman, 1995; Di Polo *et al.*, 1996). The proteins Jun (AP-1 transcription factor) and Fos recognise the AP-1 binding site and form heterodimers when bound to this motif. The transcription factors which bind to the PEA3 and GR motifs are known to interact with the Jun and Fos proteins which in turn influence the level of transcription of the gene they are associated with (reviewed in Beate, 1989 and Latchman, 1995). Thus, it was very interesting to note the conservation of the GR binding sequence within the promoters of all the six species compared in this study (Figure 5.10) and the presence of a PEA3 site in the extended sequence obtained for both the human and porcine 5′ flanking regions (Figure 5.11 and 5.12). Whether or not the PEA3 site is present within the other species requires further sequencing of 5′ flanking region of their respective BCP genes. The Ets protein product of the *ets* oncogene binds in a sequence-specific manner to the PEA3 site. The Fos and Jun factors coordinate with the Ets protein to bring about a dramatic increase of gene expression from these promoters. The GR factor, when activated in the presence of glucocorticoid, can bind to the Jun and Fos proteins via protein-protein interaction. This binding of GR to the Jun/Fos proteins inhibits the activation of the gene by the action of AP-1 binding factors. Although no direct evidence exists to indicate that expression of photoreceptor genes can be induced by phorbol esters or are influenced by the action of hormones, the fact that the GR, AP-1 and PEA3 binding sites have been maintained in the BCP promoter regions suggests that the Jun, Fos, GR and Ets protein factors may play a role in regulating the enhancement of expression. Whether or not any of the potential transcription factor binding sites identified by the TESS program are involved in the expression of mammalian BCP genes requires experimental investigation.

Potential transcription factor binding sites were not identified in most of the long stretches of 5′ flanking sequence, found to be conserved between all the species, by the computer program TESS. Thus, this

comparison of the mammalian BCP gene 5′ flanking regions has identified sequences that may bind novel protein factors, and a subset of these may be retinal-, photoreceptor- or even cone-specific.

The E-box-like motif (5′-ACTTTGT-3′) was identified on the basis of its similarity to an E-box element (Section 5.2.43). An E-box sequence 5′-CATTTG-3′ (consensus 5′-CANNTG-3′) was found within the promoter region of the human rod cGMP-phosphodiesterase β-subunit gene and this sequence was shown to be protected under DNAse I footprinting analysis (Di Polo *et al.*, 1996). Thus it was interesting to note the conservation of an E-box motif (at position -406 to -401) within all species (indicated in green in Figure 5.10). The E-box motif was first identified within the immunoglobulin promoters (Church *et al.*, 1985) and proteins with helix-loop-helix structures, such as MyoD and E2A, form homodimers to bind to the E-box sequence (Blackwell and Weintraub, 1990). Different transcription factor species, with helix-loop-helix domains, preferentially bind to the 5′-CANNTG-3′ motif, with the selectivity of binding of each protein arising from the differences within the internal and surrounding nucleotide sequence (Blackwell and Weintraub, 1990). The consensus 5′-CANNTG-3′ motif is found within the regulatory regions, and associated with the expression, of immunoglobulin- and muscle-specific genes (Pettersson *et al.*, 1990; Redondo *et al.*, 1990; Blackwell and Weintraub, 1990). It is possible that the sequences either side and within the E-box motif present within the 5′ flanking region of the mammalian BCP genes (5′-CAGA/GTG-3′) may be associated with conferring tissue or cell-specific expression.

Other highly conserved regions of 5′ flanking sequence were identified (overlined in Figure 5.10). Most of these sequences show characteristic features usually associated with potential protein binding sites, such as the repeated sequence 5′-AGGTTTAGG-3′ (Section 5.2.43). The conservation of these sequences between all the six species analysed in this study suggests that they are maintained for a particular biological function. It is possible that some of these newly identified conserved sequences may confer cone cell-specific expression. To determine whether any of these conserved motifs bind proteins, fragments of DNA containing these particular motifs need to be studied further, using methods of electrophoretic band shift assay (EBSA) or DNAse I footprinting

5.3.4 Weri-Rb1 cells do express BCP transcripts

For the purpose of this study, it was necessary to attain a retinoblastoma cell line which was expressing the BCP gene. Northern blot analysis of several retinoblastoma cell lines failed to show any BCP expression, which may be due to BCP transcripts being below the limit of detection for this particular method (Bogenmann *et al.*, 1988). RT-PCR is a more sensitive method of detecting expressed transcripts than Northern blot analysis. In this study, the Weri-Rb1 cell line was shown to express cone-specific transcripts by RT-PCR (Figure 5.14 and Section 5.2.5); thus this is the first evidence for the presence of BCP transcripts in a retinoblastoma cell line. The Weri-Rb1 cells appear to be of a cone cell specific origin, supporting the data presented by Bogenmann *et al.* (1988) that retinoblastoma is a neoplasm of the cone cell line. This contrasts with the data suggesting that the Y79 retinoblastoma cells are of a rod photoreceptor-specific lineage (Di Polo *et al.*, 1995).

Retinoblastoma tumour cell cultures were originally established for the study of the molecular biology of proliferating cancer cells. However, they also provide a useful resource for the isolation and characterisation of genes that are specific to a particular photoreceptor subtype. In addition, isolation of nuclear proteins from the Weri-Rb1 cells should facilitate the characterisation of transcription factors that are involved in cone photoreceptor-specific expression.

5.3.5 Analysis of the electrophoretic band shift assays

Although a particular pattern of DNA-protein complexes obtained for a specific fragment with EBSA were not fully reproducible, the positive result obtained was that Weri-Rb1 nuclear extracts were capable of retarding fragments of the human BCP gene promoter. The binding reaction conditions requires further investigation to optimise the EBSA such that consistent protein-DNA binding is achieved. Without a consistent set-up it is difficult to conclude which of the band shifts observed in Figures 5.18 to 5.21 are Weri-Rb1 extract-specific. For instance, one of the two DNA-protein complexes observed for F3 with the Weri-Rb1 extracts (Figure 5.19, Lane 1; Figure 5.20; Lane 4) was also seen with the fibroblast nuclear proteins on one particular gel (Figure 5.20; Lane 5) which suggests that only one of the DNA-protein complexes observed for F3 is due to a Weri-Rb1-specific factor. This contradicts the earlier result for F3 with the fibroblast extract, which gave no retardation of this fragment

(Figure 5.19; Lane 2). To resolve this problem the whole EBSA presented in Figure 5.20 was repeated with the same samples (see Figure 5.21). This repeat experiment reconfirmed the band shift which was produced by the fibroblast extract on F3 (Figure 5.21; Lane 5). However, the Weri-Rb1 extract gave a different pattern of band shifts on F3 (Figure 5.21; Lane 4) compared to the previous shifts (Figure 5.19, Lane 1; Figure 5.20; Lane 4). The shifts for the other fragments on this gel F1, F7, F11 and F12 (Figure 5.21) were similar to the previous result (Figure 5.20). However, the shifts seen were of a non-specific nature and were not clean bands as seen previously (Figure 5.19). This suggests that the conditions used for EBSA in this study requires further optimisation for reproducibility of band shifts obtained for a particular fragment.

Competitive EBSA on F3 and F4 indicate that competition with annealed oligonucleotides can be used to pinpoint the *cis*-sequences within the BCP gene promoter to which Weri-Rb1 nuclear proteins bind (Figure 5.22). Thus, for F3 it seems that all the protein binding sites are within the -747/-725F primer sequence (Section 5.2.24). This result is consistent with the data obtained from the TESS program which identified potential protein binding sites within this region (Figure 5.11). Thus, the GR/PR factors are probably responsible for causing the shift of F3. Interestingly, it was noted that the sequence from -745 to -734 (5'-C<u>AGTGAACA</u>GAC-3'), which contains the potential protein binding sites within the -747/-725F primer, was repeated further downstream at position -668 to -657 within F4 (Figure 5.23). Since F4 also contains the above *cis*-sequence one may expect competitive EBSA with the competing probe (-747/-725F/R) to completely abolish all the DNA-protein complexes seen for F4. However, there is another potential protein binding site within oligonucleotide 4 which is the UCRF-L motif (Figure 5.23; 5'-ATGGCG-3'). It is possible that this site is responsible for the result obtained for F4 seen in lanes 13 and 14 in Figure 5.22. Whether or not the GR/PR or the UCRF-L motifs are the cause of these band shifts can be tested by designing competitor oligonucleotides to the exact sequence of these binding sites for competitive EBSA.

Thus, competitive EBSA can be performed on the other oligonucleotides to effectively identify the exact sequences within the human BCP promoter to which transcription factors are binding. More specifically sense and antisense oligonucleotides designed to the evolutionary conserved motifs identified in Section 5.2.43 can be used as competitors in competitive EBSA. Double stranded competition to any of

the other newly identified putative motifs, such as 5′-AGGTTTAGG-3′, 5′-TTCCT-3′, 5′-TAGGAT-3′ or the E-box-like element (5′-ACTTTGT-3′), should elucidate whether these sequences are actually binding novel, as yet unidentified, transcription factors. Also, by comparing competitive EBSA results with various tissue-specific nuclear proteins it should be possible to determine whether these factors are cone-specific.

Chapter 6

Characterisation and Localisation of the Human Blue Cone Pigment Genomic Cosmid Clone

6.1 Introduction

Isolation of the cosmid containing the human blue cone pigment (BCP) gene, initially for analysis of 5′ flanking DNA (see Chapter 5), gave rise to the opportunity to utilise this clone for the fine localisation of the BCP gene to a specific region of human chromosome 7. The clone was also investigated for the presence of microsatellite DNA. In the process of nucleotide sequence comparisons of the human and porcine BCP genes, *Alu* repeat elements were found within the human BCP cosmid clone.

6.1.1 Chromosomal assignment of the human cone opsin genes

The molecular characteristics of the human RCP and GCP genes have been investigated and these genes have been found to be tightly linked on the q-arm of the X-chromosome (Nathans *et al.*, 1986a and b). Pulsed-field gel electrophoresis and cosmid walking analysis of the RCP and GCP genes indicated that they were arranged in a tandem array within the Xq28 region of the chromosome (Vollrath *et al.*, 1988; Feil *et al.*, 1990). The BCP gene has been localised to the lower half of the long arm of chromosome 7 (Nathans *et al.*, 1986b; Donis-Keller *et al.*, 1989). However, the exact localisation of the human BCP gene to a specific region of the chromosome has yet to be determined.

6.1.2 Repetitive DNA sequences

The human genome contains about 30-40% of repetitive DNA (Strachan and Read, 1996). There are many types of repeated DNA sequences of which only 2 will be mentioned in this chapter, namely microsatellite repeats and *Alu* repeats.

Microsatellites are tandemly arrayed repetitive sequences. Each basic repeat unit of a microsatellite is normally between 1-4 bp in length (Strachan and Read, 1996). These microsatellites are significantly polymorphic within the human population, that is the number of repeat units at a particular locus varies between individuals, and thus these repeats serve as very good markers to follow the inheritance of certain genetic traits. One of the most common polymorphic microsatellites is the

dinucleotide CA. *Alu* repeats are members of a class of repetitive sequences known as short interspersed nuclear elements (SINEs). *Alu* repeats are the most predominant form of SINEs within primate genomes (Arcot *et al.*, 1995). The human genome contains about a million copies of the *Alu* family of repeats (Schmid and Jelinek, 1982). Individual members of the *Alu* family are all approximately 300 bp in length and show an average identity of about 87% to a consensus sequence (Lewin, 1990). It is suggested that *Alu* repeats may be transposable elements since they are usually flanked by direct-repeat sequences, a feature associated with mobile genetic elements (Schmid and Jelinek, 1982). *Alu* sequences are generally situated in intergenic DNA or within introns, although there are rare exceptions where they have been found within an exon of a gene (Hobbs *et al.*, 1985).

6.2 Results

6.2.1 Localisation of the human blue cone pigment gene to chromosome band 7q31.3-32

The following results presented in Section 6.2.1 have been published (Fitzgibbon *et al.*, 1994) and the paper is included at the back of the thesis. The work was carried out in collaboration with Dr J Fitzgibbon.

The isolation and conformation of the cosmid clone containing the human BCP gene has already been described (see Section 5.2.1). This cosmid was then labelled by biotin and used as probe for fluorescence *in-situ* hybridisation (FISH) on to human metaphase chromosome spreads, (biotin labelling and FISH analysis was performed by S Gayther and D Wells at the Galton Laboratory, University College London). Figure 6.1 shows the fluorescence signals obtained from the hybridisation of the cosmid to chromosome 7. Thus, the human BCP gene was localised specifically to the junction of bands 7q31.3-32 (Figure 6.2). This result refines the previous localisation of this gene between bands 7q31 to 7q35 (Donis-Keller *et al.*, 1989; Figure 6.2).

6.2.2 Dinucleotide and trinucleotide repeats found within the human BCP cosmid

A vectorette library of the human BCP cosmid had already been constructed for the isolation of 5′ flanking DNA (see Sections 5.2.1 and 2.2.4). Thus, this vectorette library was utilised to detect for any

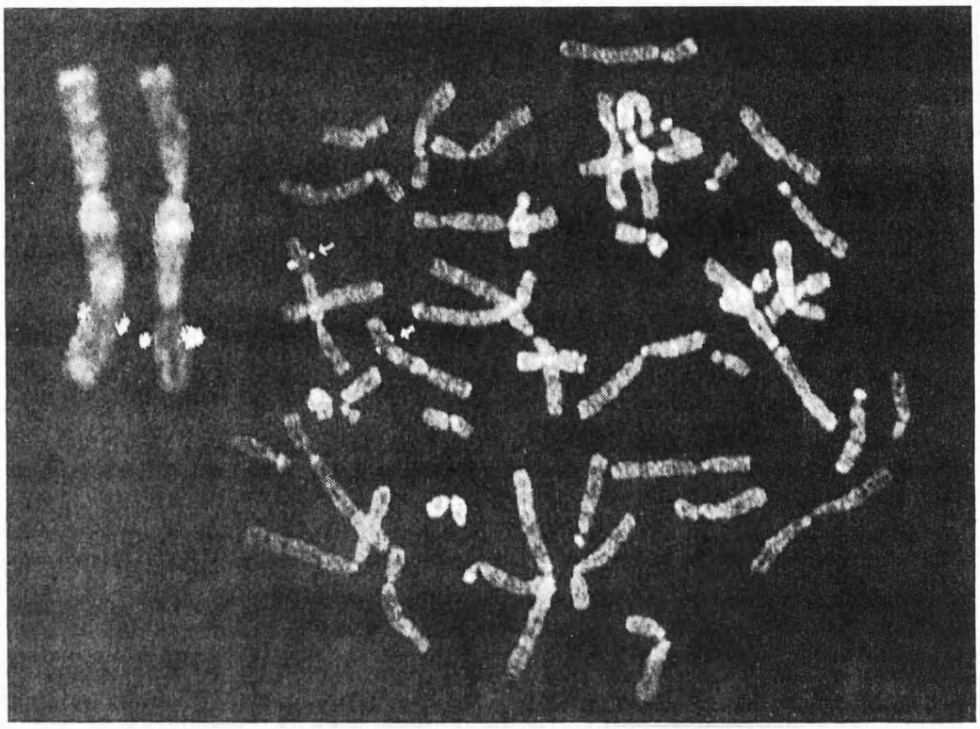


Figure 6.1 Localisation of the human BCP gene to the long arm of chromosome 7, at the junction of bands q31.3 and q32. The figure shows two pairs of signals (indicated with arrows) and includes the separation and enlargement of both copies of chromosome 7 (indicated on the left).

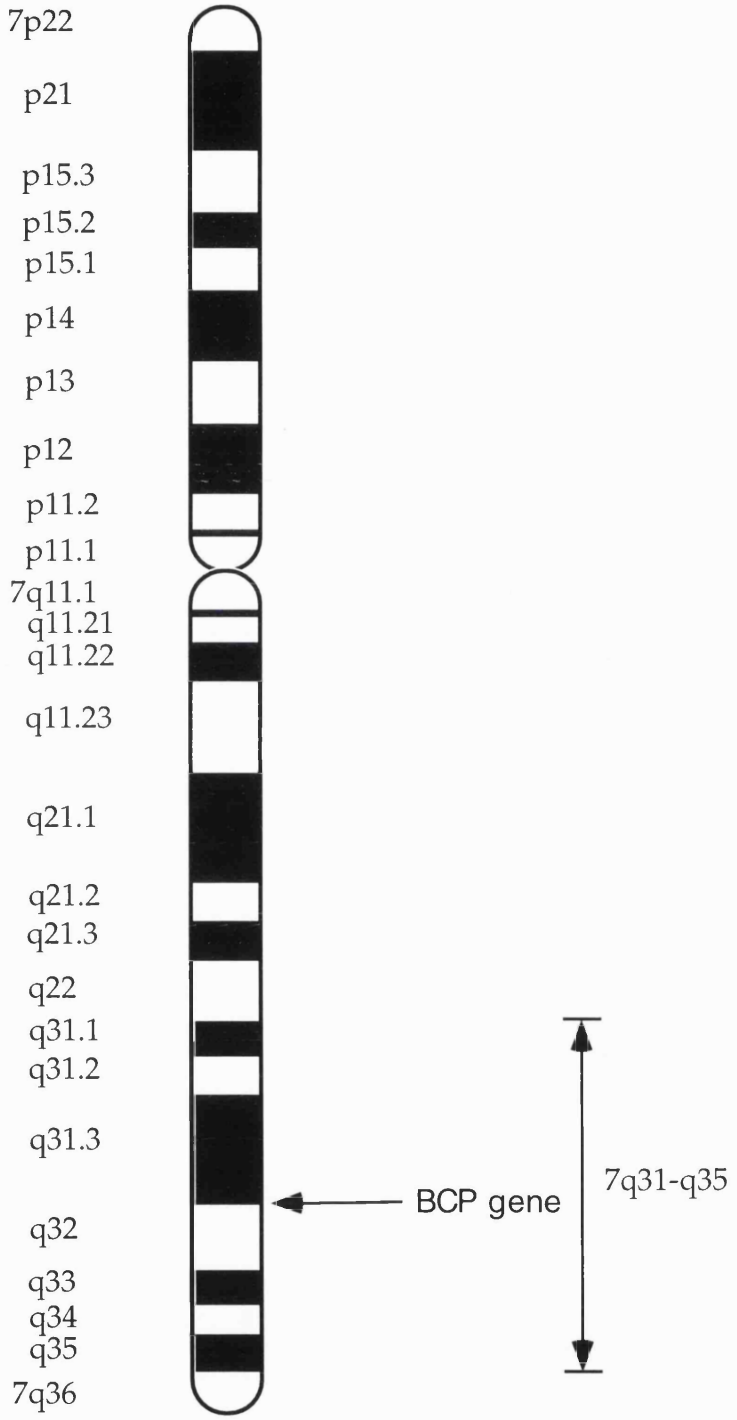


Figure 6.2 An ideogram of human chromosome 7 (adapted from Patel *et al.*, 1995). The arrow indicates the region to which the BCP gene has been localised by flourescence *in-situ* hybridisation. The previous region of localisation of the BCP gene is shown to the right (Donis-Keller *et al.*, 1989).

dinucleotide and trinucleotide microsatellite repeats within the BCP cosmid.

A *Hae*III vectorette library of the BCP cosmid was used as a template for PCR amplification with various triplet and dinucleotide repeat primers (Table 2.5) in conjunction with the vectorette primer (as described in 2.2.41). Of the 7 repeat primers used (Table 2.5) only one of these, namely the triplet repeat $(GGT)_6$, gave a PCR product which was reproducible with the vectorette protocol (Figure 6.3). This PCR product was cloned into a plasmid vector and sequenced (as described in Sections 2.5 to 2.8 and 2.12). Sequencing of this PCR product showed that it was 507 bp in length (Figure 6.4). The fact that the $(GGT)_6$ primer gave a PCR product indicates that this oligonucleotide was annealing to a complement of this triplet repeat lying within the BCP cosmid. To ascertain how many GGT triplet units were actually present, it was necessary to isolate a DNA fragment containing this repetitive sequence from the cosmid.

A new PCR primer (BCP Repeat+) was designed from the antisense strand of the 507 bp obtained previously (Table 2.5). This primer was situated downstream of the GGT triplet repeats (Figure 6.4, in red text). The BCP Repeat+ primer was used in conjunction with the vectorette primer and the BCP cosmid *Hae*III vectorette template to generate a PCR fragment of approximately 650 bp (Figure 6.5). This PCR product was cloned and partially sequenced to determine the exact number of the triplet repeat (GGT) units present (Figure 6.6).

There were 4 (GGT) triplet repeats within the BCP cosmid (Figure 6.6, underlined in bold lettering). It was noted that the 4 (GGT) triplet repeats are flanked on both sides by the sequence GAT (Figure 6.6). In addition there was a GAT triplet repeat of 4 units which was 6 bp 5′ to the GGT repeat unit (Figure 6.6, in green lettering).

6.2.3 *Alu* repeat elements situated within the BCP cosmid

Two *Alu* sequences were discovered within the BCP cosmid. One of the *Alu* repeats was discovered by design in the hope of obtaining unknown 5′ flanking DNA from the BCP gene (see Section 5.2.1). The other was found quite serendipitously from a comparison of the human and porcine BCP genes.

A consensus *Alu* repeat primer (Alu1) was used in conjunction with an antisense primer to exon 1 (Ex1-529/514-) of the human BCP gene to amplify a PCR fragment containing the 5′ proximal promoter of the BCP gene (as described in Section 5.2.1). Sequencing of this PCR product

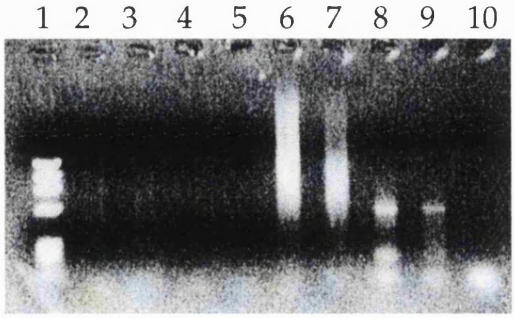


Figure 6.3 A 1% agarose gel showing the products obtained after amplification of the BCP vectorette library with the vectorette primer and the following dinucleotide and trinucleotide repeat primers:-

Lane 1	DNA ladder ϕX174 *Hae*III
Lane 2	$(ACG)_6$ primer
Lane 3	$(CAG)_6$ primer
Lane 4	$(GGC)_6$ primer
Lane 5	$(GGA)_6$ primer
Lane 6	$(CT)_{15}$ primer
Lane 7	$(CT)_{15}$ primer
Lane 8	$(GGT)_6$ primer
Lane 9	$(GGT)_6$ primer
Lane 10	Negative control

5'**GGTGGTGGTGGTGGTGGT**GATAATGGTGAGGATGACACTTCCTGGGGTTAGAAAG
AAGGGGAAACTAAAATACTAGATAGAAATATGGGAGATGGTGGAGAAAGAAGAACAGAG
TTAACATGTTCTAACATCTATTATTCAGAAAAAGTATAGTGATATTGATTCACTGTAGA
GCTAAGTAAGAAGTTTAGAAATATATAAAGAGTAATCACTCAGTTCAGTGTGTCCATCC
TGGTAGAGATGCACACCTTCCTGGGTTAGAAGAAGGGAACTAAATCATAGATAGAAATA
TGGGAGATGGTGGAGAAAGAAGAACAGAGTTAACATGTTCTAGTCTATTATTCAGAAAA
AGTATAGTGATATTGATTCACACTGTAGAGCTAAGTAAGAATGTTTAGAAATATATAAA
GAGGTAATCACTGAAAGAGTAGAAATAGAATGTGTAACTTCCAAATCAGTGGGGG**AAGG**
AGAGGACAGCGATTCTCGTACGAACGGTTACGATTCG 3'

Figure 6.4 Sequence of the 507 bp PCR product generated from BCP cosmid *Hae*III vectorette template using the triplet primer $(GGT)_6$ in conjunction with the vectorette primer. The $(GGT)_6$ primer is in bold and underlined at the 5′ end, and the complement of the vectorette primer and linker sequence is given in bold lettering at the 3′ end. The sequence in red text indicates the complement of the new primer (BCP Repeat+) which was designed from this sequence.

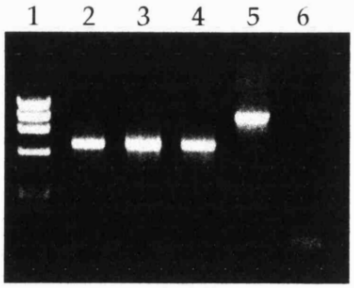


Figure 6.5 A 1% agarose gel showing the products obtained after amplification of the BCP cosmid *Hae*III vectorette library with the vectorette primer and BCP Repeat+ primer.

Lane 1: DNA ladder ØX174 /*Hae* III

Lanes 2-4: BCP Repeat+/vectorette PCR

Lane 5: Positive control (Ex4+ to Ex5- PCR of human BCP cosmid)

Lane 6: Negative control

5'...AGAGAAATTTATAAATATGTTAGTAGATCTAAACAAGCCTTGAGACAGATGATGATGATAAAGA
T**GGTGGTGGTGGT**GATAATGGTGAGGATGACACTTCCTGGGGTTAGAAAGAAGGGGAAACTAAAATA
CTAGATAGAAATATGGGAGATGG...3'

Figure 6.6 Partial sequence of the PCR product generated from BCP cosmid *Hae*III vectorette template using the BCP Repeat+ primer in conjunction with the vectorette primer. The complement of the BCP Repeat+ primer sequence is in red text at the 3′ end. Sequencing showed that there are only four GGT triplet repeat units (shown in bold lettering and underlined). Additional GAT triplet repeats situated upstream of the GGT triplets are shown in green.

indicated that the *Alu* repeat began 916 bp 5′ to the ATG start codon (at position -916) and continued further upstream of the human BCP gene (Figure 6.7). This PCR fragment only contained 47 bp of the *Alu* repeat (indicated in red; Figure 6.7). This 47 bp includes 25 bp of the Alu1 primer (underlined; Figure 6.7) and the remaining sequence must extend upstream of this primer. The full length sequence of the *Alu* repeat was obtained by direct sequencing from the cosmid on an ABI automated sequencer, with the primer -677/-707R (as described in Section 2.12.2). This *Alu* repeat was found to be 279 bp in length and showed 88.2% identity to a consensus *Alu* repeat sequence (Figure 6.8).

The second *Alu* repeat found within the human BCP cosmid was within intron 3 of the BCP gene. From a comparison of the intron sizes between the human and porcine BCP genes it was observed that intron 3 in human was approximately 300 bp larger than the porcine intron 3 (see Figure 3.7 in Section 3.2.3). This variation in size was considerably greater than the differences between the other introns of porcine and human BCP genes (Figure 3.7). The average size of an *Alu* repeat is about 300 bp (Schmid and Jelinek, 1982). Thus, it was proposed that the extra sequence within human intron 3 may be due to an insertion of an *Alu* repeat. Direct sequencing of intron 3 from the human BCP cosmid was performed using the BCP-specific primers Ex3+ 935/954F and Ex4- 1276/1254R (see Table 2.1 and Section 2.12.2). Inspection of this sequence indicated that there was indeed an *Alu* repeat present within intron 3. This *Alu* repeat of 276 bp was situated 236 bp downstream of exon 3 and 94 bp upstream of exon 4 within intron 3 (Figure 6.9). The intron 3 *Alu* repeat showed 80% identity to the consensus *Alu* sequence (Figure 6.10). When the two BCP gene *Alu* repeats were aligned, they were found to share a 75.7 % sequence identity with each other (Figure 6.11). The *Alu* within intron 3 was the smaller of the two *Alu* repeats (Figure 6.11).

6.3 Discussion

A genomic clone containing the human BCP gene was utilised to map this gene to human chromosome 7q31.1-33 (Figure 6.2). This result refines the previous localisation of this gene between bands 7q31 to 7q35 (Donis-Keller *et al*, 1989; Figure 6.2). Mutations within rod opsin are known to cause various forms of retinal degeneration (reviewed by Daiger *et al.*, 1995). Although mutations within the cone pigments are generally associated with colour blindness and not retinal degeneration, it is interesting that an autosomal dominant retinitis pigmentosa locus has

```
5'GCCTCCCAAAGTGCTGGGATTACAGCCTTGAGCCACCACGCCC   - 920
GGCCAACTCATCTATATATTTATGATCTGCTGAGGTGAAAAGATT   - 875
CATTCATAAAGCCTAAAGTGTAGTACTGAGTACAAGTGCAGGTGG   - 830
AACTTAAAGCAGAGTTGGTGTCACTGGGGTGGGCTGCACAGAAGA   - 785
GGTGTTTGTGAGGTAGGCTGTGAGGGCTAGATAAGACTTCAGTGA   - 740
ACAGACAGGAGGAAGACTGTGGTGACTTGTTGAGCTAAGGGTAGG   - 695
GGTAGGCATGAACATGGCGTAAGGACAGTGAACAGACCCACCTCC   - 650
CTGGCTGGAGGTCTGAGTAGGAAAGGCACACTCTTTGAGTAGATT   - 605
CAACACAAAGAAATGGAGGCAGCCTGGCTCAGGGATGCCCTAGAC   - 560
CTCTAGGCATTGTCAAGTTGCCTAAGTCCTGTTCCATCAAGGCTG   - 515
TTTACTGATGTGCTTCCAGGGCACTCCCCACTCCCAGCCCTTTCC   - 470
TGCAGCCCAGGGCTGGTTCCTAGCCTCTCAGCAGACTTAAGATGG   - 425
GCACCTTCCACAAAGGGGCAGATGAGTTGAGGAAAACTTAACTGA   - 380
TACAGTTGTGCCAGAAGCCAAAATAAGAGGCGTGCCCTTTCTATA   - 335
GCCCCATTAAAAGAACAAAAAAGTGGAAGCATCTTCAGTGAATAT   - 290
GGGTCAGCACCTCCCAGACCTCAGGGAGTCCACTTCTGTTCATCC   - 245
CAGCACCCAGCATTGCATATCCAGATTATTTGAGCCCAATCTCTT   - 200
ATCCTCTGAAGAACACAATCGGCTTTGGGGCCACAGAAGGTTTAG   - 155
GTAGTGGTTTAGGGATTTCTAATCCCAAACTTTGTCCTTGGGAGG   - 110
TTTAGGATTAGTATTGATCATTCACAGAGCCCAAGTGTTTTTAGA    - 65
GGAGGGGTTTTGTGGGGTGGGAGGATCACCTATAAGAGGACTCAG    - 20
AGGAGGGTGTGGGGCATCCATG 3'                         - 1
```

Figure 6.7 962 bp of the human blue cone pigment gene 5' flanking region

The first 47 bp at the 5′ end (in red) represents sequence at the 5′ end of an *Alu* repeat element. The first 25 bp underlined indicates the consensus *Alu* repeat primer (Alu1) sequence. Alu1 was used to generate this upstream region (see Section 5.2.1 and Figure 5.2). The ATG start codon for the blue cone pigment gene is also given and shown in bold lettering. The numbering system refers to the nucleotide position upstream of the start site, -1 being the first bp 5′ to the ATG start codon. The complement of the -677/-707R primer sequence, which was used to obtain the remainder of the *Alu* repeat situated 5′ to the Alu1 primer, is shown in green.

```
5'T-ATATGTAG TCTGGTTCTG TCCCACAGGC TGGAGTGCAG TGGAGTGATC 49
5'TGAGACGGAG TCTCGCTCTG TCGCCCAGGC TGGAGTGCAG TGGCGCGATC 50
  * * * * ** *** * **** **** ***** ********** *** * ****

TCGGCTCACT GCAGGCTCCA TCTCCCGGGT TCACGCCATTT CTCCTGCCTC 100
TCGGCTCACT GCAACCTCCA CCTCCCGGGT TCAAGCGATT- CTCCTGCCTC 100
********** ***  *****  ********* *** ** ***  **********

AGCCTCCCAA GTAGCTGGGA TTACAGGCGC CCACCACCAC  ACCCAGCTAA 150
AGCCTCCCGA GTAGCTGGGA TTACAGGCGC GCGCCACCAC  GCCCGGCTAA 150
******** * ********** **********  * *******   *** *****

TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACC--GTTAG  CCAGGATGGT 198
TTTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCATGTTGG  CCAGGCTGGT 200
********** ********** ********** ***  ***    ***** ****

CTCGATCTCC TGACCTT--G TGATCCACCC ATCTCAGCCT  CCCAAAGTGC 246
CTTGAACTCC TGACCTCAGG TGATCCACCC ACCTCGGCCT  CCCAAAGTGC 250
** ** **** ******   * ********** * *** ****  **********

TGGGATTACA GGCTTGAGCC ACCACGCCCG GCC  3'                279
TGGGATTACA GGCGTGAGCC ACCACGCCCA GCC  3'                283
********** *** ****** *********  ***
```

Figure 6.8 Alignment of the *Alu* repeat found upstream of the BCP gene to the complement of a consensus *Alu* sequence (Nelson *et al.*, 1989). The *Alu* sequence found upstream is shown in red lettering, the consensus *Alu* repeat is in plain text. The consensus *Alu* repeat primer (Alu1) sequence is underlined. Identical nucleotides between the two sequences are undermarked with an asterix (*). The upstream *Alu* repeat shows 88.2% identity to the consensus *Alu* sequence.

5'GTGAGTGGCATTTGATAGTCAGGGGAGAAGGGGTTCGGGGCCACATGAGAAGAAG
AGTGCTCTGAAACATAGATGCCTGGAATGTCTAGCCCAGAGAGGGTATCTAAAGCAG
CAAGAAGTTAGAGGAGGGAGAATGATGGAAAATCCTAAAGGAATTAGCCGGAGATGG
GACAAAGAGGCAATTCAGGAGTGGAATGCCCCCTTCCCCAATCCCAACAAAAATCAT
CCAGAGGACCGGGGCGCCAGTGGCTAACGCCTTATAATCCCAGCAATTTTGGGGAGG
CCGAAAGCAAGACGGATCACCTGAGGTCAATAGTTCCAGACCAGCTTGGCCTACATG
GCAAAATGCTAAAATGCGAAAATTAGCTGGGCATGGTGGTGTGTGCCTGTAACCCCA
GCTACTCAGGAGGCTGAGACAAAAGAATCACTTGAACCCGGGAGGTGGAGGTTCAGT
GAGCCGAGATTGCACCACTGCAATTCCAGCCTGGGCAACAGAGCGAGACTTGGTCTC
AAAAAGAAAATTAAAAATTAAAAATTAAAAAAAAACAAAAAAAGCATCAGAGGGCCA
GGAAAAAGAGAGATGTGATGCTCCGTGCTCCACCCCAG 3' ->Exon4

Figure 6.9 Complete sequence of intron 3 from the human blue cone pigment gene (606 bp). The 276bp of the *Alu* repeat sequence is shown in blue lettering. This *Alu* repeat is situated 236 bp downstream of exon 3 and 94 bp upstream of exon 4. The underlined region indicates the 'adenine rich tail' normally associated at the 3' end of an *Alu* repeat (Schmid and Jelinek, 1982).

```
5'GG-GGCGC-C AGTGGCTAAC GCCTTATAATC CCAGCAATTTT GGGGAGGCCGA  51
5'GGCTGGGCGT GGTGGCTCAC GCCT-GTAATC CCAGC-ACTTT -GGGAGGCCGA  50
  **  * **    ****** ** ****  ***** ***** * ***  **********

AAGCAAGACGGA TCACCTGAGG TCAATAGTTC  CAGACCAGCT  TGGCCTACAT  103
--GGTGGGTGGA TCACCTGAGG TCAGGAGTTC  AAGACCAGCC  TGGCCAACAT  100
  *   *  *** ********** ***  *****   ********   ***** ****

GGCAAAA---   ------TGCT -AAAATGCGA  AAATTAGCTG  GGCATGGTGG  143
GGTGAAACCC   CGTCTCTACT AAAAATACAA  AAATTAGCCG  GGCGTGGTGG  150
**  ***            * **  ***** * *  ******** *  *** ******

TGTGTGCCTG   TAACCCCAGC TACTCAGGAG  GCTGAGACAA  AAGAATCACT  193
CGCGCGCCTG   TAATCCCAGC TACTCGGGAG  GCTGAGGCAG  GAGAATCGCT  200
 * * *****   *** ****** ***** ****  ****** **    ****** **

TGAACCCGGG   AGGTGGAGGT T-CAGTGAGC  CGAGATTGCA  CCACTGCAATT  243
TGAACCCGGG   AGGTGGAGGT TGCAGTGAGC  CGAGATCGCG  CCACTGCA-CT  250
**********   ********** * ********  ****** **   ********  *

CCAGCCTGGG   CAACAGAGCG AGACTTGGTC  TCA 3'                   276
CCAGCCTGGG   CGACAGAGCG AGACTCCGTC  TCA 3'                   283
**********   * ******** *****  ***  ***
```

Figure 6.10 Alignment of the *Alu* repeat found within intron 3 of the BCP gene to a consensus *Alu* sequence (Nelson *et al*, 1989). The intron 3 *Alu* sequence is shown in light blue lettering, the consensus *Alu* repeat is in plain text. The complement of the consensus *Alu* repeat primer (Alu1) sequence is underlined. Identical nucleotides between the two sequences are undermarked with an asterix (*). The intron 3 *Alu* repeat shows 80% identity to the consensus *Alu* sequence.

```
5'GG-GGCGC-C AGTGGCTAAC GCCTTATAATC CCAGCAATTTT GGGGAGGCCGA  51
5'GGCCGGGCGT GGTGGCTCAC GCCT-GTAATC CCAGC-ACTTT -GGGAGGCTGA  50
  **  * **    ****** ** ****  ***** ***** * ***  ******* **

AAGCAAGACGGA TCACCTGAGG TCAATAGTTC  CAGACCAGCT  TGGCCTACAT  103
--GATGGGTGGA TCAC--AAGG TCAGGAGATC  GAGACCATCC  TGGCTAAC--   96
  *   *  *** ****   *** ***  ** **   ****** *   ****  **

GGCAAAA---   ------TGCT -AAAATGCGA  AAATTAGCTG  GGCATGGTGG  143
GGTGAAACCC   CGTCTCTACT AAAAATACAA  AAATTAGCTG  GGTGTGGTGG  146
**  ***            * **  ***** * *  **********  **  ******

TGTGTGCCTG   TAACCCCAGC TACTCAGGAG  GCTGAGACAA  AAGAA-TCAC  193
TGGGCGCCTG   TAATCCCAGC TACTTGGGAG  GCTGAGGCAG  GAGAAATGGC  196
** * *****   *** ****** ****  ****  ****** **    **** *  *

TTGAACCCGG   GAGGTGGAGG TT-CAGTGAG  CCGAGATTGC  ACCACTGCAAT 243
GTGAACCCGG   GAGATGGAGC CTGCAGTGAG  CCGAGATCAC  TCCACTGCA-C 246
 *********   *** *****   * *******  *******  *   ********

TCCAGCCTGG   GCAACAGAGC GAGACTTGGT  CTCA 3'                 276
TCCAGCCTGT   GGGACAGAAC CAGACTACAT  AT-A 3'                 279
*********    *  ***** *  *****   *   * *
```

Figure 6.11 Alignment of the two *Alu* repeats found in the human BCP gene cosmid. The sequence in blue lettering is the intron 3 *Alu* repeat. The sequence in red lettering is the complement of the upstream *Alu* repeat. Identical nucleotides between the two sequences are undermarked with an asterix (*). The upstream *Alu* repeat shows 75.7% identity to the intron 3 *Alu* sequence. The complement of the consensus *Alu* repeat primer (Alu1) sequence is underlined.

been linked to the 7q region of the chromosome (Jordan *et al.*, 1993). Thus, the precise localisation of the BCP gene facilitates the consideration of this gene as a candidate for this disease.

Two triplet repeats, $(GGT)_4$ and $(GAT)_4$, were found to be situated within the BCP cosmid clone, in close proximity to each other. These triplet repeats may be polymorphic within the human population. However, microsatellite repeats of less than 12 units are generally not investigated for polymorphism, since they are less likely to be polymorphic than repeat units of greater than 12 (Weissenbach *et al.*, 1992). The triplet repeats, $(GGT)_4$ and $(GAT)_4$, found in this study require further analysis to conclusively state whether or not they are polymorphic within the human population. If these triplet repeats are polymorphic they can be used as markers for genetic analysis of the 7q31.1-7q33 region.

The *Alu* repeat situated within intron 3 of the BCP gene showed a lower percentage identity to the consensus *Alu* sequence than the *Alu* repeat situated upstream of the BCP gene. This percentage identity was slightly lower than expected, since individual members of the *Alu* family have an average homology of 87% with the consensus sequence (Lewin, 1990). When the two BCP gene *Alu* repeats were aligned with each other they only shared a 75.7% identity.

The main difference between the intron 3 *Alu* and the consensus and upstream *Alu* repeats was a deletion of 9 bp within the intron 3 *Alu*. This 9 bp deletion corresponded to nucleotides 108-116 within the consensus *Alu* (Figure 6.10). The intron 3 *Alu* repeat is associated with a string of 35 adenine residues (poly(A)), interspersed with 3 sets of the dinucleotide TT, a single cytosine and a single guanine residue, at its 3′ end (Figure 6.9). The *Alu* repeat sequence found upstream is not associated with a poly(A) tract and thus belongs to the rare class (7%) of *Alu* sequences within the human genome which do not contain a poly(A) tract (Economou *et al*, 1990). Thus, these two BCP *Alu* sequences are presumably from different branches of the *Alu* repeat family. Investigation of the poly(A) tracts associated with three separate *Alu* repeats, on different human chromosomes, indicated that two of these were polymorphic in length, with one of these possessing six different alleles within the population (Economou *et al.*, 1990). Thus, the poly(A) tract of the intron 3 *Alu* repeat may be polymorphic. Like the triplet repeats, the poly(A) tract requires further analysis to determine whether or not this is indeed the case.

Chapter 7

7.1 Conclusions and Future Prospects

The opsin encoded by the BCP gene of the domestic pig (*Sus scrofa*), characterised in this study, is a member of the S group of visual pigments (see Section 3.2.2). Prior to this study the murine and bovine BCP genes were the only non-primate mammalian members of the S group characterised (Chiu *et al.*, 1994). A comparison of the gene sequences which encode the S group of pigments, which have been characterised to date, should allow the design of consensus BCP-specific PCR primers. Consensus BCP-specific primers can be used to amplify other mammalian, and maybe non-mammalian vertebrate, opsin genes which encode for proteins belonging to the S group. Characterisation of new S group opsins, which have λ_{max} of less than 440 nm, will increase our limited knowledge of this particular family of pigments.

Comparison of the S group of opsins with other visual pigments resulted in the identification of amino acids which are specific to this family of opsins ('S group-specific' residues; see Section 3.2.4). Amino acids at 16 positions have been identified as putative spectral tuning sites which may contribute to maintaining the short wavelength sensitivity observed in the S group of opsins (Table 3.9a and b). Site-directed mutagenesis experiments with *in vitro* expression and measurement of the absorption spectra of mutant BCP opsins need to be carried out to ascertain whether the amino acids identified in this study are involved in the blue shift of λ_{max}. Phe-86, a residue specific to the mouse BCP and the goldfish ultraviolet opsin, is a strong candidate for a spectral tuning site which may contribute to the shift of the λ_{max} of members of the S group of pigments into the ultraviolet region of the spectrum (see Section 3.2.54). Electroretinogram (ERG) data suggest that other rodents, such as the rat (*Rattus norvegicus*), gopher (*Thomomys bottae*) and gerbil (*Meriones unguiculatus*), also possess an opsin which is maximally sensitive in the ultraviolet (UV) region of the spectrum (Jacobs *et al.*, 1991). It would be interesting to obtain the sequence of a number of rodent BCP genes (using the BCP gene-specific PCR primers mentioned above) and determine, from the predicted amino acid sequence, whether the residue Phe-86 is also present in these species. As yet, there is no evidence from direct absorption readings of artificially expressed mouse or rat BCP to show that the S opsin is responsible for the UV sensitivity observed in these rodents.

Thus, site-directed mutagenesis studies, *in vitro* expression and analysis of rodent BCPs will determine whether these pigments confer UV specificity, and may provide the first direct evidence for a mammalian UV sensitive opsin.

1194 bp of the proximal 5′ flanking region of the human BCP gene was sequenced, of which 279 bp at the 5′ end represents an *Alu* repeat element (see Section 6.2.3). This *Alu* repeat shares a 75.7% nucleotide identity to another *Alu* repeat which was found within intron 3 of the human BCP gene (see Section 6.2.3 and Figure 6.11). 949 bp, 573 bp and 570 bp of the 5′ flanking regions from the BCP genes of porcine, talapoin monkey and capuchin monkey were sequenced, respectively (see Sections 5.2.1 and 5.2.2). A comparison of these isolated sequences with the published bovine and murine 5′ flanking BCP gene sequences (Chiu and Nathans, 1994b) has identified evolutionary conserved motifs (see Sections 5.24 and 5.25). The alignment of sequences from a number of different species is a powerful tool for identifying conserved DNA elements. The addition of new 5′ flanking sequences, from other mammalian BCP genes, to such an alignment will determine whether the motifs identified in this study are conserved in other BCP genes.

The Weri-Rb1 cell line was shown to express cone-specific transcripts by RT-PCR (Figure 5.14 and Section 5.2.5). This has provided the first evidence for the presence of BCP transcripts in a retinoblastoma cell line. It was shown that nuclear proteins isolated from the Weri-Rb1 cells could be used for electrophoretic band shift assay (EBSA) experiments on fragments of the 5′ flanking region of the human BCP gene (see Section 5.2.7). Competitive EBSA, using short double-stranded oligonucleotides designed to sequences within the human BCP gene 5′ flanking region, demonstrated that specific regions within the BCP promoter to which proteins were binding could be identified using this method (see Section 5.2.71). Future DNA footprinting experiments on the human BCP gene promoter utilising the Weri-Rb1 and other tissue-specific nuclear extracts should determine which motifs the proteins are actually binding to and which are Weri-Rb1-specific. The identification of transcription factors which confer cone or photoreceptor-specific gene expression would be of great interest. The 5′ flanking regions of the BCP genes isolated in this study can also be used to screen retinal expression libraries to identify the DNA-binding proteins, and the genes which encode the transcription factors, which bind to the proximal promoter of photoreceptor-specific genes (as described by Singh *et al.*, 1988). Not only will the characterisation

of these newly identified genes and proteins increase our understanding of cone and photoreceptor-specific gene expression, but the genes themselves, depending on where they are situated within the human genome, will be good candidates for cone and cone-rod dystrophies which have been identified (Bresnick *et al.*, 1989; Went *et al.*, 1992; Evans *et al.*, 1994; Hong *et al.*, 1994; Balciuniene *et al.*, 1995).

Abbreviations

ATP	adenosine triphosphate
BCP	blue cone pigment
BSA	bovine serum albumin
bp	base pair(s)
°C	degrees centigrade
CEBSA	competitive electrophoretic band shift assay
cDNA	complementary deoxyribonucleic acid
D	dalton
dATP	deoxy-adenosine-5′-triphosphate
dCTP	deoxy-cytidine-5′-triphosphate
ddATP	dideoxy-adenosine-5′-triphosphate
ddCTP	dideoxy-cytidine-5′-triphosphate
ddGTP	dideoxy-guanosine-5′-triphosphate
ddTTP	dideoxy-thymidine-5′-triphosphate
DEPC	diethyl pyrocarbonate
dGTP	deoxy-guanosine-5′-triphosphate
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxy-thymidine-5′-triphosphate
EBSA	electrophoretic band shift assay
EDTA	ethylene-diamine-tetra acetic acid
fmols	fentomoles
g	gram(s)
GCP	green cone pigment
kb	kilobase(s)
kD	kilodalton(s)
M	molar
MD	minimal dextrose
μg	microgram
min	minute(s)
μl	microlitre
ml	millilitre
mM	millimolar
MM	minimal methanol
MSP	microspectrophotometry

NE	sodium chloride/EDTA
ng	nanogram
nm	nanometre
OD	optical density
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
RCP	red cone pigment
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse transcriptase-polymerase chain reaction
SDS	sodium dodecyl sulphate
sdH_2O	sterile distilled water
SS	sequence-specific
SSC	sodium chloride/sodium citrate buffer
ssDNA	single-stranded DNA
TAE	tris/acetate/EDTA buffer
TBE	tris/borate/EDTA buffer
TEMED	N, N′, N′, N′-tetramethylethylenediamine
TESS	transcription element sequence software
UPW-PCR	unpredictably primed walking-PCR
UV	ultraviolet
WP	walking primer
WWW	world wide web
X-gal	5-bromo-4-chloro-3-indoyl-B-D-galactoside

Appendix

List of amino acids: Three and one letter codes are given with a brief description of their properties (Stryer, 1988). Residues are grouped according to the properties of their side chains.

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

PUBLICATIONS

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

2. Fitzgibbon J, Pilz A, Gayther S, **Appukuttan B**, Dulai KS, Delhanty JD, Helkamp GM, Yarbrough LR and Hunt DM (1994) Localisation of the gene encoding human phosphatidylinositol transfer protein (PITPN) to 17p13.3: a gene showing homology to the *Drosophila* retinal degeneration B gene (*rdgB*). *Cytogenet Cell Genet* **67**: 205-207.

3. Hunt DM, Cowing JA, Patel R, **Appukuttan B**, Bowmaker JK and Mollon JD (1995) Sequence and evolution of the blue cone pigment gene in Old and New World primates. *Genomics* **27**: 535-538.

4. **Appukuttan B** and Hunt DM (1996) Evolution of the S group of visual pigments and identification of potential spectral tuning sites involved in the blue shift. In preparation.

5. **Appukuttan B**, Dulai KS and Hunt DM (1996) Characterisation of the 5′ flanking proximal promoter region of mammalian cone opsin genes. In preparation.

ABSTRACTS

1. Hunt DM, Cowing JA, Bowmaker JK, Patel R, **Appukuttan B** and Mollon JD (1995) Sequence and spectral tuning of blue cone photopigments in Old and New World primates. *Investigative Ophthalmology and Visual Science* **37**: 5889 (suppl).

2. **Appukuttan B**, Dulai KS and Hunt DM (1996) Isolation and characterisation of protein binding sites within the 5′ flanking region of blue cone pigment genes. *Investigative Ophthalmology and Visual Science* **37**: 1547 (suppl).

References

Ali MA and Klyne MA (1985) Vision in vertebrates. New York, Plenum Press.

Alkorta I and Du P (1994) Sequence divergence analysis for the prediction of seven-helix membrane protein structures: A 3-D model of human rhodopsin. *Protein Engineering* **7**(10): 1231-1238.

Arcot S, Wang Z, Weber JL, Deininger PL and Batzer MA (1995) *Alu* repeats: A source for the genesis of primate microsatellites. *Genomics* **29**: 136-144.

Asenjo AB, Rim J and Oprian DD (1994) Molecular determinants of human red/green color discrimination. *Neuron* **12**: 1131-1138.

Balciuniene J, Johansson K, Sandgren O, Wachtmeister L, Holmgren G and Forsman K (1995) A gene autosomal dominant progressive cone dystrophy (CORD5) maps to chromosome17p12-p13. *Genomics* **30**: 281-286.

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

Berg JM (1992) Sp1 and the sub family of zinc finger proteins with guanine-rich binding sites. *Proc Nat Acad Sci* **89**: 11109-11110.

Blackwell TK and Weintraub H (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**: 1104-1110.

Bobola N, Hirsch E, Albini A, Altruda F, Noonan D and Ravazzolo R (1995) A single *cis*-acting element in a short promoter segment of the gene encoding the interphotoreceptor retinoid-binding protein confers tisssue-specific expression. *J Biol Chem* **270**: 1289-1294.

Bogenmann E, Lochrie MA and Simon MI (1988) Cone cell-specific genes expressed in retinoblastoma. *Science* **240**: 76-78.

Bönigk W, Altenhofen W, Müller F, Dose A, Illing M, Molday RS and Kaupp UB (1993) Rod and cone photoreceptor cells express distinct genes for cGMP-gated channels. *Neuron* **10**: 865-877.

Bowmaker JK (1983) How is colour perceived: The visual pigments of human cones. *Trans Opthal Soc UK* **103**: 373-379.

Bowmaker JK (1984) Microspectrophotometry of vertebrate photoreceptors. *Vision research* **24**(11): 1641-1650.

Bowmaker JK (1991) Photoreceptors, photopigments and oil droplets. In: *The Perception of Colour* (Gouras P, eds) Vision and Visual Dysfunction. Vol **6**: 108-127, Macmillan, London.

Bowmaker JK, Astell S, Hunt DM and Mollon JD (1991a) Photosensitive and photostable pigments in the retinae of Old World monkeys. *J Exp Biol* **156**: 1-19.

Bowmaker JK, Thorpe A and Douglas RH (1991b) Ultraviolet sensitive cones in the goldfish. *Vision Research* **31**: 349-352.

Bresnick GH, Smith VC, Pokorny J (1989) Autosomal dominantly inherited macular dystrophy with preferential short wavelength sensitive cone involvement. *Am J Ophthalmol* **108**: 265-276.

Bunge S, Wedemann H, David D, Terwilliger DJ, Van Den Born LI, Aulehla-Scholz C, Sammans C, Horn M, Ott J, Schwinger E, Schinzel A, Denton MJ and Gal A (1993) Molecular analysis and genetic mapping of the rhodopsin gene in families with autosomal dominant retinitis pigmentosa. *Genomics* **17**: 230-233.

Buratowski S, Hahn S, Sharp PA and Guarente L (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* **334**: 37-42.

Cachon-Gonzalez B (1991) Linkage analysis in familial adenomatous polyposis families in the United Kingdom, and a search for highly polymorphic markers. PhD Thesis, University College London.

Chan T, Lee M and Sakmar TP (1992) Introduction of hydroxyl-bearing amino acids causes bathochromic spectral shifts in rhodopsin. *J Biol Chem* **267**: 9478-9480.

Chang BSW, Crandall KA, Carulli JP and Hartl DL (1995) Opsin Phylogeny and evolution: A model for blue shifts in wavelength regulation. *Molecular Phylogenetics and Evolution* **4**(1): 31-43.

Chen J, Tucker CL, Woodford B, Szel A, Lem J, Gianella-Borradori A, Simon MI and Bogenmann E (1994) The human blue cone opsin promoter directs transgene expression in short wave and cones bipolar cells in the mouse retina. *Proc Natl Acad Sci* **91**: 2611-2615.

Chiu MI and Nathans J (1994a) A sequence of the mouse blue visual pigment gene directs blue cone-specific transgene expression in mouse retinas. *Visual Neuroscience* **11**: 773-780.

Chiu MI and Nathans J (1994b) Blue cones and cone bipolar cells share transcriptional specificty as determined by expression of human blue visual pigment-derived transgenes. *The Journal of Neuroscience* **14**: 3426-3436.

Chiu MI, Zack DJ, Wang Y and Nathans J (1994) Murine and bovine blue cone pigment genes: Cloning and characterisation of two new members of the S family of visual pigments. *Genomics* **21**: 440-443.

Church GM, Ephrussi A, Gilbert W and Tonegawa S (1985) Cell-type-specific contacts to immunoglobulin enhancers in nuclei. *Nature* **313**: 798-801.

Church GM and Gilbert W (1984) Genomic sequencing. *Proc Natl Acad Sci USA* **81**: 1991-1995.

Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K and Romanas MA (1991a) High-level expression of Tetanus Toxin Fragment c in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* **9**: 455-460.

Clare JJ, Romanas MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K and Henwood CA (1991b) Production of Epidermal Growth Factor in yeast: High-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene* **105**: 205-212.

Clutton-Brock J (1987) A natural history of domesticated mammals. Cambridge University Press, British Museum (Natural History).

Cobbs WH, Barkdoll AE and Pugh EN (1985) Cyclic GMP increases photocurrent and light sensitivity of retinal cones. *Nature* **317**: 64-66.

Creighton Thomas E (1993) Proteins: structures and molecular properties. WH Freeman and Company (second edition).

Daiger SP, Sullivan LS and Rodriguez JA (1995) Correlation of phenotype with genotype in inherited retinal degeneration. *Behavioral and Brain Sciences* **18**: 452-467.

Dartnall HJA, Bowmaker JK and Mollon JD (1983) Human visual pigments: microspectrophotometric results from the eyes of seven persons. *Proc. R. Soc. London* **220**(B): 115-130.

Dignam JD, Lebovitz RM and Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acid Res* **11**: 1475-1489.

Di Polo A and Farber DB (1995) Rod photoreceptor-specific gene expression in human retinoblastoma cells. *Proc Natl Acad Sci USA* **92**: 4016-4020.

Di Polo A, Rickman CB and Farber DB (1996) Isolation and initial characterisation of the 5′ flanking region of the human and murine cyclic guanosine monophosphate-phosphodiesterase β-subunit genes. *Inv Opht Vis Sci* **37**: 551-560.

Dominguez O and López-Larrea C (1994) Gene walking by unpredictably primed PCR. *Nucleic Acids Research* **22**: 3247-3248.

Donis-Keller H, Helms C, Green P, Rietman H, Ramachandra S, Falls K, Bowden DW, Weiffenbach B, Keith T, Stephens K, Cannizzaro LA, Shows TB, Stewart GD, Van Keuren M (1989) A human genome linkage map with more than 500 RFLP loci and average spacing of 6 centiMorgans. *Cytogenet Cell Genet* **51**: 991.

Dulai KS (1996) Molecular genetics of the middlewave and longwave sensitive opsin genes of higher primates. PhD Thesis. University College, London.

Dratz EA and Hargrave PA (1983) The structure of rhodopsin and the rod outer segment disk membrane. *Trends in Biochemical Sciences* **8**: 128-131.

Dryja TP, Hahn LB, Cowley GS, McGee TL and Berson EL (1991) Mutation spectrum of the rhodopsin gene among patients with autosomal dominant retinitis pigmentosa. *Proc Natl Acad Sci USA* **88**: 9370-9374.

Dynan WS and Tjian R (1985) The promoter specific transcription factor Sp1 binds to upstream sequences in the SV40 promoter. *Cell* **35**: 79-87.

Economou EP, Bergen AW, Warren AC and Antonarakis SE (1990) The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* **87**: 2951-2954.

Evans K, Fryer A, Inglehearn C, Duvall-Young J, Whittaker JL, Gregory CY, Butler R, Ebenezer N, Hunt DM and Bhattacharya S (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. *Nature Genetics* **6**: 210-213.

Farber DB (1995) From mice to men: The cyclic GMP phosphodiesterase gene in vision and disease. *Inv Opht Vis Sci* **36**: 263-275.

Feil R, Aubourg P, Heiling R and Mandel JL (1990) A 195 kb cosmid walk encompassing the human Xq28 colour vision genes. *Genomics* **6**: 367-373.

Feinberg AP and Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**: 6-13.

Ferreira PA, Hom JT and Pak WL (1995) Retina-specifically expressed novel subtypes of bovine cyclophilin. *J Biol Chem* **270**: 23179-23188.

Fortini ME and Rubin GM (1990) Analysis of *cis*-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organisation to rhodopsin promoters in *Drosophila melanogaster*. *Genes & Development* **4**: 444-463.

Franke RR, Sakmar TP, Oprian DD and Khorana HG (1988) A single amino acid substitution in rhodopsin (lysine-248-leucine) prevents activation of transducin. *J Biol Chem* **263**: 2119-2122.

Franke RR, Sakmar TP, Graham RM and Khorana HG (1992) Structure and function in rhodopsin: Studies of the interaction between the rhodopsin cytoplasmic domain and transducin. *J Biol Chem* **267**: 14767-14774.

Fukuda MN, Papermaster DP and Hargrave PA (1979) Rhodopsin carbohydrate: Structure of small oligosaccharides attached at two sites near the NH_2 terminus. *J Biol Chem* **254**: 8201-8207.

Gerber HP, Seipel K, Georgiev O, Hofferer M, Hug M, Rusioni S and Schaffner W (1994) Transcriptional activation modulated by homopolymeric glutamine and proline residues. *Science* **263**: 808-811.

Gerster T, Mathias P, Thali M, Jarciny J and Schaffner W (1987) Cell type-specificity of the immunoglobulin heavy chain gene enhancer. *EMBO Journal* **6**: 1323-1330.

Gidoni D, Dynan WS and Tjian R (1984) Mutiple specific contacts between a mammalian transcription factor and its cognate promoters. *Nature* **312**: 409-413.

Goldstein EB (1989) Sensation and perception. Wadsworth Publishing Company (Third edition).

Hargrave PA, McDowell JH, Curtis DR, Wang JK, Juszczak E, Fong SL, Mohana Rao JK and Argos P (1983) The structure of bovine rhodopsin. *Biophys Struct Mech* **9**: 235-244.

Higgins DG and Sharp PM (1988) CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-244.

Hisatomi O, Satoh T, Barthel LK, Stenkamp DL, Raymond PA and Tokunaga F (1996) Molecular cloning and characterisation of the putative ultraviolet-sensitive visual pigment of goldfish. *Vision Res* **36**: 933-939.

Hobbs HH, Lehrman MA, Yamamoto T and Russell DW (1985) Polymorphism and evolution of *Alu* sequences in the human low density lipoprotein receptor gene. *Proc Natl Acad Sci USA* **82**: 7651-7655.

Hong FD, Huang H-JS, To H, Young L-JS, Oro A, Bookstein R, Lee EY-HP and Lee W-H (1989) Structure of the human retinoblastoma gene. *Proc Natl Acad Sci USA* **86**: 5502-5506.

Hong H-K, Ferrell RE and Gorin MB (1993) Clinical diversity and chromosomal localisation of X-linked cone dystrophy (COD1). *Am J Hum Genet* **55**: 1173-1181.

Hubbard R and Sperling L (1973) The colours of the visual pigment chromophores. *Exp Eye Res* **17**: 581-589.

Hunt DM, Cowing J, Patel R, Appukuttan B, Bowmaker JK and Mollon JD (1995) Sequence and evolution of the blue cone pigment gene in Old and New World primates. *Genomics* **27**: 535-538.

Hunt DM, Fitzgibbon J, Slobodyanyuk SJ and Bowmaker JK (1996) Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in lake Baikal. *Vision Res* **36(9)**: 1217-1224.

Hurwitz RL, Bunt-Milam AH, Chang ML, Beavo JA (1985) Cyclic GMP phosphodiesterase in rod and cone outer segments of the retina. *J Biol Chem* **260**: 568-573.

Ibbotson RE, Hunt DM, Bowmaker JK and Mollon JD (1992) Sequence divergence and copy number of the middle- and long-wave photopigment genes in Old World monkeys. *Proc R Soc London* **247**(B): 145-154.

Innis MA and Gelfand DH (1990) PCR protocols: A guide to methods and applications. Academic Press, Inc London.

Invitrogen® (1994) A instruction manual of methods for expression of recombinant proteins in *Pichia pastoris*. Version 1.5.

Jacobs GH and Neitz J (1987) Inheritance of colour vision in a New World monkey (*Saimiri sciureus*). *Proc Natl Acad Sci USA* **84**: 2545-2549.

Jacobs GH, Neitz J and Deegan JF (1991) Retinal receptors in rodents maximally sensitive to ultraviolet light. *Nature* **353**: 655-656.

Jacobs GH (1993) The distribution and nature of color vision among mammals. *Biol Rev* **68**: 413-471.

Jacobs GH, Neitz M, Deegan JF and Neitz J (1996) Trichromatic colour vision in New World monkeys. *Nature* **382**: 156-158.

Johnson PF and McKnight SL (1989) Eukaryotic transcriptional regulatory proteins. *Annual Review of Biochemistry* **58**: 799-839.

Johnson RL, Grant KB, Zankel TC, Boehm MF, Merbs SL, Nathans J and Nakanishi K (1993) Cloning and expression of goldfish opsin sequences. *Biochemistry* **32**: 208-214.

Jordan SA, Farrar GJ, Kenna P, Humphries MM, Sheils DM, Kumar-Singh R, Sharp EM, Soriano N, Ayuso C, Benitez J and Humphries P (1993) Localisation of an autosomal dominant retinitis pigmentosa gene to chromosome 7q. *Nature Genet* **4**: 54-58.

Karnik SS, Sakmar TP, Chen HB and Khorana HG (1988) Cysteine residue-110 and residue-187 are essential for the formation of correct structure in bovine rhodopisn. *Proc Natl Acad Sci USA* **85**: 8459-8463.

Karnik SS and Khorana HG (1990) Assembly of functional rhodopsin requires a disulfide bond between cysteine residues 110 and 187. *Proc Natl Acad Sci USA* **265**: 17520-17524.

Khorana HG (1992) Rhodopsin, photoreceptor of the rod cell. *J Biol Chem* **276**: 1-4.

Kikuchi T, Raju K, Breitman ML and Shinohara T (1993) The proximal promoter of the mouse arrestin gene directs gene expresssion in photoreceptor cells and contains an evolutionarily conserved retinal factor-binding site. *Mol Cell Biol* **13**: 4400-4408.

Kikuchi T and Shinohara T (1994) Photoreceptor conserved element (PCE1) associated with photoreceptor-specific promoter activity were found in arrestin genes of -12 to -24 of the mouse and the first intron of bovine. *Inv Opht Vis Sci* **35** :(S) 1578.

Konig B, Arendt A, McDowell JH, Kahlert M, Hargrave PA and Hofmann KP (1989) Three cytoplasmic loops of rhodopsin interact with transducin. *Proc Natl Acad Sci USA* **86**: 6878-6882.

Kosower EM (1988) Assignment of groups responsible for the "opsin shift" and light absorptions of rhodopsin and red, green, and blue iodopsins (cone pigments). *Proc Natl Acad Sci USA* **85**: 1076-1080.

Kropf A and Hubbard R (1958) The mechanism of bleaching rhodopsin. *Ann N Y Acad Sci* **74**: 266-280.

Latchman DS (1993) Transcription factors a practical approach. Oxford University Press, (first edition).

Latchman D (1995) Gene regulation: a eukaryotic perspective. Chapman and Hall, (second edition).

Lee W-H, Bookstein R, Hong FD, Young L-J, Shew J-Y and Lee EY-HP (1987a) Human retinoblastoma susceptibility gene: cloning, identification and sequence. *Science* **235**: 1394-1399.

Lee W-H, Shew J-Y, Hong FD, Sery TW, Donoso LA, Young L-J, Bookstein R and Lee EY-HP (1987b) The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. *Nature* **329**: 642-645.

Lem J, Applebury ML, Falk JD, Flannery JG and Simon MI (1991) Tissue-specific and developmental regulation of rod opsin chimeric genes in transgenic mice. *Neuron* **6**: 201-210.

Lerea CL, Somers DE, Hurley JB, Klock IB and Bunt-Milam AH (1986) Identification of specific transducin α subunits in retinal rod and cone photoreceptors. *Science* **234**: 77-80.

Lerea CL, Bunt-Milan AH and Hurley B (1989) α transducin is present in blue-, green- and red-sensitive cone photoreceptors in the human retina. *Neuron* **3**: 367-376.

Levitan IB and Kaczmarek LK (1991) The neuron: cell and molecular biology. Oxford University Press, (First edition)

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

Liou GI, Matragoon S, Overbeek PA and Fei Y (1994) Identification of a retina-specific footprint within the retina-specific regulatory region of the human interphotoreceptor retinoid-binding protein gene. *Biochem Biophys Res Commun* **203**: 1875-1881.

Long KO and Fisher SK (1983) The distribution of photoreceptors and ganglion cells in the California ground squirrel (*Spermophilus beechevi*). *Journal of Comparative Neurology* **221**: 329-340.

Maden BEH (1995) Opsin genes: *Essays in Biochemistry* **29**: 87-111.

Mathies R and Stryer L (1976) Retinal has a highly dipolar vertically excited singlet state: implications for vision. *Proc Natl Acad Sci USA* **73**: 2169-2173.

McFall RC, Sery TW and Makadon M (1977) Characterisation of a new continuous cell line derived from a human retinoblastoma. *Cancer Research* **37**: 1003-1010.

McFall RC, Nagy RM, Nagle BT and McGreevy LM (1978) Scanning electron microscopic observation of two retinoblastoma cell lines. *Cancer Research* **38**: 2827-2835.

Merbs SL and Nathans J (1992) Absorption spectra of human cone pigments. *Nature* **356**: 433-35.

Miller WH and Snyder AW (1977) The tiered vertebrate retina. *Vision Res* **17**: 239-255.

Mismer D, Michael MW, Laverty TR and Rubin GM (1988) Analysis of the promoter of the Rh2 opsin gene in *Drosophila melanogaster*. *Genetics* **120**: 173-180.

Mitchell PJ and Tjian R (1989) Transcriptional regulation in mammalian cells by sequence specific DNA binding proteins. *Science* **245**: 371-378.

Mollon JD (1991) Hue and the heptahelicals. *Nature* **351**: 696-697.

Mollon JD, Bowmaker JK and Jacobs GH (1984) Variations of color-vision in a New World primate can be explained by polymorphism of retinal photopigments. *Proceedings of the Royal Society of London Series (B-Biological Sciences)* **222**: 373-399.

Morabito MA, Yu X and Barnstable CJ (1991) Characterisation 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**: 9667-9672.

Nakayama TA and Khorana HG (1990) Orientation of retinal in bovine rhodopsin determined by crosslinking using a photoactivatable analog of 11-*cis*-retinal. *J Biol Chem* **265**: 15762-15769.

Nakayama TA and Khorana HG (1991) Mapping of the of the amino acids in membrane-embedded helices that ineract with the retinal chromophore in bovine rhodopsin. *J Biol Chem* **266**: 4269-4275.

Nathans J (1987) Molecular biology of visual pigments. *Ann Rev Neurosci* **10**: 163-194.

Nathans J (1992) Rhodopsin: Structure, function and genetics. *Biochemistry* **31**: 4923-4931.

Nathans J and Hogness DS (1983) Isolation, sequence analysis and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* **34**: 807-814.

Nathans J and Hogness DS (1984) Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc Natl Acad Sci USA* **81**: 4851-55.

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

Nathans J, Piantanida TP, Eddy RL, Shows TB and Hogness DS (1986b) Molecular genetics of inherited variation in human color vision. *Science* **232**: 203-202.

Nathans J, Weitz CJ, Agarwal N, Nir I and Papermaster DS (1989) Production of bovine rhodopsin by mammalian cell lines expressing cloned cDNA: spectrophotometry and subcellular localisation. *Vision Research* **29**: 907-914.

Neitz J and Jacobs GH (1989) Spectral sensitivity of cones in an ungulate. *Visual Neuroscience* **2**: 97-100.

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

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH and Caskey CT (1989) *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci USA* **86**: 6686-6690.

Nie Z, Chen S, Kumar R and Zack DJ (1996) RER, an evolutionary conserved sequence upstream of the rhodopsin gene, has an enhancer activity. *J Biol Chem* **271**: 2667-2675.

Occleston N (1996) Cellular and molecular basis of wound healing: Effects of antimetabolite induced growth arrest on ocular fibroblast behaviour. PhD Thesis, University College, London.

Okano T, Kojima D, Fukada Y, Shichida Y and Yoshizawa T (1992) Primary structures of chicken cone visual pigments: Vertebrate rhodopsins have evolved out of cone visual pigments. *Proc Natl Acad Sci USA* **89**: 5932-5936.

Ong OC, Yamane HK, Phan KB, Fong HKW, Bok D, Lee RH and Fung BKK (1995) Molecular cloning and characterisation of the G protein γ subunit of cone photoreceptors. *J Biol Chem* **270**: 8495-8500.

Oprian DD, Molday RS, Kaufman RJ and Khorana HG (1987) Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc Natl Acad Sci USA* **84**: 8874-8878.

Oprian DD, Asenjo AB, Lee N and Pelletier SL (1991) Design, chemical synthesis, and expression of genes for the three human colour vision pigments. *Biochemistry* **30**: 11367-72.

Partridge JC, Shand J, Archer SN, Lythgoe and van Groningen-Luyben WAHM (1989) Interspecific variation in the visual pigments of deep-sea fishes. *J Comp Physiol A* **164**: 513-529.

Patel RJ, Keen TJ, Grzeschik K-H, Nierman WC, Hayes P, Bhattacharya SS and Inglehearn CF (1995) Regional assignment of 30 expressed sequence tags on human chromosome 7 using a somatic cell hybrid panel. *Genomics* **30**: 112-114.

Pettersson S, Cook GP, Brüggeman M, Williams GT and Neuberger MS (1990) A second B cell-specific enhancer 3' of the immunoglobulin heavy chain locus. *Nature* **344**: 165-168.

Piatigorsky J and Wistow GJ (1989) Enzyme/crystallins: Gene sharing as an evolutionary strategy. *Cell* **57**: 197-199.

Piriev NI, Viczian AS, Ye J, Kerner B, Korenberg JR and Farber DB (1995) Gene structure and amino acid sequence of the human cone photoreceptor cGMP-phosphodiesterase α subunit (PDEA2) and its chromosomal localisation to 10q24. *Genomics* **28**: 429-435.

Pittler SJ, Huang X, Feinstein DS and Oliveira L (1996) Characterisation of the upstream region of the PDEA gene. *Inv Opht Vis Sci* **37** (S): 1541.

Pokorney J, Graham CH and Lanson RN (1968) Effect of wavelength on foveal grafting acuity. *Journal of the Optical Society of America* **58**: 1410-1414.

Postel EH, Berberich SJ, Flint SJ and Ferrone CA (1993) Human *c-myc* transcription factor PuF identified as nm23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. *Science* **261**: 478-480.

Provencio I, Loew ER and Foster RG (1992) Vitamin A_2-based visual pigments in fully terrestrial vertebrates. *Vision Res* **32**: 2201-2208.

Rajmondjean M, Cereghini C and Yaniv M (1988) Several distinct 'CCAAT' box binding proteins coexist in eukaryotic cells. *Proc Natl Acad Sci USA* **85**: 757-761.

Rayer B, Naynert M and Stieve H (1990) Phototransduction: different mechanisms in vertebrates and invertebrates. *Journal of Photochemistry and Photobiology* **7**: 107-148.

Redondo JM, Hata S, Brocklehurst C and Krangel MS (1990) A T cell-specific transcriptional enhancer within the human T cell receptor δ locus. *Science* **247**: 1225-1229.

Reid TW, Albert DM, Rabson AS, Russel P, Craft J, Chu EW, Tralka TS and Wilcox JL (1974) Characteristics of an established cell line of retinoblastoma. *J Natl Cancer Inst* **53**: 347-360.

Ridge KD, Bhattacharya S, Nakayama TA and Khorana HG (1992) Light-stable rhodopsin. *J Biol Chem* **267**: 6770-6775.

Röhlich P, van Veen TH and Szél A (1994) Two different visual pigments in one retinal cone cell. *Neuron* **13**: 1159-1166.

Rosenfeld PJ and Kelley TJ (1986) Purification of nuclear factor 1 by DNA recognition site affinity chromatography. *J Biol Chem* **261**: 1398-1408.

Rupp RAW, Kruse U, Multhaup G, Govel U, Beyreuther K and Sippel AE (1990) Chicken NF1/TGGCA proteins are encoded by at least three independent genes: NF1-A, NF1-B and NF1-C with homologues in mammalian genomes. *Nucleic Acids Research* **18**: 2607-2616.

Rushton WAH (1960) Rhodopsin measurement and dark adaptation in a subject deficient in cone vision. *Journal of Physiology* **156**: 193-205.

Saibil H (1990) Cell and molecular biology of photoreceptors. *Seminars in the Neurosciences* **2**: 15-23.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA and Arnheim N (1985) Enymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.

Sakmar TP, Franke RR and Khorana HG (1989) Glutamic acid-113 serves as the retinylidene Schiff base counterion in bovine rhodopsin. *Proc Natl Acad Sci USA* **86**: 8309-8313.

Sambrook T, Fritsh EF and Maniatis T (1989) *Molecular cloning: A laboratory manual.* **1-3**: CSH Lab Press USA.

Sanford C and Mukai S (1993) Genetics of retinoblastoma. *International Ophthalmology Clinics* **33**: 57-65.

Sanger F, Nicklen ST and Coulsen AR (1977) DNA sequencing with chain termination inhibitors. *Proc Natl Acad Sci USA* **74**: 5463.

Santoro C, Mermod N, Andrews PC and Tijian R (1988) A family of human CCAAT box binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature* **334**: 218-224.

Schertler GFX, Villa C and Henderson R (1993) Projection structure of rhodopsin. *Nature* **362**: 770-772.

Schertler GFX and Hargrave PA (1995) Projection structure of frog rhodopsin in two crystal forms. *Proc Natl Acad Sci USA* **92**: 11578-11582.

Schmid CW and Jelinek WR (1982) The Alu family of dispersed repetitive sequences. *Science* **216**: 1065-1070.

Sheshberadaran H and Takahashi J (1994) Characterisation of the chicken rhodopsin promoter: identification of retina-specific and glass-like protein binding domains. *Molecular and Cellular Neurosciences* **5**: 309-318.

Singh H, Sen R, Baltimore D and Sharp PA (1986) A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319**: 154-158.

Singh H, Le Bowitz JH, Baldwin AS and Sharp PA (1988) Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. *Cell* **52**: 415-429.

Small KW, Syrquin M, Mullen L and Gehrs K (1996) Mapping of autosomal dominant cone degeneration to chromosome 17p. *American Journal of Ophthalmology* **121**: 13-18.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-518.

Strachan T and Read AP (1996) Human molecular genetics. Bios Scientific Publishers Ltd, (First edition).

Stryer L (1988) Biochemistry. WH Freeman and Company, (third edition).

Tovee MJ, Bowmaker JK and Mollon JD (1992) The relationship between cone pigments and behavioural sensitivity in a New World monkey (*Callithrix jacchus jacchus*). *Vision Res* **32(5)**: 867-878.

Ts'o MOM, Fine BS and Zimmerman LE (1970) The nature of retinoblastoma I, photoreceptor differentiation: a clinical and histopathological study. *American J Ophthalmology* **69**: 339-349.

Vogt K and Kirschfeld K (1984) Chemical identity of the chromophore of fly visual pigment. *Naturewissenchaften* **71**: 211-213.

Vollrath D, Nathans J and Davies RW (1988) Tandem array of human visual pigment genes at Xq28. *Science* **240**: 1669-1672.

Wang JK, McDowell JH and Hargrave PA (1980) Site of attachment of 11-*cis* -retinal in bovine rhodopsin. *Biochemsitry* **19**: 5111-5117.

Wang Y, Macke JP, Merbs SL, Zack DJ, Klaunberg B, Bennet J, Gearhart J and Nathans J (1992) A locus control region adjacent to the human red and green visual pigment genes. *Neuron* **9**: 429-440.

Wang Z, Asenjo AB and Oprian DD (1993) Identification of the Cl^- binding site in the human red and green color vision pigments. *Biochemistry* **32(9)**: 2125-2130.

Weatherall DJ (1991) The new genetics and clinical practice. Oxford University Press, (Third edition).

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G and Lathrop M (1992) A second-generation linkage map of the human genome. *Nature* **359**: 794-801.

Weitz CJ, Went LN and Nathans J (1992) Human tritanopia associated with a third amino acid substitution in the blue-sensitive visual pigment. *Am J Hum Genet* **51:** 444-446

Went LN, Schooneveld MJ va, Oosterhuis JA (1992) Late onset dominant cone dystrophy with early blue cone involvement. *J Med Genet* **29**: 295-298.

Wilden U and Kuhn H (1982) Light dependent phosphorylation of rhodopsin: number of phosphorylation sites. *Biochemistry* **21**: 3014-3022.

Wilden U, Hall SW and Kuhn H (1986) Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kDa protein of rod outer segments. *Proc Natl Acad Sci USA* **83**: 1174-1178.

Williams AJ, Hunt DM, Bowmaker JK and Mollon JD (1992) The polymorphic photopigments of the marmoset: spectral tuning and genetic basis. *The EMBO Journal* **11**(6): 2039-2045.

Wirth T, Staudt L and Baltimore D (1987) An octamer oligonucleotide upstream of a TATA motif is sufficient for lympoid-specific promoter activity. *Nature* **329**: 174-178

Yau KW (1994) Phototransduction mechanism in retinal rods and cones. *Invest Ophthalmol Vis Sci* **35**: 9-32.

Yokoyama S (1994) Gene duplication and evolution of the short wavelength-sensitive visual pigments in vertebrates. *Mol Biol Evol* **11**: 32-39.

Yokoyama T, Liou G, Caldwell RB and Overbeek PA (1992) Photoreceptor-Specific Activity of the Human Interphotoreceptor Retinoid-Binding Protein (IRBP) Promoter in Transgenic Mice. *Exp Eye Res* **55**: 225-233.

Young RW (1971) The renewal of rod and cone outer segments in the rhesus monkey. *J. Cell. Biol.* **49**: 303-318.

Young T (1801) On the theory of light and colours. *Philosophical Transactions of the Royal Society of London* **92**: 12-48.

Zack DJ, Bennett J, Wang Y, Davenport C, Klaunberg B, Gearhart J and Nathans J (1991) Unusual topography of bovine rhodopsin promoter-*lacZ* fusion gene expression in transgenic mouse retinas. *Neuron* **6**: 187-199.

Zhukosky EA and Oprian DD (1989) Effect of carboxylic acid side chains on the absorption maximum of visual pigments. *Science* **246**: 928-930.

ADDENDUM

Crescitelli F (1972) The visual cells and visual pigments of the vertebrate eye. In: *Photochemistry of Vision* (Dartnall HJA eds) Handbook of Sensory Physiology. 245-363. Volume VII/1, Springer, Berlin.

Kito Y, Partridge JC, Seidou M, Narita K, Hamanaka T, Michinomae M, Sekiya N and Yoshihara K (1992) The absorbance spectrum and photosensitivity of a new synthetic "visual pigment" based on 4-hydroxyretinal. *Vision Research* **32**: 3-10.

Zubay G (1993) *Biochemistry*: Chapter 36. Wm C Brown Publishers, (third edition).

INSTITUTE OF OPHTHALMOLOGY

Hum Genet (1994) 93:79–80

human genetics

© Springer-Verlag 1994

Localisation of the human blue cone pigment gene to chromosome band 7q31.3–32

Jude Fitzgibbon[1], Binoy Appukuttan[1], Simon Gayther[2], Dagan Wells[2], Joy Delhanty[2], David M. Hunt[1]

[1] Department of Molecular Genetics, Institute of Opthalmology, University of London, Bath Street, London ECIV 9EL, UK
[2] The Galton Laboratory, University College London, University of London, 4 Stephenson Way, London NW1 2HE, UK

Received: 30 July 1993

Abstract. Blue cone pigment (BCP) is one of three types of cone photoreceptors responsible for normal colour vision. In this study, the BCP gene has been localised to chromosome 7q31.3–32 by fluorescent in situ hybridisation of cosmid clones containing the gene. This is consistent with previous mapping of the BCP gene to chromosome 7q31–35.

Introduction

Retinitis pigmentosa (RP) is a group of heterogeneous disorders in which loss of peripheral vision precedes any central involvement; mutations in the rod photoreceptor pigment gene that codes for the protein component of the visual pigment rhodopsin are known to be a major cause of autosomal dominant RP (Dryja et al. 1990; Inglehearn et al. 1991; Sheffield et al. 1991; Sung et al. 1991). Rod photoreceptors are important for scotopic vision, whereas normal photopic colour vision depends on the presence of three types of cone photoreceptor that contain either a red-, green-, or blue-sensitive pigment. The opsin component of each pigment is coded by a separate gene, the red and green opsins by adjacent genes on the X chromosome and the blue opsin by a gene (BCP) mapped to chromosome 7. Analagous mutations in the cone opsin genes to those in rhodopsin would be expected to cause retinal degeneration, with central involvement preceeding any peripheral loss of vision. So far, however, only a single mutation of this type has been described (Reichel et al. 1989), although there are two reports of cone dystrophy with early blue cone involvement (Bresnick et al. 1989; Went et al. 1992). The cone pigment genes are candidates therefore for particular types of retinal degeneration. The red and green opsin genes have been located to Xq28 (Feil et al. 1990) but the mapping of the BCP gene to the long arm of chromosome 7 has been less accurately determined (Nathans et al. 1986a; Donis-Keller et al. 1989). In this paper, we report the precise chromosomal localisation of the BCP gene.

Correspondence to: J. Fitzgibbon

Materials and methods

Preparation of probes by the polymerase chain reaction

Four oligonucleotides corresponding to different parts of the coding region of the BCP gene sequence were used (Nathans et al. 1986b): primer 1^+ (5′ ATGAGAAAAATGTCGGAGGAAG 3′) at position 411–432, primer 2^- (5′ TGGGATGGAGACGCCAATAC 3′) at position 911–892, primer 4^+ (5′ GTTGCAGCTCAGCAGCAGGA 3′) at position 1098–1117 and primer 5^- (5′ GTTGGGGCCAACTTGGGTAG 3′) at position 1454–1435. Genomic DNA (200 ng) was amplified using the polymerase chain reaction (PCR) in 25 µl mixtures that included 10 m*M* TRIS-HCl pH 8.8, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 200 µmol of each dNTPs, 50 ng of each primer and 0.5 U *Taq* polymerase (Northumbria Biologicals). The reactions were incubated for 3 min at 94°C, followed by 35 cycles at 94°C for 30 s, at 56°C for 1 min and at 72°C for 90 s.

Isolation of genomic clones

A human cosmid library (Cachon-Gonzales 1991) consisting of approximately 5×10^5 independent recombinants was screened with the [α^{32}P] dCTP-labelled 786 bp PCR fragment obtained by amplification of human DNA with primers 1^+ and 2^-. Cosmid DNA was isolated from purified positive clones and verified by PCR using the 4^+ and 5^- primer pair.

Fluorescent in situ hybridisation

The procedures for fluorescent in situ hybridisation were adapted from the original description by Pinkel et al. (1986). DNA probes were prepared from 10 ml bacterial cultures, grown overnight at 37°C, using the Magic Miniprep DNA purification system (Promega). Biotin-labelled probes were hybridised to human chromosome spreads prepared from normal lymphocyte cultures, at 37°C overnight, followed by signal detection with avidin-fluorescein isothiocyanate (avidin-FITC). For R-banding, chromosomes were counterstained with propidium iodide and diamidophenylindole (DAPI), and visualised under UV light with an FITC filter for simultaneous viewing of signals and chromosomes. Approximately 20 metaphase chromosome spreads were analysed for accurate assessment of signal localisation.

Results and discussion

A probe for the BCP gene was generated using the primer pair 1^+ and 2^-. Amplification of genomic DNA yielded a

Fig. 1. Localisation of the human BCP gene to the long arm of chromosome 7, bordering bands q31.3 and q32. The figure demonstrates two pairs of signals (*arrows*) and includes the separation and enlargement of both copies of chromosome 7 (*left*)

specific PCR product of 786 bp. A human cosmid library was screened with this probe and two clones were identified. Their authenticity was confirmed by PCR amplification using the primer pair 4^+ and 5^- giving a specific PCR product of 1344 bp.

In situ hybridisation of biotinylated cosmid DNA to human metaphase chromosome spreads localised the BCP gene to the long arm of chromosome 7, band q31.3–32 (Fig. 1). This is consistent with previous mapping data localising the gene to chromosome 7q31–35 (Nathans et al. 1986a; Donis-Keller et al. 1989). Specific signals were observed on each chromatid of the chromosome 7 homologues in all the metaphases analysed.

In view of the strong homology between the gene for rhodopsin, implicated in approximately 25% of all autosomal dominant RP cases (Sung et al. 1991), and the BCP gene (Nathans et al. 1986b), it is interesting that a new form of autosomal dominant RP has recently been linked to this region of chromosome 7 (Jordan et al. 1993). Our precise assignment may allow its appraisal as a candidate gene for this form of RP. The absence however of a physical assignment for the markers used in this study serves to emphasise the importance of integrating the genetic and physical maps of the genome in order to utilise a candidate gene approach towards the identification of disease loci.

Acknowledgements. The confocal imaging equipment was provided by the Medical Research Council as part of the Human Genome Mapping project. B.A. is the recipient of a Wellcome prize studentship.

References

Bresnick GH, Smith VC, Pokorny J (1989) Autosomal dominant inherited macular dystrophy with preferential short wavelength sensitive cone involvement. Am J Ophthalmol 108:265–276

Cachon-Gonzalez B (1991) Linkage analysis in familial adenomatous polyposis families in the United Kingdom, and a search for highly polymorphic markers. Ph. D. Thesis, University College, London

Donis-Keller, Helms C, Green P, Riethman H, Ramachandra S, Falls K, Bowden DW, Weiffenbach B, Keith T, Stephens K, Cannizzaro LA, Shows TB, Stewart GD, Van Keuren M (1989) A human genome linkage map with more than 500 RFLP loci and average spacing of 6 centiMorgans. Cytogenet Cell Genet 51:991

Dryja TP, McGee TL, Reichel E, Hahn IB, Cowley GS, Yadell DW, Sandberg MA, Berson EL (1990) A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature 343: 361–366

Feil R, Aubourg P, Heilig R, Mandel JL (1990) A 195 kb cosmid walk encompassing the human Xq28 colour vision genes. Genomics 6:367–373

Inglehearn CF, Bashir R, Lester DH, Jay M, Bird AC, Bhattacharya SS (1991) A 3-bp delection in the rhodopsin gene in a family with autosomal dominant retinitis pigmentosa. Am J Hum Genet 48:26–30

Jordan SA, Farrar GJ, Kenna P, Humphries MM, Sheils DM, Kumar-Singh R, Sharp EM, Soriano N, Ayuso C, Benitez J, Humphries P (1993) Localisation of an autosomal dominant retinitis pigmentosa gene to chromosome 7q. Nature Genet 4: 54–58

Nathans J, Piantanida TP, Eddy RL, Shows T, Hogness DS (1986a) Molecular genetics of inherited variations in human colour vision. Science 232:203–210

Nathans J, Thomas D, Hogness DS (1986b) Molecular genetics of human colour vision: the genes encoding blue, green, and red pigments. Science 232:193–202

Pinkel D, Straume T, Gray JM (1986) Cytogenetic analysis using quantitative high sensitivity fluorescent hybridisation. Proc Natl Acad Sci USA 83:2934–2938

Reichel E, Bruce AM, Sandberg MA, Berson EL (1989) An electroretinographic and molecular genetic study of X-linked cone degeneration. Am J Ophthalmol 108:885–895

Sheffield VC, Fishman GA, Beck JS, Kimura AE, Stone EM (1991) Identification of novel rhodopsin mutations associated with retinitis pigmentosa by GC-clamped denaturing gel electrophoresis. Am J Hum Genet 49:699–706

Sung C, Davenport CM, Hennessey JC, Maumenee IH, Jacobson SG, Heckenlively JR, Nowakowski R, Fishman G, Gouras P, Nathans J (1991) Rhodopsin mutations in autosomal dominant retinitis pigmentosa. Proc Natl Acad Sci USA:6481–6485

Went LN, Schooneveld MJ van, Oosterhuis JA (1992) Late onset dominant cone dystrophy with early blue cone involvement. J Med Genet 29:295–298