

**Towards positional cloning of the gene responsible for
autosomal dominant retinitis pigmentosa mapped to
chromosome 19q13.4.**

**Eranga N. Vithana
B.Sc.**

A thesis submitted to the University of London for the
Degree of Doctor of Philosophy

Department of Molecular Genetics
Institute of Ophthalmology
University College London
University of London
Bath Street
London EC1V 9EL

July 1998

ProQuest Number: 10010381

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 10010381

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

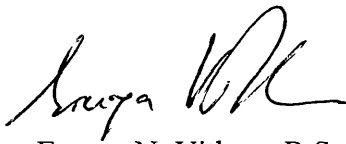
Retinitis pigmentosa (RP) is an inherited group of retinal degenerations that are both clinically and genetically heterogeneous. Autosomal dominant form of RP is caused by mutations in at least 10 loci. The research described in this thesis is focused upon the adRP locus on chromosome 19q13.4 (RP11), which was found by a genome wide linkage analysis on a large British pedigree in our laboratory. At the start of this project RP11 was confined by markers D19S180 and D19S926 to a genetic interval of approximately 11cM and 3 adRP families was linked to RP11 (ADRP5, ADRP29 and RP1907). To further refine this localisation and in order to obtain an estimate of the frequency of RP caused by mutations at this locus 12 more adRP families were genotyped for markers from the disease interval. This led to the identification of two additional RP11 families (ADRP2 and ADRP11) and haplotype analysis in these two families confined the disease region to a 5.8 cM interval between D19S921 and D19S418 and the frequency of the RP11 locus was estimated at approximately 20%.

The gene for protein kinase gamma (*PRKCG*) is localised in the refined RP11 interval and was considered as a positional candidate gene worthy of a thorough investigation. Therefore the genomic organisation of this gene was determined and exon-intron structure was characterised. *PRKCG* gene was found to consist of 18 exons and a genomic span of at least 16 kb. Following its characterisation this gene was subjected to a thorough mutation analysis. The exons, 5' promoter region, and 3'UTR of this gene was screened in 6 families linked to RP11 locus (ADRP5, ADRP29, ADRP2, RP1907, ADRP24 and a Japanese family). In this analysis a mutation was identified in two adRP families (RP1907 and ADRP24) and two sporadic RP cases, which were subsequently identified as being founded upon a common ancestor and therefore a single mutation event rather than several independent events. This mutation resulted in the non-conservative replacement of a basic arginine with an uncharged serine at codon 659. Arg659Ser mutation was not found in over 500 normal chromosomes indicating it to be a rare polymorphism. Unfortunately the absence of *PRKCG* mutations in 7 other RP11 linked pedigrees led to the exclusion of this gene.

In order to further refine the RP11 locus novel microsatellite markers were isolated from the distal end of the RP interval using cosmid clones FISH mapped to the region between D19S927 and D19S418, and analysed in the pedigree ADRP2, which presented the distal flanking cross over with D19S418. In this endeavour a total of four microsatellite markers and 5 useful STSs were generated. One of the novel markers, D19S781.2N, recombined with the disease phenotype in ADRP2 and superseded D19S418 as the distal flanking marker. At this stage another group linked three additional families of North American origin to 19q13.4 and refined the locus proximally replacing D19S921 with D19S572 as the novel proximal flanking marker of the RP11 disease interval. Therefore through the combined genetic data generated both at our laboratory and elsewhere, the RP11 locus was confined between the markers D19S572 (proximal) and D19S781.2N (distal). According to the physical mapping data generated at the Lawrence Livermore National Laboratory (LLNL) Genome Centre the physical distance between these two markers approximate 0.7 Mb. Concurrent to the work described above construction of a YAC contig was also initiated across the RP11 interval, which includes YAC clones from three different YAC libraries. However STS content mapping revealed deletions in several YAC clones that were meant to provide coverage across the RP11 interval, thus indicating a region of instability or a region unclonable in YACs. Therefore PAC clones have replaced YACs as the cloning vector of choice and the current on going work of the project involves the construction of a complete PAC contig across the RP11 interval. Future work will involve the mapping of retinally or brain expressed ESTs, already placed in the 19q13.4 region by the EST mapping consortium (Schuler *et al.*, 1996), to the RP11 critical interval to facilitate the identification of the RP11 gene.

Declaration

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



Eranga N. Vithana B.Sc.

Dedication

I dedicate this thesis to my dearest Amma and Thatha. To my Thatha for sacrificing everything and going without on many an occasion to provide us with the very best and my Amma for all her goodness that has been my inspiration.

I love you.



Acknowledgements

First and foremost I would like to thank my supervisor, Prof. Shomi Bhattacharya for all the guidance, advice, and encouragement given to me throughout these PhD years (especially for all the ‘primer order forms’!). Most of all I am deeply grateful for his understanding and forbearance during the last stages of writing. I also thank Mai (my partner in crime) for all the help in the beginning and setting me on my feet (although it was achieved in a short time you did a good job of it!!). Thanks for being supportive during the ups and downs especially during the ‘is it the gene (or not)’ stage.

I extend my gratitude to my family, my amma & thatha (the most wonderful in the world), sister & brother (the very best) and my adopted sister and brother/BIL (Shanika and Sanjeeva) for all the support and encouragement given to me during the final stages of writing ...deepest thanks from the ‘camel’! Without you this thesis could not have been completed (Special thanks go to Sanjeeva and Sachi for all their help during the very last stages...I will say no more! ☺). Words are not enough to express the gratitude I feel for you all, but I would say this ‘I owe my sanity to you’. I am also grateful to all the many good friends I made at Roseberry and Jennie (flat mate and long suffering friend). I also would like to thank my three ‘S’ friends at the Institute, Sana, Sin, and Su (the order of reference is random!) who have seen me through the best and the worst.... you have been like sisters to me and I am deeply grateful for all your support, love and understanding given to me over a period of seven to four years! I also thank Neil (big brother and ‘agony uncle’), Resh & Debi (the rest of the ‘gang’), Naheed, Kamal (Comp. & ABI wizard AND the Lab Bully!), Zara, Ali, Dawn, David Bessant, Steph, Jill, Vinny and Annette for all your help, friendship encouragement and also for critical analysis of my thesis. Last but not least I would like to thank the special friends I made in Florida, USA. Thank you Ralph and Debs for keeping me company on many a sleepless nights at the Institute and showing me a wonderful time in Florida! I will never forget you all. Thank you very much.

Publications

1. **Vithana, E.**, Al-Magthteh, M., Bhattacharya, S.S. and Inglehearn, C.F. (1998) RP11 is the second most common locus for dominant retinitis pigmentosa. *J Med Genet.* **35**, 174-175.
2. Al-Magthteh, M., **Vithana, E.**, Inglehearn, C.F., Moore, T., Bird, A.C. and Bhattacharya, S.S. (1998) Segregation of a PRKCG mutation in two RP11 families. *Am. J Hum. Genet.* **62**, 1248-1252.
3. Al-Magthteh, M., **Vithana, E.**, Tarttelin, E., Jay, M., Evans, K., Moore, T., Bhattacharya, S.S and Inglehearn, C.F. (1996) Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. *Am J Hum Genet.* **59**, 864-871.
4. Inglehearn, C.F., Tarttelin, E.E., Plant, C., Peacock, R.E., Al-Magthteh, M., **Vithana, E.**, Bird, A.C. and Bhattacharya, S.S (1998) A linkage survey of 20 dominant retinitis pigmentosa families: frequencies of the nine known loci and evidence for further heterogeneity. *J Med Genet.* **35**, 1-5.
5. **Vithana, E.**, Al-Magthteh, M., Lennon, G., Bhattacharya, S.S. and Inglehearn, C.F. (1996) A common locus for autosomal dominant retinitis pigmentosa on chromosome 19q and investigation of a candidate gene. *Am. J. Hum. Genet* **59**, 1382.
6. **Vithana, E.**, Al-Magthteh, M., Bhattacharya, S.S. and Inglehearn, C.F. (1997) Physical map spanning the adRP locus on chromosome 19 and investigation of candidate genes. *Invest. Ophthalmol. Vis. Sci.* **38**, 3684.
7. Al-Magthteh, M., **Vithana, E.**, Bhattacharya, S.S. and Inglehearn, C.F. (1998) Positional cloning of the RP11 gene: exclusion of 3 candidate genes and a mutation in *PRKCG*. *Invest. Ophthalmol. Vis. Sci.* **39**, 1353.
8. **Vithana, E.**, Al-Magthteh, M., Bhattacharya, S.S. and Inglehearn, C.F. (1998) Genomic organisation of human protein kinase C gamma (*PRKCG*) gene. (manuscript in prep.)

Table of Contents

	Page No.
<i>Abstract</i>	2
<i>Declaration</i>	4
<i>Dedication</i>	5
<i>Acknowledgements</i>	6
<i>Publications</i>	7
CHAPTER 1	21
GENERAL INTRODUCTION.	
1.1 THE HUMAN GENOME PROJECT	21
1.2 MAPPING GENES FOR SINGLE GENE DISORDERS BY LINKAGE ANALYSIS	22
1.2.1 <i>Markers used in linkage analysis</i>	25
1.2.1.1 <i>RFLPs and mini satellites</i>	25
1.2.1.2 <i>Microsatellite markers [Short Tandem Repeat polymorphisms (STRP)]</i>	26
1.2.1.3 <i>Single nucleotide polymorphisms (Biallelic markers)</i>	27
1.2.1.4 <i>Evolution of genetic maps and future prospects</i>	28
1.3 MAPPING OF GENES FOR RECESSIVE DISORDERS.....	30
1.3.1 <i>homozygosity mapping</i>	30
1.4 FROM LINKED MARKER TO THE GENE (PHYSICAL MAPPING).....	31
1.4.1 <i>Low resolution physical mapping</i>	32
1.4.1.1 <i>Cytogenetic techniques [rodent-hybrid cell lines and (FISH)]</i>	32
1.4.1.2 <i>Pulse-field Gel Electrophoretic mapping</i>	33
1.4.2 <i>High resolution physical mapping</i>	34
1.4.2.1 <i>Choice of vector</i>	34
1.4.2.2 <i>Contig assembly</i>	36
1.4.2.3 <i>Radiation hybrid mapping</i>	40
1.4.2.4 <i>Physical maps and the current progress of the Human Genome Mapping Project</i>	42
1.5 IDENTIFICATION OF GENES	43
1.5.1 <i>Identifying the gene of interest in positional cloning</i>	43
1.5.2 <i>Development of the human transcript map</i>	46
1.5.3 <i>Cross species comparison for gene identification/ Comparative genomics</i>	47
1.6 RETINAL GENETICS.....	49
1.7 STRUCTURE OF THE RETINA	50
1.7.1 <i>The sensory (Neural) retina</i>	50
1.7.1.1 <i>Photoreceptor cells</i>	53
1.7.2 <i>Retinal pigment epithelium</i>	56

1.8	PHOTOTRANSDUCTION	57
1.9	REGULATION OF PHOTOTRANSDUCTION AND THE RETINAL PROTEINS INVOLVED.....	60
1.9.1	<i>Regulation by Ca²⁺ - binding proteins</i>	61
1.9.2	<i>Regulation by non Ca²⁺ binding proteins</i>	63
1.10	INHERITED RETINAL DEGENERATIONS.....	65
1.10.1	<i>Retinitis pigmentosa</i>	65
1.10.1.1	<i>Autosomal dominant RP (adRP)</i>	67
1.10.1.2	<i>Autosomal recessive RP (arRP)</i>	74
1.10.1.3	<i>X-linked RP (xLRP)</i>	81
1.10.1.4	<i>Congenital stationary night blindness (CSNB)</i>	82
1.10.1.5	<i>Digenic RP</i>	84
1.10.1.6	<i>Syndromic RP</i>	84
1.10.2	<i>Central retinal dystrophies</i>	85
1.10.2.1	<i>Cone-rod dystrophy</i>	85
1.10.2.2	<i>Cone dystrophy</i>	87
1.11	ANIMAL MODELS OF RETINAL DEGENERATION.....	88
1.11.1	<i>rd Mouse</i>	88
1.11.2	<i>rds Mouse</i>	89
1.11.3	<i>rho knockout Mouse:</i>	90
1.11.4	<i>Apoptosis: a final common pathway of photoreceptor cell death</i>	91
1.12	AIMS OF THIS STUDY	92
CHAPTER 2		94
MATERIALS & METHODS.		
2.1	PREPARATION OF DNA	94
2.1.1	<i>DNA extraction from peripheral blood lymphocytes</i>	94
2.1.2	<i>Extraction of cloned DNA (plasmid and cosmid DNA) by alkaline lysis mini prep method</i>	95
2.1.3	<i>Isolation of total yeast DNA in solution. (YAC DNA preparation)</i>	96
2.1.4	<i>Preparation of YAC DNA plugs in agarose</i>	97
2.2	PURIFICATION OF DNA.....	97
2.2.1	<i>Phenol chloroform extraction and ethanol precipitation</i>	97
(a)	<i>Phenol chloroform extraction</i>	97
(b)	<i>Ethanol precipitation</i>	98
2.2.2	<i>Use of Sephacryl microspin columns (Sephacryl-S200 and S-400 HR columns, Pharmacia, UK)</i>	98
2.2.3	<i>Use of Centricon 100 spin columns (Centricon, Princeton, USA)</i>	99
2.3	RESTRICTION ENZYME DIGESTS OF DNA	99
2.4	AMPLIFICATION OF DNA BY PCR	99
2.5	FRACTIONATION OF DNA	100

2.5.1	Agarose gel electrophoresis	100
2.5.2	Denaturing polyacrylamide gel electrophoresis	101
2.5.3	Pulse field gel electrophoresis (PFGE).....	102
2.5.4	Heteroduplex analysis	103
2.6	RADIO-ISOTOPIC LABELLING OF DNA.....	104
2.6.1	5' End-labelling of primers	104
2.6.2	Labelling of DNA probes for hybridisation.....	104
(a)	Random prime method.....	104
2.7	SOUTHERN BLOTTING.....	105
2.8	HYBRIDISATION OF LABELLED PROBE	105
2.8.1	cDNA probe for identification of specific gene sequence.....	105
2.8.2	G ₄ (GT) ₁₃ probe for identification of CA(n) repeats.....	106
2.8.3	Human Cot-1 DNA probe for identification of YAC DNA.....	106
2.8.4	Removal of non-specific hybridisation.....	106
2.8.5	Stripping the filters of all radioactivity.....	107
2.9	FLUORESCENT <i>IN SITU</i> HYBRIDISATION (FISH).....	107
2.10	DNA SEQUENCING.....	108
2.10.1	Direct sequencing of PCR products	108
2.10.2	Sequencing of cloned DNA.....	109
2.10.3	Automated DNA sequencing.....	109
(a)	Preparation of DNA for sequencing.....	110
(b)	Cycle sequencing.....	110
(c)	Electrophoresis and automated analysis	111
2.10.4	Computational analysis of DNA sequence	111
2.11	DNA CLONING (PCR PRODUCTS).....	111
2.11.1	Competent <i>E. coli</i> cells.....	111
2.11.2	Ligation in to specific vectors	112
(a)	Use of pGEM-T vector and pTAG vector	112
2.11.3	Transformation of competent <i>E. coli</i> cells and selection of recombinants.....	112
2.12	Screening of genomic DNA library cloned in Lorist B cosmid vector.....	113
(a)	Preparation of master filters for primary screening.....	113
(b)	Replica filters and preparation for hybridisation.....	114
(c)	Isolation of positive colonies	115
(d)	Secondary and tertiary screening for purification of clones	115
2.13	SUB-CLONING OF COSMIDS INTO PLASMID VECTORS (PBLUE SCRIPT (PBS) AND PUC18)	115
(a)	DNA digestion and purification.....	116
(b)	Dephosphorylation of vector.....	116
(c)	Ligation to plasmid vectors.....	116

(d) Transformation into competent <i>E. coli</i> cells and isolation of positive clones.....	117
2.14 MICROSATELLITE ANALYSIS	117
2.14.1 Radio-isotope detection by 5' end labelling or incorporation of radiolabel in to the PCR product`	117
CHAPTER 3	120
GENETICS OF RP11 AND REFINEMENT OF THE LOCUS.	
3.1 INTRODUCTION	120
3.1.1 Autosomal dominant retinitis pigmentosa locus linked to chromosome 19q13.4 (RP11)	120
3.1.2 Clinical phenotype of RP11 families and partial penetrance.....	121
3.1.3 Possible causes for the reduced penetrance of RP11.....	124
3.1.4 Linkage history between RP11 and chromosome 19q 13.4 markers	127
3.1.5 Aims of study	127
3.2 MATERIALS AND METHODS	130
3.2.1 Clinical assessment and DNA samples.....	130
3.2.2 Microsatellite analysis	130
3.2.3 Linkage analysis.....	130
3.3 RESULTS	131
3.3.1 Analysis of new 19q markers in previously linked families ADRP5, ADRP29 and RP1907	131
3.3.2 Identification of three additional families: ADRP2, ADRP11 and ADRP24 and refinement of the locus.....	136
3.3.3 Founder effect in RP1907 and ADRP24	138
3.3.4 Analysis of sib pairs in RP11 families to see evidence of 'allelic effect'	144
3.3.5 Exclusion of families	147
3.3.5.1 Families linked to other RP loci: US family and ADRP 8.....	147
3.3.5.2 Families with tentative evidence of partial penetrance: ADRP 2228, ADRP 83 and ADRP 2752	147
3.3.5.3 Families excluded from RP11: ADRP33, ADRP710, Bristol family and ADRP 730	149
3.4 DISCUSSION	154
3.4.1 Refinement of the RP11 locus.....	154
3.4.2 Allelic effect and refinement of the RP11 disease penetrance influencing locus	154
3.4.3 Frequency of the RP11 locus.....	155
CHAPTER 4	157
GENOMIC ORGANISATION AND MUTATION ANALYSIS OF A CANDIDATE GENE: PROTEIN KINASE C GAMMA (PRKCG).	
4.1 INTRODUCTION	157
4.2 POSITIONAL CANDIDATES FOR RP11	157
4.3 PROTEIN KINASE C (PKC).....	159
4.3.1 Retinal protein kinase C.....	161

4.3.2	<i>Protein kinase C and Phototransduction</i>	161
4.3.2.1	<i>Invertebrate phototransduction</i>	163
4.3.2.2	<i>Regulation of vertebrate phototransduction and PKC</i>	165
4.3.3	<i>Protein kinase C gamma (PRKCG)</i>	166
4.4	<i>PRKCG: A CANDIDATE FOR RP11?</i>	167
4.5	<i>GENOMIC ORGANISATION OF THE PRKCG GENE</i>	168
4.5.1	<i>Materials and methods</i>	168
4.5.1.1	<i>The cDNA sequence of PRKCG and genomic clones of the PRKCG gene</i>	168
4.5.1.2	<i>Elucidation of PRKCG genomic structure</i>	170
4.5.2	<i>Results</i>	174
4.5.2.1	<i>Identification of splice sites conserved between Drosophila dPKC53E gene and PRKCG</i>	174
4.5.2.2	<i>Identification of PRKCG specific splice sites</i>	179
4.5.2.3	<i>Confirmation of splice junctions using PRKCG genomic clones</i>	181
4.5.3	<i>Discussion</i>	183
4.5.3.1	<i>Genomic organisation of the PRKCG gene</i>	183
4.5.3.2	<i>Comparison of structure between PRKCG and Drosophila dPKC53E genes</i>	189
4.6	<i>MUTATION ANALYSIS OF THE PRKCG GENE</i>	198
4.6.1	<i>Methods and materials</i>	198
4.6.1.1	<i>Primers and mutation detection methods employed</i>	198
4.6.2	<i>Results</i>	199
4.6.2.1	<i>Mutation in the family RP1907 and ADRP24 segregates with disease</i>	199
4.6.2.2	<i>Mutation in the Japanese family</i>	203
4.6.2.3	<i>Polymorphisms in the PRKCG gene</i>	203
4.6.2.4	<i>Screening of the 5' untranslated region of the PRKCG gene</i>	207
4.6.2.5	<i>Investigation of large rearrangements within the gene</i>	208
4.6.2.6	<i>Mutations in Drosophila eye-PKC gene (inaC mutants)</i>	208
4.7	<i>DISCUSSION</i>	209
4.7.1	<i>The efficiency of mutation detection methods employed</i>	209
4.7.2	<i>The codon 647 mutation in exon 18 of PRKCG</i>	210
4.7.3	<i>Is PRKCG the gene for RP11?</i>	210
4.7.4	<i>Allelic effect, modifier genes and PRKCG</i>	212
CHAPTER 5		214
YAC CONTIG OF THE RP11 INTERVAL AND ISOLATION OF NOVEL MICROSATELLITE MARKERS FOR FURTHER LOCUS REFINEMENT.		
5.1	<i>INTRODUCTION</i>	214
5.2	<i>PHYSICAL MAPPING</i>	215
5.2.1	<i>Cloning in YACs</i>	215
5.2.2	<i>STS content mapping and chromosome walking for contig establishment</i>	217

5.2.3	<i>The genomic region of chromosome 19q13.4</i>	218
5.2.4	<i>Physical maps of chromosome 19</i>	223
5.2.5	<i>Aims of study</i>	224
5.3	ISOLATION AND CHARACTERISATION OF YACs IN THE RP11 REGION OF CHROMOSOME 19Q13.4	225
5.3.1	<i>Materials and methods</i>	225
5.3.1.1	<i>YAC libraries</i>	225
5.3.1.2	<i>Isolation of YACs by PCR screening of YAC libraries</i>	225
5.3.1.3	<i>Isolation of YAC insert end sequences for chromosome walking by Alu-Vector arm PCR</i>	227
5.3.2	<i>Results</i>	228
5.3.2.1	<i>Verification of the integrity of mega-YACs obtained from pre existing contigs</i>	228
5.3.2.2	<i>Isolation of novel YACs by PCR screening of ICI and ICRF YAC libraries</i>	232
5.3.2.3	<i>Investigation of chimaerism in critical YACs spanning the RP11 interval by FISH</i>	239
5.3.2.4	<i>Closure of gaps by chromosome walking</i>	239
5.4	ISOLATION, CHARACTERISATION AND GENETIC MAPPING OF MICROSATELLITES FROM COSMIDS IN THE RP11 REGION.....	249
5.4.1	<i>Results</i>	251
5.4.1.1	<i>Obtaining cosmids from the proximal end of RP11</i>	251
5.4.1.2	<i>Screening cosmids for the presence of (CA)_n repeat sequences</i>	252
5.4.1.3	<i>Construction of cosmid mini libraries</i>	254
5.4.1.4	<i>Isolation of (CA)_n positive subclones and sequencing</i>	256
5.4.1.5	<i>Determining the heterozygosity of novel microsatellites</i>	263
5.4.1.6	<i>Verification of genetic location of novel microsatellites by haplotype analysis in the RP11 linked families</i>	266
5.4.1.7	<i>Genetic analysis of novel microsatellites in ADRP2 for locus refinement</i>	267
5.5	INTEGRATION OF NOVEL STSS IN TO THE YAC CONTIG.....	271
5.5.1	<i>STS content mapping with the novel markers</i>	271
5.5.2	<i>Isolation of PAC clones with the new markers to establish a deeper contig across the RP11 interval</i>	273
5.6	GENES AND ESTs LOCATED IN THE RP11 REGION.....	274
5.7	DISCUSSION.....	279
5.7.1	<i>The establishment of a YAC contig in the RP11 and verification of integrity</i>	279
5.7.2	<i>Isolation of novel microsatellites and refinement of the RP11 locus</i>	281
5.7.3	<i>The size of the RP11 interval</i>	282
5.7.4	<i>Use of PACs in preference to YACs</i>	283
5.7.5	<i>Future endeavours</i>	284

CHAPTER 6	287
GENERAL DISCUSSION AND FUTURE PROSPECTS.	
<i>6.1 Overview of the work presented.....</i>	<i>287</i>
<i>6.2 Allelic effect Vs Modifier gene in RP11: implications for future research.....</i>	<i>289</i>
<i>6.3 Progress made in RP genetics and its future.....</i>	<i>291</i>
<i>6.4 Therapeutics in RP research.....</i>	<i>292</i>
REFERENCES	295
REPRINTS OF PUBLISHED WORK	328

List of Figures and Tables

List of figures

Fig. No:		Page No:
1.1	Morphological organisation of the retina.	52
1.2	Diagram of a generalised photoreceptor in the human retina.	54
1.3	Schematic representation of photoreceptors of the human retina.	54
1.4	Vertebrate phototransduction pathway.	59
1.5	Electroretinogram (ERG) of a normal retina.	67
3.1	Fundus image of a normal, and a RP affected individual from a RP11 family.	122
3.2a	A representation of the genetic map from the RP11 region of chromosome 19q13.4.	128
3.2b	A representation of a sub-section of the 1996 CEPH/Genethon genetic map of 19q13.4.	128
3.3a	Haplotype analysis of 12 markers from 19q13.4 in ADRP5.	132
3.3b	Haplotype analysis of 10 markers from 19q13.4 in ADRP29.	134
3.3c	Haplotype analysis of 9 markers from 19q13.4 in RP1907.	135
3.4a	Haplotype analysis of 12 markers from 19q13.4 in ADRP2.	141
3.4b	Haplotype analysis of 10 markers from 19q13.4 in ADRP11.	142
3.4c	Haplotype analysis of 9 markers from 19q13.4 in ADRP24.	143
3.5	Haplotypes of the six RP11 linked families.	143
3.6	Excluded families with haplotypes constructed using markers from RP11 region of 19q13.4	152
4.1	The genetic region of RP11 showing the localisation of candidate genes relative to the RP11 region.	158
4.2	Domain structure of PKC isoenzymes.	160
4.3	Invertebrate phototransduction pathway.	164
4.4	The cDNA sequence and the 5' promoter sequence of <i>PRKCG</i> .	169
4.5	The alignment between amino acid sequences of dPKC, human, mouse, rat and bovine PKC γ , human PKC α and PKC β .	172
4.6	The alignment of the cDNA sequence of PKC α , β and γ for the design of PKC γ specific primers.	175
4.7	Agarose gel photographs showing the PCR products produced with primer pairs designed, according to the dPKC gene structure, to capture putative <i>PRKCG</i> introns.	177

4.8	Schematic diagram of the structure of PKC γ showing the intronic primers designed to verify identified exons.	180
4.9	Amino acid sequence alignment between dPKC and mammalian PKCs showing the conserved splice site junctions and those specific for <i>PRKCG</i> gene.	185
4.10	Comparison of structure between dPKC and <i>PRKCG</i> genes.	188
4.11	The entire sequence of <i>PRKCG</i> gene showing the structural features of this gene as well as all the primers designed for mutation analysis.	191
4.12a	Comparison of sequence data from a normal (a), and a heterozygous mutant (b), <i>PRKCG</i> exon 18 sequence, from the family RP1907.	201
4.12b	Sequence of exon 18 showing the <i>AciI</i> restriction sites eliminated due to mutation at codon 659.	201
4.13	Pedigrees of RP1907 (a) and ADRP24 (b), the segregation of codon 659 mutation is shown by the <i>Aci I</i> restriction assay on exon 18.	202
4.14	MDE gel photographs showing the heteroduplex produced by the codon 647 mutation of exon 18 in the entire Japanese family.	204
4.15	Comparison of sequence data from a normal (a), and a heterozygous mutant (b), <i>PRKCG</i> exon 18 sequence, from the Japanese family.	205
4.16	Examples of polymorphisms identified during the mutation analysis of <i>PRKCG</i> .	206
5.1	A portion of the LLNL integrated physical map of chromosome 19q13.4 showing the RP11 genomic region and the existing YAC coverage.	219
5.2	Comparison of mouse chromosome 7 and human chromosome 19 showing the relative rearrangement that had occurred within syntenic regions.	222
5.3	The STS content map of the CEPH mega-YACs from the RP11 region.	229
5.4	Agarose gel depicting the results of the PCR screening of the ICI YAC library with marker D19S926.	233
5.5	The STS content map of the isolated ICI and ICRF YACs from the RP11 region.	235
5.6	STS content mapping result obtained with <i>PRKCG</i> exon 8 in a selection of YACs from RP 11 region.	236
5.7	The distribution of <i>PRKCG</i> gene in YAC, PAC and cosmid clones from the RP11 region.	237
5.8a	Pulse field gel photograph showing the fractionation of a selection of YACs from RP11 region.	238
5.8b	Autoradiograph of the pulsed field gel after probing with Cot I DNA.	238
5.9	Fluorescent <i>in situ</i> Hybridisation (FISH) of YACs to metaphase chromosomes.	240
5.10	Agarose gel photograph showing the Alu-PCR results obtained with selected YACs.	243

5.11	Sequence generated from the unique products obtained by Alu-PCR (YAC end clones).	244
5.12	Agarose gel depicting the PCR results obtained with the end clone STS 23L.	247
5.13	Mapping of the novel STSs generated from YAC termini in all the YACs located in the RP11 interval showing the gap closure achieved.	248
5.14	Adaptation of the LLNL physical map of the RP11 region showing the cosmid contigs utilised for the isolation of (CA) _n markers.	250
5.15	Autoradiographs of the 11 cosmids belonging to 'contigs' D19S268, D19S781, D19S785, D19S774, D19S767 digested to completion with <i>HaeIII</i> , <i>MspI</i> , <i>Sau3AI</i> and <i>BamHI</i> .	253
5.16	G ₄ (GT) ₁₃ probed replica filters of the cosmid 20019 pBS mini library.	257
5.17	Selected plasmid subclones following restriction enzyme digestion and hybridisation with G ₄ (GT) ₁₃ probe.	258
5.18	Electropherograms, sequence data generated using ABI automated sequencer, depicting the (CA) _n repeats obtained from cosmids 20019 and 30020.	261
5.19	Polymorphic nature of PRKCG and D19S781.2N	264
5.20a	Haplotype analysis with novel markers in ADRP2.	268
5.20b	Haplotype analysis with novel markers in ADRP11.	269
5.20c	Haplotype analysis with novel markers in RP1907.	270
5.20d	Haplotype analysis with novel markers in ADRP24.	270
5.21	STS content mapping results obtained with the novel STS D19S767.2N.	272
5.22	The most up to date YAC (and PAC) contig of the RP11 interval.	277

List of Tables

Table No:		Page No:
1.1	The evolution of genetic maps.	30
1.2	Cloned or mapped genes causing autosomal recessive retinitis pigmentosa.	80
3.1	Two point lod scores between 19q13.4 markers (shown in order from proximal to distal) and disease phenotype in ADRP2 and ADRP11.	140
3.2	Phenotype/Genotype correlation of sib pairs with at least one discordant pair.	146
3.3	Two point lod scores between 19q13.4 markers (shown in order from proximal to distal) and disease phenotype in adRP families described in section 3.2.2.	149
4.1	Occurrences of PKC isoenzymes in retinas of different species as described in Wood <i>et al</i> (1997).	162
4.2	Methods used to elucidate the structure of <i>PRKCG</i> organisation.	182

4.3	Intron-exon boundary sequences of the human <i>PRKCG</i> Gene.	184
4.4	Intron-exon boundary comparison between <i>dPKC</i> and <i>PRKCG</i> genes.	187
4.5	Polymorphisms identified during the mutation analysis of <i>PRKCG</i> gene.	207
5.1	The human YACs libraries available from the HGMP resource centre.	226
5.2	Features of CEPH mega YACs obtained from chromosome 19 maps.	231
5.3	Characteristics of the ICI and ICRF YACs.	241
5.4	Details of YAC end recovery.	246
5.5	Identities of cosmid clones chosen for isolation of (CA) _n markers.	251
5.6	Details of mini library construction of cosmid clones used for isolation of (CA) _n repeats.	255
5.7	Presentation of all polymorphic markers (P) and STSs (S) isolated and characterised from cosmids located in the distal RP11 region.	262
5.8	The characteristics of all polymorphic markers identified from cosmids in the RP11 region chromosome 19q13.4.	265
5.9	Characteristics of the ESTs assayed within the RP11 contig.	276

Abbreviations

adRP	autosomal dominant retinitis pigmentosa
APS	Ammonium Per Sulphate
arRP	autosomal recessive retinitis pigmentosa
bp	base pair
cDNA	Complementary deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra-acetic acid
FISH	Fluorescent <i>in situ</i> hybridisation
kb	kilo base
kd	kilodalton
mRNA	messenger ribonucleic acid
ng	nanogram
PAC	P1 derived artificial chromosomes
PCR	Polymerase chain reaction
PFGE	Pulse field gel electrophoresis
pmole	picomole
ROS	Rod outer segment
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
SDS	Sodium dodecylsulphate
TBE	Tris-Borate-EDTA
TE	Tris-Cl-EDTA
TEMED	N,N,N', N'-tetra methyl ethylene diamine
UV	Ultra violet light
X-GAL	5-Bromo-4-chloro-3-indolyl- β -galactoside
xLRP	X-linked retinitis pigmentosa
YAC	Yeast artificial chromosome

Amino acid codes:

Amino acid	Three letter abbreviation	One-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

CHAPTER 1

GENERAL INTRODUCTION

The work presented in this thesis concerns the positional cloning of a gene responsible for a hereditary retinal degeneration called retinitis pigmentosa (RP). Therefore in the following sections an overview of the genetic concepts behind the project will be described with more emphasis given to strategies employed for identification of disease genes through positional cloning. The reader will also be introduced in to the structure and physiology of the retina, the photochemistry of visualtransduction, retinal disorders and advances made in the field of retinal degeneration research with the advent of recombinant DNA technology.

All positional cloning projects such as the one described here contribute to our understanding of the human genome and biology. Successful identification of a gene predisposing to human illness represents the traversal of a major bottleneck in understanding of that disease and allows the development of diagnostic and therapeutic advances of immense potential medical benefit. At present all positional cloning projects are essentially in symbiosis with the collaborative international effort known as the Human Genome-Mapping Project and have been accelerated due to many technologies that have arisen from this mammoth project. In fact it could be argued that the greatest medical justification of the Human Genome Project has been the success of many positional cloning ventures.

1.1 The Human Genome Project

The Human Genome Project is the first internationally co-ordinated effort in the history of biological research. It aims to determine the complete nucleotide sequence of the nearly 3 billion base pairs of DNA that constitutes the human genome, and in doing so identify the 100 thousand or so genes that define the human genome. The project was initiated in the late 80's and was designed as a three-step program to produce genetic maps, physical maps and finally the complete sequence of the

genome by the year 2005. The first two mile stones have essentially been reached and sequencing projects have now begun (Lander, 1996).

In the process of achieving these goals the human genome project has been responsible for the development of many molecular biology techniques and reagents. Polymorphic DNA markers, dense genetic maps, variety of cloning vectors and physical maps that were developed as prerequisite to obtaining the entire human genomic sequence have been essential for many positional cloning projects and also constitute their driving forces. Moreover this information is all the more beneficial for being easily accessible via the Internet. With the advent of the genome project numerous electronic databases have been established to record the genetic and physical mapping information that is being generated. The central data resource for the human gene mapping effort is the Genome Data Base (GDB), which collects, organises, stores and distributes human genome mapping information and also offer interaction with other data bases maintained by major laboratories i.e. Genethon/CEPH and Co-operative Human Linkage Centre (CHLC).

Considering the fact that work presented in this thesis constitutes a positional cloning project, it is only befitting to analyse the progress of the Genome Project. Therefore in the following sections the concept of linkage analysis, the use of polymorphic markers to localise disease genes, the evolution of polymorphic DNA markers, genetic maps and physical maps will be described. This will also provide an insight in to the rapid progress made in the field of human molecular genetics in the last decade.

1.2 Mapping genes for single gene disorders by linkage analysis

Classical linkage analysis:

In classical linkage analysis the arrangement of genes on the chromosome of an organism was determined by following the segregation pattern of two variable phenotypic traits through successive generations and seeing how often the different forms of the two traits are co-inherited. From this we can infer whether the genes for the two traits are on the same chromosome (linked) or located on separate chromosomes entirely. For the latter case Mendels law of independent assortment applies, resulting in a 50% recombination rate of phenotypic traits since the

chromosomes on which the genes reside are inherited independently. Linked pairs of genes or loci are co-inherited and there is a deviation from the independent assortment of traits, and any recombinants observed are due to events called cross overs. These events occur during meiosis when the homologous chromosomes align allowing the exchange of corresponding segments of DNA between non-sister chromatids. This process leads to formation of gametes that possess chromosomes containing new combinations of alleles, or recombinant chromosomes. Any progeny resulting from these gametes have the recombinant phenotypic traits.

Recombination fraction (θ), genetic distance and order:

Cross over events are more likely to occur between loci that are far apart than between two loci close together on a particular chromosome. Therefore the recombination fraction (θ), which is the probability of producing a recombinant, is related to the genetic distance separating the loci of the linked genes. These distances are measured in map units referred to as 'centi-Morgans' (cM), where 1 cM is equal to a 1% chance of recombination between two loci during a meiotic event. This simple relationship only applies when the genetic interval is too small to allow multiple cross over events, which in reality do occur between loci that are further apart. Therefore mapping functions that relate recombination fraction with genetic distance have been derived to take in to account such multiple cross over events and also the phenomenon of *interference*. Interference is the negative effect an already existing crossover has on the formation of another in its vicinity. The Kosambi mapping function (1944), which takes interference and multiple cross over events in to account, has been widely used in human genetic calculations since it produces more realistic map distance values.

Recombination fractions derived for linked pairs of loci can also be used to produce linkage groups. A linkage group is a set of gene pairs, each of which has been linked to at least one other member in the set. Since recombination fraction increases with the distance separating the loci of the two gene pairs, it can be used to derive the order of linked loci and produce a linkage map for a given chromosome.

Polymorphic DNA marker used as a tool for linkage analysis:

Today linkage analysis is employed predominantly for the mapping of inherited disease genes, which are prominent among the variable single-gene traits of humans. It differs from classical linkage analysis in that at the end linked loci are also given a “genomic address” or an actual physical location in the genome, as well a linkage order. In order to do so, linkage analysis employs the use of a large number of polymorphic DNA markers of known genomic localisation, in a random search throughout the genome in a particular family within which the disease segregates. Polymorphic markers are variable regions in the DNA that exhibit a limited number of sequence variations (or alleles) among the population. The purpose of such a genome scan is to detect co-segregation between the disease locus, represented by the disease phenotype, and an allele of a DNA marker. Establishing such segregation of a known marker locus with the disease phenotype will infer linkage between the two loci and therefore localise the disease to the region where the marker maps. Once linkage has been established, more DNA markers that map at close proximity to the disease locus are typed in the family to establish a haplotype, which is the combination of alleles at the linked locus. It is easy to identify a haplotype for the disease phenotype, as they should co-segregate. Moreover any recombination events that occur during meiosis in the affected parent would be reflected in the haplotype of the resulting offspring. This would lead to refinement of the disease locus provided that the offspring still has the disease phenotype.

Statistical evaluation of linkage:

The conventional approach to evaluate whether two given loci are linked is to apply the maximum likelihood analysis (Morton, 1955). This estimates the “most likely” value of the recombination fraction (θ) as well as the odds in favour of linkage versus non-linkage. The ratio of likelihood or the *odds ratio* is the ratio of the probability of observing a segregation pattern of two loci if they are linked at a given recombination value (θ) where $\theta < 0.5$, versus the probability of such segregation if they are not linked ($\theta = 0.5$). This ratio is calculated at a series of recombination values (from $\theta = 0$ to $\theta = 0.5$). The logarithm of the odds ratio called the lod score (Z) is what geneticists normally use to report results of linkage analysis. These calculations can be summarised in the following equation

$$Z(\theta) = \text{Log}_{10}[L(\theta)/L(0.5)]$$

Where:

$L(\theta)$ = the likelihood of obtaining the data if the two loci are linked and have a recombination fraction of θ

$L(0.5)$ = the likelihood of obtaining data when the two loci are unlinked

The recombination value that gives the maximum lod score (Z) is the best estimate of the degree of linkage between two loci. Positive lod scores suggest the presence of linkage and a lod score of 3, which corresponds roughly to 1000 to 1 odds that two loci are linked, or higher is considered definitive evidence for linkage. Similarly a negative lod score of -2 at a given θ value is regarded as evidence against linkage within an interval equal to θ from either side of the marker locus.

1.2.1 Markers used in linkage analysis

The merits of polymorphic markers and their usefulness for mapping genes were discussed in the previous sections. To conduct a cost effective linkage analysis it is essential to have genetic maps, which are dense in informative DNA markers. The nature of naturally occurring DNA polymorphism used in linkage analysis has changed dramatically over the years. Before the advent of recombinant DNA technology, linkage studies were performed using blood group and other biochemical (protein) polymorphisms.

1.2.1.1 RFLPs and mini satellites

Recombinant DNA technology enabled the definition of a variety of scorable DNA polymorphisms. The first such DNA polymorphisms to be detected were differences in the length of DNA fragments after digestion with sequence specific restriction endonucleases, the so called restriction fragment length polymorphisms (RFLPs) (Botstein *et al.*, 1980). RFLPs were based on a variety of polymorphisms at the sequence level such as single nucleotide changes, insertions and deletions. Then came the variable number of tandem repeat polymorphisms (VNTRs), also known as minisatellites, which are due to variation in number of head to tail repeats, usually $>10\text{bp}$, at a given locus (Jeffreys *et al.*, 1985). RFLPs and VNTRs were both assayed

by southern blotting (Southern *et al.*, 1975) followed by hybridisation with the specific DNA probe.

The most useful markers in linkage analysis are those that are very polymorphic with a high heterozygosity value. Heterozygosity of a marker is directly proportional to the number of alleles each marker exhibits and the frequency of each allele in the general population (see section 1.2.1.2). In this the multi-allelic VNTRs have the advantage over RFLPs, which are usually bi-allelic. But VNTRs do not have as even a distribution in the genome as RFLPs, as they are more frequent at telomeric ends of chromosomes. The first RFLP map of the human genome produced by Donis-Keller *et al.* (1987) was a landmark publication. It characterised 393 RFLP loci with an average spacing of 10 cM. Given that the human genome is 4000 cM in length, the distance between each RFLP was 10Mb on average, which was too great to be of use for gene isolation. Nevertheless, these markers made human molecular genetics a reality and led to the mapping of number of important Mendelian diseases. To overcome the limitations posed by RFLPs new generation of markers, with high incidence and greater informativeness were identified. These markers known as microsatellites not only fulfil both criteria required of an ideal DNA marker for linkage analysis, they could also be assayed by PCR thus obviating the need for laborious and time consuming blotting techniques.

1.2.1.2 Microsatellite markers [Short Tandem Repeat polymorphisms (STRP)]

Microsatellites are tandem repeats of simple sequence that occur abundantly and at random throughout most eukaryotic genomes with the exception of yeast (Weber and May, 1989; Litt and Luty, 1989). In human 76% of repeats are of the types A, AC, AAAN, AAN or AG, in decreasing order of abundance (Beckmann and Weber, 1992). Of these AC is the most commonly used type of microsatellite and they occur every 30 kb. The CA repeats are equally distributed in the 5'- and 3'-untranslated regions and introns. The informativeness of a microsatellite marker is indicated by its polymorphic information content (PIC). The PIC value of a marker is calculated from the number of alleles and their frequencies in the population and is related to the mean repeat length of the marker (Weber, 1990). To be of much use in linkage analysis the PIC of a microsatellite need to be >0.7 . The equation to calculate the PIC value is given below, where p_i and p_j are population frequencies of the i^{th} and the j^{th} alleles.

$$\text{PIC} = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 p_i^2 p_j^2$$

$$\text{Heterozygosity} = 1 - \left(\sum_{i=1}^n p_i^2 \right)$$

Their abundance and ease of typing by PCR made microsatellites the ideal genetic markers for linkage analysis. Moreover, the use of fluorescent labelling technology enabled the typing of several markers in a single run (multiplex PCR) leading to lower costs and a greater degree of automation of marker typing (Ziegle *et al.*, 1992). Several genetic maps composed of microsatellites have been published, and the publication of the final Genethon map in 1996 marked the end of genetic mapping phase of the Genome mapping Project (Dib *et al.*, 1996) (see section 1.2.1.4 for more detail of these maps).

Even though microsatellites are seen as the best DNA markers, the need to run electrophoretic gels to separate microsatellite marker products by size makes it hard to fully automate the genotyping process. Improvements can still be made in mapping technology and the answer may be provided by the use of single nucleotide polymorphisms, which could be scored by DNA chip technology, which may enable complete automation of linkage analysis.

1.2.1.3 Single nucleotide polymorphisms (Biallelic markers)

Recent attention has focused on the use of single nucleotide polymorphisms (SNPs) as genetic markers. As the name implies the polymorphism is based on a single nucleotide change at the sequence level. The SNPs only have two alleles, hence the term biallelic, therefore genotyping them only requires a plus/minus assay. Biallelic markers vary in their polymorphism rates, the common allele could range in frequency from 50% to nearly 100% (Kruglyak, 1997). At first glance it appears to be a step back in to the days of low polymorphism rates characteristic of RFLPs. However modern technology is said to allow efficient assays of SNPs in sufficient

numbers to overcome the low polymorphism rates (Kruglyak, 1997). The main advantage offered by SNPs over microsatellites is their great abundance, which is estimated to be 1 every kilobase pair, and also the fact that the assay technique is more amenable for automation. To overcome the lower polymorphism rates of SNPs, maps based on SNPs need to be denser than maps of microsatellite markers (Kruglyak 1997). In general it is said that maps of biallelic markers need to be about 2.25-2.5 times the density of microsatellites to provide a comparable information content.

Even though facts as yet do not argue in favour of SNPs, with improvements in technology, screening of denser biallelics may prove to be more cost effective than the use of microsatellites. Such dense maps would be useful for homozygosity mapping (see section 1.3) and would also enable genome scans for linkage disequilibrium (LD) and association required for the analysis of complex genetic disorders. A more promising outlook for biallelics is the fact that functional biallelic polymorphisms, found within coding regions of genes, can be included for direct association studies for both monogenic and polygenic disorders.

1.2.1.4 Evolution of genetic maps and future prospects

Genetic markers and maps of human chromosomes are essential for localising Mendelian traits and disease-susceptibility genes to precise chromosomal regions. Since 1987 several genetic maps have been published. The “genetic composition” of these maps has changed with the evolution of polymorphic markers. The early maps had uneven coverage, with larger gaps especially at the telomeric and centromeric regions of most chromosomes. However with the development of more markers, the marker density gradually increased to meet the final goal set by the Human Genome Mapping Project, which was to have a genetic map of 1cM density. Table 1.1 summarises the evolution of these genome maps.

Early maps were composed of RFLP or VNTR polymorphisms detected by DNA hybridisation (Donis-Keller *et al.*, 1987). Some maps integrated RFLP and VNTR markers with microsatellites (STRPs) (NIH/CEPH Collaborative Mapping Group 1992). The most recent publications of genetic maps of the human genome have been composed entirely of microsatellites or STRPs. The three genome maps published by Genethon stand as the best example (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994;

Dib *et al.*, 1996). Apart from this, other groups have also have created genome maps from independent collection of markers (The Utah Marker Development Group, 1994).

The lack of integration was the major draw back with maps created independently. In principle marker data from individual laboratories can be used to integrate genetic maps, provided that they have been genotyped using a common set of reference families. The most widely used set of reference families is that distributed through CEPH (Centre d'Etude du Polymorphisme Humaine). In the beginning two integrated maps were published by NIH/CEPH collaborative mapping group (1992) and Cooperative Human linkage Centre (CHLC) using the CEPH reference panel (Buetow *et al.*, 1994). However these maps became outdated as soon as they appeared in print due to the rapid pace of new marker development within the human genetic community. Integration of markers has now become easier with the increased physical map information for these markers from independent sets. Since markers from different sources are being extensively integrated in to a physical map by means of radiation hybrid panels and overlapping YAC clones (see sections 1.4.2.3 and 1.4.2.4), more integrated maps are now available for each chromosome through electronic sources such as the World Wide Web (WWW) (see section 1.1). CEPH/Genethon (<http://www.genethon.fr/genethon-cu.html>) Whitehead Institute for Biomedical Research/MIT Centre for Genome Research (<http://www-genome.wi.mit.edu/>), Sanger Centre (<http://webace.sanger.ac.uk/HGP/>) and Stanford Human Genome Centre (SHGC) (<http://shgc-www.stanford.edu>) can be sited as being some of the main Genome Centres, which offer collated mapping information most essential to gene hunters.

Year	Marker type	Average resolution (cM) and Total map distance	Total no Of markers	Mapping group	Reference
1987	RFLP+VNTR	10 cM	403 (393RFLP)	Collaborative Research	Donis-Keller <i>et al.</i> , 1987
1992	(CA/TG) _n	4.4 cM 3576 cM	814	Genethon	Weissenbach <i>et al.</i> , 1992
1992	RFLP/(CA) _n	<5 cM	1416 (279 genes)	NIH/CEPH	NIH/CEPH Collaborative Mapping Group, 1992
1994	(CA/TG) _n	4.9 cM 4798.3 cM	1123	CHLC	Buetow <i>et al.</i> , 1994
1994	(CA/TG) _n	2.9 cM 3690cM	2066	Genethon	Gyapay <i>et al.</i> , 1994
1996	(CA/TG) _n	1.6 cM 3699 cM	5264	Genethon	Dib <i>et al.</i> , 1996

Table 1.1

The evolution of genetic maps.

The evolution of these dense and highly informative genetic maps has been most beneficial for mapping of susceptibility loci involved in complex traits such as juvenile onset diabetes, schizophrenia and hypertension. They also provide a key tool for positional cloning of disease genes by providing ready access to all chromosomal regions, more over such a map is a necessity in order to obtain the complete genomic sequence. Even though the genome-mapping phase is essentially complete, marked by the publication of the last Genethon map. With the development of SNPs, the genetic mapping arena still appears to be dynamic. So the next question is whether a high-density map based on SNPs will replace the current genetic map?

1.3 Mapping of genes for recessive disorders

1.3.1 homozygosity mapping

Mapping of autosomal dominant disorders through traditional linkage analysis has been discussed previously. For mapping of recessive disorders a different approach is used, which obviates the use of traditional family mapping, referred to as homozygosity mapping (Lander and Botstein, 1987). It requires polymorphic markers of high heterozygosity and DNA of affected children from consanguineous marriages. The method essentially involves the detection of the disease locus by virtue of the fact that the adjacent region will preferentially be homozygous by descent in such inbred children. The main advantage of homozygosity mapping is that it provides a way to

map a recessive disease even when families with multiple affecteds are scarce. When affected individuals from different families are used in the analysis, the disease is assumed to be homogeneous i.e. caused by mutation at a single locus.

Given the fact dense maps of highly polymorphic markers are now available they facilitate the detection of homozygosity by descent, by detecting regions in which a contiguous stretch of markers are homozygous. Specifically homozygosity mapping would be performed as follows. To test the hypothesis that the disease gene maps to a given interval, the DNA of each affected inbred child is examined for homozygosity at say six consecutive loci, three on either side of the interval. Then for each possible outcome, the probability of it occurring due to linkage (P_1) and chance (P_2) is calculated. In which case the odds ratio in favour of linkage is $P_1 : P_2$. Computer programs called HOMMAP has been develop to do these calculations (Lander and Green, 1987)

1.4 From linked marker to the gene (physical mapping)

In the previous sections, various genetic mapping techniques that are employed to localise genes responsible for monogenic disorders to specific chromosomal regions were described. But the path from the closely linked genetic marker to the gene is arduous, and requires laborious physically mapping techniques. A physical map is a representation of the locations of identifiable landmarks on DNA. Physical mapping technologies can be categorised according to the level of resolution achieved with each technique, the complete sequence of the genome being the physical map of the highest possible resolution or greatest molecular detail. The physical mapping phase of the Human Genome Project has been responsible for the development and improvement of many of these techniques, therefore this section will be dedicated to the detailed description of a variety of these methods giving close attention to those techniques of more recent evolution.

1.4.1 Low resolution physical mapping

The traditional methods of physical mapping to complement genetic localisation have been cytogenetic techniques, which include the use of somatic cell hybrids (rodent-human hybrid cells) and *in situ* hybridisation. Although considered traditional both techniques have been improved considerably.

1.4.1.1 Cytogenetic techniques [rodent-hybrid cell lines and (FISH)]

Somatic cell hybrids are formed by fusing the cells of the two different species and applying conditions that select against the two donor cells. If hybrid cells are grown under non-selective conditions after the initial selection, chromosomes from one of the parents tend to be lost more or less at random. In the case of human/rodent fusions, which are the most common, the human chromosomes are preferentially lost. In this manner rodent cell lines containing one or a few human chromosomes have been constructed. Such rodent-human hybrid cell lines that contain individual human chromosomes in whole or part can localise a marker of interest to the level of a cytogenetic band on a chromosome and provide mapping at a resolution of about 10 Mb. Individual chromosomes can also be tagged with selectable genes (i.e. thymidine kinase gene that allow growth in HAT medium) (Tunnacliffe *et al.*, 1983; Saxon *et al.*, 1985), fractionated by separation into microcells, and fused with rodent cell lines. Selection allows for the recovery of hybrid cell lines with a choice of intact chromosome (Warburton *et al.*, 1990). These chromosomes maintain their integrity and can be propagated indefinitely and can also be subfractionated by the introduction of interstitial or terminal deletions by radiation (Dowdy *et al.*, 1990). Development of chromosome or chromosomal region-specific libraries and *in situ* hybridisation analysis are some of the uses of hybrid cell lines. Further extensions of hybrid cells are the radiation hybrids used to generate genetic/physical maps at increasing levels of resolution (see section 1.4.2).

In situ hybridisation allows comparison to be made between high-resolution physical map such as contigs assembled from cloned DNA and the genetic linkage map. Provided that hybridisation to repetitive sequences is suppressed and that no DNA chimaeras are present, a cloned fragment or restriction fragment should anneal to a single location on the cytogenetic map. Furthermore, the physical map order should match that found by *in situ* hybridisation. Where genetic markers have been

located on the cytogenetic map by *in situ* hybridisation they also can be positioned on the physical map. Conventional fluorescent *in situ* hybridisation (FISH) as applied to metaphase chromosomes has a resolving power of 1~ Mb. For higher resolution mapping, FISH has been applied to interphase cell nuclei where DNA is less condensed than in metaphase chromosomes, which permits resolution in the 50-100 kb range (Trask *et al.*, 1989,1991). Simultaneous *in situ* hybridisation of cosmid clones labelled with different coloured fluorescent dyes on interphase nuclei have been used to order cosmid clones and have been the basis of the metric map of chromosome 19 physical map developed by the Lawrence Livermore National Laboratories (LLNL) (Ashworth *et al.*, 1995). A new technique has been developed that allows even greater resolution (Parra and Windle, 1993) combination of this procedure with differential fluorescent labelling of individual probes enables the arrangement of DNA probes along an extended strand of DNA to be observed. The images observed are recorded through a fluorescent microscope and used to construct a direct visual hybridisation (DIRVISH) DNA map. DIRVISH mapping has been used to determine orientation and distance between two different cosmid clones and have facilitated chromosome walking.

1.4.1.2 Pulse-field Gel Electrophoretic mapping

Physical maps have been constructed in the mega base range with the use of pulsed-field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984), which allows the separation of DNA molecules as large as 10 Mb, and rare cutting restriction enzymes. Markers which hybridise to a common restriction fragment of DNA separated on a PFGE gel are assumed to be linked and a restriction map of a region can be constructed in this manner provided that the region of interest has an adequate supply of markers and the cutting sites of the enzyme are conveniently spaced. Such long range restriction maps have successfully been constructed of the 12 Mb CF region (Fulton *et al.*, 1989) and the Duchenne muscular dystrophy locus (Burmeister *et al.*, 1988). The complications associated with this technique are comigration of fragments and the partial methylation of sites for many methylation-sensitive restriction enzymes in uncloned human DNA. However the major limitation of this method is that it does not provide the DNA in the form that facilitates further study and this technique is now mostly used for further analysis of 'contigs' assembled from cloned DNA.

1.4.2 High resolution physical mapping

Physical mapping at high resolution involves the assembling of cloned DNA fragments in the same linear order as found in the corresponding chromosomal DNA and such an assembly is described as a 'contig'. In the human genome-mapping endeavour the ultimate purpose of this process is to sequence the overlapping fragments of DNA and obtain the physical map at its highest possible resolution i.e. the DNA sequence of the entire genome. However in a positional cloning endeavour the purpose of contig construction across a candidate interval is to have the genomic segment that harbours the disease gene in an accessible form for subsequent use in gene identification protocols (see section 1.5). There are several aspects to physical mapping, which include the choice of vector for cloning, the methods for assembling clones and detecting overlaps between the assembled DNA fragments.

1.4.2.1 Choice of vector

Many different types of vector have been developed to propagate DNA, to enable it to be mapped and ultimately sequenced. Yet only a few are suitable for large scale mapping projects. The early human genomic libraries were constructed in bacteriophage λ and cosmid vectors both hosted in *E. coli* with respective insert capacities of 25 and 50 kb (Murray, 1986; Whittaker *et al.*, 1988). While the cloning efficiency of these systems was high they were also associated with several disadvantages. These included the requirement of a great number of clones to cover a given human region and unclonability of certain human sequences such as repetitive sequences. The advent of the second high molecular weight DNA cloning system, Yeast Artificial Chromosomes (YACs), greatly expanded the ability to clone and characterise large segments of DNA since fragments as large as 1 Mb in size could be isolated (Burke *et al.*, 1987; Anand *et al.*, 1990). The advantages and limitations associated with YACs are described in greater detail in section 5.2.1. Despite the many advantages of YACs, low cloning efficiencies, difficulties in isolating cloned DNA and chimaeric cloning artefacts have led to the development of other large-fragment cloning systems to supplant YACs.

Alternative vectors to YACs

The bacteriophage P1 cloning system: The P1 cloning system complements cosmid and YAC cloning systems by offering an intermediate range of clonable genomic fragment sizes along with the ability to isolate from *E. coli* large amounts of plasmid DNA by standard molecular techniques. Furthermore unlike YAC clones P1 clones are also less prone to chimaerisms (Sternberg, 1994). The P1 vector system, based on the P1 bacteriophage that mediates generalised transduction, contains a packaging site (*pac*) necessary for in vitro packaging of recombinant molecules into phage particles and two *loxP* sites. These sites are recognised by the phage recombinase, the product of the host *cre* gene, and lead to the circularisation of the packaged DNA after it has been injected into an *E. coli* host expressing the recombinase. The clones are maintained in *E. coli* as low copy number plasmids by selection for a vector kanamycin-resistance marker. However, high copy number can be induced by exploitation of the P1 lytic replicon (Sternberg, 1990). The original cloning vector *pAd10* employs a unique *Bam*HI site for the insertion of genomic fragments as large as 100 kb. The second-generation vector *pAd10-sacBII* also contains rare cutting restriction sites and T7 and SP6 phage promoters that directly border the *Bam*HI cloning site. These elements are useful for the isolation, characterisation and analysis of plasmid DNA from P1 clones (Sternberg, 1990; Pierce *et al.*, 1992). The limitation of co-cloning events (found in YACs) due to the ability of P1 phages to package DNA by a headful mechanism and the recovery of cloned DNA in a pure form are other attractive features of P1 vector system. However since P1 phage amplifies in bacteria, as with cosmids, certain regions may also be difficult to clone in P1 vector.

Bacterial Artificial Chromosomes (BACs): The BAC vector (Shizuya *et al.*, 1992) is based on the single-copy sex factor F of *E. coli* and includes the λ *cosN* and P1 *loxP* sites, two cloning sites (*Hind* III and *Bam* HI) and several restriction enzyme sites for potential excision of the inserts. Similar to P1 clones the cloning site is also flanked by T7 and SP6 promoters for generating RNA probes. BACs can be transformed into *E. coli* very efficiently without the use of packaging extracts required for the P1 system. Moreover reports claim that BACs are capable of maintaining human and plant genomic fragments of greater than 300 kb for over 100 generations with a high degree of stability (Woo *et al.*, 1994). However currently

there are two disadvantages associated with BACs. First, as yet there is no method for positively selecting clones with foreign DNA inserts and secondly it is difficult to isolate large amounts of DNA.

P1- derived artificial chromosomes (PACs): PACs developed by combining features of both P1 and F-factor systems (Ioannou *et al.*, 1994) can handle inserts in the 100-300 kb range. As yet no chimaeras or clone instability have been reported for this vector. The advantages with the use of this vector are ease of manipulation and isolation of insert DNA in moderate quantities.

1.4.2.2 Contig assembly

There are several methods for assembling contigs, which include more traditional methods such as chromosome walking and restriction enzyme fingerprinting and more popular methods such as marker sequence and hybridisation assays.

Chromosome walking: This was originally developed for the isolation of gene sequences whose function is unknown but whose genetic location was known, i.e. in positional cloning ventures (Bender *et al.*, 1983). The principle of this method is as follows. For the purpose of map generation a single cloned fragment is selected and used as a probe to detect other clones in the library with which this clone will hybridise and therefore represent clones overlapping with it. The overlap could be in both direction and this single walking step is repeated many times with the newly identified clones to extend the contig. A potential problem with chromosome walking is created by the existence of repeated sequences, which will lead to the isolation of non-contiguous fragments with a clone that contain such sequences. Hence probes used for stepping from one genomic clone to the next must be a unique sequence clone. Furthermore use of large fragments of DNA for walking also minimises the number of steps required to achieve coverage of a given region. Although inherently attractive, chromosome walking is too laborious and time consuming to be of value for mapping whole genomes. Therefore other methodologies have been developed as alternatives to chromosome walking.

Restriction enzyme fingerprinting: The principle of restriction enzyme fingerprinting was originally developed for the nematode *Caenorhabditis elegans* (Coulson *et al.*, 1986) and yeast (Olson *et al.*, 1986) and has been successfully used for the mapping of the *E. coli* genome (Kohara *et al.*, 1987). The original fingerprinting method devised by Coulson *et al.* (1986) is as follows. Cloned DNA is first digested with a 6 base pair cutter that leaves staggered ends (e.g. *Hind*III) and labelled by end-filling. After deactivation of the original enzyme the fragments are re-cleaved with a restriction enzyme with tetranucleotide recognition sequence (e.g. *Sau*3A), followed by separation on a high-resolution gel and detection by autoradiography. The fingerprint is prepared by determining the size of each band from each clone. In the fingerprinting technique used for mapping *E. coli*, each clone within the genomic library of *E. coli* was restriction mapped using eight different restriction enzymes and the resulting restriction fingerprint of each clone was compared with that of all other clones. Overlapping clones were detected on the basis of a match of at least five consecutive cleavage sites. In the two previous methods two clones had to overlap at least 50% in order to declare with a high degree of certainty that the clones do indeed overlap and was a significant rate limiting factor in map completion. It was soon realised that increasing the information content in each clone fingerprint would enable the detection of small overlaps and therefore increase the rate at which the contigs increase in length. As a result in the whole human genome approach of Bellanné-Chantelot *et al.* (1992) used a different method for assembling contigs. To enable small overlaps to be detected YAC clones were digested by *Pvu*II separated by gel electrophoresis and a fingerprint was prepared by hybridisation with probes prepared from a LINE-1 repeated sequence and an *Alu* consensus sequence. Overlaps between YACs are identified by observation of fragments of the same size in the two YACs.

However there are a number of disadvantages with the fingerprinting approach. It is highly labour intensive and involves extensive handling of clones. Although the method generates large number of small contigs it is increasingly difficult to extend and join these contigs. Finally the method only works well with phage clones but gives poor results with YACs and is susceptible to chimaera problems. Therefore restriction fingerprint mapping was superseded by sequence tagged sites (STSs) mapping.

STS content mapping: The concept of sequence tagged site was first developed by Olson *et al.* (1989) in an attempt to systematise landmarking of the genome. An STS is a short region of DNA about 200-300 bases long whose sequence is unique to the target region. The idea behind STS content mapping is simple, if two or more clones contain the same STS then they must overlap and the overlap must include the STS. A major advantage of this strategy is that it provides an ordered set of markers that can be stored as sequence information in a data base allowing reconstitution of contigs from different clone libraries. Operationally, an STS is defined by the sequence of the two primers that make its production possible and can be created from any unique sequence from the desired region. Polymorphic DNA markers with their specific amplimers (primers) are also classified as STSs, but with dual use since they are capable of acting as landmarks both in a genetic and a physical map. Expressed sequence tags (ESTs) which are partial cDNA sequences can serve the same purpose as the random genomic STSs but have the added advantage of pointing directly to an expressed gene. A physical map can easily be established if the region contains a high density of markers or STSs. Initially clones that span the region are isolated through screening of a genomic YAC, PAC or a BAC library by PCR using the primers for these STSs, then by the STS content of each isolated YAC (PAC or a BAC) overlaps between clones can also be detected. The use of STSs to generate physical maps has been authenticated by a number of workers. Green and Olson (1990) used 16 STS to isolate and overlap 30 YACs spanning 1.5 Mb of the region containing the cystic fibrosis trans-membrane regulator (*CFTR*) gene on chromosome 7, Chumakov *et al.* (1992) also constructed a continuous array of overlapping YAC clones that covered the entire chromosome 21q using the STS strategy.

Alu PCR: This is another rapid method of detecting overlaps between YAC clones (Nelson *et al.*, 1989). Alu PCR utilises consensus primers derived from Alu repeat sequences and can be applied directly to YAC colonies. This generates an average of 10 bands between 0.1 and 2.0 kb per YAC. This method has been used to identify overlap between YACs in intron 1 of the dystrophin gene, where no STSs were available (Coffey *et al.*, 1992), and in a modified strategy to confirm overlap between two YAC clones in the class II region of MHC (Ragoussis *et al.*, 1991).

Physical maps that are composed entirely of overlapping YAC clones still need further characterisation to enable DNA sequencing or if one is to have the DNA in form that will enable easy purification essential for many a gene identification protocols. The required in-depth characterisation can be achieved through methods such as genome sequence sampling (GSS) and hybridisation mapping, which although highly laborious result in the production of a 'sequence ready' physical map.

Genome sequence sampling (GSS): This method combines elements of restriction mapping and STS mapping and generates maps with resolution of 1-5 kb (Smith *et al.*, 1994). GSS also requires a chromosome specific cosmid library that represents the genome at 20-30-fold redundancy with reasonably random distribution of clone ends, which is produced by cloning using a variety of restriction enzymes and cloning sites. Initially a YAC clone is used as a hybridisation probe to select cosmid clones included in the YAC clone, which are then restriction mapped and arranged into contigs using methods previously described. STSs are generated and sequenced from the ends of cloned insert of each cosmid and aligned on the map. If a high density of cosmid map has been prepared then it also enables the preparation of high-density sequence map. Provided that the restriction sites used for cloning were evenly spaced, the DNA sequences determined would be spaced every kilobase on average. Using GSS Smith *et al.* (1994) studied the protozoan parasite *Giardia lamblia* with a genome size of 10.5 Mb.

Hybridisation mapping: This method is similar to chromosome walking but incorporates necessary steps to avoid the isolation of non-contiguous clones from the clone used as probe. This is achieved by first blocking the genomic library with five kinds of probe, representing known repetitive sequences (centromeric, telomeric, 17S and 5S ribosomal and the long terminal repeat (LTR) of retrotransposons) to identify those clones that only contain unique sequences. A number of these clones are selected at random and used as hybridisation probes to detect overlapping clones. From these clones which do not give a positive hybridisation signal another set is selected at random for use as probes in the next round of experiments. This process is continued until all clones show positive hybridisation at least once and in practice some clones containing repetitive DNA have to be used to join contigs. However a key feature of this technique is that clones are randomly chosen as probes based on

the criterion that they have not yet given a positive hybridisation signal, which avoids a great number of redundant hybridisations. Any overlaps due to cross hybridisation between repetitive elements are avoided by demanding that all pairs of cosmid overlaps be reciprocal if both cosmids in the pair are used as probes. That is, if cosmid x hybridises to cosmid y then y must hybridise to x . The utility of hybridisation mapping has been shown by the construction of a map of the long arm of human chromosome 11 (Evans and Lewis, 1989) and a complete map of the fission yeast (*Schizosaccharomyces pombe*) (Maier *et al.*, 1992; Hoheisel *et al.*, 1993; Mizukami *et al.*, 1993).

1.4.2.3 Radiation hybrid mapping

Another powerful method for creating physical maps is the use of irradiation and fusion to gene transfer (IFGT) to produce radiation hybrid maps. This technology originally developed by Goss and Harris, (1975) demonstrated that chromosome fragments, generated by lethal irradiation of donor human cells could be rescued by fusion to rodent recipient. Radiation maps are based on breaks induced by radiation and the resolution of radiation hybrid mapping is a function of both fragment size and retention frequencies, which is the percentage of radiation hybrid cell lines containing a given chromosome marker. The fragment size can be varied by altering the radiation dose and it is possible to construct panels designed either for map continuity with few markers or for high resolution with large number of markers. Information on localisation using radiation hybrids is obtained by determining linkage to markers of the framework. Such a linkage is expressed by a distance measured in breakage frequency accompanied by a likelihood estimate expressed as a lod score. Distances between markers in the radiation hybrid maps are expressed in cR_{3000} where 1 $cR_{(N \text{ rad})}$ correspond to a 1% frequency of breakage between two markers after exposure to “N” rad of X-rays (Cox *et al.*, 1990; Boehnke *et al.*, 1991).

Cox and co-workers (1990) modified the original approach by using somatic cell hybrid containing a single human chromosome as the donor cell instead of a diploid human cell. This approach is not feasible for mapping an entire genome, as it would require over 4000 hybrids. Walter and co-workers (1994) have found a solution to this problem by reverting to the original protocol of Goss and Harris (1975) and using a diploid human fibroblast as a donor. Using 44 radiation hybrids they constructed a

map of the human chromosome 14 containing 400 ordered markers and concluded that a high resolution map of the whole genome is feasible with only a single panel of 100-200 radiation hybrids.

Radiation hybrid mapping complements both recombination maps and physical maps based on contigs. Whereas recombination maps are constrained by the recombination rate and only limited to polymorphic markers, radiation hybrid methods can be used to map both polymorphic markers and non-polymorphic STSs and ESTs. Therefore a radiation hybrid map of the human genome was developed to facilitate the completion of the YAC contig of the human genome as it presented an independent means of integrating STS with the polymorphic markers (Gyapay *et al.*, 1996). The limitations of YAC contig based maps, for example poor YAC coverage in some regions of the genome such as terminal parts of chromosomes and particularly GC-rich regions, which are more likely to be rich in genes, further emphasised the need for an independent method of mapping. The radiation hybrid map, which consisted of 168 whole-genome radiation hybrids, was constructed by irradiating donor human fibroblasts at 3000 rads, a relatively low dose to ensure continuity of the map, and fusing with recipient hamster cells. A framework map of 404 markers was constructed by typing the panel with polymorphic markers and ordering them using standard methods (Cox *et al.*, 1990; Boehnke *et al.*, 1991). Even though the order of these markers in the genetic linkage map and the radiation map was the same there were some discrepancies in distances. Some gaps that correspond to 7-10 cM in the genetic map corresponded to very short distances on the radiation map and conversely some clusters on the genetic linkage map were split on the radiation map. Nevertheless the utility of the map has been demonstrated by mapping of 374 ESTs, whose accurate localisation on the radiation map was verified by assigning these ESTs to YACs that map to the same interval.

A subset of 93 hybrids from this panel has been made widely available for genome mapping projects under the panel name Genebridge4. Since the creation of Genebridge4 panel, which classifies as a lower resolution panel, two other radiation hybrid panels of higher resolution have been constructed. The Stanford G3 panel is a medium resolution 10,000-rad panel consisting of 83 clones and the Stanford TNG is

a high resolution 50,000-rad panel consisting of 90 clones, the details of these panels are available from <http://shgc-www.stanford.edu>.

1.4.2.4 Physical maps and the current progress of the Human Genome Mapping Project

The physical mapping phase of the human genome project was officially completed with the landmark publication of the final YAC contig map of the human genome, which provided 75% coverage of the entire human genome (Chumakov *et al.*, 1995). This map although a major achievement, still fell short of the requirement set by the Human Genome Project (HGP) for a 'sequence ready' physical map. Therefore, since reaching this goal efforts have been directed at further map construction, integration and validation to meet the final goal set by HGP, which is to produce a physical map of the human genome containing 30,000 unique markers ordered with respect to each other and spaced on average every **1000**,000 base pairs (1 Mb). The physical map constructed by Hudson *et al.* (1995) at the Whitehead Institute for Biomedical Research goes more than halfway to meet the criteria set by the human Genome Project. This map includes a large number of STSs previously generated and mapped by other groups. Approximately 7000 STSs on the map represent genetic markers that have been previously placed on meiotic linkage maps (Gyapay *et al.*, 1994; Dib *et al.*, 1996). Finally STSs developed from random genomic DNA and Expressed Sequence Tags (ESTs) from dbEST database (division of GenBank allocated for expressed sequences) have also been incorporated into the Whitehead physical map. Therefore this map integrates the locations of a large number of genes with meiotic linkage maps of the human genome. The two independent means used to order these STSs were radiation hybrid mapping (RH) using the Genebridge4 Radiation Hybrid panel (see section 1.4.2.3) and YAC-STS content mapping. Combining RH and YAC-STS content mapping information Hudson *et al.* (1995) estimate that they have achieved an effective resolution of 1 Mb and that their map provides RH coverage of 99% of the genome and physical coverage of 94% of the genome. Even though the Whitehead map is of immense use to disease gene hunters, in its present form this map does not provide adequate scaffold for sequencing the human genome. Presently, human DNA cloned as bacterial artificial chromosomes (BACs), with an average insert size of 100 kb appears to be the most likely source of template for large scale DNA sequencing. As such a contiguous map containing an

ordered marker every 100 kb on average is required to isolate and order efficiently the 100-kb BAC sequencing templates. However as far as ordering of STSs is concerned, physical maps such as the Whitehead map with STSs ordered at an effective resolution of 1 Mb still represent a giant step forward in reaching this ultimate goal. The most up to date physical mapping information on this map can be obtained through the Whitehead Institute for Biomedical Research/MIT Centre for Genome Research's Human Physical Mapping Project World Wide Web server: (<http://www-genome.wi.mit.edu/>). Meanwhile further characterisation and sequencing of individual chromosomes have already been initiated at various genome centres that took up the task of constructing 'sequence ready' maps of a particular chromosome at the very outset of the human genome project. Information on these genome centres can be obtained from the World Wide Web server: (<http://webace.sanger.ac.uk/HGP/>). The most up to date physical mapping information on these chromosomes is available thorough the electronic genomic databases maintained by the respective genome centres. For example a high-resolution cosmid based physical map of chromosome 19 was constructed using restriction mapping and FISH at the Lawrence Livermore National Laboratories (LLNL), USA, and information can be retrieved from their world wide web server at: <http://www-bio.llnl.gov/genome/html/chrom-map.html> (Trask *et al.*, 1992; Ashworth *et al.*, 1995).

1.5 Identification of genes

1.5.1 Identifying the gene of interest in positional cloning

In a positional cloning venture the physical characterisation of a disease candidate interval is usually followed by the isolation of transcripts to identify the disease causative gene. The methodologies used for such gene identification from cloned DNA can be categorised as either traditional or novel. Traditional methods include analysis of zoo blots, detection of CpG islands and newer approaches include exon amplification, cDNA selection, genome sequencing and computer analysis and analysis of regionally mapped candidate genes.

Zoo blot analysis is based on the conservation of coding regions between species (Monaco *et al.*, 1986). In this method hybridisation of a human genomic DNA clone to a southern bolt containing genomic DNA samples of a variety of species at low

stringency is used to identify homologous genes present within the cloned DNA. However the genes that are species specific will not be detected by this method. By means of such zoo blots Sedlacek *et al.* (1993) isolated genomic DNA species conserved between humans, mice and pigs.

Identification of CpG islands is another traditional approach and has been instrumental in the localisation and subsequent isolation of many genes involved in human disease and mouse mutations (Gessler *et al.*, 1990; Herrman *et al.*, 1990). In mammals such islands are thought to mark transcriptionally active DNA sequences and are often located at the 5' end of house keeping genes (Craig and Bickmore, 1994). CpG islands are short stretches of DNA often 1 kb or less, containing CpG dinucleotides which are unmethylated and which are present at the expected frequency, unlike in the remainder of the genome. CpG islands can be recognised by diagnostic patterns of short and long-range restriction maps. For example the enzyme *HpaII* cleaves at the recognition site CCGG but will not cleave CC^MGG. Thus *HpaII* tend to cut selectively at CpG islands and as a consequence these islands are often referred to as HTF islands (*HpaII* tiny fragments). In mammals CpG islands are generally also G + C rich and so are principle sites of cleavage for methylation sensitive enzymes with G + C rich recognition sequences such as BssHII, EagI and SacII. Rare cutter enzymes such as NotI (GCGGCCGC) cut almost exclusively within islands. Clustering of such sites thus is commonly associated with CpG islands.

Conserved DNA fragments identified from Zoo blots and CpG island containing fragments can be tested for expression by hybridisation to Northern blots of RNA isolated from foetal and adult tissues. If a transcript is identified then the genomic fragment subsequently can be hybridised to cDNA libraries. Although Northern blotting and analysis of conserved DNA fragments and CpG islands have been particularly effective in the isolation of many human disease genes they are very labour intensive. An alternative approach is the direct selection method (Monaco, 1994) where whole YAC or cosmid inserts are directly hybridised to cDNA libraries on filters after suppression of repeated sequences in the radioactively labelled probe. Although associated with technical difficulties this method has enabled many new genes to be isolated.

In cDNA selection (Lovett *et al.*, 1991; Parimoo *et al.*, 1991), recognised as one of the newer approaches to gene isolation, an amplified cDNA library is hybridised to immobilised YAC or comid clones that cover part of the genome of interest. Of those cDNAs isolated at least one should correspond to the desired gene. As with direct selection this process is technically difficult (Lovett, 1994) due to simultaneous selection of cDNAs with homology to pseudogenes and low copy number repeat sequences.

In exon amplification (Buckler *et al.*, 1991) random segments of chromosomal DNA are inserted into an intron present within a mammalian expression vector. After transfection cytoplasmic mRNA is screened by PCR amplification for the acquisition of an exon from the genomic segment. The amplified exon is derived from the pairing of unrelated vector and genomic splicing signals. A number of human genes have successfully been isolated through this method (Monaco, 1994).

Sequencing of genomic segments from the region of interest and analysis of the assembled sequences for potential Open Reading Frames (ORFs) is another novel approach. This approach has been applied to the positional cloning of the human gene for Kallmann syndrome (Legouis *et al.*, 1991) and the identification of *RPGR* gene causing X-linked retinitis pigmentosa (Meindel *et al.*, 1996).

The success of Legouis *et al.* (1991) in identifying the Kallmann syndrome gene from DNA sequence data led to an alternative approach designated 'candidate gene' or 'positional candidate' approach for gene identification (Ballabio, 1993). Unlike functional and positional cloning this approach does not require the isolation of novel genes but relies on the availability of information regarding function and map position from previously isolated genes, which could have originated from either genomic sequencing or expressed sequence tags (EST)/ cDNA fingerprints. In fact most of the genes that have associated with various retinopathies have been identified by the positional candidate approach (see section 1.10). The future expanding success of the positional candidate approach is predicted on the development of the increasing dense transcript map, which is described in detail below.

1.5.2 Development of the human transcript map

A comprehensive human gene map is a critical resource for the positional candidate approach to gene cloning (Wilcox *et al.*, 1991; Collins, 1995). Therefore as a prelude to the development of a gene map the tens of thousands of partial cDNA sequences of genes (or ESTs) produced by the Washington University/Merck & Co's public EST initiative and other groups that have been deposited in the public database (dbEST), were used by an international EST mapping consortium to develop a human transcript map (Schuler *et al.*, 1996; Marra *et al.*, 1998). Initially all the available ESTs in dbEST were clustered to identify novel, non-redundant mapping candidates for generating a transcript map. This 'UniGene collection' contains more than 48,000 clusters of sequences, each representing the transcription product of a distinct human gene (Schuler, 1997). With current estimates of 80,000 to 100,000 genes in the human genome, this is close to the 50% mark and the collaborative mapping effort has resulted in the placement of 15,000-20,000 of these transcripts on radiation hybrid (RH) and YAC maps. Previously other groups have also mapped cDNAs but mapping was at a limited scale and usually only to the resolution of a chromosome assignment (Wilcox *et al.*, 1991; Adams *et al.*, 1991, Khan *et al.*, 1992; Polymeropoulos *et al.*, 1992, 1993). However not all genes will be represented by ESTs, since some genes will either be poorly expressed or completely absent in the cDNA libraries analysed. These would include genes that show highly restricted tissue or cell expression patterns and genes that are expressed at certain developmental stages. Therefore it has been estimated that the EST mapping effort will only identify sequences from about 80% of all the human genes. The remaining genes are to be identified as coding sequences in genomic DNA clones.

The UniGene collection and the transcript maps are an important resource for many investigators. Gene hunters can use the transcript maps to gain valuable clues to expected gene location and density in an area of interest. However the regional assignment of the ESTs in the gene map is still limited by the resolution of the radiation hybrid maps and, most ESTs that have been mapped to large genetic interval still have to be assayed in an ordered set of clones for a finer localisation. Therefore these ESTs represent a somewhat raw resource for positional cloning endeavours, requiring a further assay within DNA contig(s) of respective disease intervals for

identification of potential candidate genes. The cDNA clones corresponding to the ESTs can be obtained from the IMAGE (integrated molecular analysis of gene expression) consortium at Lawrence Livermore National Laboratories (LLNL, USA), which maintains cDNA clones representing unique genes in gridded arrays.

UniGene clusters are also being studied to find gene polymorphisms. The Co-operative Human Linkage Centre (CHLC) has found ESTs to be a useful source of polymorphic markers for genetic mapping (Boguski and Schuler, 1995). Furthermore the recently developed techniques for assessing gene expression on a genomewide scale (e.g., microarray expression systems) take advantage of the abundance of unique EST sequences readily retrieved from GenBank (Strachan *et al.*, 1997). Finally the transcript map will be an important resource for genomic sequencing as gene-dense regions could be used as initial targets for sequencing. The gene map and the UniGene data set can be accessed through National Centre for Biotechnology Information's WWW service (<http://www.ncbi.nlm.nih.gov/>).

1.5.3 Cross species comparison for gene identification/ Comparative genomics

Characterisation and sequencing of genomes of model organisms (such as *E. coli*, *Saccharomyces cerevisiae*, *Drosophila*, *C. elegans*, *Fugu rubripes* and *Mus musculus*) is an essential part of the Human Genome Mapping effort and was undertaken to produce comparative maps to facilitate linkage predictions, disease gene identification, and studies of genome organisation and evolution. Gene identification strategies that rely on cross species comparisons are based on the belief that functionally significant regions of the genome are highly conserved during evolution. Once identified comparison between homologous genes in distantly related species also provide remarkable insight in to their function. In cases where orthologs and family members of genes mutated in human disease are cloned across several species, the experimental advantage in each organism can be applied synergistically to characterise the homologous gene products to understand the molecular mechanisms of the human disease process. Studies of the *S. cerevisiae* *MEC1* and *TEL1* genes, for example, have provided valuable insight into the function of the human *ATM* gene, which is mutated in ataxia telangiectasia, and further analysis is in progress using the yeast experimental system (Morrow *et al.*, 1995).

Traditionally gene identification based on cross-species comparison relied on the two major approaches, low stringency hybridisation of cDNA or genomic DNA libraries and PCR using degenerate oligonucleotides, both of which are technically difficult and time consuming. Therefore in an alternative approach the proliferating model organism sequence data and thousands of human partial cDNA sequences of genes (or ESTs) that have been deposited in the public database (dbEST), have been utilised to accelerate the identification of genes mutated in human diseases. This approach described as 'Genome cross referencing' is involved in systematically mapping novel human ESTs that show significant sequence similarity with model organism proteins (Bassett *et al.*, 1997). These mapped cDNAs cross-reference model organism protein function data with positions on the mammalian maps and allow application of the positional candidate approach to identify the genes mutated in human diseases that map to the same location in the genome. Through the cross-reference database (XREFdb) on the World Wide Web (<http://www.ncbi.nlm.gov/XREFdb/>), which also integrates the vast resource of EST mapping data generated by the Radiation hybrid mapping consortium (Schuler *et al.*, 1996), investigators can identify human homologues of model organism genes and establish links between model organism genes and human disease states, using either a protein query or/and map position-based query tool. Using this genome cross-referencing concept Banfi and co workers (1996) identified 66 human cDNAs with significant homology to previously identified *Drosophila* genes involved in the generation of mutant phenotypes and mapped these ESTs using FISH and radiation hybrid mapping. Preliminary data available for *D. melanogaster* (5% genome sequenced) and *C.elegans* (65% genome sequenced) suggest that the cross-referencing method will become increasingly valuable as more genome sequence from model multi-cellular eukaryotes becomes available (Bassett *et al.*, 1997). At present only *S. cerevisiae* and *E. coli* genomes have been completely sequenced (Goffeau *et al.*, 1996; Blattner *et al.*, 1996), and have already proved their utility by leading to the identification and functional analysis of several genes mutated in human disease (Hieter *et al.*, 1996; Goffeau *et al.*, 1996).

Similar to cross referencing a technique known as ‘homology probing’ is also capable of identifying genes mutated in human diseases and relies on the gene sequence and function data of the model organism as well as a robust collection of ESTs (Yahraus *et al.*, 1996). In this method a process or pathway is identified in humans that when defective result in disease. The genes that participate in this pathway are then identified in model organisms, and the EST database is screened for human cDNAs representing homologues of these model organism genes, which become candidates for genes mutated in the human disease. The potential of this technique is exemplified by the identification of human genes PXR1 (Yahraus *et al.*, 1996) and PXAAA1 (Dodt *et al.*, 1996) involved in peroxisome biogenesis and responsible for inherited defects in this process (Zellweger syndrome and related peroxisome biogenesis disorders).

As the techniques outlined above (and in section 1.5.2.) would indicate the future gene hunting process will soon be reduced to a mere computer exercise. When a new disease is assigned to a specific map position it may be possible to interrogate the genome database for that portion of the chromosome and obtain a list of all the genes assigned to the same region. By comparing the features of the gene with the features of the disease it will be easy to select the most likely candidate gene(s).

1.6 Retinal Genetics

Inherited eye diseases constitute a substantial proportion of diseases classified as genetic defects (McKusick, 1992). Amongst these diseases retinal dystrophies constitute a major cause of blindness. Most of these diseases were originally defined based on the fundus appearance as seen through the ophthalmoscope, through histopathologic examination of autopsied eyes, and more recently through visual function assessments. Now biochemical studies and molecular genetic analyses have provided even newer approaches for, detecting these diseases, finding the underlying genetic defect, and defining pathogenetic mechanisms. The rapid development in gene cloning techniques that has been witnessed in recent years has led to the cloning of many of the genes associated with retinal degenerations and disorders (<http://www.sph.uth.tmc.edu/RetNet/>) leading to a greater understanding of the visual process and the functions of the retina. Moreover, through the identification of disease causative genes and mutations previously unrecognised associations between different

clinical entities that share common gene mutations (gene sharing/ allelic effect) as well as distinctly different molecular alterations within the spectrum of what traditionally was believed to be the same disease (locus heterogeneity) has been revealed. Genetic heterogeneity suggests that retinal degeneration is a common end point for many biochemical abnormalities and allelic heterogeneity suggests that abnormalities at different sites within the same gene can have either similar or dissimilar consequences, depending upon the functional roles of these sites. More importantly however the identification of the precise genetic defect in a hereditary retinal disease helps to define its cause, allows new insights into pathogenesis and provides a framework for considering means of treatment.

All of the above observations are particularly applicable for one representative hereditary retinal disease- retinitis pigmentosa (RP), the retinal degeneration under investigation in this study. However prior to embarking upon the detailed description of molecular genetics of RP, it is important to first understand the normal anatomy and physiology of the retina and the photochemistry of visual transduction.

1.7 Structure of the retina

The retina, which is one of the best-characterised parts of the central nervous system, in the most comprehensive sense, includes all the structures that are derived from the optic vesicle. Namely the sensory layers of the retina, the pigmented epithelium as well as the epithelial lining of the ciliary body and of the iris. However put simply the retina can be said to have two components: a sensory layer (neural retina) and a pigmented layer derived from the inner and outer layers of the optic vesicle, respectively. The two layers are attached loosely to each other by the extracellular matrix that fills the space between the apical villi of the retinal pigment epithelium (RPE) and the outer segment of the photoreceptors as well as inter photoreceptor space (see fig 1.1). The retinal pigment epithelium is firmly attached to the Bruch's membrane of the choroid.

1.7.1 The sensory (Neural) retina

The retina, when seen in cross section by light microscopy, is represented by 10 layers (fig 1.1). The structure of the neural retina is best explained in the context of these layers of which, layers 2-10 represent the neural retina.

1. retinal pigment epithelium
2. outer segments of photoreceptors (rods and cones)
3. external limiting membrane
4. outer nuclear layer
5. outer plexiform layer
6. inner nuclear layer
7. inner plexiform layer
8. ganglion cell layer
9. nerve fibre layer
10. internal limiting membrane

The mature vertebrate retina possesses five major neuronal cell types that form the visual pathway from retina to brain. The primary neurons in this visual pathway are the *photoreceptors*, which constitute the layer of rod and cone outer segments, the outer nuclear layer and the outer plexiform layer. The *horizontal*, *bipolar* and *amacrine* cells constitute the second order neurons and the *ganglion cells* constitute the third order neurons in the visual pathway. Photoreceptor axons synapse with the bipolar and horizontal cells in the outer plexiform layer, which marks the junction of the first and second order neurons in the retina. The flat horizontal cells serve to modulate and transform visual information received from photoreceptors and also form a network of fibres that integrate the activity of photoreceptor cells horizontally. The inner nuclear layer consists of nuclei of bipolar cells, amacrine cells and supporting Müller cells. Müller cells are the largest cells of the retina and extend from the external to the internal limiting membranes. As the principal glial cells they conserve the structural alignment of its neuronal elements. Müller cells also have an important role in the metabolism of the retina and they contribute to the b-wave of the electroretinogram. The axons of bipolar cells synapse with amacrine cells and dendrites of ganglion cells in the inner plexiform layer, which marks the junction of the second order neurons, the bipolar cells, with the third order neurons, the ganglion cells. The ganglion cell layer is composed mainly of the cell bodies of the ganglion cells and the nerve fibre layer contains axons of these cells that join to form the optic nerve, which send the information gathered by the photoreceptors for further processing and interpretation by the visual cortex. The distribution of photoreceptors to optic nerve fibres is not uniform. In the fovea centralis approximately 200,000

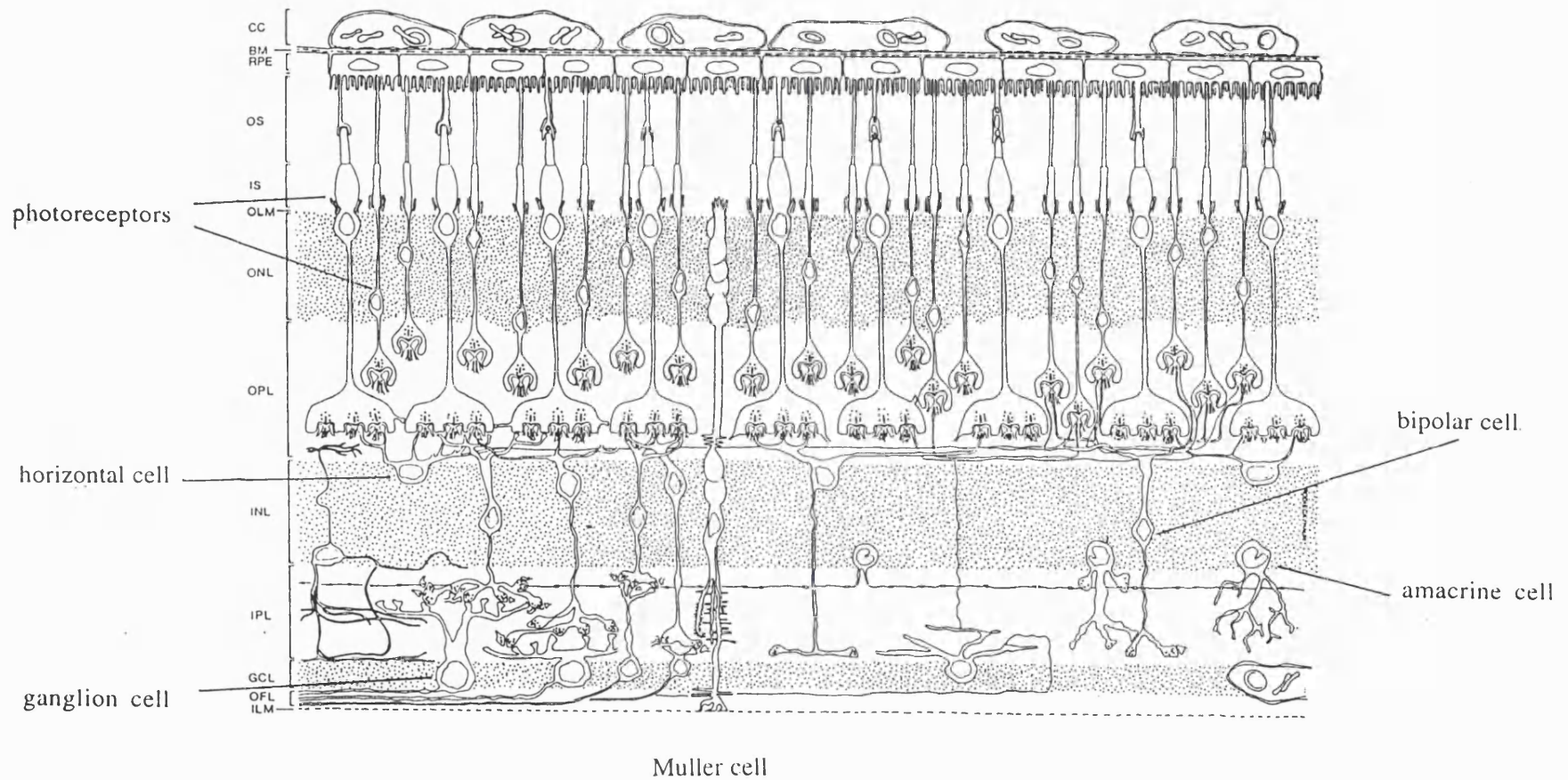


Figure 1.1

A schematic diagram of the human retina, showing most of the cell types and layers. The photoreceptors' terminals are greatly enlarged to show their connections with other cell types. CC: choroidocapillaries; BM: Bruch's membrane; RPE: pigment epithelium; OS: photoreceptor outer segments; IS: photoreceptor inner segments; OLM: outer limiting membrane; ONL: outer nuclear layer; OPL: outer plexiform layer; INL: inner nuclear layer; IPL: inner plexiform layer; GCL: ganglion cell layer; OFL: optic fibre layer; ILM: inner limiting membrane.

cones are connected to at least as many optic nerve axons. However in the peripheral retina as many as 10,000 rods may connect in clusters to a single nerve fibre, with considerable overlapping, so that a single point of light may stimulate several clusters simultaneously. In the neural retina the photoreceptors cells are the percipient elements, required for the reception and conversion of light energy to neural impulses. All other neurons constitute the associative and conductive elements, which convey the nerve impulse to the visual cortex. Photoreceptors are of particular interest in this study as they are also commonly involved in retinal degenerations.

1.7.1.1 Photoreceptor cells

The two kinds of photoreceptors found in the vertebrate retina are morphologically and functionally distinct from each other (fig 1.2 and 1.3). The rod photoreceptors possess thin cylindrical processes, function in dim light and are responsible for scotopic (night) vision and furthermore do not perceive colour. The cone photoreceptors are generally conical in shape, function in bright light (photopic) and are responsible for colour vision. There are on average 92 million rods and 4.6 million cones in the human eye. Individual variations in the density of both rods and cones occur in different regions in the eye. The greatest variability occurs near the fovea and at the very periphery of the retina (i.e. at the ora serrata).

Rods: Rods numbers far exceed that of cone photoreceptors and are at their greatest concentration in mid periphery of the retina. Rod concentration decreases as approaching the centre where they are completely absent in the fovea centralis. Extending from their cell bodies, the photoreceptors have two morphologically distinct regions: the inner and outer segments (fig 1.2). In rods the inner and outer segments are 40-60 μm long throughout the retina. The slender outer segment (25-28 μm long and 1-1.5 μm in diameter) and the slightly thicker inner segment of rods do not show much variation in morphology from the fovea to the periphery (Stryer, 1988). The major functions of the rod cell are highly compartmentalised. The function of the rod outer segment (ROS), which lies embedded in the interphotoreceptor matrix just internal to the retinal pigment epithelium, is the conversion of light energy into electrical impulses. The outer segments of the rods are cylindrical in shape and contain stacks of flattened double lamellae in the form of discs that are formed by the basal invagination of the outer segment plasma membrane. There are about 1000 discs

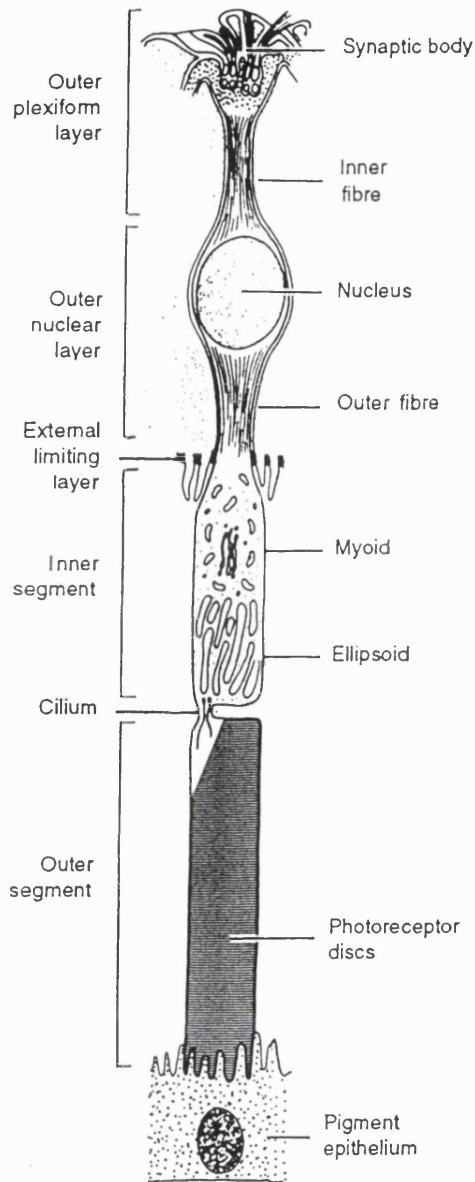


Figure 1.2

Diagrammatic representation of a generalised photoreceptor in the human retina, showing its various components and their location relative to the retinal layers. Figure adapted from Tripathi and Tripathi (1984).

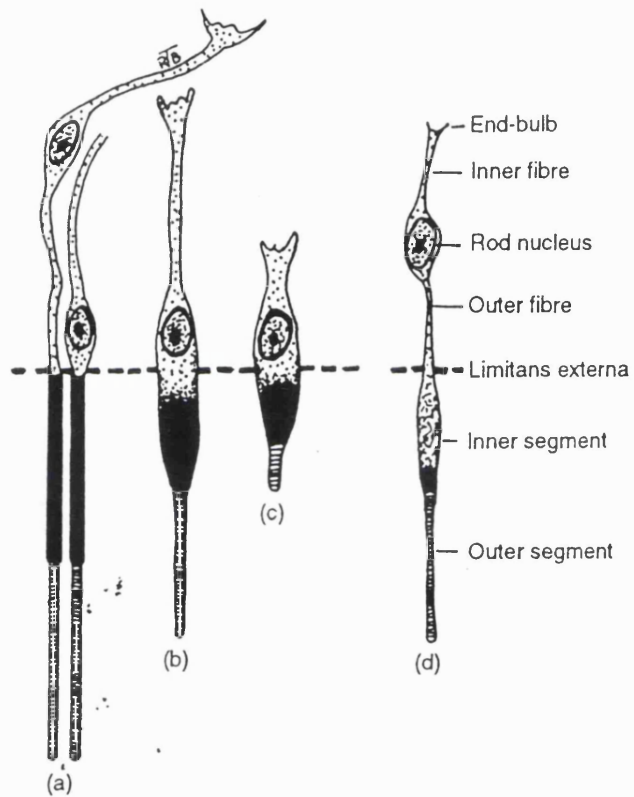


Figure 1.3

Schematic representation of photoreceptors of human retina. (a) Cones from the foveola; (b) cones from midway between the ora serrata and the optic disc; (c) cones from near the ora serrata; (d) rod. Figure adapted from Tripathi and Tripathi (1984).

in each human rod outer segment and the discs contain 90% of the visual pigment, rhodopsin, whereas the remaining pigment is scattered on the surface of the plasmalemma. Rhodopsin has the greatest sensitivity for blue-green light (λ_{\max} 493nm) and allows scotopic vision. The discs in the ROS are continually renewed throughout life by the “disc shedding” process in which the discs at the very apex of rods are removed in a light triggered rhythmic pattern and phagocytised by the RPE cells thus maintaining the outer segments at a relatively uniform length (Young and Bok, 1969; Bok, 1985). The outer segment is connected to the inner segment by a modified cilium, which functions in transmitting cellular components from the inner segment and cell nucleus to the discs and their plasma membrane. The rod inner segment is cylindrical and is composed of a finely granular cytoplasm. There are two histologically discernible regions: an outer eosinophilic ellipsoid and an inner basophilic myoid (fig 1.2). The ellipsoid contains a large number of long and slender mitochondria and the cytoplasm contains smooth endoplasmic reticulum, neurotubules, free ribosomes and glycogen granules. The myoid, which is the site of major protein synthesis for new outer segment discs, contains high concentration of free ribosomes, a rough endoplasmic reticulum, glycogen granules and a golgi apparatus. The cytoplasm also contains microtubules and microfilaments that are arranged in parallel to the long axis of the cell and extend to the level of external limiting membrane and of the synapse, respectively. A thin cytoplasmic process (the outer fibre) connects the inner segment to the cell body that contains the nucleus. The rods synapse with second order neurons i.e. the bipolar and horizontal cells, through round or oval cytoplasmic expansions 1 μm in diameter known as spherules at the synaptic body. The spherule contains numerous presynaptic vesicles filled with acetylcholine, mitochondria and neurotubules. Both horizontal and bipolar cells make contact with the rod spherule. A horizontal cell only contacts a rod spherule once, but several different horizontal cells may contact each spherule. One to four bipolar cells contact an individual spherule at separate points and each bipolar cell may contact up to hundreds at the periphery of the retina.

Cones: The density of cones is maximal, with an average of 199,000 cones per mm^2 at the fovea, where the visual resolution is at its greatest but this number reduces to 4500 per mm^2 towards the periphery. The morphology of cones differs depending on their location in the retina (Tripathy and Tripathy, 1984). Cone length is maximal (80 μm) at the fovea and gradually reduces to 40 μm at the periphery. Moreover, cones appear rod-like at the fovea and the inner region of the outer segment becomes wider towards the periphery (fig 1.3). Cones located in the periphery have a conical shaped outer segment with the apex pointing towards the RPE cell layer. Ultrastructurally, the cone outer segments have more discs than (1000-1200 per cone) than do rod outer segments and are also arranged differently to rods. Unlike the rod discs, cone discs are attached to each other as well as to the surface plasma membrane and are not detached easily. Disc shedding also takes place in the cone photoreceptor cells in a process in which the cone outer segments remodel themselves after each shedding event, since most apical cone outer segments have a constant but smaller diameter than the basal ones (Hogan *et al.*, 1974; Anderson and Fisher, 1976).

Photopic (colour) vision originates in the cones by the presence of trichromatic pigments. Each cone contains one of three different iodopsin molecules that absorb light at three distinct peaks at 440 nm (blue), 540 nm (green) and 577 nm (orange) referred to as S- (short), M- (medium), and L- (long) wavelength cones, respectively (Dacey and Lee, 1994). Cone synapses have cytoplasmic expansions known as pedicles that are larger (5 μm in the fovea) than the rod spherules and are pyramidal in shape. At the fovea only one bipolar cell contacts one cone pedicle along with two different horizontal cells.

1.7.2 Retinal pigment epithelium

The retinal pigment epithelium (RPE) consists of a single layer of approximately five million cuboidal cells firmly attached to its basal lamina, the Bruch's membrane. RPE cells are joined to one another by tight junctions and constitute an important structural and functional part of the outer blood retinal barrier. Together with the endothelium of the choriocapillaries, it effectively excludes the exchange of potentially toxic substances between the choroidal circulation and the neural retina. The basal microfibrils of RPE are functionally linked to the choroid via the Bruch's membrane whereas apical region of the RPE is loosely attached to the sensory retina.

The absence of specialised adhesion molecules in the interphotoreceptor matrix, and the lack of junctional complexes between apical microvilli of pigmented epithelial cells and the outer segments of the photoreceptor leave the sensory retina prone to detachment in pathological conditions. Despite the lack of any adherent functions the apical villi of the RPE are involved in phagocytosis of outer segment discs, and thus promote the renewal of photoreceptor cells (Young and Bok, 1969; Bok, 1985).

Another important function of the RPE is the uptake, processing, transport and release of vitamin A (retinol) and some of its visual cycle intermediates (retinoids) (Dowling, 1960; Bok, 1985). The purpose of this cycle is to regenerate 11-*cis* retinaldehyde, the retinoid that serves as the chromophore for the visual pigments in rod and cone outer segments. Retinol uptake occurs at both the basolateral and apical surfaces of RPE via a receptor mediated processes. The release of a crucial retinoid, 11-*cis* retinaldehyde (11-*cis* retinal) occurs solely across the apical membrane. Delivery of retinol across the basolateral membrane is mediated by a retinol binding protein (RBP) that is secreted by the liver as a complex with retinol and through the receptors for RBP found on the basal and lateral plasma membrane of the RPE cells. Within the cell retinol and its derivatives are solubilised by intracellular retinoid binding proteins that are selective for retinol (cellular retinol binding protein, CRBP) and 11-*cis* retinoids (cellular retinal binding protein, CRALBP). Release of 11-*cis* retinal across the apical membrane and re-uptake of retinol from the photoreceptors during the visual cycle is promoted by an intercellular retinoid binding protein (IRBP). In addition to photoreceptor renewal and recycling of vitamin A, the RPE is involved in the absorption of scattered light by the melanin granules, transport of nutrients and metabolites through this extra retinal-blood barrier, secretion of interphotoreceptor matrix (IPM) and maintenance of the photoreceptor microenvironment.

1.8 Phototransduction

A visual image enters the eye as light of different wavelengths and intensities and is captured by the two types of retinal cells, rods and cones, which exhibit distinct light sensitivities and response kinetics. The light signal captured by these cells stimulate a series of chemical reactions, called phototransduction, which is similar in both rods and cones and culminates in the generation of a neuronal signal that is

transmitted via the optic nerve to the visual cortex where perception and interpretation occurs (for review see Koutalos and Yau, 1996). The outer segments of rods and cones which are the sites of visual transduction in vertebrate eyes can easily be isolated in pure forms and maintained in isolation for biochemical and electrophysiological studies. This has led to the elucidation of the intervening biochemical steps that comprise the phototransduction pathway. The visual pigment of the rod cell, rhodopsin (R), consists of the transmembrane protein opsin, chemically linked to the chromophore 11-*cis* retinal at Lys296. The primary event in the phototransduction cascade is the light triggered isomerisation of the 11-*cis* retinal of rhodopsin to its all *trans* isomer (see fig 1.4). This isomerisation alters the geometry of retinal and results in the release of all-*trans* from its membrane bound cofactor in to the disc membrane lipid (Wald, 1968). This conformational change results in the formation of the photoactivated or photolysed rhodopsin (R*), which is catalytically active.

The photoactivated rhodopsin binds a G protein called transducin (T), initiating a signal-amplifying cascade of reactions. G proteins are signal transducing guanosine nucleotide binding proteins whose function is to transmit signals between transmembrane receptors and cellular effectors. In the inactive state, transducin is a membrane-associated complex consisting of two functional subunits T_{α} and $T_{\beta\gamma}$ (T_{β} and T_{γ}) and non-covalently bound GDP (Baehr *et al.*, 1982; Fung, 1987). Interaction of R* with transducin catalyses the exchange of GDP bound to T_{α} subunit for GTP and the subsequent dissociation of the active GTP- T_{α}^* complex from the $T_{\beta\gamma}$ heterodimer (see fig 1.4). In rods each photoactivated rhodopsin generates several hundred GTP- T_{α}^* that persist in the active state long enough to find and activate membrane associated cGMP-PDE molecules (PDE).

In the dark rod cGMP-PDE is a heterotetrameric peripheral membrane protein composed of two catalytic α and β subunits and two smaller inhibitory γ subunits (Baehr *et al.*, 1979; Deterre *et al.*, 1988; Fung *et al.*, 1990). The interaction of PDE γ subunits with GTP- T_{α}^* leads to the activation of PDE (PDE*) releasing its hydrolytic potential (see fig 1.4).

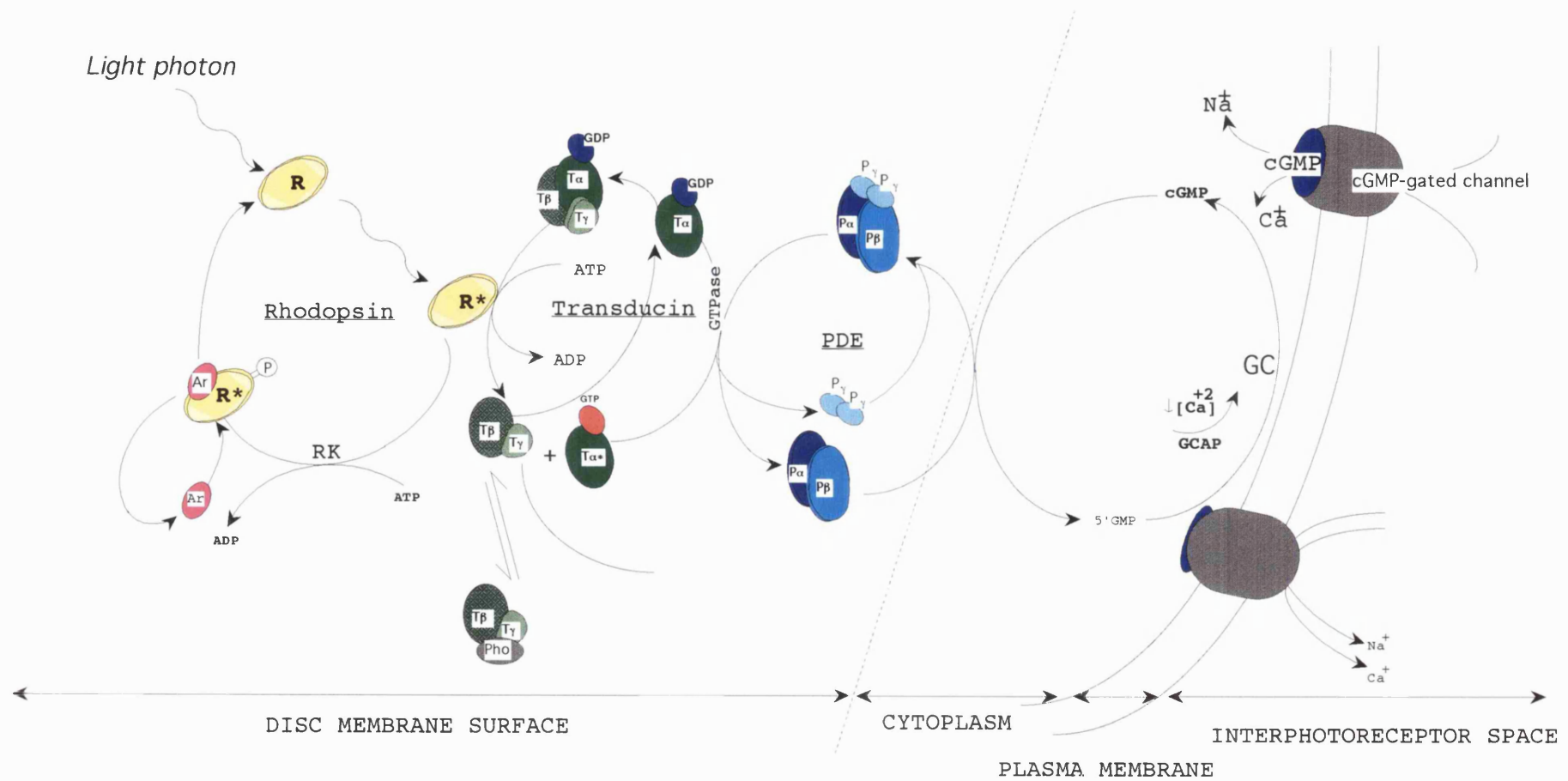


Figure 1.4:

Diagram of the biochemical events involved in the activation and inactivation of phototransduction. Ar: arrestin, Pho: phosducin, and other abbreviations are as described in text (figure modified from Farber, 1995).

PDE* hydrolyses 3', 5'-cGMP to 5'-GMP at a high rate which is limited only by the availability of the substrate through diffusion (Bourne *et al.*, 1990). PDE remains active until GTP associated with GTP-T α * is hydrolysed by the intrinsic GTPase activity of T α . Upon which T α dissociates from PDE γ and re-associates with T $\beta\gamma$ and the PDE γ subunits are released to form PDE complex once again and inhibit enzyme activity (see fig 1.4). The outer segment cation channels that control the influx of ions across the photoreceptor plasma membrane are gated directly by cGMP and the fall in cGMP that result from light triggered PDE activation changes the conformation of these cGMP activated cation channel proteins leading to channel closure. This channel protein consists of a 63 kd α subunit and 240 kd tightly linked β and γ subunits (Cook *et al.*, 1989; Molday *et al.*, 1990; Chen *et al.*, 1993; Illing *et al.*, 1994). Channel closure decreases the conductance of the plasma membrane to cations and results in the hyperpolarisation of the plasma membrane, inhibition of neurotransmitter release, and signalling of the light stimulus to adjacent neurons (Fesenko *et al.*, 1985; Stryer, 1986; Yau and Baylor, 1989).

1.9 Regulation of phototransduction and the retinal proteins involved

Mechanisms must exist in order for the photoreceptors to maintain their sensitivity and adaptability. In the dark-adapted state the photoreceptors are capable of responding to a single photon of light. Such sensitivity is achieved through signal amplification i.e. activated R* is able to activate over 500 transducin molecules and each molecule of activated PDE is capable of hydrolysing approximately 1000 cGMP molecules (Stryer, 1986). Light adaptation is a process whereby absorption of each additional photon is less effective in activating the phototransduction pathway, thus producing smaller alterations in the conductance of the cell. It is a necessary process to maintain sensitivity of the receptor to flash of light in the presence of background illumination. Moreover modulation of light adaptation is required for the photoreceptors to respond to light stimuli that vary in intensity without response saturation.

1.9.1 Regulation by Ca^{2+} - binding proteins

In photoreceptors, Ca^{2+} play a crucial role in photorecovery and adaptation (Matthews *et al.*, 1988). Calcium ions regulate several stages of the phototransduction pathway by modifying the activity of different Ca^{2+} binding proteins that in turn interact with key enzymes in the pathway. In the dark, the concentration of Ca^{2+} is maintained at 300nM as the entry of Ca^{2+} through the cGMP-gated cation channels is balanced by efflux of calcium through the $\text{Na}^+/\text{Ca}^{2+}\text{-K}^+$ exchanger (Koch and Stryer, 1988). Upon light illumination, the closure of cGMP-gated cation channels and the continued efflux of calcium through the $\text{Na}^+/\text{Ca}^{2+}\text{-K}^+$ exchanger result in the drop of Ca^{2+} concentration to <70 nM. This decrease in $[\text{Ca}^{2+}]$ stimulates the enzyme guanylate cyclase (GC), the peripherally membrane bound enzyme that catalyses the conversion of GTP to cGMP. Following illumination, this key enzyme is responsible for the synthesis of cGMP, which in turn opens cation channels in the outer segment plasma membrane and re-establishes the dark potential of the cell. Two retina-specific membrane-associated guanylate cyclases have been cloned and sequenced (retGC1 and retGC2) (Shyjan *et al.*, 1992; Margulis *et al.*, 1993; Lowe *et al.*, 1995), however only retGC1 has been localised to outer segments by immunocytochemical localisations (Dizhoor *et al.*, 1994; Liu *et al.*, 1994). These GCs are activated and regulated by specific Ca^{2+} -binding proteins known as guanylate cyclase activating proteins (GCAPs) (Gorczyca *et al.*, 1995; Dizhoor *et al.*, 1995; Palczewski *et al.*, 1994). Two GCAP proteins have been isolated from retina (GCAP1 and GCAP2), but only GCAP1 have been localised definitively to both rod and cone outer segments and purified. Ca^{2+} free form of GCAP1 has been shown to regulate the activity of ROS guanylate cyclase as well as recombinant retGC1 (Subbaraya *et al.*, 1994; Gorczyca *et al.*, 1995). Therefore in summary GCAPs mediate Ca^{2+} sensitive regulation of guanylate cyclase, which by synthesising cGMP restores the open state of the channels, thus promoting recovery of the dark state of rod photoreceptors following light exposure. The entry of Ca^{2+} then leads to decreased activity of GC and the return of the dark state. This feed back loop involving Ca^{2+} is likely to be a major contributor in the maintenance of a constant cGMP level in the dark and to recovery following illumination (Koch and Stryer, 1988).

Originally a different photoreceptor-specific Ca^{2+} -binding protein, recoverin, was thought to be the regulator of rod outer segment guanylate cyclase activity (Dizhoor *et al.*, 1991; Lambrecht and Koch, 1991), but once cloned and expressed recoverin did not alter the activity of GC under *in vitro* conditions (Hurley *et al.*, 1993; Gray-Keller *et al.*, 1993). It was also demonstrated that raising the concentration of recoverin within rod cells slows recovery from photoexcitation. Moreover, Visinin and S-modulin (Gray-Keller *et al.*, 1993) the recoverin like proteins isolated from chicken cones and frog rods, respectively, were also shown to prolong the activation of cGMP phosphodiesterase (PDE) (Kawamura and Murakami, 1991). Since the proposed function for recoverin in the regulation of GC was not supported, the precise role of recoverin in the phototransduction pathway has been intensely investigated. Recent studies of rhodopsin phosphorylation has revealed a possible function for recoverin and related proteins. S-modulin was shown to inhibit phosphorylation at elevated levels of Ca^{2+} (Kawamura, 1993; Klenchin *et al.*, 1995). Because phosphorylation of rhodopsin and subsequent binding of arrestin block further activation of transducin, thus reducing the effective lifetime of photolysed rhodopsin (see section 1.9.2), the inhibition of phosphorylation by S-modulin and recoverin would be expected to prolong the lifetime of activated rhodopsin. This is also an effect consistent with the longer lifetime of activated PDE and the prolonged photoresponse. Recoverin is now thought to function during light adaptation through its Ca^{2+} -dependent inhibition of rhodopsin kinase (Klenchin *et al.*, 1995; Gorodovikova and Philippov, 1993; Gorodovikova *et al.*, 1994), an idea further supported by the observation that recoverin binds to rhodopsin kinase in a Ca^{2+} -dependent manner (Chen *et al.*, 1995a). However the transgenic mouse in which recoverin expression has been eliminated does not show the expected response kinetics (Baylor, 1996). In addition there is disagreement about the Ca^{2+} concentration at which recoverin might regulate the phosphorylation of rhodopsin, perhaps requiring a concentration elevated beyond physiological conditions. Furthermore recoverin immunoreactivity is most prevalent at photoreceptor terminals (Polans *et al.*, 1993; Milam *et al.*, 1993), and unlike other phototransduction-specific proteins which are sequestered in the outer segments, recoverin is distributed throughout the cell, indicating that it might be involved in functions distinct from phototransduction. Therefore the precise function of recoverin in phototransduction is still controversial. Aside from studies of phototransduction, recoverin was also identified as the protein previously known as CAR (cancer

associated retinopathy) protein and recoverin has been associated with some ocular carcinomas (McGinnis *et al.*, 1995).

Calmodulin is another cytosolic Ca^{2+} -binding protein that is expressed ubiquitously and found in the rod outer segments (Nagao *et al.*, 1987). It has been suggested that the cGMP-gated channels are also responsive to concentrations of Ca^{2+} and might be regulated by Calmodulin (Hsu and Molday, 1993). *In vitro* experiments have shown that in the dark, binding of Ca^{2+} bound calmodulin to the β -subunit of the channel protein lowers the apparent affinity of the channel for cGMP. In this low affinity state any decrease in cytosolic cGMP, as which occurs upon light illumination will lead to channel closure. The resultant lowering of intracellular Ca^{2+} levels due to channel closure will in turn increase the sensitivity of the channel to cGMP by uncoupling calmodulin from the channel, thus allowing the channel to reopen at lower cGMP levels, leading to recovery of the ROS to its dark state as cGMP synthesis proceeds with the activity of guanylate cyclase. The opening of the channel will in turn restore the $[\text{Ca}^{2+}]$ to dark levels and Ca^{2+} bound calmodulin will rebind the channel. All these have been postulated upon *in vitro* experiments, however a Ca^{2+} binding protein other than calmodulin might contribute to channel sensitivity *in vivo* (Downing and Zimmerman, 1995).

Therefore in summary, the lowered Ca^{2+} levels due to photon absorption: (1) accelerate the synthesis of cGMP owing to GC stimulation by the Ca^{2+} -free form of GCAP1; (2) increase sensitivity of the channel to cGMP and accelerate the recovery of the dark current and Ca^{2+} by uncoupling calmodulin from the cGMP gated channel; and (3) shorten the life time of photolysed rhodopsin (R^*) by uncoupling recoverin from rhodopsin kinase, thus allowing the inhibitory effect of rhodopsin phosphorylation to proceed (see section 1.9.2). These Ca^{2+} sensitive steps represent the principle mechanisms of light adaptation in vertebrate photoreceptors.

1.9.2 Regulation by non Ca^{2+} binding proteins

Apart from regulation by the Ca^{2+} sensitive mechanisms described above other mechanisms exist for inactivation of photoactivated rhodopsin (R^*), which help maintain the fast physiological response of photoreceptors to light. The deactivation of the phototransduction cascade is initiated by phosphorylation of R^* by a cytosolic

protein in the rod photoreceptors known as rhodopsin kinase (RK) (Shichi and Somers, 1978). It phosphorylates the threonine and serine residues located in the C-terminal domain of rhodopsin on the cytoplasmic surface of the disc membrane (Hargrave *et al.*, 1980). Thus phosphorylated rhodopsin has a decreased ability to activate transducin and enhanced ability to bind a 48 kd protein known as arrestin (Arr). Arrestin, also known as S-antigen is a cytoplasmic protein, which under dark conditions binds specifically to photolysed and phosphorylated rhodopsin (R*-P) and completes the deactivation of R* and quenches its activation of transducin (Wilden *et al.*, 1986). Ultimately, all-*trans*-retinal is reduced to all-*trans* retinol by the enzyme retinal dehydrogenase present in the disk membranes of the outer segments. This loss of the chromophore is the final step in the quenching process, since the resulting phosphorylated opsin is incapable of binding transducin, rhodopsin kinase or arrestin.

The intrinsic GTPase activity of T_{α} subunit is another contributory factor for terminating the signal initiated by photolysed rhodopsin. Like other G-proteins T_{α} subunit is capable of hydrolysing the bound GTP to GDP and inactivate itself (Bourne *et al.*, 1990). As described earlier PDE remains active only until GTP associated with GTP- T_{α} * is hydrolysed by the intrinsic GTPase activity of T_{α} . Upon which T_{α} dissociates from PDE and re-associates with $T_{\beta\gamma}$ and the PDE γ subunits are released to form PDE complex, and once again inhibit enzyme activity.

Transducin may also be regulated by phosducin, which is a soluble 33kd phosphoprotein highly expressed in retinal photoreceptors and the pinealocytes of the pineal gland (Lolley *et al.*, 1992). Normally phosducin exists in photoreceptor cell in the form of phosducin/ $T_{\beta\gamma}$ complex and unlike transducin which is concentrated in the rod outer segment, phosducin/ $T_{\beta\gamma}$ complex is dispersed through out the cytosol of photoreceptor cells (Lee *et al.*, 1990). It has been suggested that phosducin/ $T_{\beta\gamma}$ complex may be involved in the direct regulation of transducin function via inhibition of the GTPase activity of T_{α} chain (Bauer *et al.*, 1992).

1.10 Inherited retinal degenerations

Hereditary retinal^{de}generations and dysfunctions are an extremely heterogeneous group of diseases in terms of clinical description and genetic cause. This term encompasses: diseases of the peripheral retina, such as retinitis pigmentosa and congenital stationary blindness; diseases of the central retina, such as macular degenerations (solely confined to the macula region) and those that eventually also lead to involvement of the peripheral retina; and many others in which the pattern of degeneration is complex. The following sections will be dedicated to the description of a few of these disorders with more emphasis given to retinitis pigmentosa (RP), the more relevant subgroup with regard to this research.

1.10.1 Retinitis pigmentosa

Retinitis pigmentosa (RP) describes a group of retinal degenerations characterised by night blindness (nyctalopia) and a gradual loss of peripheral visual field; with disease progression signs of macular (the central region of the retina) involvement is indicated by loss of reading acuity, abnormal colour vision and sensitivity to glare. Often in the final stage of disease, most patients become legally blind. RP is the most common amongst inherited retinal disorders affecting 1-2 in every 5000 births in the western world (Boughman *et al.*, 1980; Bundy and Crews, 1984).

Even though RP is progressive there is a variation in the age of onset and a wide range of severity. In a typical case the age of onset is in childhood; the disease progresses from night blindness to noticeable visual field loss by late twenties and then to severe visual disability with tunnel vision and night blindness in their forties. By the age of sixty or seventy the patient would have little or no functional vision (Heckenlively, 1988). Fundus examination of the retina late in the disease typically shows a pale optic nerve head, attenuated retinal vessels and bone-corporuscular-like (bone spicule) pigmentary deposits and depigmentation and atrophy of the retinal pigment epithelium (RPE) (see fig 3.1). The pigmentary deposits occur most prominently in the periphery of the retina and often in a perivascular pattern due to pigment within vessel walls. Despite the fact that the pigmentary deposition accounts for the disease name, the observed pigmentary pattern is not an exclusive feature of RP but is a feature also found in other acquired and hereditary retinal degenerations. Moreover it is still a secondary effect to the pathological process of RP.

However the term retinitis is a misnomer as there is little evidence of inflammation on histopathology (Heckenlively, 1988). In typical cases of RP the rod photoreceptors are more severely affected early in the disease and give rise to the classical early symptom of night blindness. Later in the disease cone photoreceptors degenerate as well. A progressive deterioration of retinal function is also revealed in electroretinographic analysis. The electroretinogram (ERG) is the compilation of electrical signals generated by the retina in response to a flash of light (fig 1.5). When a patch of the retina is illuminated the current generated from the photoreceptors, Müller and retinal bipolar cells flow in to the vitreous and can be recorded as an ERG outside the eye at the cornea using a contact lens electrode. A standard ERG is biphasic; the first negative downward peak called the 'a' wave represents the negative receptor potential mainly from repolarisation of photoreceptors. This wave potential falls and recovers slowly from rods while recovering rapidly from cones. The second positive wave of the ERG called the 'b' wave is a combination of potential generated from both Müller and retinal bipolar cells. In RP the electrical signals generated from rods in response to flashes of light are reduced in amplitude and delayed in response times, which may also be the case for cones. In advanced cases of RP all responses are extinguished.

The genetics of RP is complex. RP can be monogenic (caused by defects in single genes) or digenic (caused by two genes). Moreover, monogenic RP can be either syndromic, whereby other organs/tissues are also affected in addition to the retina, or non-syndromic with effects limited to the retina. Monogenic forms of RP can be autosomal dominant, autosomal recessive or X-linked. Genetic distribution of each form of RP varies in studies carried out in different populations (Boughman *et al.*, 1980; Bunker *et al.*, 1984; Bird, 1988). The complexity of RP is further accentuated by the fact that there is extensive non-allelic heterogeneity even among families that show the same mode of inheritance.

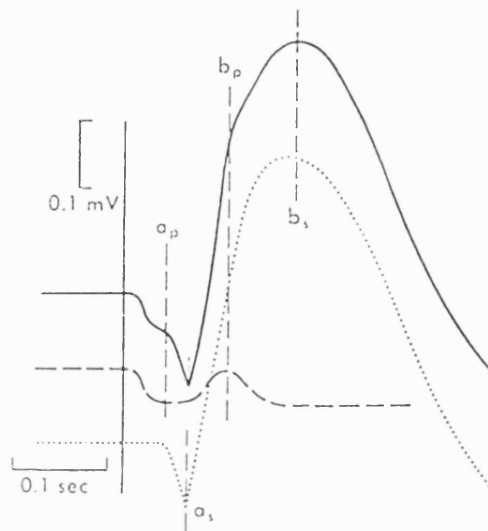


Figure 1.5. (figure and legend from Berson, 1992)

Analysis of ERG in dark-adapted human eye as the resultant of photopic (dashed line) and scotopic (dotted line) components. The a-wave is composed of photopic (a_p) and scotopic (a_s) components, and the b-wave is similarly composed of photopic (b_p) and scotopic (b_s) components.

1.10.1.1 Autosomal dominant RP (adRP)

Autosomal dominant retinitis pigmentosa (adRP) accounts for approximately 25% of all retinitis pigmentosa cases (Jay, 1982; Bundy and Crews, 1984) and is itself clinically and genetically heterogeneous. Based entirely on the clinical variation, adRP has been classified into distinct subgroups.

Clinical classification of adRP:

The most widely accepted clinical classification system of adRP discriminates between two main types based on differences in loss of rod and cone function. Type I of Massof and Finkelstein (1981) designates an early and diffuse loss of rod function followed by a progressing loss of cone function. In type II, there is a regionally progressing concomitant loss of both rod and cone function. Types I and II of the previous classification correspond well to the 'D' type and the 'R' type of Lyness *et al.* (1985). On ERG testing, D type patients have no measurable rod function, whereas R type patients usually have either normal or reduced rod signals with increased implicit peak times. Cone signals are also reduced in most D type cases and are either normal or reduced in most R type patients. Cone peak times may be normal or delayed in both types. In type I/D adRP the age of onset is quite early (with night blindness occurring before the age of 10), yet the pigmentary changes may not be evident until the second or the third decade of life and visual difficulties are usually identified 10-20 years after the onset of night blindness. In contrast for type II/R the onset of night blindness is more variable. While it could occur before the age of 10 the preponderance is towards late onset (i.e. after the age of 20). Moreover in type II/R onset of night blindness is closely followed by visual field loss (Bird, 1988). Some type II/R adRP families show variable disease severity within one pedigree (variable expressivity) and also incomplete penetrance of the disease, where some obligate disease carriers appear completely normal (Boughman *et al.*, 1980; Kaplan *et al.*, 1990; Moore *et al.*, 1993) (see chapter 3).

Sectoral RP is a third subgroup classified thus due to observation of retinal atrophy only in one part of the fundus (usually the nasal quadrant) with severe loss of visual field corresponding to that affected retinal area alone (Fishman *et al.*, 1985). Sectoral RP categorised thus has a better visual prognosis than other forms of adRP. Heckenlively *et al.* (1991) include in this subgroup progressive diseases affecting preferentially a sectorial part of the retina but eventually affecting the whole visual field.

Molecular genetics of adRP:

Rhodopsin adRP:

As mentioned earlier adRP shows evidence of high genetic heterogeneity. To date 10 loci have been identified through linkage analysis of adRP pedigrees. The first adRP locus to be identified was on chromosome 3q21 (RP4) (McWilliam *et al.*, 1989), which was soon followed by the identification of the first mutation in the rhodopsin gene (Pro-23-His), the most obvious positional candidate gene for adRP on 3q21 because of its abundance and primary role in phototransduction (Dryja *et al.*, 1990). The rhodopsin protein in man consists of 348 amino acid residues with seven transmembrane helices forming a binding pocket for the light sensitive ligand, 11-*cis*-retinal. Since the discovery of the first mutation, at least 88 different mutations (mostly missense) have been identified in patients with adRP and rhodopsin mutations are known to account for between 20 and 31% of all adRP cases in the western world (Sung *et al.*, 1991; Dryja *et al.*, 1991; Sheffield *et al.*, 1991; Inglehearn *et al.*, 1992; Gal *et al.*, 1997). Furthermore, type 1/D RP is relatively frequent in adRP patients that carry rhodopsin mutations and this encourages the molecular geneticist to search for rhodopsin mutations in families that has type 1 clinical presentation.

Rhodopsin mutations have been studied extensively to understand their biochemical consequences and relationship with disease severity. The deleterious nature of certain mutations, such as those predicting a truncated protein or a polypeptide with a considerably altered number and composition of amino acids (larger deletions or insertions) are obvious. As it is for those missense mutations which affect residues important for correct post-translational processing, transport and function of rhodopsin (reviewed in Gal *et al.*, 1997). However there have been

instances where the deleterious nature of the mutation is not as clear, indicating that these mutations affect presently unknown features or functions of rhodopsin.

In spite of understanding the negative affects of mutations, the reasons for their dominant nature and the associated pathogenic mechanism(s) have yet to be elucidated. The range of phenotypes associated with rhodopsin mutations suggests that a mutant rhodopsin may initiate different degenerative mechanisms. Haploinsufficiency is unlikely to be important in dominant rhodopsin disease since carriers of at least one apparently *null* allele (no protein produced) have been shown to be phenotypically normal (Rosenfield *et al.*, 1992). Instead most dominant rhodopsin mutants appear to be gain-of-function or dominant-negative alleles. From studies of dominant rhodopsin mutations in *Drosophila* (Colley *et al.*, 1995; Kurada and O'Tousa, 1995) it seems that degeneration (in insects at least) depends on co-expression of the wild type and mutant forms of rhodopsin and is due to a defect(s) interfering with post-translational maturation and proper intracellular transport of the wild type poly-peptide. In these mutant flies expansions of the endoplasmic reticulum were also observed which was interpreted as being due to excessive accumulation of unprocessed protein. To better understand the relationship between mutations and clinical phenotype in mammals the biochemical and physiological consequences of dominant mutations have been studied at cellular level. In one such approach Sung *et al.* (1993, 1994) introduced mutant rhodopsin gene to cultured cells (kidney cell lines) and examined translation, transport and functional activity of the resulting protein to relate the extent of abnormality observed in culture to the pathogenicity observed in photoreceptor cells *in vivo*. This method allowed mutant rhodopsin to be categorised in to three classes: those that remain in the cytoplasm and accumulate within rough endoplasmic reticulum not showing binding with the chromophore, 11-*cis*-retinal, those that are inserted into the plasma membrane but are non-functional, and those that are membrane associated and have wild-type activity. However, these mutant classes do not always correspond well to severity of the disease associated with the mutation, most probably due to differences between cells grown in culture versus cells grown *in vivo*. It was also suggested that even though in cultured cells the majority of rhodopsin mutants accumulated in the cytoplasm, the same mutants *in vivo* might interfere with intracellular rhodopsin transport or processing as observed in *Drosophila* mutants. To circumvent this problem, transgenic mice have been

constructed that carry rhodopsin mutations and indeed differences in abnormalities have been observed between the two experimental systems. For example Gln344ter mutant opsin in cell culture is transported to the plasma membrane and appears to have wild-type function whereas in the transgenic mice the mutant opsin accumulates in the cytoplasm followed by retinal degeneration (Sung *et al.*, 1994). In this model, congestion in the protein transport/maturation pathways could indeed be the primary trigger of photoreceptor cell death. However unlike in *Drosophila* the cytoplasmic accumulation of Gln344ter mutant opsin in mouse photoreceptors is not accompanied by impaired transport of the wild-type opsin, raising the possibility that there may be mechanistic differences between insects and mammals in the transport of newly synthesised opsin.

Rhodopsin comprises greater than 90% of outer segment protein (Stryer, 1988). Therefore rhodopsin may also have a role in maintaining the structural integrity of the photoreceptor besides functioning as the visual pigment. This is somewhat supported by the observation that mice with only a single functional rhodopsin allele exhibit disorganisation and progressive shortening of the outer segments (Humphries *et al.*, 1997). However this might as well be due to simple mass effect and not due to rhodopsin having any structural role, since rhodopsin accounts for a large proportion of the outer segment protein. The fact that mutations in peripherin/RDS (Farrar *et al.*, 1991; Kajiwara *et al.*, 1991), which encodes a structural protein of the outer segment discs, cause retinal degeneration also give support to the postulate that rhodopsin plays a structural role. Therefore certain missense mutations, which introduce charge or bulky amino acids into the transmembrane domain, or might change the shape of the molecule may cause membrane instability and lead to cell death. However these mutant opsin need to be expressed in the cytoplasmic membrane alongside the wild-type protein for such membrane destabilising factors to take effect.

Another pathogenic mechanism, a 'constant equivalent light' model (Fain and Lisman, 1993) has been proposed for the severe adRP phenotype associated with a rhodopsin codon 296 mutation. This hypothesis is based on the observation that exposing an animal to continuous light for long periods of time produce retinal degeneration leading to blindness (Noell *et al.*, 1966). Following this theme it has been proposed that certain molecular defects may produce electrical signals

'equivalent' to those produced by real light and trigger degeneration by the same mechanism. As lysine-296 is the site of attachment of the chromophore, 11-*cis*-retinal, to the apoprotein, opsin, its replacement by any other residue would render the protein incapable of forming a photopigment with retinal. Missense mutation in codon 296 has been shown to constitutively activate rhodopsin *in vitro* in the absence of chromophore (Robinson *et al.*, 1992) and lead to over stimulation of the phototransduction cascade: a situation similar to constant light exposure, which is known to lead to photoreceptor death. However this hypothesis has recently been challenged by the observation that the Lys296Glu opsin does not appear to continuously activate transducin *in vivo* in the transgenic mice carrying the mutated gene. Moreover, the mutated opsin is inactivated by phosphorylation and arrestin binding, therefore cell death must be mediated by a mechanism other than over stimulation (Li *et al.*, 1995a). Therefore in summary it can be said that even though the identification of various rhodopsin gene mutations as the primary cause of adRP has been very successful, the elucidation of the subsequent pathogenic mechanisms leading to photoreceptor degeneration is still incomplete and requires further study (see section 1.11).

Exemplifying the clinical diversity observed with rhodopsin, mutations in rhodopsin have also been implicated in autosomal recessive RP (Rosenfield *et al.*, 1992; Kumaramanickavel *et al.*, 1994) and autosomal dominant congenital stationary night blindness (CSNB) (Dryja *et al.*, 1993). The pathogenic mechanisms of recessive and CSNB rhodopsin mutations are discussed in detail in respective sections.

Peripherin/RDS adRP:

Once peripherin/*RDS* (retinal degeneration slow) (Travis *et al.*, 1989), the human homologue of the gene mutated in the *rds* mouse (see section 1.11), was localised to 6p21.1-cen, it became a candidate locus for human retinopathy. The subsequent linkage (Jordan *et al.*, 1992) and mutation analysis (Farrar *et al.*, 1991; Kajiwara *et al.*, 1991) in the linked adRP pedigrees implicated this gene in adRP (locus designated as RP6). Peripherin/*RDS* is a photoreceptor specific transmembrane protein expressed in the rim region of outer segment discs of both rods and cones, suggesting that it anchors the discs to the cytoskeleton of the disc and helps to maintain the rim curvature (Connell and Molday, 1990; Travis *et al.*, 1991). This protein is essential

for the biogenesis of photoreceptor outer segments, since mice homozygous for the *rds* allele fail to develop them. Several mutations have been identified in peripherin/*RDS* gene where the productions of aberrant disk structure lead to photoreceptor degeneration. Peripherin/*RDS* mutations account for 5% of total adRP cases (Wells *et al.*, 1992; Rosenfield and Dryja, 1995). However unlike rhodopsin, peripherin/*RDS* mutations have a much broader clinical spectrum and in addition to RP, peripherin/*RDS* has been implicated in a myriad of dominant central retinal dystrophies (see section 1.10.2), including cone-rod dystrophy (CRD), macular dystrophy, central areolar choroidal dystrophy and pattern dystrophy (<http://www.sph.uth.tmc.edu/RetNet/>). This is probably due to the fact that *RDS* is expressed in both rods and cones. The different clinical manifestations of mutations in peripherin/*RDS* gene suggest that certain portions of the protein are critical for rod function only (associated with RP), cone function only (central retinal degenerations) or both (reviewed in Daiger *et al.*, 1995). Other genetic and non-genetic influences, which modify the phenotype, have been proposed to explain the wide inter- and intra-familial variation in phenotype associated with particular *RDS* mutations (Weleber *et al.*, 1993).

adRP loci identified through linkage analysis:

In addition to the two genes described above further 7 adRP loci have been identified through linkage analysis. They are localised on chromosomes: 1 cen (RP18) (Xu *et al.*, 1996), 7p15-13 (RP9) (Inglehearn *et al.*, 1993), 7q31 (RP10) (Jordan *et al.*, 1993), 8q11 (RP1) (Blanton *et al.*, 1991), 17p (RP13) (Greenberg *et al.*, 1994), 17q (RP17) (Bardien *et al.*, 1995) and 19q (RP11) (Al-Magthteh *et al.*, 1994). Physical maps have been constructed (or are being constructed) for all the loci that have been assigned by linkage analysis to facilitate the positional cloning of the disease causative gene.

More recently another novel adRP locus was localised to chromosome 14q11 through linkage analysis in a single pedigree in our laboratory and a mutation that segregate with the disease was soon identified in the positional candidate gene *NRL* (Neural retina-specific leucine zipper gene) (personal comm.). Previously the *Nrl* protein was shown to interact with rhodopsin promoter and positively regulate rhodopsin gene expression, thus making it an attractive, and a plausible candidate for

adRP (Rehemtulla *et al.*, 1996). Furthermore *NRL* is also highly expressed throughout the neural retina and is possibly a continuous regulator of differentiation during development (Swaroop *et al.*, 1992; Yang-Feng and Swaroop, 1992;). Therefore in total 10 loci have been implicated in the causation of adRP, for which mutations have only been identified in three genes to date. The existence of further adRP families from which all the above mentioned loci have been excluded indicate for further genetic heterogeneity of adRP.

1.10.1.2 Autosomal recessive RP (arRP)

Clinical features:

Clinical features of arRP are very similar to those observed in adRP. However Pigmentation occurs late and increases at a slower rate than in adRP. Moreover loss of visual function is faster than adRP but slower than xLRP (Fishman, 1978; DeRouck *et al.*, 1986).

Molecular genetics of arRP:

Similar to adRP, arRP also exhibits genetic heterogeneity. In contrast to adRP there are many more causative genes identified for arRP than loci implicated through linkage analysis. This is mainly due to the fact that large recessive/consanguineous RP pedigrees are less easily accessible than large adRP pedigrees, to enable linkage analysis. Instead most of the recessive RP genes were identified through screening of candidate genes, which encode integral proteins of the phototransduction cascade. The advantage of candidate gene screening is that the study of large families is not necessary, and samples from numerous unrelated affected individuals can be tested quickly, therefore being highly appropriate for identifying recessive RP genes. In addition to the 8 identified genes (table 1.x.) 3 other arRP loci have been implicated through linkage analysis in large arRP families. Autosomal recessive account for 20% of all cases of RP, whilst sporadic RP, which is presumed to be recessive in most cases, accounts for a further 50% (Jay, 1982).

Rhodopsin arRP:

Two rhodopsin alleles (Glu150Lys and Glu249Ter) leading to arRP have been reported to date (Kumaramanickavel *et al.*, 1994; Rosenfield *et al.*, 1992). The pathogenic nature of the Glu249Ter can readily be explained by the fact that it results in a functionally inactive rhodopsin, which is missing the sixth and the seventh transmembrane domains including the 11-*cis*-retinal attachment site. In contrast the pathogenic nature of Glu150Lys cannot be explained as easily, and is also a rare example of a non-conservative amino acid exchange in rhodopsin, which is not deleterious in the heterozygous stage. It has been suggested that Glu150Lys mutants are unstable to interact with and activate transducin (Franke *et al.*, 1992), but that a 50% reduction in transducin activation is tolerable hence the reason for having no pathogenic consequences in the heterozygotes.

PDEA and PDEB arRP:

Rod cGMP phosphodiesterase (PDE) holoenzyme, which plays an integral role in the phototransduction cascade, consist of two large subunits α and β that share considerable homology to each other, and two small, identical γ subunits (see section 1.8). Many nonsense and missense mutations of the α and β subunit encoding genes (*PDEA* and *PDEB*) have been associated with arRP, the former mutations leading to *null* alleles (McLaughlin *et al.*, 1993; Huang *et al.*, 1995). Most *PDEB* mutations associated with arRP are located to the C-terminal half of the protein containing the catalytic domain and therefore directly affect the enzymatic activity. The heterozygous carriers of these mutations are normal indicating that half the amount of normal product is sufficient to prevent disease. Homozygous nonsense mutations in the murine and canine homologues of *PDEB*, producing *null* alleles, have also been found to cause retinal degeneration in *rd* mice (Bowes *et al.*, 1990; Danciger *et al.*, 1990) and Irish setter dogs (*rcd1*) (Farber *et al.*, 1992; Suber *et al.*, 1993) (see section 1.11). A missense mutation in *PDEB* have been identified in a pedigree with autosomal dominant congenital stationary night blindness (CSNB) (Gal *et al.*, 1994) (section 1.10.1.4). As yet no mutations in the PDE γ subunit gene have been associated with human retinal disease (Hahn *et al.*, 1994).

CNCG1 arRP:

The cGMP-gated cation channel protein, which resides on the rod outer segment plasma membrane, is composed of an α subunit that is functional by itself (encoded by *CNCG1* gene), a β subunit of similar primary structure that modulates the responsiveness of the channel to cGMP by interacting with calmodulin (section 1.9.1) and a γ subunit of unknown function that complexes with the β subunit. In the process of screening genes encoding components of the phototransduction pathway, which represent good candidate genes for inherited retinopathy, mutations in the *CNCG1* gene were identified in a few families with arRP (Dryja *et al.*, 1995). All of the identified mutations are either obviously *null* or encode a channel protein that functions poorly. ‘Equivalent light hypothesis’ has also been suggested as the pathogenic mechanism associated with *CNCG1* mutations (Lisman and Fain, 1995). This was postulated on the argument that a membrane absent of channels, or in which the channels were nonfunctional, would be hyperpolarised, just as if the rods had been exposed to steady bright light.

ABCR arRP:

The arRP locus on 1p21-p13 was first found by linkage analysis in a Spanish consanguineous arRP family (Martinez-Mir *et al.*, 1997) and the positional candidate gene *GNAT2* (α subunit of cone transducin) was already excluded prior to identification of a mutation in the *ABCR* gene, which encodes a novel retina-specific ATP-binding cassette transporter protein (Martinez-Mir *et al.*, 1998). The homozygous mutation (1 bp deletion) generates a frameshift early in the coding region that adds 32 new residues and a premature stop codon, and produce a putative *ABCR null* allele (Martinez-Mir *et al.*, 1998). ABCR is a member of the ABC transporter super family and is located exclusively in the disc membranes of rod outer segments (Sun and Nathans, 1997). The restricted pattern of ABCR expression supports a specific role in the ROS. It has been suggested that *ABCR* might be involved in storage or release of intracellular calcium, in interaction with the cGMP-gated channels or in the retinoid transport (Azarian and Travis, 1997). It has been suggested that mutations might lead to accumulation of debris by affecting RPE/photoreceptor recycling of retinoids or lipids.

Mutations in this gene are also associated with recessive Stargardt's disease (STGD)/fundus flavimaculatus (FFM) and age-related macular degeneration (AMD), all of which represent macular dystrophies (see section 1.10.2) (Allikmets *et al.*, 1997a and 1997b). Recently *ABCR* gene mutations were also identified in a consanguineous family with individuals showing either recessive RP or cone-rod dystrophy (Cremers *et al.*, 1998). The RP patients were homozygous for 5' splice site mutation, IVS30+G→T and the CRD patients were compound heterozygotes for the IVS30+G→T mutation and a 5' splice site mutation in intron 40 (IVS40+5G→A). The authors hypothesise that the severe phenotype in RP patients is due to the complete absence of ABCR protein and that the intron 30 splice site mutation represents a true *null* allele. Given the more severe clinical picture of RP patients compared with CRD and STGD patients, they suggest that every combination of two *null* alleles in the *ABCR* gene results in autosomal recessive RP (supported by data of Martinez-Mir *et al.*, 1998) and that a combination of an *ABCR null* allele with a mutation yielding residual ABCR function can result in CRD. Similarly in all STGD patients characterised to date at least one of the two mutated *ABCR* alleles encodes a protein that could be directed to the membrane and retain some function. Therefore rods in STGD patients may remain partly active, but the disease process leads to the accumulation of lipofuscin-like material observed in the underlying RPE (a characteristic feature of Stargardt's disease). Lipofuscin is a complex mixture of lipid, protein and often pigment which characteristically autofluoresces. The molecular basis for the variable phenotypes associated with *ABCR* mutations remains to be elucidated, and requires further functional studies at the cellular level.

RLBP1 arRP:

The gene, *RLBP1*, which encodes the cellular retinaldehyde-binding protein (CRALBP), was recently found to be responsible for arRP in a consanguineous Indian family, which showed linkage to chromosome 15q26 (Maw *et al.*, 1997). Interestingly prior to *RLBP1*, which is expressed in the RPE and Müller cells of the neural retina, the expression of all the genes implicated in RP was in, or was confined to, the photoreceptors of the neural retina (also see *RPE65* arRP). CRALBP plays an important part in the process in which all-*trans*-retinaldehyde, released during the breakdown of activated visual pigment in photoreceptors, is enzymatically processed

into 11-*cis*-retinaldehyde and returned to the photoreceptors for visual pigment regeneration (Saari, 1994) (section 1.7.2). CRALBP binds and promotes 11-*cis*-retinol oxidation to 11-*cis*-retinaldehyde by 11-*cis*-retinol dehydrogenase in the RPE. CRALBP also retards the esterification of 11-*cis*-retinol to retinyl ester thus promoting the transport of 11-*cis*-retinaldehyde into the photoreceptors rather than its storage in the RPE (Saari *et al.*, 1994). The homozygous mutation found in *RLBP1* was found to produce a protein which lacked the ability to bind 11-*cis*-retinaldehyde *in vitro*. Lack of functional CRALBP would disrupt the retinal vitamin A cycle and destabilise rod and cone opsins (which account for 90% of outer segment protein) leading to slow retinal degeneration and arRP. Alternatively, lack of visual pigment may lead to constant depolarisation of the outer segment membrane thus exhausting the photoreceptor by a high energy demand and calcium load, which would also result in retinal degeneration (Wright, 1997a). These conjectures have to be tested by further studies, possibly by using a CRALBP knock-out mouse model.

RPE65 arRP:

Loss-of-function mutations in *RPE65* have been implicated in autosomal recessive childhood onset severe retinal dystrophy (arCSR) (Gu *et al.*, 1997). arCSR can only be loosely categorised as a recessive RP, it rather designates a heterogeneous group of disorders affecting rod and cone photoreceptors simultaneously where the most severe cases are termed Lebers congenital amaurosis (LCA) (Heckenlively and Foxmann, 1988). RPE65 protein is exclusively and abundantly expressed in the RPE, where it comprises about 10% of total membrane protein (Hamel *et al.*, 1993). Even though the actual function of RPE65 is yet unknown, because its initial expression coincides with the appearance of rod outer segments, RPE65 could have a function in the visual cycle. Moreover it has been speculated that *RPE65* may actually encode retinol isomerase, the protein that converts all-trans retinyl ester to 11-*cis*-retinol (Crouch *et al.*, 1997). Two *null* mutations in *RPE65* have also been identified in a family with autosomal recessive Leber's congenital amaurosis (LCA) (Marlhens *et al.*, 1997).

TULP1 arRP:

TULP1, expressed exclusively in the retina, is a member of a family of tubby-like genes (*TULPs*) and co-localises with the arRP locus (RP14) mapped to chromosome 6p21.3 (Knowles *et al.*, 1994; Shugart *et al.*, 1995). *TULP1* was considered as a good candidate gene for RP in humans because of its homology with the murine *tub* gene, recessive mutation in which causes obesity, deafness and retinal degeneration in tubby mice (Coleman and Eicher, 1990). Subsequent mutation screening of *TULP1* in two Dominican arRP kindreds revealed one homozygous splice site mutation (IVS14+1G→A), which presumably renders the donor splice site dysfunctional and alters the C-terminal most exon of *TULP1* (Banerjee *et al.*, 1998). Interestingly a mutation at the identical donor splice site of the mouse *tub* gene produces an aberrant mRNA transcript from intron-retention resulting in phenotype that includes progressive retinal degeneration (Kleyn *et al.*, 1996; Noben-Trauth *et al.*, 1996). Unlike the *tubby* mice, none of the RP patients show hearing impairment or overt obesity, which is consistent with the absence of *TULP1* expression outside the retina. Three other missense mutations (Arg420Pro, Ile459Lys and Phe491Leu) which segregate with arRP have also been reported (Hagstrom *et al.*, 1998). These mutations affect residues near the C terminus of the protein that is conserved in all reported members of the *TULP* family suggesting an important functional region (North *et al.*, 1997). The exact function of *TULP1* gene is unknown but must play an essential role in physiology of the retina.

Three other arRP loci have also been identified at chromosome 1q32.1-31 (van Soest *et al.*, 1994), 2q33-q31 (Bayes *et al.*, 1998) and 16p12.3-p12.1 (Finckh *et al.*, 1998). Of these three loci, only linkage to the 1q33-q31 locus (RP12) has been reported in more than one pedigree. The 8 known recessive genes also only account for a small proportion of arRP cases. Therefore it is likely that 80-90% of recessive RP cases are due to defects in as yet unidentified genes.

Chromosomal Location	Gene (if known)	Allelic retinopathies	Reference(s)
1p31	<i>RPE65</i>	arRP LCA	Gu <i>et al.</i> , 1997
1p21-p13	<i>ABCR</i>	arRP Stargardt's disease (STGD1)	Martinez-Mir <i>et al.</i> , 1998
1q32.1-31	-	arRP (RP12)	Van-Soest <i>et al.</i> , 1994
2q33-q31	-	arRP	Bayes <i>et al.</i> , 1998
3q21	<i>RHO</i>	adRP adCSNB arRP	Rosenfield <i>et al.</i> , 1992
4p16.3	<i>PDEB</i>	arRP adCSNB	McLaughlin <i>et al.</i> , 1995
4p14-q13	<i>CNCG1</i>	arRP	Dryja <i>et al.</i> , 1995
5q31.2-ter	<i>PDEA</i>	arRP	Huang <i>et al.</i> , 1995
6p21.3	<i>TULP1</i>	arRP	Banerjee <i>et al.</i> , 1998; Hagstrom <i>et al.</i> , 1998
15q26	<i>RLBP1</i>	arRP	Maw <i>et al.</i> , 1997
16p12.3-p12.1	-	arRP	Finckh <i>et al.</i> , 1998

Table 1.2

Cloned or mapped genes causing autosomal recessive retinitis pigmentosa

1.10.1.3 X-linked RP (xLRP)

X-linked form of RP is the most severe in terms of onset and progression, typically presenting in the first decade of life and progressing to partial or complete blindness by the third or the fourth decade of life (Heckenlively, 1988). X linked inheritance can be determined by pedigree analysis, as the males are severely affected and there is no male to male transmission of the disease. Carrier females also show varying degrees of clinical manifestation. xLRP makes up more than 15% of total RP cases and may account for as much as 20% if a proportion of simplex cases are included (Jay, 1982). Like other types of RP described so far, xLRP is both clinically and genetically heterogeneous.

A xLRP locus was among the first to be mapped when polymorphic markers became available for linkage analysis in the early 1980's (Bhattacharya *et al.*, 1984). Subsequent linkage analysis on this locus distinguished two xLRP loci on the short arm of X chromosome rather than one (Wright *et al.*, 1987). RP2 is localised within Xp11.3-11.22 and RP3 is localised to Xp21.1. A gene, *RPGR*, responsible for RP3 has been positionally cloned through identification of microdeletions within this region (Meindl *et al.*, 1996). In fact *RPGR* was the first RP gene to be mapped by positional cloning since all the other RP genes were pre-characterised positional candidates. *RPGR* (retinitis pigmentosa GTPase regulator), which is expressed ubiquitously encodes a 90 kD protein that has a tandem repeat structure, highly similar to RCC1 (regulator of chromosome condensation) protein, in its N-terminal half, suggesting an interaction with small GTPase. The C-terminal of the protein shows means of potential membrane anchorage (Meindl *et al.*, 1996). The actual function of *RPGR* is yet unknown but its suggested function is as a regulator of a specific type of membrane transport or trafficking in which the retina or RPE is particularly active. The *RPGR* mutations identified to date are all located in the highly conserved residues, suggesting that loss of function of *RPGR* is responsible for RP3. Mutation analysis in RP3 xLRP cases has revealed that *RPGR* only account for a portion of RP3 cases, raising the possibility of alternative transcripts or yet undiscovered portion of the gene. The other possibility is the existence of another RP3 gene i.e. microheterogeneity within the RP3 locus.

1.10.1.4 Congenital stationary night blindness (CSNB)

Congenital stationary night blindness (CSNB) is a clinically and genetically heterogeneous disorder characterised by disrupted (or absent) night vision from early infancy. Although both diseases feature night blindness as a prominent symptom, CSNB differs from RP in that there is no apparent degeneration of rod or cone photoreceptors. Instead CSNB patients retain their daytime vision throughout life. In summary night blindness in CSNB is considered a pure functional defect, non-progressive, and is not accompanied by retinal degeneration. Moreover, Autosomal dominant, recessive and X-linked forms of CSNB have been reported.

The first genetic defect associated with CSNB was found in the rhodopsin gene in a single patient who carried a heterozygous missense mutation (Ala292Glu) (Dryja *et al.*, 1993) and the pathogenic mechanism was suggested as the constitutive activation of the phototransduction cascade in the absence of light. When coupled with the chromophore Ala292Glu opsin resembles wild type rhodopsin in all respects, however, Ala292Glu opsin (without 11-*cis*-retinal) is still capable of activating transducin. In a normal individual in the dark, wild type rhodopsin molecules in the ROS occasionally lose their chromophore moieties without stimulating the phototransduction cascade (Defoe and Bok, 1983). Therefore in CSNB patients, the same physiological chromophore turn over in the dark could cause mutant Ala292Glu opsin molecules to activate transducin. Consequently even in the dark when the rods are in their dark adapted state there will be a subset of opsin molecules that would mimic the effect of a steady background light, which would reduce the sensitivity of rods thus leading to the phenotype of CSNB. In fact steady light that photoisomerises only a few hundred rhodopsin molecules per second in each rod of a wild type retina is known to greatly reduce the sensitivity of rods (Baylor *et al.*, 1984).

The second autosomal dominant locus of CSNB was assigned to chromosome 4p16.3 by linkage analysis in a large Danish family and the disease causative mutation (His258Asp) was identified in the gene *PDEB* which encodes the β -subunit of rod photoreceptor cGMP-specific phosphodiesterase (β PDE) (Gal *et al.*, 1994). This mutation, which is located adjacent to the β PDE domain that interacts with the inhibitory PDE γ chain, is thought to affect binding with PDE γ and inhibit

inactivation of PDE. The authors hypothesised that the His258Asp mutation in β PDE impedes complete inactivation of PDE in dark-adapted photoreceptors causing the rod photoreceptors to be permanently hyperpolarised and unable to return to the resting (dark) state. The result is inability to transduce at low light levels, and night blindness.

A heterozygous missense mutation (Gly38Asp) in the gene *GNAT1*, localised on chromosome 3p21, encoding the α subunit of rod transducin has also been found to cause autosomal dominant CSNB in a single large French family (Nougaret kindred) (Dryja *et al.*, 1996). Gly38 is a highly conserved residue among heteromeric G proteins, and the predicted consequence of this mutation is also to produce a defective α transducin that constitutively activates the phototransduction cascade leading to desensitisation of rod photoreceptors.

Oguchi disease is a rare autosomal recessive form of congenital stationary night blindness. A typical feature of this disorder is a golden or grey white colouration of the fundus that occurs as the retina adapts to light, called the Mizuo phenomenon. Homozygous *null* mutations of both arrestin (S-antigen) (Fuchs *et al.*, 1995) and rhodopsin kinase (Yamamoto *et al.*, 1997) have been associated with Oguchi disease. As both arrestin and rhodopsin kinase play a role in quenching the phototransduction cascade their deficiency could result in prolonged activation of rod photoreceptors leading to their inability to perceive light signals at low light levels. Therefore the underlying pathogenic mechanism is the same as described for rhodopsin, β PDE and α transducin associated CSNB i.e. constitutive activation of phototransduction cascade in the absence of light. In a patient with Oguchi disease, without functional RK or arrestin, photolysed rhodopsin remains active for much longer than normal until it regains 11-*cis*-retinal as its chromophore. However the other forms of CSNB differ from Oguchi disease in that affected individuals never recover rod sensitivity even after prolonged dark adaptation as there is no mechanism by which the phototransduction cascade adjusts to continuous activation by the mutant protein.

X linked CSNB also exemplifies the genetic and allelic heterogeneity associated with retinal diseases. Four X-linked CSNB loci that may be distinct or allelic with other co-localised retinal degenerations have been identified. The first locus, CSNB1, (Bech-Hansen *et al.*, 1992) is located on Xp11.3-11.22 and could be allelic with the RP2 locus. Another CSNB locus (CSNB2) has been mapped to the same locality as CSNB1, and may very well represent another distinct locus (Musarella *et al.*, 1989). CSNB3 locus was proved as being allelic with the RP3, since a mutation in the *RPGR* gene that segregates in CSNB patients has been identified (Hermann *et al.*, 1996). Recently another novel CSNB locus (CSNB4) was localised to Xp11.4 (Hardcastle *et al.*, 1997).

1.10.1.5 Digenic RP

Digenic RP is a retinal dystrophy caused by non-complementation of heterozygous mutations of two different and unlinked genes (Kajiwara *et al.*, 1994). The presence of either mutation without the other is not pathogenic. Only one example of digenic RP has been reported so far, namely the Leu185Pro mutation of peripherin/*RDS* and a null allele of *ROM1* gene (Kajiwara *et al.*, 1994). *ROM1* (rod outer segment membrane protein 1) encodes an integral membrane protein that is abundant at the rim region of the outer segment disk and which is very similar to peripherin/*RDS*. In fact it was found that both *ROM1* and peripherin form covalently linked homodimers, and that two homodimers non-covalently interact to form a tetrameric complex (Bascom *et al.*, 1992). Even though the precise function of *ROM1* is still obscure the severe RP phenotype associated with mutations in the two genes indicate an important structural relationship between peripherin and *ROM1* in the architecture of the rod photoreceptor.

1.10.1.6 Syndromic RP

Syndromic RP describes a subset of RP where retinitis pigmentosa is invariably associated with other disease manifestations. In the Usher syndrome, hearing loss accompanies RP. In Bardet-Biedl syndrome, RP is concurrent with polydactyly, obesity, hypogonadism and mental retardation, and in Kearns-Sayre syndrome RP is associated with cardiac conduction defects.

1.10.2 Central retinal dystrophies

Central retinal dystrophies can be subdivided into those that are solely confined to the macular region, termed macular dystrophies and those that eventually lead to involvement of the peripheral retina. Examples of the latter group are cone dystrophy and cone-rod dystrophy. Some well-known examples of macular dystrophies are age related macular dystrophy (AMD), which is the most common cause of legal blindness in older patients in the developed countries, Stargardt's disease (fundus flavimaculatus), Sorsby's fundus dystrophy, North Carolina macular dystrophy, and Best's vitelliform macular dystrophy.

Central retinal dystrophies are characterised by loss of central vision (loss of visual acuity) and degeneration of the retinal pigment epithelium underlying the retina. In this group of retinopathies peripheral vision is either present indefinitely or retained long term. As with retinitis pigmentosa, central retinal dystrophies are genetically heterogeneous with autosomal dominant, recessive and X-linked inheritance patterns observed. Moreover in most subgroups there is extensive non-allelic heterogeneity even among families that show the same mode of inheritance.

1.10.2.1 Cone-rod dystrophy

Cone-rod dystrophy is a severe form of chorioretinal dystrophy characterised by loss of colour vision and visual acuity followed by night blindness and peripheral visual field loss with widespread advancing retinal pigmentation and chorioretinal atrophy of the central and peripheral retina (Moore, 1992). Autosomal dominant, X-linked and recessive modes of inheritance have been described for CRD.

Autosomal dominant forms have been associated with mutations in the *peripherin/RDS* gene on chromosome 6p21.2-cen (Nakazawa *et al.*, 1994; Nakazawa *et al.*, 1996), in the *CRX* gene on chromosome 19q13.3-q13.4 (Evans *et al.*, 1994; Gregory *et al.*, 1994; Freund *et al.*, 1997) and in the *GUCD2* gene on chromosome 17p (Kelsell *et al.*, 1997; Kelsell *et al.*, 1998).

The CRD gene on 19q13.3, *CRX*, is an *OTX*-like homeobox gene. The CRX homeodomain transcription factor regulates the expression of photoreceptor-specific genes such as rhodopsin, cone opsins, interphotoreceptor retinoid binding protein (IRBP), β -PDE, and arrestin, and in a dominant-negative form blocks photoreceptor morphogenesis (Furukawa *et al.*, 1997). The mutations identified to date include one missense mutation (Glu80Ala) and a 1-bp deletion (Glu168 [Δ 1bp]) causing a frameshift. The authors have speculated that CRD patients carrying either a Glu80Ala or (Glu168 [Δ 1bp]) *CRX* allele are most likely to have an overall reduction of *CRX* gene function, but that the loss-of-function will be greater than 50% if these alleles are dominant negative. This proposition is based on the premise that dominant negative alleles would bind to the target sequences and thus obstruct binding of normal CRX protein and other components of the transcription complex. The mechanisms by which these mutations cause premature death of photoreceptors are not known. Nevertheless it can be argued that the mutation induced reduction of CRX function would disrupt the turnover of photoreceptor outer segment discs and that this process may lead over time to the complete loss of outer segments and cell death.

Mutations in the *CRX* gene have also been associated with Leber's congenital amaurosis (LCA) (Freund *et al.*, 1998). Since *CRX* is required for biogenesis of the outer segments it is appropriate that this gene be mutated in LCA which is essentially a photoreceptor developmental defect. LCA is a clinically heterogeneous group of childhood retinal degenerations in which the affected infants have little or no retinal photoreceptor function from early infancy (see *RPE65* arRP in section 1.10.1.2).

The third autosomal dominant CRD gene to be identified was *GUCD2*, which encodes retinal guanylate cyclase (RetGC1). Retinal guanylate cyclase restores the level of cGMP to the dark levels by converting GTP to 3', 5' cGMP and is an essential component in the recovery process of photoreceptors (see section 1.9.1). The fact that cone photoreceptors are affected first in CRD is consistent with the observation that retinal guanylate cyclase is predominantly found in the cone outer segments of the retina (Polans *et al.*, 1996). Similar to *CRX*, mutations in *GUCD2* have also been identified in LCA where disease has been ascribed to an impaired production of cGMP in the retina (Perrault *et al.*, 1996).

Autosomal recessive forms of cone rod dystrophy in association with Bardet-Biedl syndrome have been linked, by homozygosity mapping to chromosome 3 (Sheffield *et al.*, 1994), 11q13 (Leppert *et al.*, 1994), 15 (Carmi *et al.*, 1995) and 16q21 (Kwitek-Black *et al.*, 1993). Recently mutations in the *ABCR* gene on chromosome 1p21-p13 have also been identified in a pedigree segregating autosomal recessive cone-rod dystrophy (see *ABCR* arRP in section 1.10.1.2). In addition to this, two sporadic cases of cone-rod dystrophy have been reported; one in association with a cytogenetically visible deletion of chromosome 18q21.1-q21.3 (Warburg *et al.*, 1991) and the other in association with neurofibromatosis, suggestive of a cone-rod dystrophy gene situated close to the *NF1* gene on chromosome 17q11.2 (Kylstra and Aylsworth, 1993).

1.10.2.2 Cone dystrophy

The cone dystrophies are characterised by progressive degeneration of the cone photoreceptors with preservation of rod function. Affected individuals suffer from photophobia, loss of visual acuity, colour vision and central visual field (Weleber and Eisner, 1988). In the early stage of disease electrodiagnostic tests are required to distinguish cone dystrophy from cone rod dystrophies and macular dystrophies.

Cone dystrophies are genetically heterogeneous with autosomal dominant cone dystrophy loci localised on 17p12-p13 (Small *et al.*, 1996; Balciuniene *et al.*, 1995), on 6q25-q26 as identified by deletion mapping (McKusick, 1992), and on 6p21.1 (Payne *et al.*, 1998). X-linked cone dystrophy locus (COD1) has also been located on Xp11.4-p11.3 (Dash-Modi *et al.*, 1996) and pedigrees with autosomal recessive cone dystrophy have also been reported but no loci have yet been assigned by linkage analysis (Krill *et al.*, 1973).

The gene for the adCD locus on 6p21.1 has been identified as the guanylate cyclase activator 1A (*GUCA1A*) (Payne *et al.*, 1998). Guanylate cyclase activator 1A (GCAP1) is an important regulatory component of the phototransduction cascade, necessary for restoration of photoreceptors back to the dark state following activation (see section 1.9.1). The disease mechanism associated with GCAP1 mutation (Tyr99Cys) in cone dystrophy is not known. However the mutation is predicted to lead to an aberrant change in the concentration of cGMP, either high or low,

depending on whether the mutation causes GCAP to be permanently activated or lose functionality altogether. The fact that Tyr99Cys mutation only affects the cone cells suggests that this mutation is more deleterious in cone cells rather than in rods or that GCAP1 is more important to cones rather than rods. The second possibility is supported by the fact that GCAP1 has a greater expression in the cones than in rods (Polans *et al.*, 1996).

1.11 Animal models of retinal degeneration

Along with strategies for cloning disease genes, techniques for manipulating the mammalian genome have also been refined, such that the modification of the mouse genome via transgenic technology has become routine practise after a disease gene has been cloned. Animal models mimicking human disease are powerful tools and can be used, to understand the functions of a disease gene in normal animals, to develop theories concerning pathological processes induced by mutations in the particular gene that has been disrupted and to test therapeutic options, especially gene therapy approaches. Naturally occurring animal models and those created by transgenic technology are used for just such purposes, and described below are some of the best-characterised animal models relevant to the study of retinal degenerations.

1.11.1 *rd* Mouse

The *rd* mouse is a naturally occurring autosomal recessive animal model for retinal degeneration that has been studied extensively. In the *rd* mouse the photoreceptors start degenerating after the second week of life and virtually all rod cells disappear by postnatal day 20. Even though cone photoreceptors survive to this stage, cones also degenerate later but at a slower rate than the rods. Before the onset of cell degeneration elevated levels of cGMP are detected in the *rd* retina, which is followed by a steep decline in cGMP levels to barely detectable levels when all photoreceptors have disappeared. The initial rise in cGMP correlates with a deficiency in rod specific cGMP PDE activity. Linkage analysis followed by mutation screening established the mouse β -PDE gene (*pdeb*) as the *rd* locus (reviewed by Farber, 1995). The mutation in the *rd* mouse is recessive, owing to a premature stop codon and an insertion of viral DNA in the *pdeb* gene, which results in no enzyme production (Bowes *et al.*, 1990; Pittler and Baehr, 1991; Farber, 1995). Similarly a homozygous nonsense mutation identified in the canine homolog of rod cGMP-PDEB

was found to be the cause of the rod/cone dysplasia type 1 (*rcd 1*) in Irish setter dogs (Suber *et al.*, 1993). The *rd* mouse is a good model for a subset of autosomal recessive RP caused by null mutations in the *PDEB* gene (see section 1.10.1.2). The increased level of cGMP has now been proposed to trigger photoreceptor death, as prior onset of cell degeneration elevated levels of cGMP are detected in the *rd* retina.

1.11.2 *rds* Mouse

The retinal degeneration slow (*rds*) mutation is phenotypically characterised by abnormal development of retinal photoreceptors followed by their slow degeneration without any of the other cell types of the retina being affected (Van Nie *et al.*, 1978). This is also a naturally occurring animal model. In *rds/rds* homozygotes the retina undergoes normal development and differentiation until the first postnatal week when the photoreceptors normally appear. While the other retinal cells continue their normal development, *rds/rds* fail to form outer segment discs even though the inner segments, including the ciliary projections, appear morphologically normal (Sanyal, 1987; Cohen, 1983). The process of photoreceptor degeneration in *rds/rds* mice, which affect both rods and cones, is first detected histologically in the third postnatal week. The rate of photoreceptor degeneration escalates then becomes more gradual with significantly reduced thickness of the outer nuclear layer; by one year of age the degeneration is complete (Sanyal, 1987). The defect in *rds/rds* mice is a pure structural defect as all the components of the phototransduction cascade are present even if at greatly reduced levels (Cohen, 1983; Reuter and Sanyal, 1984). The *rds* mutation is not recessive as originally thought since the *rds/+* heterozygote mice also exhibit mild phenotypic abnormality. In contrast to homozygotes, heterozygotes do form outer segments, which are reduced in length and contain irregularly arranged discs that appear swollen and vacuolated with very slow degeneration (Sanyal, 1987). The phenotype observed in the *rds* mouse is caused by an insertion mutation that disrupts the gene encoding the *rds*/peripherin protein, which produces a *null* allele (Travis *et al.*, 1989, 1991). Peripherin is a photoreceptor specific transmembrane protein expressed in the rim region of outer segment discs of both rods and cones. The phenotype in *rds/rds* mice show that peripherin is essential for biogenesis of photoreceptor outer segments. Furthermore due to the phenotype in *rds* mice, the production of aberrant disc structure has been proposed as means of triggering photoreceptor degeneration.

1.11.3 *rho* knockout Mouse:

Humphries and co-workers (1997) recently generated mice carrying a targeted disruption of the rhodopsin gene (*Rho*^{-/-}). These mice do not develop rod outer segments or show any ERG response after 8 weeks but lose their photoreceptors over 3 months. The heterozygous *Rho*^{+/-} mice retain majority of their photoreceptors although the inner and outer segments of these cells display some structural disorganisation, the outer segment becoming shorter in old mice. Therefore the *rho* knockout mouse appears as a good animal model for autosomal recessive RP caused by *null* mutations in the rhodopsin gene.

The *rho* mice can also be used for a myriad of experiments, which include somatic gene therapy to see if the *null* phenotype can be rescued using one of the recombinant viral delivery systems. Success of such experiments will pave for similar therapeutic intervention in the cognate human disease. Other experiments relate to the creation of transgenic mice that mimic human dominant RP disease. Previously when transgenic mice were created to develop theories concerning the pathological processes induced by certain rhodopsin mutations, such as Pro23His (Roof *et al.*, 1994), Gln344Ter (Sung *et al.*, 1994) and Pro347Ser (Li *et al.*, 1996), by necessity the mutant transgenes were placed on a wild type genetic background. Now these transgenes can be placed in the *rho* ^{+/-} mice background to generate more faithful animal models of dominant RP since humans with rhodopsin mediated dominant RP are also hemizygous for the wild type allele. Moreover the knockout mouse presents an opportunity to study the effects of mutations, especially those mutations that affect the post-translational modification of rhodopsin, without the confounding presence of wild-type rhodopsin. Such experiment carried out *in vivo* might also lead to the recognition of other proteins that interact with rhodopsin for the correct folding of rhodopsin, which by virtue of their function be candidate genes for RP. Moreover, experiments carried out on the *null* background can clarify some post-translational anomalies observed with certain rhodopsin mutations in different experimental systems (see rhodopsin adRP in section 1.10.1.1).

1.11.4 Apoptosis: a final common pathway of photoreceptor cell death

The mouse models described earlier (*rd*, *rds* and rhodopsin transgenic mice [either Pro347Ser or Gln344ter]) have also been used in two independent studies to understand the pathway leading from the primary defect (i.e. mutation in gene) to photoreceptor cell death (Chang *et al.*, 1993; Portera-Calliau *et al.*, 1994). It was observed that even though the animal models represented different basic mutations, subsequent cell death was remarkably similar and bore all the biochemical hallmarks of death by apoptosis such as cytoplasmic condensation, nuclear chromatin condensation and inter-nucleosomal DNA fragmentation seen in agarose gels. Cleavage of DNA that link multiple nucleosomes gives rise to a DNA ladder, composed of fragments that are multiples of 180-200 bp and diagnostic for cell death by apoptosis. Apoptotic cell death is distinct from cell death by necrosis or accidental cell death that produce a spectrum of DNA fragment sizes without evidence of a nucleosomal ladder (Collins *et al.*, 1992). Apoptosis is normally used by retinal cells during development to fine-tune the number of cells in the retina and their interconnections (Young, 1984).

Even though apoptosis has been recognised as the final pathway the mechanisms by which each mutation trigger apoptosis is still obscure. However some aspects of the effects of these different mutations suggests certain possibilities. Internucleosomal DNA fragmentation, a hallmark of apoptotic cell death, is thought to be mediated by a nuclear endonuclease that can be stimulated by a rise in Ca^{2+} (McConkey *et al.*, 1989). Such a rise in Ca^{2+} can be induced in the *rd* mouse where cGMP accumulates as a result of the mutation (Yau and Baylor, 1989), thus activating the endonuclease. In *rds*/peripherin mice, failure to develop an outer segment may upset some internal mechanism of the photoreceptor thus activating the endonuclease. This situation is similar to that of a cell under stress, which might cause the cell to self-destruct. In the case of rhodopsin mutations (e.g. Pro347Ser) it could be that the presence of the mutant rhodopsin alters the normal cellular pathways and disrupts normal cell-cell interactions (Huang *et al.*, 1993), leading to the transmission of an incorrect signal to the photoreceptors and thus causing them to self-destruct. An important implication of the emergence of apoptosis as the final common pathway in RP is that strategies that

interfere with the cells' ability to carry out apoptosis could be appropriate as therapeutic interventions.

1.12 Aims of this study

This study describes the positional cloning endeavours undertaken to identify the gene responsible for autosomal dominant retinitis pigmentosa localised by prior linkage analysis to chromosome 19q13.4 (RP11). The ultimate purpose of this study being to contribute to the understanding of the physiology of the retina and mechanisms that cause retinal degenerations.

At the start of the project linkage to this locus was already established on chromosome 19q13.4 by a total genome search undertaken in a large British pedigree known as ADRP5 and the disease locus was refined to a 11 cM interval through analysis of microsatellite markers in the region. Further linkage analysis using markers in the disease region had also identified two other families linked to the RP11 locus, which implied that this locus could be responsible for a significant proportion of adRP. To test this premise further analysis of adRP families were undertaken to estimate the proportion of adRP families in which the disease phenotype is caused by mutations at 19q13.4, and also to refine the disease interval by finding new recombinant individuals in the linked families (chapter 3).

Candidate gene screening is a well-established method of identifying disease genes and has led to the identification of numerous retinal genes. The gene (*PRKCG*) encoding protein kinase C gamma isozyme has been mapped to 19q13.4 and evidence suggested it to be a feasible candidate gene for RP. Therefore the analysis of this gene was undertaken, which involved the genomic characterisation of this gene to establish the exon-intron structure of the gene followed by mutation screening in all the adRP families linked to the 19q13.4 (chapter 4).

Concurrent to mutation screening of candidate genes, the characterisation of the RP11 genetic interval was also undertaken to establish a physical map of the region in the form of an overlapping YAC contig. After coverage was obtained with YAC clones PAC clones were also isolated to further substantiate the contig. As part of characterising the RP11 region, novel microsatellite markers were also isolated from

cosmids mapping to the distal region. The novel markers were analysed in the RP11 linked families to establish their genetic localisation, order in relation to other markers in the region and also to pursue further possibilities of locus refinement (chapter 5).

CHAPTER 2

MATERIALS & METHODS

2.1 Preparation of DNA

2.1.1 DNA extraction from peripheral blood lymphocytes

Blood samples were collected in 10ml tubes that contained sodium EDTA and were extracted immediately upon arrival in the laboratory, or stored at -80°C until required. Prior to DNA extraction samples were thawed at room temperature and transferred to 50ml sterile falcon tubes. 40ml of Reagent A (10mM Tris-HCl pH8.0, 320mM sucrose, 5mM MgCl_2 , 1% Triton X-100) was added and mixed by inverting. The tubes were centrifuged at 4000g for 10 minutes, with the resulting supernatant discarded the pellet was resuspended in 2ml of Reagent B (400mM Tris-HCl pH8.0, 60mM EDTA, 150mM NaCl, 1% SDS). The suspension was transferred to 5ml screw capped tubes and 500 μl of 5M sodium perchlorate was added and mixed in a rotary mixer for 15 minutes at room temperature, followed by 25 minutes at 65°C . Afterwards the tubes were cooled on ice and 2 ml of chloroform and 300 μl of silica suspension (Nucleon DNA extraction Kit from scotlab) were added and rotary mixed for further 5 minutes prior to centrifuging at 1400g for 6 minutes. The supernatant containing DNA was then carefully aspirated to a universal and two volumes of absolute ethanol added to precipitate DNA. DNA now clearly visible as thread like strands were removed with a sterile needle and rinsed with 70% ethanol. After which the DNA pellet was air dried and dissolved in suitable amount of (300-1000 μl) sterile distilled water depending on the size of pellet. These DNA stocks were stored at -80°C while 1/10 dilutions were kept at 4°C for routine usage. On average the yield of DNA from 10ml of whole blood was between 250-300 μg .



2.1.2 Extraction of cloned DNA (plasmid and cosmid DNA) by alkaline lysis mini prep method

This method is a modification of the method by Birnboim & Doly (1979). *E.Coli* cells transformed with recombinant plasmid or cosmid vector (usually kept in the form of a bacterial stab or as a glycerol stock at -80°C) were first streaked on to LB agar supplemented with the appropriate antibiotic and the plates incubated overnight at 37°C. Plasmids were grown in the presence of ampicillin (50µg/ml) and cosmids in the presence of kanamycin (25µg/ml). Single colonies, chosen from the overnight growth were inoculated in 10ml of LB-broth supplemented with the same antibiotic and grown overnight at 37°C with agitation. Glycerol stocks can be made at this stage by adding 0.5ml of 40% sterile glycerol to 0.5ml of culture in a sterile 1.5ml eppendorf., these were stored at -80°C for future use. The cultures were centrifuged at 3000g for 10 minutes and the resulting pellet was resuspended in 200ml of cell resuspension buffer (50mM Tris-HCl, pH 7.5, 10mM EDTA, 100µl/ml RNase A). The bacterial suspension was transferred to a 1.5 ml eppendorf and 200ml of cell lysis buffer (0.2m NaOH, 1%SDS) was added and the tubes inverted until the lysate appeared clear. The cell lysis stage is followed immediately by the addition of 200 ml of neutralising buffer (1.32M potassium acetate, pH 4.8) which results in a precipitate. The DNA is in danger of being nicked if left too long in the lysis solution. The precipitate, consisting of protein, bacterial chromosomal DNA and cell debris was removed by centrifugation at 13,000 rpm for 10 minutes. The clear DNA containing supernatant was aspirated into a clean eppendorf and 1ml of DNA purification resin (6M GuHCl, 50mM Tris-HCl, 20mM EDTA, 5% Celite analytical filter aid (BDH, UK)) added. The DNA-resin mixture was applied to the top of a 3ml open syringe attached at its base to a mini column, which was itself attached to a vacuum pump. The resin solution was eluted through the mini column by vacuum suction with the retention of the resin bound DNA within the mini column. The column was re-eluted with 2ml of wash solution (20mM NaCl, 20mM Tris-HCl and 5mM EDTA). The mini column was then attached to a 1.5ml eppendorf. Applying 50µl pre heated (80°C) sterile distilled water and centrifuging for 2 minutes at 13,000 rpm eluted the DNA. The plasmid/cosmid DNA was kept at 4°C.

2.1.3 Isolation of total yeast DNA in solution. (YAC DNA preparation)

YAC clones provided as stabs in agar were first streaked out on YEPD agar plates supplemented with ampicillin (50µg/ml) and incubated at 30°C for 48 hours. The pink recombinant colonies were first tested for the presence of the YAC insert by 'colony PCR'. Half of a pink colony was suspended in 10µl of sterile distilled water whilst the rest was inoculated in 50µl of YEPD broth with the appropriate antibiotic. The inoculated distilled water was heated to 95°C for 5 minutes and spun down at 13,000Xg for 1 minute in a bench top centrifuge to precipitate the cell debris. 2µl of the clear supernatant was used for PCR using with the appropriate STS primers (section 2.4). If the desired product was given in the PCR indicating that the YAC insert was indeed present, the rest of the colony was inoculated into 10 ml of sterile YEPD broth (containing ampicillin at 50mg/ml) and incubated with shaking at 30°C for 24 hours. This culture was seeded in to 100ml of fresh YEPD broth and was incubated with shaking for a further 36 hours at 30°C until an O.D₆₀₀ of 1.5-2.0 corresponding to 3.3×10^6 cell/ml was reached. These cultures were centrifuged at 3000g for 10 minutes and the lone cell pellet was re-suspended in 5ml of solution I (0.9M sorbitol, 20mM EDTA and 14mM β-mercaptoethanol). Lyticase was then added at 4µg/ml to the suspension and incubated at 37°C for 1 hour with shaking. After centrifugation at 1000g for 10 minutes the supernatant was discarded and the pellet re-suspended in 5 ml of solution II (4.5 GuHCl, 0.1M EDTA, 0.15M NaCl and 0.05% sarcosyl, pH8) and incubated at 65°C for 10 minutes. On cooling to room temperature, equal volume of cold absolute ethanol was added and the DNA precipitated by centrifuging at 2000g for 10 minutes. The pellet was re-suspended in 2 ml of sterile 1X TE and on the addition of RNase (100µg/ml) the mixture was allowed to digest for 30 minutes at 37°C. This was followed by treatment with proteinase K (200mg/ml) and a further incubation at 65°C for 1 hour. After cooling to room temperature, a single phenol/chloroform extraction and ethanol precipitation was performed (section 2.2.1). The purified YAC DNA was dissolved in 200µl of sterile distilled water and 1/10 dilutions were used for PCR analysis. This protocol yields DNA in the size range 50-200 kb.

2.1.4 Preparation of YAC DNA plugs in agarose

This method enables the isolation of YAC DNA as intact yeast chromosomes and prevents mechanical shearing by immobilising the DNA in agarose. The YAC plugs are useful for restriction enzyme digests and sizing of YACs by pulse field gel electrophoresis (PFGE) (section 2.5.3). A single yeast colony was used to prepare a 100ml culture as described in section 2.1.3 and centrifuged at 4000g for 10 minutes after halving the culture in to two 50ml falcon tubes. The supernatant was discarded and the pellets were re-suspended in 20 ml of SCE solution (1M sorbitol, 0.1M sodium citrate and 10mM EDTA) and re-centrifuged as before. Having discarded the supernatant, the pellets were re-suspended in 800µl of SCEM (SCE solution containing β-mercaptoethanol at 30mM) and pooled together to approximate a cell density of 2.9×10^9 cells/ml. In to this 120µg/ml lyticase was added and incubated at 30°C for 1 hour to digest the yeast cell wall and form spheroblasts. An equal volume of liquefied agarose (55°C) (1% low melting point agarose in 1M sorbitol) was mixed with the cell suspension and dispensed into pre-cooled plastic moulds and allowed to set for 30 minutes. The plugs were then removed in to 15 ml proteinase K solution (0.5M EDTA, 1% N-lauryl sarcosine and 2 mg/ml proteinase K) and incubated at 50°C for 48 hours. The plugs were rinsed 3-4 times in 1 X TE solution and stored at 4°C until further use. PCR can be performed on YAC DNA in agarose, for this ¼ of a plug was removed into an eppendorf with 100µl of sterile water and melted at 65°C for 10 minutes. 2-5 µl of this was sufficient for a single PCR reaction.

2.2 Purification of DNA

2.2.1 Phenol chloroform extraction and ethanol precipitation

(a) Phenol chloroform extraction

This procedure removes impurities such as proteins and salts from DNA solutions. Prior to purifying small samples of DNA such as DNA digests, volume was first increased to 200µl by the addition of sterile 1 X TE to prevent DNA loss. Large volumes of DNA were extracted directly. An equal volume of phenol chloroform isoamyl (25:24:1) was added and mixed by inversion prior to centrifuging at 6000 rpm for 3 minutes. The top aqueous layer was then carefully removed to a clean 1.5ml eppendorf and an equal volume of chloroform was added, mixed and centrifuged as

before. The phenol extraction was repeated 2-3 times prior to the chloroform stage if the protein content of the sample was high. After the chloroform extraction the top DNA containing aqueous layer was removed in to a clean 1.5ml eppendorf and subjected to ethanol precipitation.

(b) Ethanol precipitation

Two volumes of cold absolute ethanol and 1/10 th volume of 3M sodium acetate was added and the DNA solution was placed at -80°C for 30 minutes to allow DNA precipitation. (The addition of salt and freezing at -80°C to facilitate DNA precipitation was only observed when purifying small quantities of DNA such as restriction digests of DNA, PCR products and plasmid/cosmid DNA extractions). After freezing, the DNA precipitate was centrifuged at 13,000 rpm for 15 minutes. The DNA pellet was rinsed in 70% ethanol then air-dried and dissolved in the required volume of sterile distilled water for subsequent use.

2.2.2 Use of Sephacryl microspin columns (Sephacryl-S200 and S-400 HR columns, Pharmacia, UK)

These columns were routinely used for removal of salt, exchange of buffer, removal of primers from PCR products and removal of labelled and unlabelled nucleotides from radio labelling reactions. The columns were also used to purify plasmid DNA after mini preparations. The Sephacryl ® HR resin (Sephacryl equilibrated in TE buffer, pH 7.6) was first resuspended in the column by tapping. The screw cap of the column was then loosened and the base snapped before placing the column inside an open 1.5ml eppendorf. After centrifuging at 3000 rpm for 1 minute to compact the resin, the column was removed to another clean open eppendorf and 20-40 ml of DNA sample was carefully applied to the top of the resin bed before centrifuging for a further 2 minutes to elute out the DNA. S200 columns were routinely used to remove unincorporated nucleotides from labeling reactions prior to hybridisation whereas S400 columns were used for the removal of excess primers, unincorporated dNTPs, and “primer dimers” from PCR products prior to sequencing.

2.2.3 Use of Centricon 100 spin columns (Centricon, Princeton, USA)

These columns were routinely used to purify PCR products and mini preps of cosmids and plasmids for ABI automated sequencing. The columns were first assembled according to manufacturer's guidelines and 50 µl of DNA sample with 2 ml of sterile distilled water was applied to the upper reservoir of each column. These were centrifuged at 1000g for 15 minutes, allowing DNA to be retained by the membrane while all unincorporated primers and dNTPs pass through to the bottom reservoir of the column. After the bottom reservoir was emptied the column was inverted and centrifuged for 5 minutes at 300g allowing the DNA eluate to be collected.

2.3 Restriction enzyme digests of DNA

Restriction enzyme digests were performed on genomic DNA, cosmid/plasmid cloned DNA and on PCR products. In genomic DNA digests, 6-10 µg of DNA was used while 1-3 µg of DNA was used in plasmid/cosmid digests. The amount of DNA used was dependent upon what was subsequently done with the digests, the quantity stated above was sufficient for visualisation by gel electrophoresis (section 2.5.1) before southern blotting (section 2.7). Restriction digests were normally set in a total volume of 40ml containing DNA, 1 X appropriate restriction buffer and 2-10 units of the appropriate enzyme. Some enzymes required the addition of Bovine Serum Albumin (BSA) at 1 X concentration for enhanced activity. The reactions were then incubated for a minimum of 1 hour or overnight at the specified temperature for the enzyme (usually 37°C). After heat inactivation the products were visualised by agarose gel electrophoresis (section 2.5.1). Restriction digests of PCR products were normally performed using 10 µl of the purified product (section 2.2.2), in a smaller reaction volume (20 µl). Enzymes were obtained from a range of manufactures; Pharmacia (UK), New England Biolabs (UK), Promega (UK) and GIBCO, BRL (UK).

2.4 Amplification of DNA by PCR

PCR was performed routinely in this study. The parameters and reaction conditions given in this protocol are standard, the relevant alterations to this protocol are stated where necessary. PCR reactions were typically carried out on ~100ng of

template DNA in 50 µl volumes consisting of 1 X PCR buffer (10 X NH₄ buffer, Bioline), 1.5 mM MgCl₂ (Bioline), 0.2 mM of each dNTP, 25 pmoles of each primer and 0.5 units of Taq polymerase. The reaction mixture was overlaid with one drop of mineral oil to prevent evaporation. The temperature cycling profile consisted of an initial denaturing step at 94°C for 3 minutes followed by 30 cycles of denaturation at 94°C for 30 seconds, annealing for 30 seconds and extension at 72°C for 30 seconds. The reaction was completed with a final extension step at 72°C for 2 minutes. Annealing temperature for the PCR is determined by the melting temperature of the two primers and is generally 1 or 2°C less than the lower melting temperature of either of the two primers. The melting temperature (T_m) was determined by the nucleotide sequence of the primer and was calculated using the following equation: $T_m = 4(C+G) + 2(A+T)$. A Hybaid Omnigene Thermal Cycler was used for all PCR analysis.

2.5 Fractionation of DNA

2.5.1 Agarose gel electrophoresis

This method was appropriate for separation of DNA fragments between 0.5 – 25 kb and was widely used to visualise PCR products, restriction enzyme digests and neat DNA. The agarose concentrations and the range of resolution achieved are as follows:

Agarose {% (w/v)}	Range of resolution of linear DNA (kb)
0.3	5.0 - 60
0.6	1.0 - 20
1.0	0.5 - 10
1.5	0.2 - 6.0
2.0	0.1 - 2.0
3.0	0.05 - < 0.1

An agarose (Biorad) mixture of an appropriate concentration and volume was prepared in 1 X TAE buffer. The mixture was melted in a microwave oven and cooled to about 50°C before adding Ethidium bromide to a final concentration of 0.5µg/ml. The cooled agarose was poured into a sealed loading tray that included a comb at one end and allowed to polymerise. After having removed the seal and the comb, DNA samples, prepared by adding an appropriate amount of 10X loading dye (2µl of dye

for 20 μ l samples) were loaded on to wells. An appropriate DNA molecular weight size marker was also included. The gel was then placed in an electrophoresis tank containing sufficient 1 X TAE buffer, with the wells positioned near the cathode end. Electrophoresis was carried out until the required resolution was achieved. Care was taken to avoid overheating. Genomic DNA and cosmid DNA digests were resolved by overnight runs (16 hrs) at 20V whereas shorter times at higher voltages were sufficient to resolve PCR products and plasmid DNA digests (e.g. 80V for 30 mins-1hr). Gels were photographed on an UV transilluminator using a Polaroid MP4 camera with an orange filter and Kodak plus- X film.

2.5.2 Denaturing polyacrylamide gel electrophoresis

This was routinely used to resolve DNA sequences and microsatellite markers. First the two gel plates were washed and wiped with 100% ethanol. The back plate containing the buffer reservoir was silicanised (Sigmacote, Sigma) prior to assembling the gel plates including spacers of appropriate thickness (0.8 mm). The assembled gel plates were then placed within a casting tray to form a plug. The concentration of acrylamide that should be prepared to provide maximum resolution of single stranded DNA fragments is given below:

Acrylamide (%)	Size range resolved (bp)	Comigration of Xylene cyanol marker (bp)	Comigration of bromophenol Blue marker (bp)
3.5	500-1000	230	50
5.0	40-250	130	32
8.0	30-200	80	22
12.0	20-100	35	10
15.0	12-75	30	8
20.0	3-50	22	6

6% gels were used routinely, 200ml of 6% gel solution was prepared with 132 ml 8.3M Urea, 48 ml acrylamide solution (Sequagel concentrate, National Diagnostics) and 20 ml of 10X TBE for a standard Biorad 48 X 68 cm apparatus. 50 ml of this solution was mixed with 300 μ l of 25% ammonium persulphate (APS) and 100 μ l of

TEMED (Sigma) and poured in to the gel casting tray to form a plug. Once the plug had set, the remaining 150 ml of gel solution was mixed with 600 μ l of 25% APS and 60 μ l of TEMED and poured carefully between the gel plates without forming air bubbles. Once the comb was inserted, the plates were rested horizontally with slight elevation and allowed to polymerise for 45-60 minutes. The polymerised gel was placed in a buffer tank after removing the casting tray. The buffer reservoirs of the apparatus were filled with 1X TBE and the apparatus connected on to a power supply. The gel was pre-run for 30 minutes at 100W prior to removing the comb and rinsing out the wells. 2-3 μ l of heat denatured DNA sample, which also included formamide loading dye was then loaded in to the wells. Electrophoresis was carried out at 100W for the required length of time. The duration of electrophoresis is depended on the size of DNA fragment being resolved, Bromophenol blue and Xylene cyanol both which are present within the formamide loading dye act as indicators of resolution (see table above). On completion the plates were separated and the gel, which adheres to the non- silicanised front plate, was fixed in 10% methanol and 10% acetic acid solution for 5-10 minutes. The fixed gel was then transferred onto a 3MM Whatmann paper, wrapped in cling film and dried under a vacuum at 80°C for 1-2 hours. The dried gel was placed within a cassette that contains intensifying screens and exposed to XAR-5 Kodak or Fuji X-ray film. The length of exposure is dependent on the strength of the radioactive signal, gels with weaker ³²P signals were autoradiographed at -80°C to enhance the signal.

2.5.3 Pulse field gel electrophoresis (PFGE)

DNA fragments up to 10Mb in size, such as yeast chromosomes and YACs, can easily be resolved by PFGE, in which the electric field is alternated periodically between two spatially distinct pairs of electrodes. **Alternating angle electrophoresis** used in this study is based on the CHEF (clamped homogeneous electric field) technique (Carle and Olsen, 1984) using the CHEF-DR II apparatus (Biorad). First 100 ml of 1% low melting point agarose (molecular biology certified, Biorad) gel was prepared in 0.5 X TBE and poured in to casting tray containing a comb and allowed to set for 60 minutes. YAC plugs (section 2.1.4) were then placed inside the wells and sealed by overlaying melted 1% low melting point agarose (Sigma, UK) prepared in 0.5X TBE. PFGE was carried out at 4°C. The Electrophoresis tank was first rinsed with 2 litres of distilled water after which 2 litres of 0.5 X TBE was added and

circulated via the pump. The gel was then removed from the casting tray and placed within the chamber. For separation of fragments between 100 kb and 2000 kb the following parameters were used routinely, an initial A time of 60 seconds, a final time of 90 seconds, a start ratio of 1.0 and a run time of 22 hours at 185 volts. Once these settings were established the buffer flow was set to allow gentle circulation and the gel subjected to electrophoresis. Afterwards the gel was stained for 30 minutes in ethidium bromide solution (0.5mg/ml) and photographed as described in section 2.5.1. If the YAC is of a different size to the background yeast chromosomes, it can be distinguished and the size can be deduced accordingly. In other cases, the gel was subjected to a southern blot and hybridised with human Cot-1 DNA to identify the YAC amongst the yeast chromosomal background.

2.5.4 Heteroduplex analysis

Heteroduplex analysis enables the detection of DNA mutations in PCR products. Mutation detection is based upon different migration rates of the heteroduplex and the homoduplex as each adopt different conformation. After standard PCR amplification, of both the normal and mutated allele, the PCR product was denatured and allowed to re-anneal at room temperature. Heteroduplexes are formed when a normal DNA strand anneals with a mutated complementary strand of DNA. These heteroduplexes migrate at a slower rate on acrylamide than the corresponding homoduplexes.

MDE gels (J.T. Baker, USA) were routinely used for the detection of heteroduplexes, optimal resolution is achieved with 100-400bp PCR fragments. The electrophoresis apparatus (40cm X 20cm glass plates with 1mm spacers) was vertically assembled according to manufacture's instructions (J.T. Baker, USA) and clamped within the casting tray. A 100ml of gel solution was prepared by the addition of 50ml MDE gel solution, 6ml 10X TBE and 44ml of sterile distilled water. Prior to the addition of APS and TEMED 2ml of gel solution was removed, to which 30 μ l of 10% APS and 12 μ l of TEMED added. This was pipetted between the plates to form a plug at the base. 400 μ l of 10% APS and 40 μ l of TEMED was added to the remaining gel solution, mixed and poured between the gel plates avoiding the formation of air bubbles. After an appropriate comb was inserted and clamped into place the gel was allowed to polymerise for 1 hour. The comb was then removed from the gel and the wells rinsed thoroughly in 0.6X TBE. 10 μ l of DNA sample containing 100-200ng of

PCR product in 2 μ l of sucrose loading buffer (40% sucrose, 0.25% each of Orange G, Xylene cyanol and Bromophenol blue dye.) was loaded in to each well. A positive control containing a known heteroduplex DNA was also run on every gel. The gel cassette was mounted on the electrophoresis apparatus and sufficient 0.6X TBE was added to the upper and lower buffer chambers. Electrophoresis was allowed to occur for ~ 16 hours at 800 volts, Xylene cyanol marker that co-migrates with a 230 bp fragment was used as an indicator. The gels were stained with ethidium bromide (0.5 μ g/ml) and UV photographed.

2.6 Radio-isotopic labelling of DNA

2.6.1 5' End-labelling of primers

A typical end labelling reaction contained 20 pmoles (100ng) of primer, 1X reaction buffer (One-Phor-All Buffer *plus*, Pharmacia), 10 μ Ci [γ ³²P] ATP (6000 Ci/mmole, Amersham) and 5 units of T4 polynucleotide kinase (Pharmacia) in a total reaction volume of 20 μ l. The reaction was incubated at 37°C for 40-50 minutes and used immediately afterwards or stored at -20°C (no more than a few hours) until required.

2.6.2 Labelling of DNA probes for hybridisation

(a) Random prime method

Pharmacia oligolabelling kit was routinely used to label ~ 50-100 ng of DNA. The total volume of DNA was first increased to 34 μ l with 1X sterile TE, denatured at 95°C for 2-3 minutes and then placed on ice for 2 minutes. After centrifuging briefly, the following were added to the DNA; 10 μ l of Reagent mix (Buffered aqueous solution containing dATP, dGTP, dTTP and random hexadeoxy-ribonucleotides), 50 μ Ci [α ³²P] dCTP and sterile distilled water to a total volume of 49 μ l. After adding 1 μ l of klenow fragment the reaction mixture was incubated at 37°C for 50 minutes. Afterwards the labelled reaction mix was first spun through an S-200 sephacryl microspin column to remove the unincorporated label and then subjected to Trichloroacetic acid (TCA) test to investigate the degree of incorporation. The test was carried out as follows; 1 μ l of labelled probe was placed on a 2.5cm Whatman glass microfibre filter and a geiger monitor was positioned above at such a height to

read 100 counts. The filter was then washed under vacuum with 10 ml of 15% TCA and re-monitored, the count read at this stage was an indication of the extent of incorporation. An incorporation of > 50% was sufficient for most purposes. The purified probe was denatured at 95°C for 2-4 minutes, cooled on ice and then put in the hybridisation mix.

2.7 Southern blotting

First described by Southern in 1975, this method enables the identification specific DNA sequences through hybridisation with a homologous DNA probe. In this study this technique was routinely used to analyse YACs, cosmid and plasmid DNA. Electrophoresis of YACs (PFGE), cosmid and plasmid digests were carried out until the required resolution was achieved (section 2.5.1 and 2.5.3). The complex nature of YAC and cosmid DNA require that these gels be depurinated first in 0.25M HCl for 30 minutes, followed by 30 minutes each in denaturing solution and neutralising solution. The treated gel was rinsed in distilled water and placed with DNA side up on a 3MM Whatman paper assembled as a wick over a reservoir of 10 X SSC solution. A Hybond N+ membrane (Amersham) cut to size was placed over the gel in one action, taking care not to introduce air bubbles. Two additional Whatman papers soaked in 10X SSC were placed on the membrane followed by paper towels and a small weight (1kg). DNA transfer was allowed to proceed through capillary action for 4-16 hours. The membrane was briefly rinsed in 2X SSC and dried at 80°C for 10 minutes before UV cross-linking (70,000 μ J/cm²) to fix the DNA. The membrane was then used for hybridisation (section 2.8).

2.8 Hybridisation of labelled probe

2.8.1 cDNA probe for identification of specific gene sequence

cDNA of the gene *PRKCG* was first used as a probe to screen a genomic DNA library to isolate cosmid clones that contained the *PRKCG* gene (see section 4.5.1.1). It was then used on replica filters of subclones made from *PRKCG* cosmids. This probe was also used to identify *PRKCG* specific sequences, from southern blots made of digested genomic, cosmid and plasmid DNA.

While the labelling reaction remained the same (section 2.6.2 (a)) different hybridising mixtures were used for each hybridisation. Genomic DNA blots were pre-hybridised in 25 ml solution containing 5X SSC, 5X Denhardt's solution (2% BSA, 2% Ficoll™ and 2% polyvinylpyrrolidone) and 0.5% SDS. 0.5ml (1mg/ml) of denatured sonicated non-homologous DNA was also added to the prehybridisation solution for greater sensitivity. Replica filters made of cosmid/plasmid clones were hybridised in 20 ml of pH 7.2 Church's solution (7% SDS, 0.125 M NaPi (68.4ml Na₂HPO₄ and 31.6 ml NaH₂PO₄ per 100ml of solution), 1mM EDTA). Southern blots of digested cosmid and plasmid DNA were also hybridised in 20 ml of Church's solution. Hybridisation was carried out in bottles with rotation in an oven. All filters were pre-hybridised at 58°C for 2-4 hours before adding the purified denatured probe (section 2.6.2). Hybridisation was usually allowed to occur for 6-16 hours at 58°C.

2.8.2 G₄(GT)₁₃ probe for identification of CA(n) repeats

100ng of G₄ (GT)₁₃ oligonucleotide was 5' end-labelled with [γ^{32} P]- ATP as described in section 2.6.1. The filters that needed to be hybridised were pre-hybridised in 20ml of Church's solution at 65°C for 2-4 hours and hybridisation was carried out with the denatured probe as described in section 2.8.1 at 65°C.

2.8.3 Human Cot-1 DNA probe for identification of YAC DNA

100ng of human Cot -1- DNA was labelled by random priming method described in section 2.6.2 and hybridised with the YAC DNA fixed on filter (section 2.5.3 and 2.7) in 20 ml of church's solution at 60°C as described in section 2.8.1.

2.8.4 Removal of non-specific hybridisation

The filters were subjected to post hybridisation washes to remove non-specific hybridisation. The initial low stringency washes were carried out at room temperature with 2X SSC and 0.1%SDS solution. The filters were then monitored for radioactivity. In cases of high background signal, the stringency of the wash was increased by decreasing the SSC concentration to 0.2 X SSC while keeping the SDS concentration constant at 0.1%. The washing temperature was also increased to ~ 60°C. Washed filters were sealed in polythene bags and autoradiographed for the required length of time using Kodak or Fuji X-ray film.

2.8.5 Stripping the filters of all radioactivity

When filters were needed for re-hybridisation, they were first stripped of the previous probe entirely to avoid confusion of hybridisation signals. The used filters need to be kept moist to achieve complete removal of radioactivity. The filters were washed in solution A (warm 0.4M NaOH solution warmed to 50°C) then transferred to solution B (0.1X SSC, 0.1% SDS and 0.2M Tris-HCl pH7.5) until minimal radioactivity was detected. The filters were rinsed in sterile distilled water, resealed in bags and exposed to Kodak or Fuji X-ray film over night to ensure the success of the strip.

2.9 Fluorescent *in situ* hybridisation (FISH)

FISH was carried out at University College London (by Ms M. Fox) by modification of the technique first described by Pinkel *et al* (1986) and Gillet *et al.* (1993).

YAC DNA was labelled with biotin-14-dATP by nick translation (Bionick kit, Gibco-BRL Life Technologies) and the labelled probe was precipitated together with Cot-1-DNA (to suppress repetitive DNA) and herring sperm DNA (to act as a carrier). The pre-annealed probe was resuspended in a hybridisation mixture that contained 50% formamide, 10% dextran sulfate and 2X SSPE, pH 7.

Human metaphase chromosomes were obtained from lymphocyte cultures synchronised by addition of thymidine to block DNA synthesis. After removal of the block, 5-bromo-deoxyuridine (BrdU) was incorporated prior to harvest. Standard cytogenetic techniques were used for colcemid arrest, hypotonic treatment, fixing of the cells in methanol:acetic acid, and slide preparation. The slides were pre-treated with RNase (10µg/ml) and proteinase K (0.035µg/ml) and prefixed in 1% formaldehyde (in PBS/5% MgCl₂). This is followed by dehydration in an ethanol series, denaturation in 70% formamide in 2X SSC for 5 minutes at 75°C and re-dehydration in an ice-cold ethanol series.

The separately denatured probe, which had been pre-annealed for 30 minutes at 37°C, was placed on the slide, covered with a cover slip and hybridised overnight by incubation at 37°C. Slides were washed in 50% formamide in 2X SSC at 42°C (three times for 5 minutes each). Followed by stringency washes in 2X SSC (two times for 2.5 minutes each), and in 0.1X SSC (two times for 2.5 minutes each) at 42°C. The rest of the protocol was carried out at room temperature. Slides were blocked in 5% Marvel non-fat milk/4X SSC for 20 minutes and washed in 0.05% Tween 20 detergent/4X SSC. The fluorescent signal was detected by incubation in fluorescein isothiocyanate (FITC)-conjugated avidin (5µg/ml in Marvel/4X SSC) (Vector laboratories), and amplified in biotinylated anti-avidin (5µg/ml in Marvel/4X SSC) (Vector laboratories), followed by a second round of FITC-avidin. Each incubation step lasted for at least 20 minutes and slides were washed in between each, with detergent mixture. Chromosomes were identified by the banding made visible with fluorescent counterstains diamino phenylindole (DAPI) and propidium iodide which were added to the antifade mounting medium (Vectashield, Vector laboratories).

The slides were examined under a Nikon Optiphot fluorescence microscope and images were captured using Confocal laser microscopy (MRC 600). YAC probes were also analysed using a cooled CCD camera (Photometrics) attached to Zeiss Axiophot fluorescence microscope equipped with appropriate filters and using Smartcapture software (Digital Scientific)

2.10 DNA sequencing

(Dideoxy chain termination method, Sanger *et al*, 1977)

2.10.1 Direct sequencing of PCR products

All reagents used for sequencing were obtained from the T7 Sequencing kit from Pharmacia, UK.

(a) PCR products were first purified through S400 microspin columns (section 2.2.2). 10 pmoles of sequencing primer was end labelled in a 10µl reaction using 10 µCi [$\gamma^{32}\text{P}$] ATP (section 2.6.1) and 2 µl of this was added to 10 µl of the purified DNA (~100ng). Following denaturation at 95°C for 5-10 minutes the contents were snap frozen on ice. While on ice the following were added, 2 µl of labelling mix A and 2 µl of T7 DNA polymerase (diluted 1:4 in enzyme dilution buffer) to the annealed

template and mixed by gentle pipetting. The samples were transferred immediately into the chain termination mixes as described below

(b) 2.5 μ l of each dideoxy chain termination mix (A, C, G and T) were dispensed into separate tubes and pre-warmed at 37°C for 1-2 minutes. 4 μ l of the prepared sequence mix from part (a) was then added to each tube, gently mixed and incubated at 37°C for 10 minutes. Sequencing reactions were terminated with the addition of 4 μ l of formamide dye. The sequence was analysed by electrophoresis on a 6% denaturing polyacrylamide gel.

2.10.2 Sequencing of cloned DNA

Recombinant Plasmid DNA was extracted using the alkaline lysis method described in section 2.1.2. The concentration of DNA for sequencing was adjusted so ~ 1-2 μ g was present in a 32 μ l volume (10 μ l of mini prep DNA with 22 μ l of sterile distilled water). 8 μ l of freshly made 2M NaOH was added, briefly vortexed and centrifuged. The mixture was then left at room temperature for 10 minutes to allow the DNA to denature. The addition of 11 μ l of 2M Na-Acetate (pH 4.8) and 120 μ l of 100% ethanol precipitated the denatured DNA. The DNA precipitation was completed as described in section 2.2.1(a). The purified DNA pellet was dissolved in 10 μ l of sterile distilled water followed by the addition of 2 μ l of sequencing primer (5 μ mol/ μ l) and 2 μ l of annealing buffer. The contents were mixed, centrifuged and incubated for 5 minutes at 65°C followed by 10 minutes at 37°C, afterwards the tubes were left at room temperature for a further 5 minutes. To the annealed template 3 μ l of labelling mix, 1 μ l of [α^{35} S]-dATP (6000 Ci/ μ l) and 2 μ l of T7 DNA polymerase (diluted 1:4 in enzyme dilution buffer) were added, mixed by gentle pipetting and left at room temperature for 5 minutes. The protocol was then followed as described for direct sequencing of PCR products (section b).

2.10.3 Automated DNA sequencing

In automated sequencing fluorescent dyes are used to detect sequencing ladders instead of radionucleotides. Lasers are employed within the design of the apparatus to promote fluorescence, which is detected and translated to a graphic image on a computer. The ABI 373a DNA sequencer was used routinely in this study. Further information can be obtained from the suppliers (ABI cycle sequencer from Perkin

Elmer, Foster City, California). Two types of kits were available for use with the ABI 373a DNA sequencer. One utilised fluorescently tagged dideoxy terminators, permitting the use of any primer and the other incorporated a fluorescently labelled primer, where the template DNA was needed to provide a binding site for that primer. A kit from the former category was used in this study, ABI PRISM™ dye terminator cycle sequencing ready reaction kit with AmpliTaq® DNA polymerase FS, (Part number 402078). This kit was used for the direct sequencing of *PRKCG* exons, generated through PCR, for mutation screening purposes. It was also used for the sequencing of cosmids, plasmids and YAC endclones.

(a) Preparation of DNA for sequencing

PCR products were purified either through Centricon 100 spin columns or S400 spin columns as described in sections 2.2.2 and 2.2.3. 5-10µl (1µg) of the purified PCR product was used for sequencing. Plasmid (2µl) or cosmid (10µl) mini prep DNA was used directly.

(b) Cycle sequencing

The cycle sequencing reaction contained 8µl of reaction mix (FS kit), 3.2µM sequencing primer and ~1µg of purified DNA in a combined reaction volume of 20µl. Sequencing was performed on a Perkin-Elmer Cetus 2400 or 9600 machine, in 25 cycles with the following temperature cycling profile.

Max ramp

96°C for 10 sec Max ramp

50°C for 5 sec Max ramp

60°C for 240 sec

The resultant product was transferred to 0.5ml eppendorf and purified to remove excess fluorescent dye by a standard ethanol precipitation using 80µl of 95% ethanol and 3µl of 2M sodium acetate (pH4.5). The lyophilised samples were used immediately or placed at -20°C for a maximum period of 7 days prior to analysis. Samples were prepared for loading by adding of 3-6µl of formamide/EDTA loading dye (Deionised formamide : 0.05M EDTA (5:1 v/v)) to the pellet, vortexing and briefly centrifuging. These were denatured at 90°C for 2 minutes and snap frozen on ice prior to loading.

(c) Electrophoresis and automated analysis

The 373a DNA sequencer was set up and run accordance with the supplier's instructions. Sequencing runs involved electrophoresis (of up to 38 samples) for 10-13 hours. An Apple Macintosh computer linked to the sequencer was programmed to analyse the data and present it in the form of two files: the text file and an analysis file. The latter file contained the raw data, analysed data, and detailed sequence information. The quality of the sequence can be ascertained by the electropherogram (scan) data that is present within this file

2.10.4 Computational analysis of DNA sequence

Analysis of nucleic acid sequences was mainly performed using **Geneworks**™ (version 4.45). This software, developed for Apple Macintosh computer, was used to align sequences of different clones/PCR products in order to identify overlapping regions and also to identify restriction sites in sequences. BLAST/ FASTAN/ FASTA programs (Pearson and Lipman, 1988; Altschul *et al.*, 1990) accessible through the Internet site of HGMP, were consistently used to compare cloned sequences with entries in the GenBank data base for homologies. This was of utmost importance when designing primers, since primers with homologies to repetitive sequences had to be avoided, primer sequences that did not have any homology to sequences present in the database most often proved to bind only at the desired target region.

2.11 DNA cloning (PCR products)

All the vectors used in this study for the cloning of PCR products contain 3'-T overhangs at the insertion site, and take advantage of the non-template dependent addition of a single deoxyadenosine to the 3'-end of the PCR products by certain Taq polymerases.

2.11.1 Competent *E.coli* cells

In all cases the ligated vector was transformed in to *E.coli* DH5 α host competent cells with genotype *supE44*, Δ *lacU169*, (\emptyset 80*lacZ* Δ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*. The vectors used (pGEM-T and pTAg) contain *LacZ* α peptide sequence which when functionally produced complements the N-terminal truncated *LacZ* peptide synthesised in the competent cells. The resulting enzyme, β -galactosidase, cleaves X-Gal to give blue colonies. Insertional inactivation of the

LacZ α peptide produce white recombinant colonies thus providing the basis for blue-white screening of recombinants.

2.11.2 Ligation in to specific vectors

Plasmid vectors used for direct cloning were pGEM-T, 3.0 kb (pGEM-T Vector Systems, Promega) and pTAg, 3.8 kb (The LigATor kit, R&D Systems). Both vectors were pUC derivatives and contain T overhangs at the cloning site.

(a) Use of pGEM-T vector and pTAg vector

pGEM-T vector carry the gene β -lactamase gene for Ampicillin resistance allowing the selection of transformed cells, when grown in the presence of the antibiotic. The plasmid also contain T7 RNA and SP6 RNA polymerase promoters flanking the a multiple cloning region within the α -peptide coding region of the enzyme β -galactosidase. Insertional inactivation of the α -peptide allows recombinant clones to be directly identified by blue white selection, and SP6 and T7 primers can be used to sequence in to the insert. The pTAg vector carried the Kanamycin phosphotransferase gene in addition to the β -lactamase gene permitting growth in the presence of Ampicillin and Kanamycin. Like the pGEM-T vector selection of recombinants was based on blue-white screening. Special pTAg SEQ 5' primer and pTAg SEQ 3' primer was used to sequence the insert.

Ligation was carried out according to the recommended protocols using reagents provided with each kit (pGEM-T Vector Systems, Promega and The LigATor kit, R&D Systems). In each case 50ng of vector was ligated to an appropriate amount of purified PCR product (section 2.2.2 and 2.2.3), such that the molar ratio of the vector to the insert was either 1:1 or 1:3. Ligations were allowed to occur overnight (or 2 hours) at the recommended temperatures of 15°C (Promega) and 16°C (R&D Systems).

2.11.3 Transformation of competent *E.coli* cells and selection of recombinants

DH5 α competent cells (supplied by Invitrogen or Gibco BRL or made in-house) were transformed according to the manufacturer's recommended protocols. Generally 2 μ l of the ligation reaction was added in to 50 μ l of thawed competent cells. The contents were placed on ice for 20-30 minutes and the cells were transformed by heat

shocking at 42°C for 40-50 seconds. This was followed immediately by incubation on ice for 2-3 minutes. 200µl of LB broth was added to the tubes and incubated at 37°C for 1 hour with shaking. 50-100µl of this was plated out on LB/Amp (50µg/ml)/ IPTG / X-Gal plates and incubated over night at 37°C. These plates were prepared beforehand by spreading 100µl of 100mM IPTG (isopropylthiol-β-D-galactoside) and 20µl of 50mg/ml X-Gal (5-bromo-4-chloro-3 indoyl-β-galactoside) on LB/Amp agar plates which were then allowed to dry at 37°C for 30 minutes prior to use. When using the pTAg vector the agar plates were also supplemented with 15µg/ml Tetracyclin. Appropriate positive and negative controls were used to test the efficiency of transformation. The recombinant colonies appeared white after the overnight growth and were larger than the blue colonies; storage at 4°C also facilitated the blue/white screening.

The white recombinant colonies were first tested for the presence of the desired insert by 'colony PCR'. Half of the white colony was suspended in 10µl of sterile distilled water whilst the rest was inoculated in 50µl of LB broth with the appropriate antibiotic. The inoculated distilled water was heated to 95°C for 5 minutes and spun down at 13,000Xg for 1 minute in a bench top centrifuge to precipitate the cell debris. 2µl of the clear supernatant was used for PCR using appropriate vector primers. If the desired product was given in the PCR, glycerol stocks were made of the recombinant clones as described in section 2.1.2.

2.12 Screening of genomic DNA library cloned in Lorist B cosmid vector

A human cosmid library (Cachon-Gonzalez, 1991) consisting of approximately 5×10^5 independent recombinants was obtained from Dr. J. Fitzgibbon, Dept. of Molecular Genetics, Institute of Ophthalmology, London. The Lorist B human cosmid library was generated following a partial *Mbo*I restriction digest of human genomic DNA (Cachon-Gonzalez, 1991), with an estimated titre of 1.0×10^9 cfu/ml.

(a) Preparation of master filters for primary screening

The human cosmid library was plated out on two 20cm X 20cm charged nylon filters (Hybond-N+, Amersham). Approximately 2×10^5 colonies can be accommodated on each filter, therefore 2µl of the amplified cosmid library corresponding to about 2×10^6 colonies was used per filter. Ethanol soaked and flamed

forceps were used for all manipulation of filters and gloves were worn at all times. The numbered filters were placed on megaplates of LB agar containing 50µg/ml of Kanamycin. 2µl of a glycerol stock of the library was first diluted in 2ml of LB/Kanamycin broth, after mixing by gentle inversion four aliquots, each of 0.5ml, were dispensed on to the filter. Using a sterile glass spreader the cultures were evenly dispersed over the surface leaving a border, 2-3mm wide, at the edge of the filter free of bacteria. The mega plates were sealed with Parafilm and incubated at 37°C overnight until single colonies (0.1-0.2 mm diameter) appeared.

(b) Replica filters and preparation for hybridisation

Two replica filters were made from each master filter. Four LB/Kanamycin agar mega plates were poured, set and Hybond-N+ filters were placed on the agar (numbered side down) to pre-wet the filters. The master filter was peeled of the storage plate and placed colony side up on a dampened pad of sterile 3MM Whatman paper. The damp replica filter was placed on the master filter without trapping air bubbles between the two filters. Care was taken not to move the filters relative to each other once contact has been made. A sterile 3MM paper was placed on top, and the two filters were pressed firmly together with a heavy glass plate. After making a series of orientation holes with a sterile needle the filters were peeled apart. The first replica filter was replaced on its agar plate, whilst the master filter was used to create the second replica in the same manner. The replica plates were incubated at 37°C for 6-8 hours whilst the master plate was incubated for 1 hour at 37°C prior to being stored at 4°C.

To prepare the colony filters for hybridisation, the replica filters were first placed on a bed of 3MM paper soaked in denaturing solution for 5 minutes followed by 5 minutes in neutralising solution. This procedure lysed the colonies and allowed the DNA to bind to the filter. The filter surface was carefully wiped with a folded tissue under 2X SSC to remove colony debris. The filters were UV cross-linked as described in section 2.7. Hybridisation of the filters was carried out as described in section 2.8.1.

(c) Isolation of positive colonies

Authentic positive signals were identified as those that were present on both of the X-ray films of the replica filters. These were marked along with the orientation marks on overlying acetate that was then secured on a light box. A few ml of LB Broth/Kanamycin was placed on the acetate to help keep the master filter on place, which was oriented according to the orientation marks. A sterile plastic loop was used to pick all colonies in a 5mm radius centred on the positive signal in to 50µl of LB broth/Kanamycin. These were incubated at 37°C with agitation (225 rpm) for 1 hour to prepare for the next stage of screening.

(d) Secondary and tertiary screening for purification of clones

A serial dilution (10^2 , 10^3 and 10^4) was made for each positive pool in LB/Kanamycin and 100µl of each dilution was plated out on to a 82 mm filter on LB agar/Kanamycin plate and grown overnight at 37°C. The particular dilution giving approximately 300 or so colonies with good separation was chosen for secondary screening. These filters were used to make replica filters and these filters were processed as described in the previous section, the only difference being the usage of 90mm bioassay plates instead of the mega plates. Positive signals were picked as **single** colonies into 50µl of broth and treated in the same way as their primary screen counterparts.

Tertiary screening was performed in the same manner as the secondary screen. The density of colonies was reduced to about 50 on each plate, allowing the isolation of well defined pure clones. Glycerol stocks of these clones were made as described in section 2.1.2. and stored at -80°C.

2.13 Sub-cloning of cosmids into plasmid vectors (pBlue Script (pBS) and pUC18)

Cosmids clones were subcloned in to smaller sequence ready-vectors such as pBlue Script (pBS) and pUC18, twice in this study. In the first instance the cosmid (i.e. 20019) harbouring the gene *PRKCG* was subcloned in order to deduce the genomic organisation of a portion of this gene, next series of cosmids from the RP11 disease interval were subcloned to isolate novel microsatellite/ (CA)_n markers. The

procedure used for subcloning was similar in both instances except for the ligation vector and restriction enzymes used for digestion of cosmids. Cosmid 20019 was digested with *Bam*HI and then ligated to pUC18 vector using Ready-To-Go™ *Bam*HI/ BAP + ligase kit (Pharmacia, UK), while the other cosmids were digested with two separate restriction enzymes *Sau*3A1 and *Hae*III which were then respectively ligated to pBS vector digested with *Bam*HI and *Eco*RV.

(a) DNA digestion and purification

40µl of the cosmid mini prep DNA (~2µg of DNA) was digested overnight with 20 units of the specified restriction enzyme (*Sau*3A1, *Hae*III or *Bam*HI), in a 50µl reaction volume containing 1X BSA. 1.5 µg of pBluescript® II KS (+/-) vector (pBS) was separately digested in a similar manner with 15 units of *Bam*HI and 15 units of *Eco*RV enzymes. A fraction of each digest (5µl) was analysed by agarose gel electrophoresis and purified by phenol/chloroform extraction and ethanol precipitation (section 2.2.1). After resuspending in 20µl of sterile 1X TE, DNA recovery was investigated by agarose gel electrophoresis.

(b) Dephosphorylation of vector

The vector was dephosphorylated to prevent self-ligation. 10µl of purified and digested pBS vector DNA (~500ng) was dephosphorylated in a 20µl reaction volume using 1.5 units of calf intestinal alkaline phosphatase (Pharmacia, UK) in 1X buffer. The reaction was incubated at 37°C for 2 hours and heat denatured at 85°C for 20 minutes to stop enzyme activity.

(c) Ligation to plasmid vectors

*Sau*3A1 digested cosmids were ligated to *Bam*HI cut vector and *Hae*III digested cosmids were ligated to *Eco*RV cut vector as follows. 16µl (~500ng) of purified cosmid DNA was ligated with 4 µl of (100ng) of purified, dephosphorylated pBS vector DNA using Ready-To-Go™ T4 DNA ligase kit (Pharmacia, UK). Each tube within the kit contained the necessary components for ligation and the reaction volume had to be exactly 20 µl. After adding the DNA the tubes were left at room temperature for 5 minutes, followed by gentle mixing and incubation overnight at 16°C. The following procedure was carried out for cosmid 20019, which harboured the *PRKCG* gene. 20 µl of purified *Bam*HI digested cosmid 20019 was directly

ligated into pUC18 vector using Ready-To-Go™ *Bam*HI/ BAP + ligase kit (Pharmacia, UK) instructions for its use was identical to that described above.

(d) Transformation into competent *E.coli* cells and isolation of positive clones

The transformation procedure for cosmid ligations is as described in section 2.11.3. The only difference being that transformed cells were plated onto Hybond N+ gridded membranes overlaid on LB/Amp (50µg/ml)/ IPTG / X-Gal agar plates also prepared as described in section 2.11.3. The plates were incubated overnight at 37°C. Replica filters were made of the master filter and prepared for hybridisation as described in section 2.12.b. To isolate subclones with (CA)_n repeat sequences the replica filters were hybridised with labelled G₄ (GT)₁₃ probe (section 2.8.2). Subclones of cosmid 20019 were screened with the cDNA probe of *PRKCG* gene (section 2.8.1). Isolation of positive clones was as described in section 2.12.c, however the selected positive colonies were inoculated in 10 ml of LB-broth supplemented with 50µg/ml ampicillin (instead of kanamycin). After overnight growth the DNA (of positive clones) was isolated using the alkaline lysis mini prep method. ~ 1 µg of DNA was digested with an appropriate enzyme(s), southern blotted and re-probed with the appropriate DNA probe to confirm the positive nature of the selected clones.

2.14 Microsatellite analysis

2.14.1 Radio-isotope detection by 5' end labelling or incorporation of radiolabel in to the PCR product'

Chromosome 19q13.4 specific, microsatellite markers were chosen from a range of genome maps and the associated primers obtained commercially (Cruachem or Genosys, UK) or from then HGMP Resource centre. The markers were first tested non-radioactively by standard PCR amplification (section 2.4) to determine the product size and to ensure the success of the PCR at the specified conditions. For the **5' end-labelling method**, either the forward or the reverse primer was end labelled as described in section 2.6.1 prior to the addition to the PCR reaction mix. Unlike for the non-radioactive PCR (section 2.4) far less primer is required (2-3pmoles/reaction) and the total reaction volume was 10 µl. The **incorporation method** of labelling involved the incorporation of [α^{32}]-dATP or [α^{32}]-dCTP directly into the PCR reaction mix.

The protocol was essentially the same as above with the exception of the dNTP concentration used. The concentration of the unlabelled (dATP (or dCTP) was 0.02mM as opposed to the usual 0.2mM. The concentration of the other 3 dNTPs was maintained at 0.2mM. After the PCR was complete, 3 μ l of loading dye was added to each tube prior to heat denaturing and electrophoresis (section 2.5.2) for the resolution of the allele system.

General solutions and media used

LB-broth (1 litre)

10g NaCl

10g Bactotryptone

5g Yeast extract

LB-agar

LB-broth with 15g/L Bactoagar

YEPD broth

2% glucose

2% Bactotryptone

2% Yeast extract

YEPD agar

YEPD broth and 2% bactoagar

1X TAE buffer

40mM Tris acetate

2mM EDTA

1X TBE

89mM Tris base

89mM boric acid

2mM EDTA

20X SSC

3M NaCl

0.3M Na₃citrate.H₂O (pH adjusted to 7.0 with 1M HCl)

TE buffer

10mM Tris-HCl (required pH)

1mM EDTA (pH 8.0)

Neutralising solution

3M NaCl

0.5M Tris

Denaturing solution

1.5M NaCl

0.5M NaOH

Depurinator

0.25M HCl

Phenol-chloroform-isoamyl

25ml equilibrated phenol

24ml chloroform

1ml isoamyl alcohol

Formamide loading dye

80% formamide

10mM NaOH

0.1% Bromophenol blue

0.1% Xylene cyanol

CHAPTER 3

GENETICS OF RP11 AND REFINEMENT OF THE LOCUS

3.1 Introduction

3.1.1 Autosomal dominant retinitis pigmentosa locus linked to chromosome 19q13.4 (RP11)

Retinitis pigmentosa (RP) is an inherited retinal degeneration affecting approximately one in 4-5000 people (Boughman *et al.*, 1980, Bunker *et al.*, 1984). It can be inherited as an X linked, dominant, recessive or digenic disorder and at least nineteen different genetic loci have been implicated. The autosomal dominant form of retinitis pigmentosa (adRP) is itself genetically heterogeneous and can be caused independently by mutations in genes encoding rhodopsin and RDS (Dryja *et al.*, 1990; Farrar *et al.*, 1991; Kajiwara *et al.*, 1991). More recently a mutation was also identified in the Neural Retina-specific Leucine zipper gene *NRL*, which is localised on chromosome 14q11, in a single adRP pedigree (see section 1.10.1.1). Furthermore through linkage analysis, 7 yet unidentified genes localised to chromosome 1 cen (RP18) (Xu *et al.*, 1996), 7p13-15 (RP9) (Inglehearn *et al.*, 1993), 7q31 (RP10) (Jordan *et al.*, 1993), 8q11 (RP1) (Blanton *et al.*, 1991), 17p (RP13) (Greenberg *et al.*, 1994), 17q (RP17) (Bardien *et al.*, 1995) and 19q (RP11) (Al-Magthteh *et al.*, 1994) have also been implicated for causing adRP.

The linkage to chromosome 19q was originally obtained in our laboratory in a large British family named ADRP5 (Al-Magthteh *et al.*, 1994) and the locus was given the number RP11 (MIM 600138; McKusick 1992). Subsequently two other families, ADRP29 and RP1907, with phenotypes similar to that of ADRP5 were linked to this locus. Haplotype analysis in these families excluded the possibility of a founder effect. Reporting of a Japanese family (Xu *et al.*, 1995) and three other

American families (McGee *et al.*, 1997) also linked to 19q suggest RP11 to be the second most common adRP locus after rhodopsin based on linkage studies.

3.1.2. Clinical phenotype of RP11 families and partial penetrance

Phenotypic variability is an established fact for autosomal dominant retinitis pigmentosa as different clinical presentations have been observed along with genetic heterogeneity. In accordance with this fact the clinical features observed in adRP families linked to chromosome 19q are also quite distinct and unique (Moore *et al.*, 1993 and Evans *et al.*, 1995). The adRP phenotype of RP11 has been classified as R type (regional/ typeII) due to the presence of matching areas of both cone and rod functional loss (Lyness *et al.*, 1985) but it is the presence of asymptomatic obligate carriers indicating partial penetrance of the disease trait in RP11, which sets this locus apart from all other adRP loci. (Moore *et al.*, 1993 and Evans *et al.*, 1995).

Both symptomatic and asymptomatic members from two of the largest RP11 families, ADRP5 and ADRP29, have been subjected to extensive clinical, electrophysiological and psychophysical studies. Autosomal dominant inheritance in these families is evident with symptomatic individuals in each generation and male to male transmission. All symptomatic individuals had severe and early onset of disease. Typical fundus features of extensive peripheral degeneration were observed in even the youngest examined symptomatic patient who was 24 years of age. Older symptomatic individuals also showed macular atrophy, oedema and secondary cataract formation. The asymptomatic (obligate) disease haplotype carrier individuals had normal fundus appearance and minimal or no psychophysical or electroretinographic abnormalities. Figure 3.1 depicts typical fundi photographs of an affected and normal individual from a RP11 linked family.

Difference in clinical presentation between members of the same family, as shown in families linked to RP11, is known as variable expressivity. Variable expressivity and 'R type' adRP are also features of the dominant retinitis pigmentosa locus linked to chromosome 7p (RP9) (Kim *et al.*, 1995). However the variable expressivity

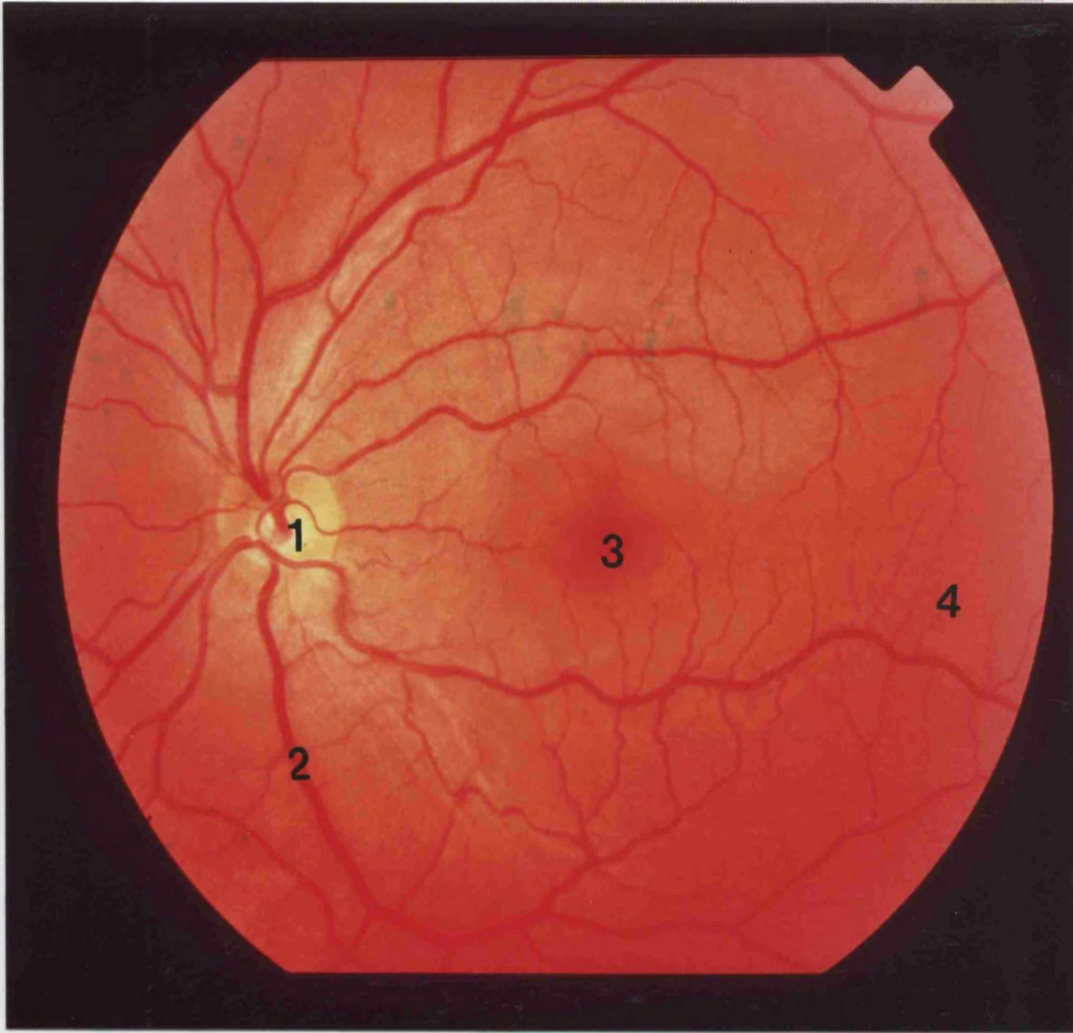


Figure 3.1a

The fundus image of a normal retina (adapted from Paton *et al.*, 1976)

- 1- Optic nerve demarcating the site of entry of the retinal blood vessels and the site of departure of the optic nerve.
- 2- Major retinal vessels.
- 3- The macular region.
- 4- Peripheral retina.



Figure 3.1b

The fundus image of a late stage RP patient (from a RP11 family) showing the typical features of RP i.e. a pale optic nerve head, attenuated retinal vessels and bone-corporuscular like (bone spicule) pigmented deposits, which occur most prominently in the periphery of the retina, and depigmentation and atrophy of the retinal pigment epithelium.

phenotype described for RP9 is a graded disease severity among affected members of the same family, where the clinical presentation of the disease range from mild, moderate to severely affected and independent of age. This contrasts with the unusual polarity of phenotype seen in RP11 where individuals with the disease gene seem either to be severely afflicted or are essentially asymptomatic. Both RP9 and RP11 therefore depict two different types of incomplete penetrance, the 'all or nothing' form of incomplete penetrance shown in RP11 led to the description of the RP11 phenotype as 'bimodal expressivity'. Further pedigree analysis also suggested that although the offspring of parents with the disease haplotype were at 50% risk of having the genetic defect, the risk of being symptomatic was only 35%. The disease penetrance defined as the proportion of affected individuals among the susceptible disease gene carriers, for RP11 was estimated at 0.7. This value was derived from the haplotype data of the complete pedigree of ADRP5. All individuals less than 20 years of age who possessed the disease haplotype yet asymptomatic at the time of clinical examination, were excluded from the penetrance estimation since the disease status could not be determined for certainty at that an early age.

3.1.3 Possible causes for the reduced penetrance of RP11

The basis for the reduced penetrance phenotype in RP11 is yet unknown, however several models for reduced penetrance have been evaluated and are discussed below.

Environmental influence as a cause has been ruled out since it is more likely to produce a graded rather than the all or nothing phenotype typical of RP11 (Evans *et al.*, 1995). Moreover, environmental factors have not been proved to have that great an influence on the disease severity of RP (Heckenlively, 1988).

Anticipation and genomic imprinting are two genetic mechanisms associated with variable expression of disease phenotype. Anticipation describes the phenomenon whereby severity of disease increases with successive generations in a family concurrent with the early onset of disease. The genetic cause of anticipation has been attributed to increasing trinucleotide repeat expansions within genes as seen in fragile X syndrome (Richards *et al.*, 1992), myotonic dystrophy (Tsilfidis *et al.*, 1992) and numerous other hereditary neurological disorders. However, anticipation with

severity related to position in a pedigree was not seen in any of the RP11 pedigrees. The sex of the disease gene-carrying parent nor the sex of the carrier individual influence the likelihood of a disease gene carrier being symptomatic, hence excluding genomic imprinting and the possibility that penetrance is sex related.

Digenic describes the mode of inheritance in which the simultaneous segregation of mutations at two unlinked genes give rise to the disease phenotype (see section 1.10.1.5). This mode of inheritance has already been implicated in at least one example of RP (Kajiwara *et al.*, 1994), where heterozygous mutations in peripherin/RDS and ROM1 genes together cause retinal degeneration. Digenic inheritance has been suggested to explain the RP11 phenotype, needing a second mutation/polymorphism in a second locus for the expression of the disease phenotype. If this was so then symptomatic individuals are significantly more likely to have symptomatic children than disease haplotype carriers, however this has not been observed in the RP11 linked pedigrees. In fact the opposite was observed when the pedigree data of ADRP5 and ADRP29 was compiled. A higher percentage of symptomatic children were born to asymptomatic rather than to symptomatic individuals (76% as opposed to 50%). However the difference in percentage was not significant (Evans *et al.*, 1995). This observation can be explained if the second mutation was allelic to the 19q primary mutation. Allelic effect was initially ruled out as a possibility due to the presence in ADRP5 of two siblings [IV-23 and IV-25, and V-1 and V-4 (see figure 3.3a)] who had identical genotype at the RP11 locus yet had different disease phenotype (one was symptomatic while the other was asymptomatic).

However results of a recent study (McGee *et al.*, 1997) does support the hypothesis that wild-type alleles at the RP11 locus or a closely linked locus (i.e. a modifier gene) inherited from the non-carrier parent (i.e. the unaffected/normal parent) are a major factor influencing the penetrance of disease alleles at this locus. They studied the correlation between inheritance of alleles from the non-carrier parents and presence of disease in carrier offspring, by sib pair analysis with data derived from the pedigrees ADRP5, ADRP29 and the Japanese family. They found that pairs of carrier siblings with the same phenotype almost always inherited the same haplotype from the non-carrier parent. Among pairs of carrier siblings with

contrasting phenotypes (i.e., one affected and one unaffected) most inherited different haplotype from the unaffected parent. Even though the correlation was statistically significant it was imperfect due to the same sib pairs in ADRP5 (IV-23 and IV-25, and V-1 and V-4) who previously stood as exception to the allelic effect. If the allelic effect is true then possible explanations for the contrasting phenotypes in these individuals, include the existence of other factors modifying penetrance (e.g. other genes in the genetic background) or the miscategorisation of the phenotype of one member of a pair.

In hereditary elliptocytosis (or haemolytic anaemia) the disease penetrance is specified by wild type alleles in *trans* at a disease locus and it stands as a good example for allelic effect. Hereditary elliptocytosis is caused by dominant mutations in the gene coding for the α subunit of spectrin (Gratzer, 1994), which is a cytoskeletal protein essential for the normal morphology of red blood cells. Similar to the RP11 phenotype, carriers of dominant spectrin mutation could either be severely affected or be unaffected. The penetrance of mutations is specified by a high-frequency, otherwise silent polymorphism at this locus which determines the relative level of expression of α -spectrin alleles (Wilmotte *et al.*, 1993). An individual with a high-expressing wild type allele in *trans* of a pathogenic mutation will be unaffected or very mildly affected whereas an individual with a low-expressing wild type allele in *trans* would have severe haemolytic anemia unless the pathogenic mutation by chance is also on a low-expressing allele (Randon *et al.*, 1994).

The biochemical mechanisms by which the wild type alleles at the RP11 locus or a closely linked locus influence severity will only be known once the RP11 gene has been cloned, till then the actual cause of partial penetrance phenotype in RP11 will remain a mystery.

3.1.4 Linkage history between RP11 and chromosome 19q 13.4 markers

Following a genome wide search, linkage was established between the RP phenotype in ADRP5 and markers on chromosome 19q13.4 (Al-Magthteh *et al.*, 1994). Extensive genotyping of 19q13.4 microsatellite markers [obtained from all available genetic maps at the time (Gyapay *et al.*, 1994; NIH/CEPH, 1992; Weber *et al.*, 1993)] in ADRP5 followed by haplotype analysis enabled the refinement of the RP11 locus to an 8 cM interval between D19S180 and D19S605 (Al-Magthteh *et al.*, 1996). However, in the most recent CEPH/ Généthon map this genetic distance is estimated at 13 cM (Dib *et al.*, 1996). This interval contained four other markers D19S572, PRKCG [RFLP marker associated with the gene protein kinase C gamma (*PRKCG*)], D19S926 and D19S418, which showed very tight linkage to the disease locus with no recombination (see figure 3.2a). In subsequent studies two other families (ADRP29 and RP1907) with similar clinical features to ADRP5 were identified and linkage of these families to the RP11 locus was confirmed by haplotype analysis with markers localised to the RP11 interval. Whilst no critical recombination events were detected in RP1907 with the markers flanking the disease interval, in ADRP29 a second proximal cross over was detected with D19S180 along with another distal recombination event with D19S926. This further refined the RP11 locus, replacing the previous distal flanking marker D19S605 with the more centromeric D19S926 (AFMc001yb1). Therefore at the outset of this study the RP11 locus was confined to a genetic interval of approximately 11cM between the markers D19S180 and D19S926 (see figure 3.2a). This genetic distance is as according to the most recent CEPH/ Généthon map (Dib *et al.*, 1996).

3.1.5 Aims of study

Once a gene has been localised to a specific chromosomal region through linkage analysis, it is imperative that maximum genetic refinement is achieved prior to initiating a positional cloning venture. Families such as ADRP29 in which recombination events have occurred between available markers and the disease are an invaluable resource for finding the minimum genetic distance within which the disease gene is localised. As new markers become available analysis of segregation in such cross over families may define new boundaries for the disease interval.

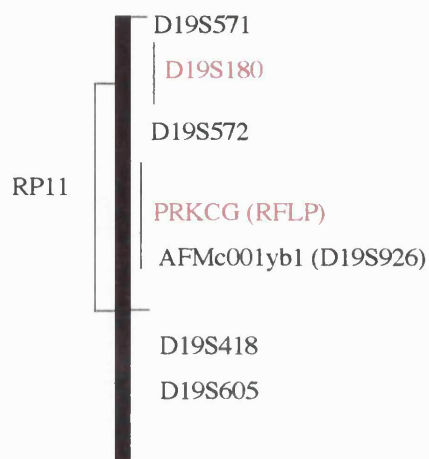


Figure 3.2 a

A representation of the genetic map from the RP11 region of chromosome 19q13.4 showing the order of markers and the refinement of the RP11 locus at the onset of this study (Al Maghethieh *et al.*, 1996). This map includes NIH/CEPH markers (highlighted in red) and microsatellite markers obtained from Weissenbach and coworkers (CEPH/Genethon, France), prior to the publication of the 1996 CEPH/Genethon microsatellite map (Dib *et al.*, 1996) depicted below. The genetic distances between markers are not drawn to scale.

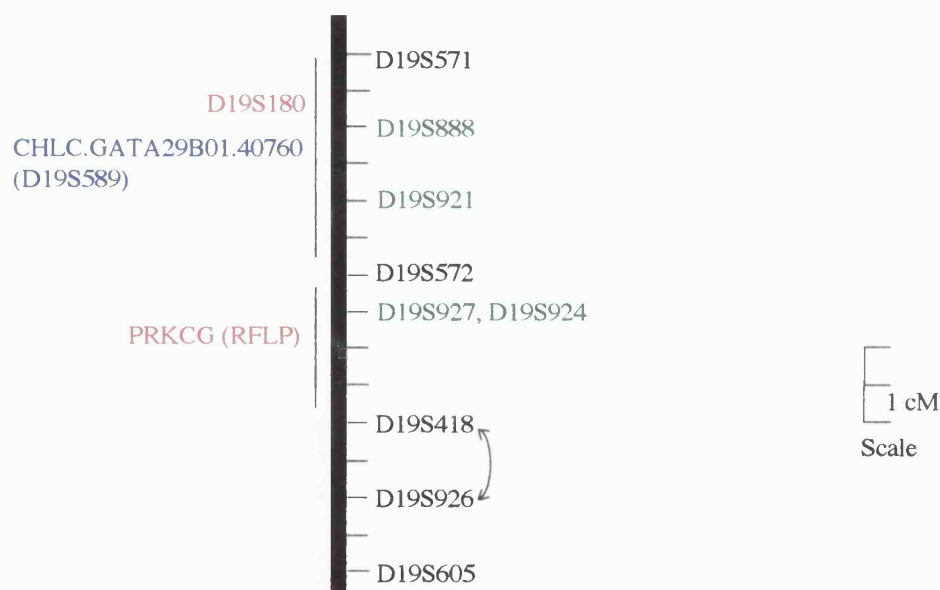


Figure 3.2 b

A representation of a sub-section of the 1996 CEPH/Genethon genetic map of 19q13.4 (Dib *et al.*, 1996) highlighting the reorientation of certain markers (indicated by arrow) in the RP11 region from that observed in figure 3.2a. The novel CEPH/Genethon markers (shown in green) as referred to in section 3.1.5 need to be integrated with the two NIH/CEPH markers (figure 3.2a) and the tetranucleotide marker (shown in blue) obtained from an integrated marker map of chromosome 19 generated at Cooperative Human Linkage Centre (CHLC) (<http://www.chlc.org/data/CHLCMaps/>).

With the publication of the updated 1996 CEPH/ Généthon map (Dib *et al.*, 1996) such an opportunity was provided to further refine the RP11 disease interval. The new map placed two new microsatellite markers D19S888 and D19S921 between D19S571 and D19S572 where the recombination events had been observed in ADRP5 and ADRP29. Therefore further locus refinement at the proximal end of the disease interval was a possibility. The map had re-located the markers D19S418 and D19S926 such that D19S418 was now more centromeric, and two other markers D19S924 and D19S927 were also placed between D19S572 and D19S418 (see figure 3.2b). In addition to these new Généthon markers, a novel tetranucleotide-repeat containing marker CHLC.GATA29B01.40760 (D19S589), which was localised to a genetic region corresponding to the proximal half of the RP11 interval by the Co-operative Human Linkage Centre (CHLC) (<http://www.chlc.org/data/CHLCMaps/>), was also ascertained for analysis

Focus was thus placed on refining the RP11 interval at the proximal end by analysing these new markers identified from updated genome maps in the families ADRP5 and ADRP29. Furthermore haplotype analysis in these families was expected to help integrate markers from different maps into the current Généthon map. These other markers included D19S180, a marker that had been cytogenetically mapped to 19q13.4 (Weber *et al.*, 1993), an RFLP marker associated with the gene *PRKCG* (NIH/CEPH, 1992) and the CHLC marker, D19S589 mentioned above (see figure 3.2b).

The recombination event in ADRP29, which refined the RP11 locus distally, had occurred between D19S926 and D19S418. With all the new microsatellite markers placed between D19S418 and D19S572 (and none placed between D19S418 and D19S926), it was not possible to achieve further locus refinement at the telomeric boundary with the recombinants in the available RP11 families. Therefore it was imperative to identify more RP11 linked families with possible new proximal and distal recombination events in order to achieve further locus refinement. With this in mind newly ascertained adRP families were haplotyped with markers mapping to the RP11 region of chromosome 19q13.4. Screening of more adRP families for this locus was also expected to give a more accurate estimate of the frequency of RP11.

3.2 Materials and methods

3.2.1 Clinical assessment and DNA samples

Most of the adRP patients/families were clinically examined at Moorfields Eye Hospital. In instances where a member of the family was unable to attend the clinical examination, blood samples were collected through their GPs and the disease status was verbally established without a clinical examination. The disease status of deceased members of the family was established through other members of the family. The mode of inheritance in each family was determined by the family history provided by the patients. DNA was extracted as described in section. 2.1.1.

3.2.2 Microsatellite analysis

This was carried out as described in section 2.14.1. Microsatellite markers located within the RP11 interval [as described in Généthon maps 1993-1994 (Gyapay *et al.*, 1994) to 1996 (Dib *et al.*, 1996) and Généthon data base at <http://www.genethon.fr>] were genotyped in 15 adRP families. A few control individuals from ADRP5 were included to maintain consistent numbering and sizing system of the alleles.

3.2.3 Linkage analysis

For all adRP pedigrees pairwise linkage analysis was performed between the disease phenotype and various marker loci located within the RP11 interval between D19S571 and D19S926. Data files were prepared on the LINKSYS data management package (Attwood and Bryant, 1988) and then transferred to the two-point linkage analysis program MLINK from the LINKAGE package version 5.1 (Lathrop and Lalouel, 1984) for lod score calculations. The genetic model used in the linkage analysis for the disease locus was of autosomal dominant inheritance with reduced penetrance estimated at 0.7 for the susceptible genotype based on the pedigree ADRP5 (see section 3.1.2). The disease gene frequency was set at 0.0001. Marker allele frequencies were calculated from 30 unrelated married-in individuals of the other adRP pedigrees used in this study.

3.3 Results

3.3.1 Analysis of new 19q markers in previously linked families ADRP5, ADRP29 and RP1907

Initially the new markers (D19S888 and D19S921) placed between D19S571 and D19S572 were tested in ten individuals from the branch of ADRP5 family, which presented the northern flanking crossover with D19S180 (see Fig 3.3A). These individuals (III-15, IV-31, IV-33 and V-36 of ADRP5) showed recombination with D19S888 but not with D19S921. Since D19S888 appeared to be genetically indistinguishable from the previous northern flanking marker, D19S180, analysis of these new proximal markers in this branch of ADRP5 did not refine the RP11 interval any further. Nevertheless this analysis localised D19S921 south of D19S180. Next, the tetranucleotide-repeat marker D19S589 was analysed in the complete ADRP5 pedigree to map it in relation to other 19q markers near the RP11 interval. D19S589 showed recombination with RP phenotype in individuals who were also recombinant with D19S180. Therefore as with D19S888, D19S589 could not be placed in relation to D19S180. However other recombination events in the family localised D19S589 between D19S571 and D19S921.

The markers D19S927 and D19S924 were also typed in the complete pedigree of ADRP5 to extend the haplotype within the RP11 interval and also to map these two markers in relation to other markers, such as the RFLP marker associated with the gene *PRKCG*. This marker has previously been typed in ADRP5 and mapped between D19S572 and D19S418 (Al-Maghteh *et al.*, 1996). As expected the markers D19S927 and D19S924 did not show recombination with the RP phenotype. However the genotyping of all these markers in ADRP5 enabled the allocation of the 19q markers within the RP11 region in the order, D19S571, (D19S888/180/589), D19S921, D19S572, (D19S924/927), *PRKCG*, D19S418 and D19S926 with D19S571 lying most centromeric. The markers D19S180, D19S888 and D19S589 were genetically indistinguishable from each other, as were D19S927 and D19S924. All new markers were also analysed in the families ADRP29 and RP1907 to confirm their genetic localisation and to explore further possibilities of locus refinement.

ADRP5

I

II

III

IV

V

- 1. D19S571
- 2. D19S180
- 3. D19S888
- 4. D19S589
- 5. D19S921
- 6. D19S572
- 7. D19S924
- 8. D19S927
- 9. PRKCG(RFLP)
- 10. D19S418
- 11. D19S926
- 12. D19S605

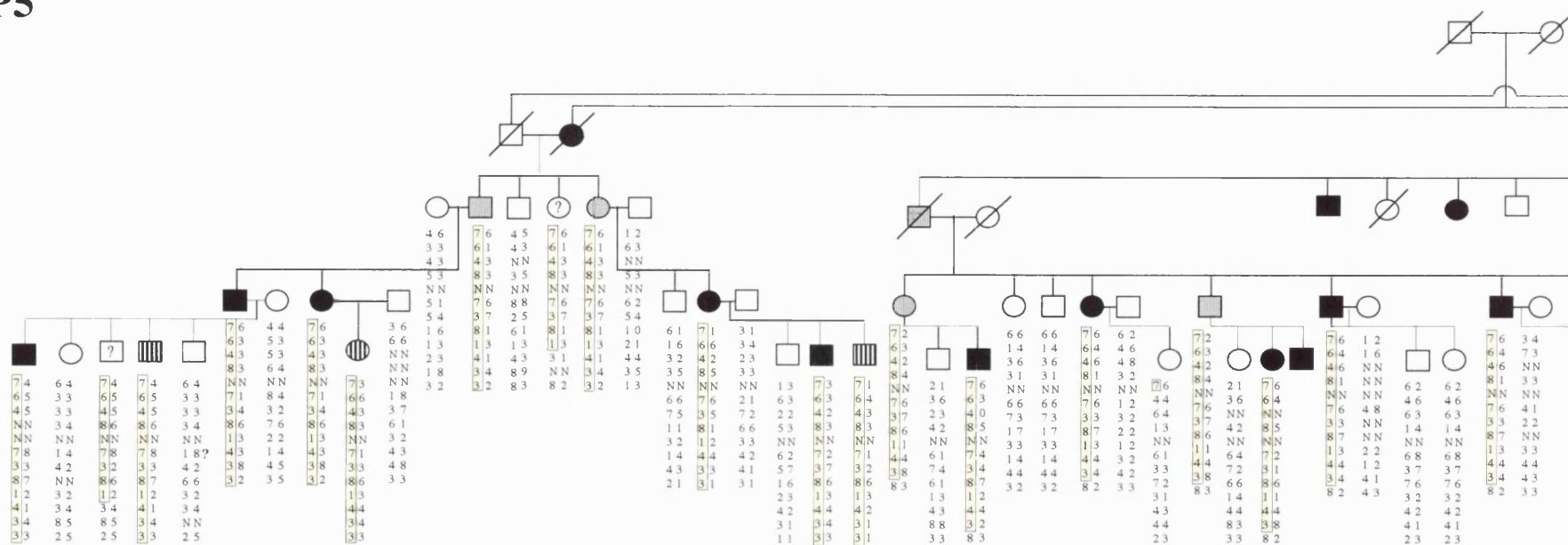


Figure 3.3 a

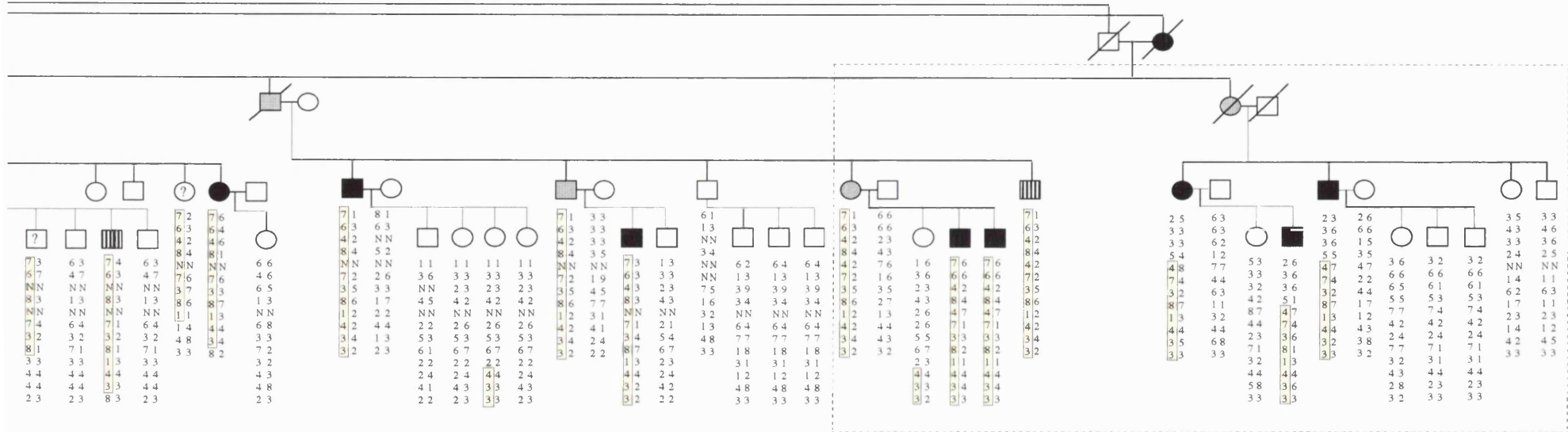
Haplotype analysis of 12 markers from the 19q13.4 in ADRP5, the disease haplotype is boxed in yellow. The marker D19S921 was only typed in the branch of ADRP5 that presented the cross over with D19S180, individuals genotyped are enclosed within the box. The figure legend described below is applicable for all pedigree presented henceforth.

FIGURE LEGEND

- ○ normal male and female symbols respectively.
- ● affected symbols.
- ◻ ◐ obligate disease gene carriers.
- ▨ ▩ disease haplotype carriers.
- ? ○ not known if normal or disease haplotype carriers since recombination break point took place within the RP11 interval.
- NN Genotype not known.

(Crossed symbols represent deceased family members).

Figure 3.3 a (contd)



ADRP29

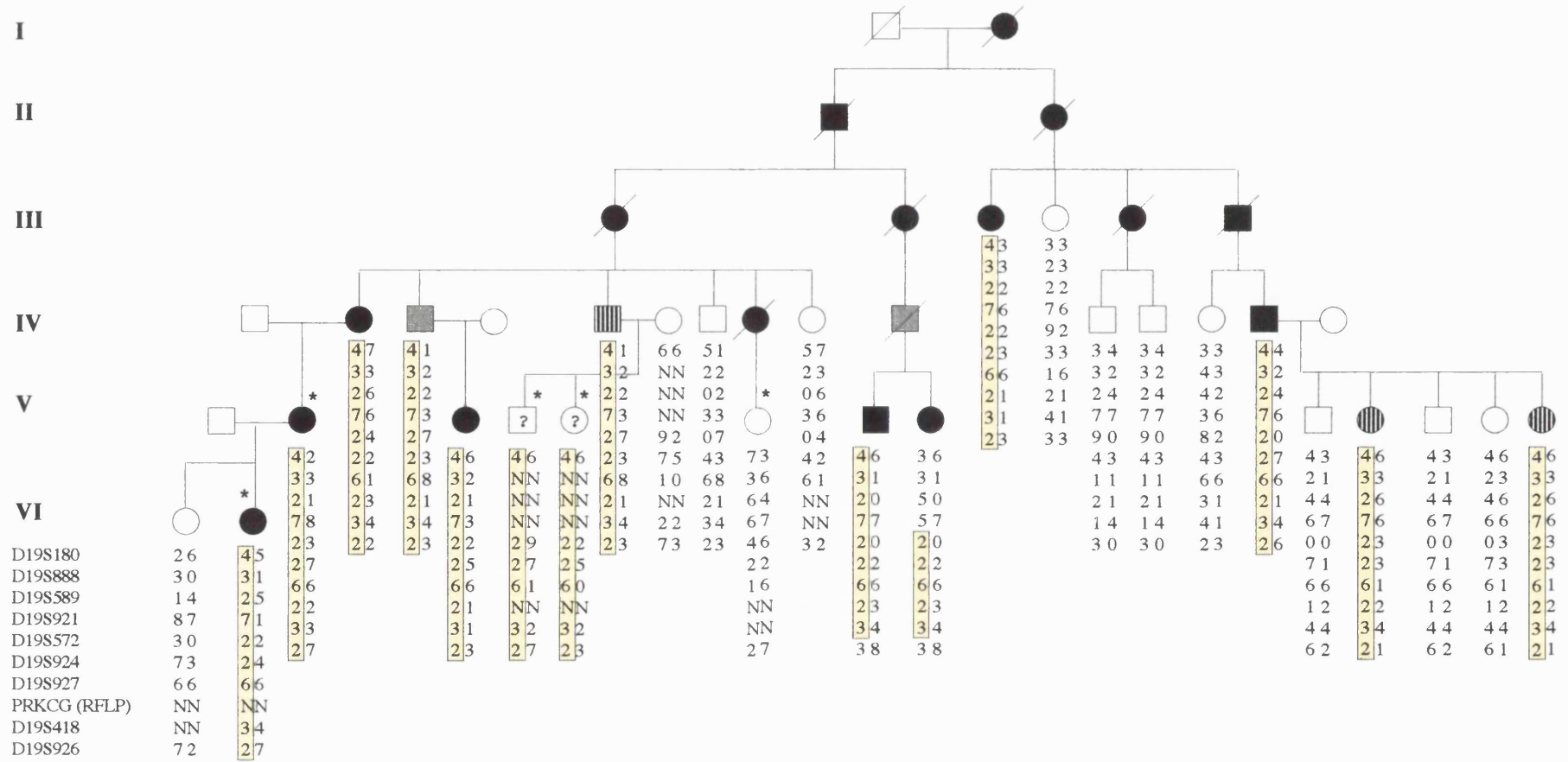


Figure 3.3b

Haplotype analysis of 10 markers from the 19q13.4 in ADRP29. Recombinant individual V-8 present a proximal crossover with marker D19S921 and a distal crossover with D19S926. Newly recruited individuals of the pedigree are indicated by (*).

RP1907

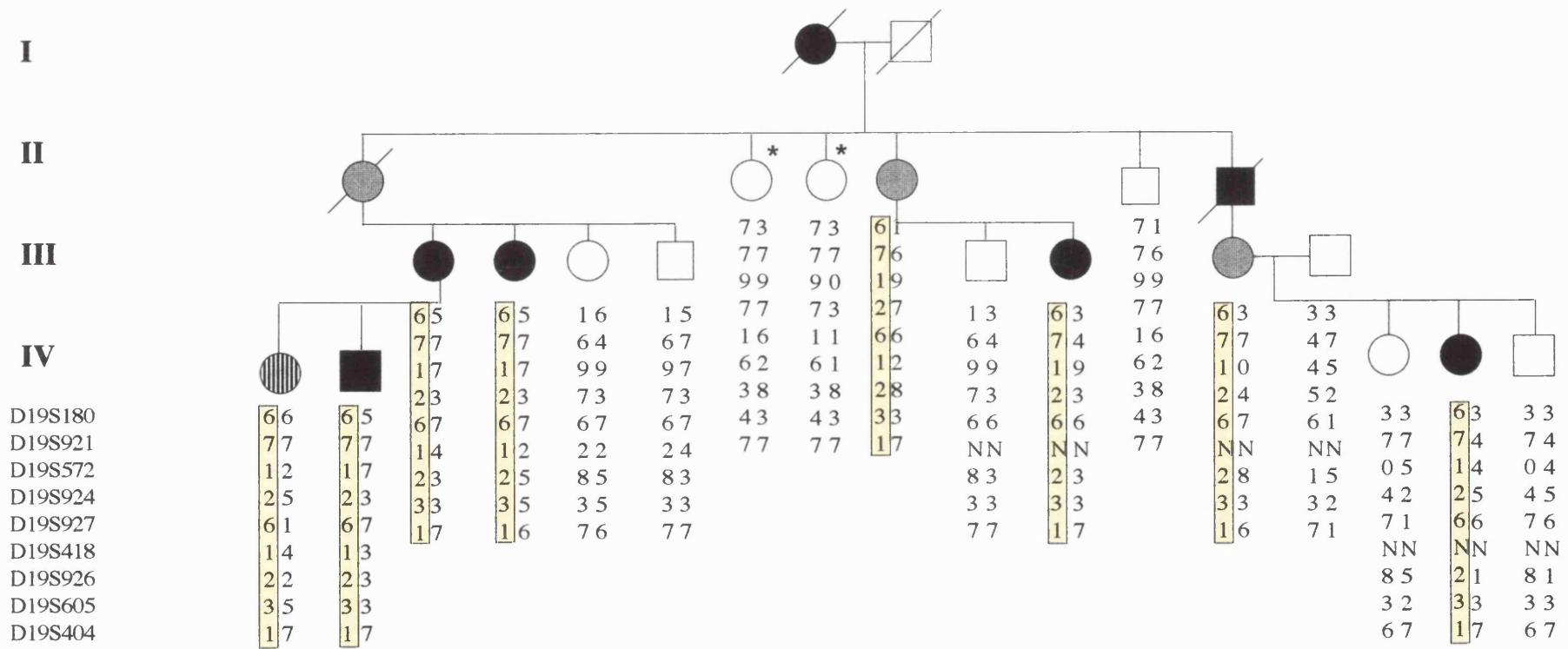


Figure 3.3c

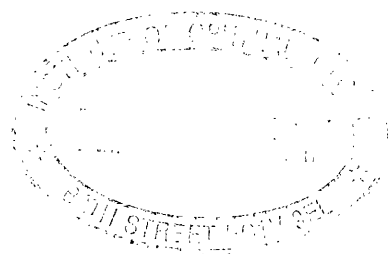
Haplotype analysis of 9 markers from 19q13.4 in RP1907. No recombinations were observed in this pedigree with markers that flank the RP11 interval. (*) indicate newly recruited individuals.

In ADRP29, markers D19S888, D19S589 and D19S921 all showed recombination with the RP phenotype in individual V-8 who presented the previous proximal crossover with D19S180. This recombination event, which had occurred between D19S921 and D19S572, further refined the RP11 locus proximally and changed the proximal flanking marker from D19S180 to D19S921. All these 19q13.4 markers were also typed in any newly recruited member of the families ADRP29 and RP1907. Such individuals represent new meiotic events and it is possible to identify new locus refining recombination events through haplotype analysis of these individuals. Unfortunately none of the new members (marked by a * in the pedigree figures 3.3b and c) presented further recombination events.

3.3.2 Identification of three additional families: ADRP2, ADRP11 and ADRP24 and refinement of the locus

In order to further refine the RP11 locus and to estimate the frequency of adRP caused by mutations at the RP11 locus (and other adRP loci), adRP families from Moorfields Eye Hospital genetic register and other sources were genotyped for markers between D19S571 and D19S418. In total 12 adRP families, which represented a mixture of both type I/D and type II/R adRP with or without incomplete penetrance were screened. Other members of the laboratory simultaneously screened these families for all previously mapped loci for adRP. Three families included in this study, ADRP2, ADRP11 and ADRP24 (fig. 3.4) were of particular interest as they exhibited partial penetrance, which is a characteristic feature of RP11 phenotype. On haplotype analysis the linkage of these three families to 19q13.4 became obvious.

ADRP2: Pairwise lod score analysis in ADRP2 demonstrated linkage to markers D19S571 and D19S572, with lod scores of 2.2 and 2.6, respectively, with no recombination ($\theta = 0$). A lod score of 2 is generally accepted as sufficient evidence of linkage of a disease to a previously known locus with similar phenotype, an understanding that is often referred to as “prior probability” (Terwilliger and Ott, 1994). Therefore with the partial penetrance phenotype in ADRP2, a lod score value of 2.6 was sufficient to prove linkage to RP11.



Some affected members of ADRP2 had pronounced macular degeneration, which is not unusual with the progression of adRP. However, the location of a cone rod dystrophy (CRD) locus proximal to the RP11 locus, prompted an investigation to see if ADRP2 excludes this CRD region, even though the clinical diagnosis of ADRP2 as an adRP family was unequivocal. CRD is located proximal to the RP11 locus on chromosome 19q13.3 between the markers D19S219 (proximal) and D19S246 (distal). CRD is phenotypically quite distinct from RP. It is initially characterised by loss of colour vision and visual acuity followed by night blindness and peripheral visual field loss accompanied by widespread retinal pigmentation and chorioretinal atrophy of the central and peripheral retina (Yagasaki and Jacobson, 1989; Szyrak *et al.*, 1993; Evans *et al.*, 1995). When the markers D19S246 and D19S604 that map distal to the CRD locus were typed in ADRP2 an informative recombination event was observed in individual III-10 with the marker D19S604 excluding ADRP2 from the CRD locus (see figure 3.4a). Further analysis of haplotype data of ADRP2 also revealed a recombination event between the disease and the marker D19S418 in the affected individual II-8, which changed the telomeric flanking marker for RP11 from D19S926 to D19S418. This recombination event instigated the typing of the RFLP marker (an *MspI* polymorphism) associated with the *PRKCG* gene, which has been genetically localised proximal to D19S418, in ADRP2 to see if this gene could also be excluded genetically. However the haplotype analysis did not reveal a recombination event between *PRKCG* and disease in the individual II-8, indicating that the cross over has occurred distal to *PRKCG*. Therefore the *PRKCG* gene was placed within the RP11 interval and it was to be regarded as a candidate gene (see figure 3.4a). Similar to *PRKCG*, the gene for human interleukin-11 (*IL11*) first localised to 19q13.3-13.4 region by *in situ* hybridisation (McKinley *et al.*, 1992), was also fine mapped by analysing an intragenic marker of *IL11*, developed by a colleague (Bellingham *et al.*, 1997), in ADRP2. The segregation analysis of this intragenic *IL11* marker in the recombinant individuals of ADRP2, localised *IL11* distal to the marker D19S572 (see figure 3.4a).

ADRP11: In family ADRP11 maximum lod scores of 4.24 and 3.47 were obtained with markers D19S924 and D19S927, respectively, at $\theta = 0$ (Table 3.1). Haplotype analysis in this family (fig 3.4b) revealed another recombination event with the marker D19S921 in affected individual III-9 (in addition to that observed in ADRP29). ADRP11 also had several normal recombinant individuals useful for marker localisation, therefore the *IL11* marker was also typed in ADRP11. The segregation analysis in one such normal recombinant individual helped localise *IL11* gene distal to the distal flanking marker, D19S418, thus excluding this gene from the RP11 interval. Unification of haplotype data from ADRP2, ADRP11 and ADRP29 refine the RP11 locus to a 5.8cM genetic interval between D19S921 and D19S418. Furthermore, association of different haplotypes with the disease in ADRP2 and ADRP11 exclude the possibility of the two families being related to each other (founder effect) (fig. 3.5).

3.3.3 Founder effect in RP1907 and ADRP24

The clinical information on ADRP24, family 2 described by Moore *et al.* (1993), suggested this family as a good candidate for the RP11 locus. ADRP24 had previously been analysed with 19q markers but linkage analysis had given inconclusive results due to the small size of this family. On reanalysis of ADRP24 with new 19q markers and haplotype analysis, linkage to RP11 was suggested due to the segregation of a common haplotype among the affected individuals. However linkage analysis still gave poor evidence of linkage to RP11 (Table 3.1). A founder event was then discovered between ADRP24 and RP1907, a family already linked to RP11.

The first indication that these two families might be related came with the discovery, in both ADRP24 and RP1907, of a rare nucleotide substitution in the gene *PRKCG* located within the RP11 disease interval (see section 4.6.2.1). Following a haplotype comparison between ADRP24 and RP1907 this relationship was confirmed with the identification of a shared haplotype. Ordinarily when genotyping adRP families with microsatellite markers from the RP11 interval, control individuals from ADRP5 was included as a matter of routine, to ensure a consistent numbering system of alleles and also to investigate the possibility of founder effect(s) in families linked to the RP11 locus. Unfortunately in the earlier analysis of the family ADRP24 such

control individuals were not included, hence the founder relationship between ADRP24 and RP1907 escaped detection, prior to the discovery of mutation in *PRKCG*. Figure 3.5 depicts the five different haplotypes that segregate with the RP phenotype, in each of the five RP11 linked families including the shared haplotype of RP1907 and ADRP24. The probability of the allele combination in ADRP24 (and RP1907) occurring by chance was calculated by multiplying together the individual frequencies (in the general population) of these alleles shared between D19S572 and D19S926. The value thus obtained was 0.000011, which meant that the probability of this allele combination occurring just by chance was 1 in 100,000. This made the converse, a founder effect between ADRP24 and RP1907, the more likely possibility even though no genealogical link has been found within the last 160 years. The shared haplotype of RP1907 and ADRP24 extends from D19S921 to D19S404 (proximal to distal). As the shared haplotype does not extend proximally beyond D19S921 it provided evidence for another proximal recombination event at D19S921 in addition to the one described in the family ADRP11 in the previous section. Affected individuals of ADRP29 also have identical alleles for both D19S924 and D19S927 as RP1907 (and ADRP24). With the frequencies of allele 2 of D19S924 and allele 6 of D19S927 being 0.16 and 0.30, respectively, the probability of this allele combination occurring in the general population by chance is 0.05. However the haplotype between ADRP29 and RP1907 differ either side of these two markers, therefore the most likely possibility in this case is that these two families are unrelated and that haplotype sharing is a chance occurrence.

(1) ADRP2

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4	Z_{\max}
D19S571	2.18	2.14	2.01	1.82	1.40	0.91	0.38	2.18
D19S180	0.87	0.85	0.78	0.70	0.52	0.35	0.17	0.87
D19S572	2.60	2.56	2.40	2.18	1.67	1.09	0.46	2.60
D19S924	1.98	1.95	1.83	1.65	1.25	0.79	0.29	1.98
D19S927	0.87	0.85	0.79	0.70	0.52	0.35	0.17	0.87
PRKCG (RFLP)	0.74	0.73	0.69	0.62	0.43	0.23	0.06	0.74
D19S418	- 1.90	0.55	1.11	1.21	1.07	0.73	0.31	1.21
D19S926	- 1.91	0.55	1.11	1.21	1.06	0.73	0.31	1.21

(2) ADRP11

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4	Z_{\max}
D19S571	- 3.30	- 1.16	- 0.45	- 0.17	0.03	0.06	0.03	0.06
D19S589	- 5.97	- 1.45	- 0.26	0.08	0.19	0.12	0.03	0.19
D19S180	- 3.14	- 2.33	- 1.07	- 0.56	- 0.16	- 0.02	0.02	0.02
D19S921	- 4.96	- 0.52	0.66	0.97	0.94	0.62	0.02	0.97
D19S572	2.22	2.19	2.03	1.82	1.35	0.86	0.39	2.22
D19S924	4.24	4.17	3.87	3.47	2.60	1.66	0.70	4.24
D19S927	3.47	3.41	3.14	2.78	2.03	1.24	0.50	3.47
D19S418	1.13	1.11	1.01	0.88	0.59	0.31	0.09	1.13
D19S926	2.75	2.70	2.47	2.17	1.54	0.88	0.30	2.75

(3) ADRP24

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4	Z_{\max}
D19S180	- 0.07	- 0.06	- 0.03	- 0.01	0.01	0.01	0.00	0.01
D19S921	- 0.14	- 0.13	- 0.09	- 0.05	- 0.01	0.00	0.00	0.00
D19S572	0.57	0.56	0.53	0.49	0.36	0.23	0.10	0.57
D19S924	0.13	0.13	0.14	0.14	0.11	0.07	0.02	0.14
D19S927	0.1	0.10	0.11	0.12	0.10	0.06	0.02	0.12
D19S418	0.24	0.24	0.27	0.28	0.25	0.18	0.09	0.28
D19S926	0.22	0.23	0.25	0.26	0.24	0.17	0.08	0.26

Table 3.1.

Two point lod scores between 19q13.4 markers (shown in order from proximal to distal) and disease phenotype in ADRP2 and ADRP11. Lod scores were calculated at a disease penetrance of 0.7.

ADRP2

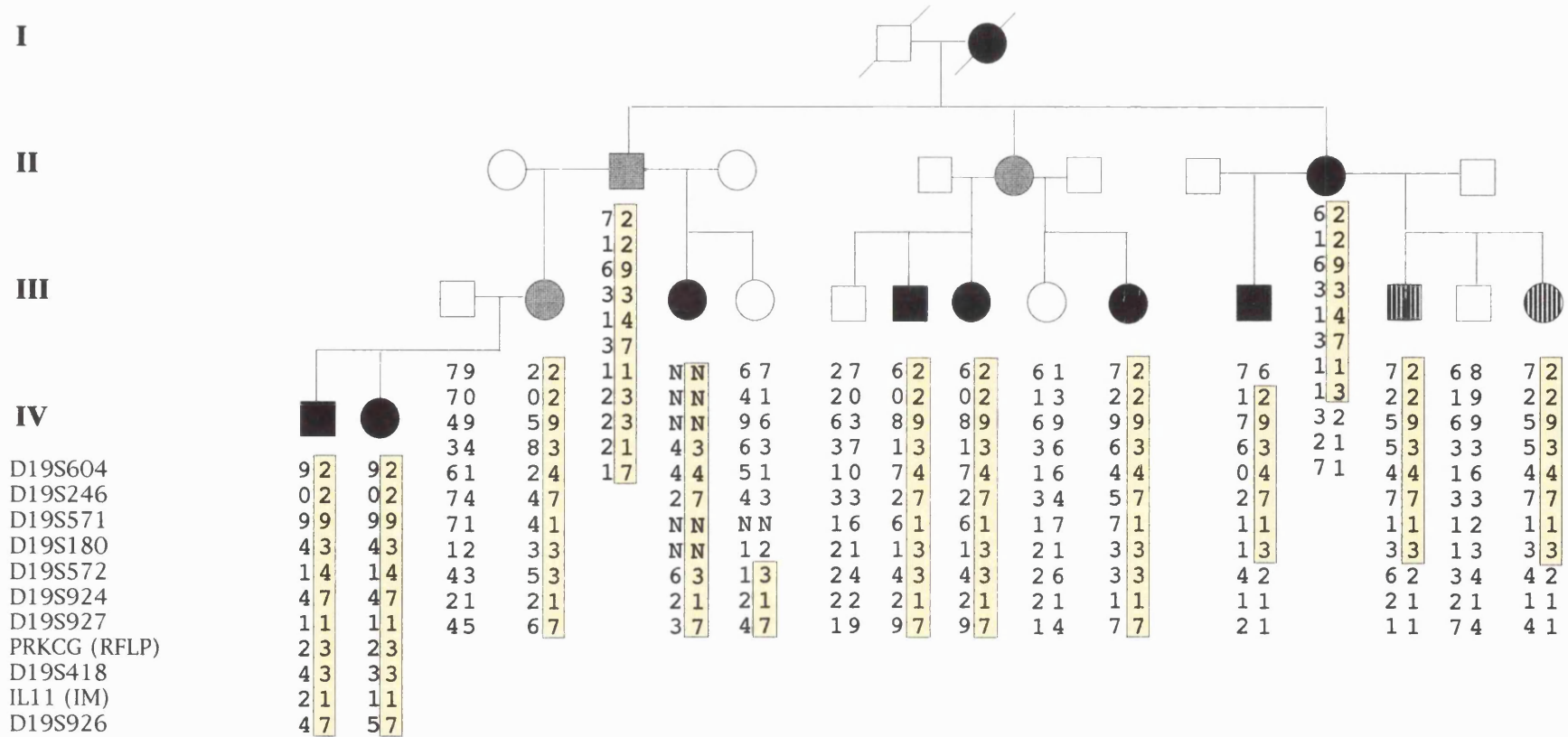


Figure 3.4 a

Haplotype analysis of 11 markers from 19q13.4 in ADRP2. According to recombination events observed in this pedigree the disease gene is localised between D19S246 (proximal) and D19S418 (distal) hence excluding the CRD locus. The affected haplotype is boxed. The generalised legend for RP11 pedigrees is presented with figure 3.3a.

ADRP24

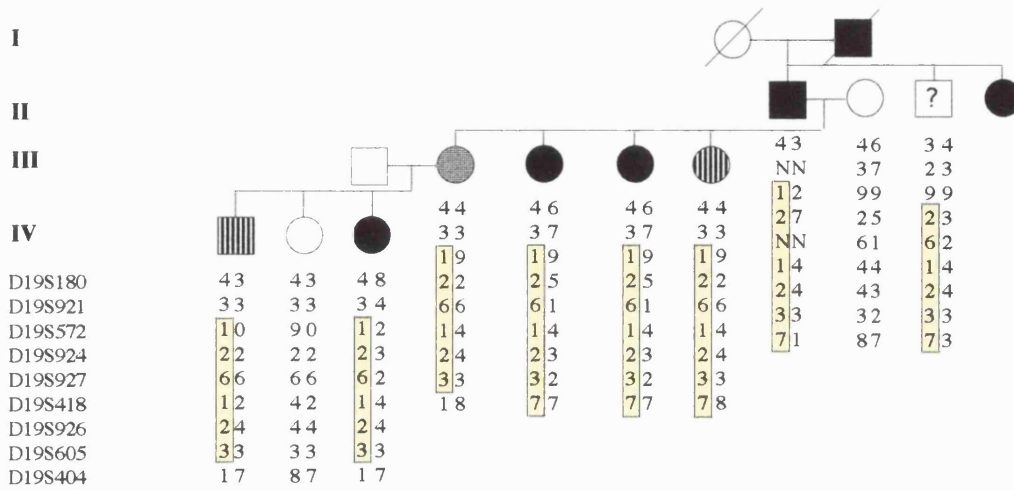


Figure 3.4 c

Haplotype analysis of 9 markers from 19q13.4 in ADRP24. The affected haplotype found in common with RP1907 is boxed in yellow thus indicating a crossover event with D19S921.

RP11 FAMILY \ MARKER	ADRP24	RP1907	ADRP29	ADRP5	ADRP2	ADRP11
D19S571	N	N	1	7	9	6
D19S180	4	6	4	6	3	6
D19S888	N	N	3	4	N	N
D19S589	N	N	2	8	N	4
D19S921	3	7	7	4	N	2
D19S572	1	1	2	7	4	0
D19S924	2	2	2	3	7	5
D19S927	6	6	6	8	1	7
PRKCG (RFLP)	N	N	2	1	3	N
D19S418	1	1	3	4	3	1
D19S926	2	2	2	3	7	4
D19S605	3	3	N	3	N	N
D19S404	7	1	N	N	N	N

Figure 3.5

Haplotypes of the six RP11 linked families constructed using a range of markers from 19q13.4. The different haplotypes exclude founder effect except in ADRP24 and RP1907. The shared haplotype is shaded. 'N' indicate instances where the genotype for a given marker is not known.

3.3.4 Analysis of sib pairs in RP11 families to see evidence of ‘allelic effect’

In section 3.1.3, allelic effect was suggested as the possible genetic mechanism for the reduced penetrance of mutations in the RP11 locus. The sib pair analysis carried out by McGee *et al* (1997) showed statistically significant correlation between the inheritance of the wild type allele from the non-carrier parent and presence of disease in carrier offspring. The data for this analysis was derived from pedigrees ADRP5, ADRP29, RP1907, which are described in this study, and the previously reported Japanese family (Xu *et al.*, 1995). With the discovery of two new RP11 families, ADRP2 and ADRP11 and pedigree extension of RP1907 (ADRP24), additional haplotype data was available and they were analysed to further test the phenomenon of allelic effect. If allelic effect does determine the disease status, the non-carrier parent of sibships in which there were at least one affected carrier and one unaffected carrier would necessarily be heterozygous for any wild-type alleles that specify the penetrance of the pathogenic RP11 allele inherited from the carrier parent. Moreover, among the offspring who inherit the pathogenic allele from the carrier parent, the symptomatic siblings would all inherit one RP11 allele from the noncarrier parent and the asymptomatic siblings would inherit the other (McGee *et al.*, 1997). To test this, all sibships from the pedigrees ADRP5, ADRP29, RP1907, ADRP2 and ADRP11, in which there were at least one affected and one unaffected carrier were placed into four categories according to the concordance or discordance of phenotype (symptomatic vs. asymptomatic) and of genotype (inheriting the same vs. different haplotype from the noncarrier parent) (table 3.2). The tabulated results show the non-random distribution of sib pairs. All 14 sib pairs in which both members had the same phenotype inherited the same haplotype from the noncarrier parent; 20 of 24 sib pairs with different phenotypes inherited different haplotypes. Two more sib pairs with contrasting phenotype but identical genotype was identified in the extended pedigree of ADRP5 in addition to the two sib pairs mentioned in section 3.1.3 (see fig 3.3a). The total number of sib pairs analysed in this study was greater than the numbers analysed in the previous study (McGee *et al.*, 1997), with a still greater number of sib pairs conforming to the model. To evaluate the statistical significance of data, a standard 2 X 2 chi-squared test (Siegal and Castellan, 1988) was conducted to see if there was a phenotype genotype correlation (i.e. an allelic effect). The computed X^2

value (with a degree of freedom of 1) corresponded to a probability value of $P \ll 0.001$, therefore strongly supporting the proposed allelic effect model.

There was an added impetus for this analysis. If the allelic effect is true and the RP11 linked marker haplotype inherited from the noncarrier parent determines the disease status, then a comparison of this haplotype inherited from the noncarrier parent in sibships in which all carriers had the same disease status, could also fine localise the region which harbours the modifier gene (if a closely linked locus to RP11 influence the disease status of RP11). However if the penetrance is influenced by the otherwise silent alleles at the RP11 locus itself, then such an analysis could lead to further locus refinement of RP11. To investigate this possibility those sibships from all RP11 linked families in which all the carriers had the same disease status were analysed, and the desired sibships were identified in ADRP11 and RP1907. In the symptomatic siblings III-9 and III-10 of ADRP11 the haplotype inherited from their noncarrier parent was only identical telomeric of the marker D19S927 (see fig 3.4b and fig 5.20b). Therefore if the disease status is indeed influenced by the otherwise silent homologous allele at the RP11 locus, the disease gene must lie distal to D19S927. In RP1907 the siblings III-1 and III-2 who were also symptomatic shared an identical noncarrier haplotype centromeric of the marker D19S418 (see figure 3.3c and figure 5.20c). Therefore application of the same premises localise the disease gene proximal to D19S418. This localisation is also in agreement with the previous RP11 locus refinement achieved with recombination events in ADRP2 (section 3.3.2). The compilation of this data localise the 'penetrance influencing locus' of RP11 between D19S927 and D19S418, which is a narrower refinement than what was achieved by recombination analysis within affected individuals of RP11 linked families.

Family and Sibpair Phenotype Genotype	Phenotype/Genotype correlation			
	Same Same	Same Different	Different Same	Different Different
ADRP5				
IV-8/IV-11				X
IV-8/IV-13	X			
IV-8/IV-14				X
IV-8/IV-16				X
IV-8/IV-21				X
IV-11/IV-13				X
IV-11/IV-14	X			
IV-11/IV-16	X			
IV-11/IV-21	X			
IV-13/IV-14				X
IV-13/IV-16				X
IV-13/IV-21				X
IV-14/IV-16	X			
IV-14/IV-21	X			
<u>IV-16/IV-21</u>	X			
IV-23/IV-25			X	
IV-23/IV-28			X	
IV-23/IV-30			X	
IV-25/IV-28	X			
IV-25/IV-30	X			
<u>IV-28/IV-30</u>	X			
V-1/V-4			X	
V-8/V-9				X
ADRP29				
IV-2/IV-3				X
IV-2/IV-5				X
IV-3/IV-5	X			
ADRP2				
II-2/II-8				X
ADRP11				
III-1/III-3				X
III-1/III-6	X			
III-3/III-6				X
ADRP24				
III-2/III-3				X
III-2/III-4				X
III-2/III-5	X			
III-3/III-4	X			
III-3/III-5				X
<u>III-4/III-5</u>				X
IV-1/IV-3				X
RP1907				
IV-1/IV-2				X
Total no	14	0	4	20

Table 3.2.

Phenotype/Genotype correlation of sib pairs with at least one discordant pair.

3.3.5 Exclusion of families

Of the 12 adRP families, which represented a mixture of both type I/D and type II/R adRP with or without incomplete penetrance, subjected to haplotype analysis, only three families ADRP11, ADRP2 and ADRP24 showed linkage to 19q. Since ADRP24 was related to RP1907 only two new RP11 families were identified from this survey. The rest of the families were either excluded and/or subsequently linked to another RP locus (Inglehearn *et al.*, 1998). A few families had evidence of partial penetrance phenotype yet the affected individuals did not have a common haplotype nor give a significant lod score value on linkage analysis. The results of the haplotype and linkage analysis (see table 3.3) are given below for each excluded family categorised as follows.

3.3.5.1 Families linked to other RP loci: US family and ADRP 8

As mentioned earlier, testing of these 12 adRP families for all known adRP loci were conducted in parallel, therefore some families were linked to other RP loci while they were being tested for linkage to 19q. Consequently these families were not haplotyped for many markers from the RP11 interval. Two adRP families known as the US family and ADRP8 were linked to other RP loci. The US family was linked to the rhodopsin locus and disease causative mutation was identified as CCC to CAC (Pro-His) change in codon 23 of rhodopsin gene (Inglehearn *et al.*, 1998). The family ADRP8 was linked to the recently identified adRP locus on chromosome 1cen (Xu *et al.*, 1996; Inglehearn *et al.*, 1998).

3.3.5.2 Families with tentative evidence of partial penetrance: ADRP 2228, ADRP 83 and ADRP 2752

Of the 12 adRP families included in the study, ADRP2, ADRP11 and ADRP24 were not the only families to show indications of partial penetrance, even though the clinical data available for ADRP2, ADRP11 and ADRP24 was more detailed and complete than that available for the families described here. The presence of partial penetrance was deemed to exist in these families due to the presence of at least one obligate carrier, who was clinically asymptomatic. In some cases partial penetrance was inferred according to information given by a family member.

In family ADRP2228 individuals II-2, II-4 and III-5 were clinically asymptomatic and partial penetrance was suggested to explain the pattern of inheritance. Even though this may have suggested ADRP2228 to be a family linked to RP11, some affected individuals had very mild retinitis pigmentosa, which is unlike the typical clinical phenotype described for RP11, where the affected individuals generally have severe and early onset of disease. Haplotype analysis of this family revealed that individual II-2 and II-4 transmit two completely different haplotypes to their affected offsprings. However since the haplotype of II-2 was inferred, exclusion on this basis is not satisfactory, moreover linkage analysis gave poor exclusion of RP11 locus (lod score of < -2 was only given for the marker D19S572 at θ value of 0). Since the data is inconclusive, this family should be tested for mutations once the RP11 gene has been cloned.

In family ADRP2752, individual II-1 appears to be an asymptomatic carrier. It is possible that this individual also had very mild form of RP since there is no clinical information available on this individual. This family was excluded on the basis of affected individual II-4 who was recombinant for the entire RP11 region.

ADRP83 is a Spanish pedigree with incomplete clinical information, once again partial penetrance was thought to exist on the basis of individual II-2. Similar to ADRP2752, ADRP83 was excluded from RP11 locus on the basis of the recombinant individual III-4. In RP83, individuals III-1, IV-2 and IV-8 also had the affected haplotype, yet in these individuals the disease status was not known. Hence they cannot be classified as normal recombinant individuals. Even so the presence of normal recombinants cannot be argued as basis of exclusion due to the partial penetrance phenotype of RP11. Linkage analysis gave unsatisfactory results since a majority of markers tested gave lod scores of -2 at θ values of 0.01 or less. As with ADRP2228 this family should also be tested for mutations once the RP11 gene has been cloned.

3.3.5.3 Families excluded from RP11: ADRP33, ADRP710, Bristol family and ADRP 730

These families were also excluded on the basis of the presence in these families of at least one affected cross over. The linkage analysis gave unsatisfactory exclusions due to the small size of the families (see table 3.3). ADRP33 has been excluded from all known adRP loci and provide evidence for further genetic heterogeneity of adRP.

Table 3.3.

Two point lod scores between 19q13.4 markers (shown in order from proximal to distal) and disease phenotype in adRP families described in section 3.2.2. Lod scores were calculated at a disease penetrance of 0.7.

(1) US family

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S572	- 4.64	- 1.20	- 0.50	- 0.21	0.02	0.08	0.07
D19S418	- 3.57	- 2.69	- 1.39	- 0.86	- 0.40	- 0.20	- 0.08
D19S926	0.49	0.49	0.50	0.50	0.45	0.35	0.20

(2) ADRP8

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S572	- 6.20	- 2.25	- 0.95	- 0.46	- 0.11	- 0.00	0.01
D19S927	- 3.19	- 1.40	- 0.72	- 0.44	- 0.19	- 0.08	- 0.02

(3) ADRP2228

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S180	- 0.14	- 0.14	- 0.12	- 0.1	- 0.06	- 0.03	- 0.01
D10S921	- 0.80	- 0.72	- 0.51	- 0.35	- 0.17	- 0.07	- 0.02
D19S572	- 3.67	- 1.05	- 0.40	- 0.17	- 0.02	0.01	0.01
D19S924	0.28	0.27	0.25	- 0.16	0.08	0.03	0.00
D19S927	- 0.63	- 0.57	- 0.40	- 0.28	- 0.13	- 0.05	- 0.01
D19S418	0.46	0.44	0.42	0.30	0.17	0.08	0.02
D19S926	- 3.55	- 1.06	- 0.78	- 0.19	- 0.03	0.00	0.00

(4) ADRP2752

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S571	- 3.20	- 1.70	- 0.98	- 0.64	- 0.30	- 0.12	- 0.03
D19S180	- 3.51	- 1.89	- 1.13	- 0.77	- 0.38	- 0.16	- 0.04
D19S572	- 3.68	- 1.89	- 1.13	- 0.77	- 0.38	- 0.16	- 0.04
D19S418	- 0.00	- 0.00	- 0.00	- 0.00	- 0.00	0.00	0.00
D19S926	0.12	0.12	0.13	0.14	0.14	0.12	0.07

(5) ADRP83

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S589	- 3.21	- 0.65	0.03	0.30	0.43	0.35	0.17
D19S180	- 0.00	- 0.00	- 0.00	- 0.00	- 0.00	- 0.00	- 0.00
D19S572	- 2.70	- 0.19	0.45	0.66	0.71	0.55	0.27
D19S924	- 3.15	- 1.05	- 0.34	- 0.04	0.19	0.24	0.17
D19S927	- 3.39	- 1.25	- 0.55	- 0.27	- 0.04	0.03	0.04
D19S926	- 2.35	- 1.69	- 0.46	0.04	0.38	0.40	0.26

(6) ADRP33

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S571	- ∞	- ∞	- ∞	- ∞	- ∞	- ∞	- ∞
D19S180	- 0.45	- 0.43	- 0.36	- 0.26	- 0.11	- 0.03	0.00
D19S572	- 3.24	- 1.00	- 0.33	- 0.08	0.07	0.08	0.05
D19S418	0.49	0.48	0.43	0.35	0.20	0.08	0.01
D19S926	- 3.23	- 1.11	- 0.46	- 0.23	- 0.07	- 0.03	- 0.02

(7) ADRP710

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S180	- 0.69	- 0.66	- 0.55	- 0.42	- 0.23	- 0.11	- 0.03
D19S572	- 0.39	- 0.37	- 0.31	- 0.24	- 0.14	- 0.07	- 0.02
D19S927	- 0.48	- 0.45	- 0.36	- 0.27	- 0.14	- 0.06	- 0.02

(8) Bristol family

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S571	- 3.46	- 1.38	- 0.66	- 0.35	- 0.06	0.05	0.06
D19S180	- 3.17	- 1.07	- 0.43	- 0.18	0.02	0.08	0.07
D19S572	- 3.02	- 1.07	- 0.42	- 0.17	0.03	0.08	0.07
D19S418	- 2.82	- 0.70	- 0.05	0.19	0.34	0.33	0.21
D19S926	0.75	0.75	0.71	0.66	0.55	0.41	0.23

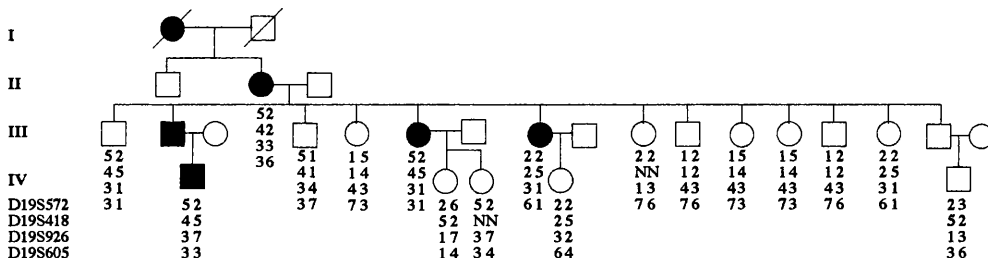
(9) ADRP 730

Marker	$\theta = 0$	0.01	0.05	0.1	0.2	0.3	0.4
D19S571	- 3.65	- 1.45	- 0.76	- 0.47	- 0.21	- 0.08	- 0.02
D19S180	- 3.64	- 1.05	- 0.40	- 0.16	0.02	0.07	0.06
D19S572	0.32	0.31	0.30	0.27	0.21	0.15	0.08
D19S418	- 3.91	- 1.44	- 0.72	- 0.40	- 0.12	- 0.01	0.02
D19S926	- 3.26	- 0.69	- 0.06	0.15	0.26	0.22	0.12

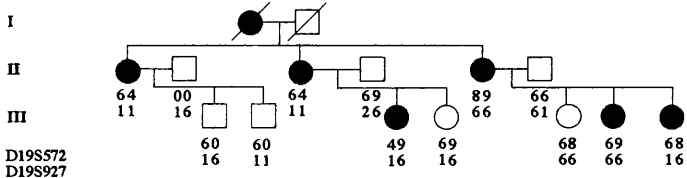
Figure 3.6

Excluded families with haplotypes constructed using markers from RP11 region of 19q13.4. For families that have not yet been linked to a known locus (e.g all families apart from US family and ADRP8) the individual that recombine for the whole of RP11 region is indicated by (*).

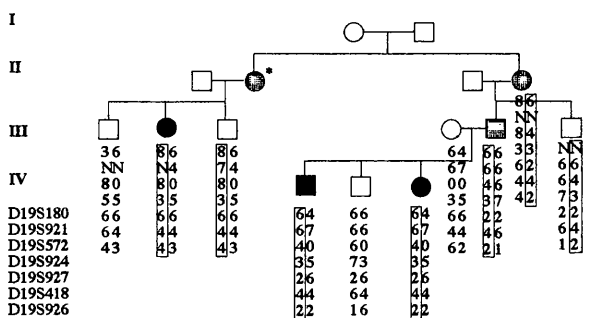
US -Family



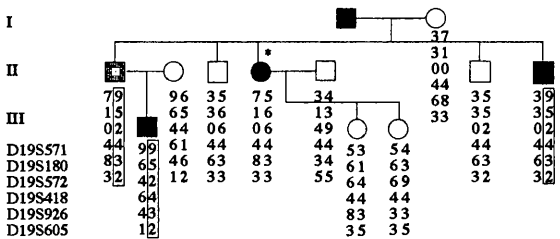
ADRP8



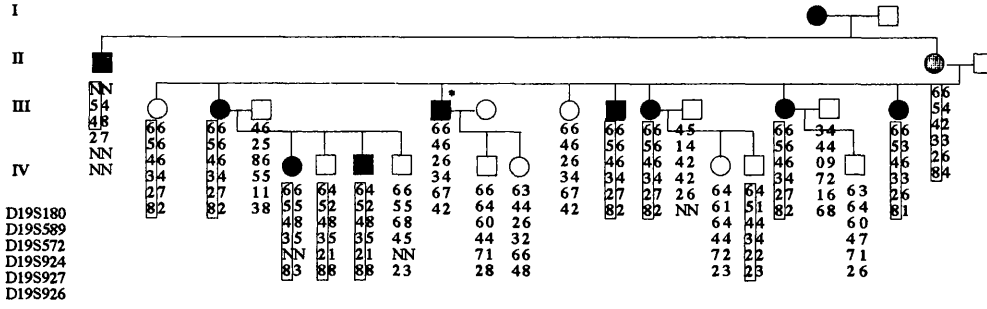
ADRP 2228



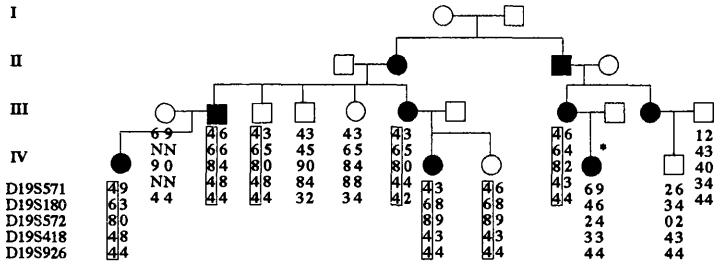
ADRP 2752



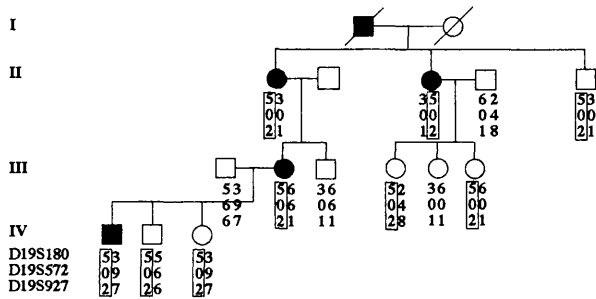
ADRP83



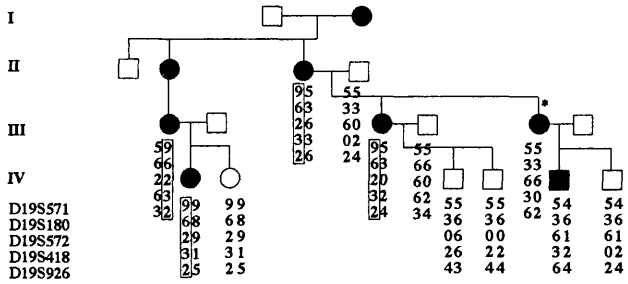
ADRP33



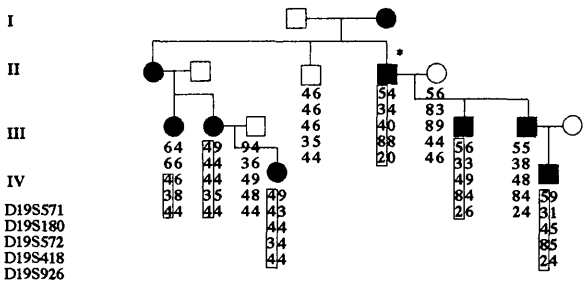
ADRP 710



Bristol RP



ADRP 730



3.4 Discussion

3.4.1 Refinement of the RP11 locus

Locus refinement based on linkage analysis of ADRP5 and ADRP29 had previously placed the RP11 gene between D19S180 and D19S926 (Al-Magthteh *et al* 1996). Haplotype analysis of new microsatellite markers in ADRP29 then constrained the RP11 interval further at its centromeric boundary, and defined D19S921 as the new proximal flanking marker. Another proximal recombination event with D19S921 was also discovered in ADRP11, one of the two newly identified families (Vithana *et al.*, 1998). ADRP2 the other newly discovered RP11 family, introduced a distal crossover at marker D19S418 refining the RP11 interval to a 5.8cM genetic distance between D19S921 and D19S418. Recently McGee *et al* (1997) reported a further refinement, defining D19S572 as the proximal flanking marker of the RP11 interval. Collating all this data now confines the RP11 interval to a 4.1cM genetic distance between the markers D19S572 and D19S418. The genetic distance between markers is the same as published by Dib *et al.* (1996). This refinement, while excluding the gene *IL11* from the RP11 interval, placed the gene for protein kinase C gamma (*PRKCG*) within the RP11 interval. The analysis of this gene as a candidate for RP11 is described in chapter four. The physical distance between D19S572 and D19S418 has been approximated at 1.5 Mb by the Lawrence Livermore National Laboratories (LLNL, USA), who were involved in the construction of a physical map of chromosome 19 in its entirety (Ashworth *et al.*, 1995). Therefore sufficient genetic refinement of RP11 interval has been achieved to enable the construction of our own YAC contig spanning the RP11 region.

3.4.2 Allelic effect and refinement of the RP11 disease penetrance influencing locus

The simplistic sib pair analysis conducted in this study corroborated the findings of the previous study (McGee *et al.*, 1997) in which a statistically significant correlation was observed between the development of RP in a carrier and the inheritance of the region around RP11 from the non-carrier parent. It also inferred that this penetrance-determining factor might be a closely linked modifier gene to RP11 or even be the otherwise silent alleles of the RP11 locus it self. In order to constrain this

penetrance-determining locus, the noncarrier haplotype of sibships in which all the carriers had the same disease status was analysed for locus refining recombination events. Such recombination events in two sibships from ADRP11 and RP1907 constrained this region between D19S927 and D19S418 to a genetic interval of 2.8 cM. If the penetrance of RP11 is indeed influenced by the RP11 allele inherited from the non carrier parent then this refinement also constrain the primary disease mutation between D19S927 and D19S418.

3.4.3 Frequency of the RP11 locus

Autosomal dominant retinitis pigmentosa (adRP) has been accounted for 19% of all RP in a US survey (Bunker *et al.*, 1984) and 22% in a UK survey (Bundy and Crews 1984) (see section 1.10.1.1). adRP it self is heterogeneous and mutations in the Rhodopsin gene have been estimated to account for between 20 and 31% of adRP (Sung *et al.*, 1991; Dryja *et al.*, 1991; Inglehearn *et al.*, 1992), while RDS/peripherin mutations account for approximately 5% (Wells *et al.*, 1992; Rosenfield and Dryja, 1995). To date mutations in the *NRL* gene have only been identified in a single adRP pedigree (personal. comm.). Of the remaining six adRP loci mapped by linkage analysis, the loci 7q, 17p and 19q appear to be more frequent. However, there is evidence to suggest RP11 to be the second most common locus after Rhodopsin. Apart from the five RP11 pedigrees identified in our laboratory, four other RP11 pedigrees have also been reported (Xu *et al.*, 1995; McGee *et al.*, 1997). Recently another adRP pedigree of Russian origin has been linked to 19q13.4 (pers. comm.), which brings to a total of 10 the number of RP11 pedigrees reported in the literature, well above the number identified for the six other linked loci. Partial penetrance has consistently been observed in all the families linked to RP11. Therefore RP11 locus is suggested for any new adRP family with evidence of non-penetrance although the 7p and 8q adRP loci do also exhibit variation in severity of disease (Inglehearn *et al.*, 1993; Blanton *et al.*, 1991).

In a previous survey carried out in our laboratory of 20 adRP pedigrees, some of which are also described in this study (ADRP5, ADRP29, ADRP2, RP1907, US family and RP33), frequency of RP11 locus was estimated at 20% (Inglehearn *et al.*, 1998). This thesis describes nine more adRP pedigrees to the twenty reported earlier

with an additional family (ADRP11) linked to the RP11 locus. Frequency estimate for RP11 as a proportion of all adRP is still approximately 20%, but with increased significance resulting from a slightly larger sample size. Since all the adRP families included in the survey were of European origins, the frequency of RP11 derived from this study is better quoted for this geographical population. Some of the adRP families included in this study with evidence of partial penetrance (i.e. ADRP83 and ADRP2228) and excluded on the basis of one individual could still be linked to RP11, in which case the frequency of RP11 would be greater than the estimated value. However at this stage these families stand excluded from the RP11 locus and further testing of their linkage to RP11 would follow once the RP11 gene has been cloned.

At present with the gene yet unknown frequency estimates can only be derived from relatively large adRP pedigrees through haplotype analysis and linkage analysis. The relative rarity of large adRP pedigrees means that significantly better estimates are unlikely to become available until all dominant RP genes have been cloned. Once genes have been cloned very small pedigrees and sporadic cases can also be screened for mutations and true estimates of loci frequencies can be derived. The partial penetrance phenotype of RP11 raises the interesting possibility that a significant proportion of sporadic or apparently recessive RP may also derive from mutations at this locus. The disease could then appear with any apparent family history, confusing the diagnosis. Sporadic RP accounts for ~50% of all cases and is often assumed to be recessive in origin (Jay, 1982). Therefore true estimate of the frequency of RP11 locus would be possible only after the RP11 gene has been cloned and this frequency is likely to be second only to rhodopsin with a high incidence both in the UK and in other populations.

CHAPTER 4

GENOMIC ORGANISATION AND MUTATION ANALYSIS OF A CANDIDATE GENE: PROTEIN KINASE C GAMMA (*PRKCG*)

4.1 Introduction

The analysis of genes, located within the genetic interval of an unidentified disease gene is described as the study of **positional candidates**. Unlike positional cloning, a positional candidate gene approach does not involve the laborious task of isolating new genes but involves a survey of the disease interval to find the most plausible candidate gene for the disease. The screening of such a suitable candidate could either lead to the identification of a disease causing mutation or exclusion of the gene from disease causation. The positional candidate gene approach has been successfully used to identify many disease genes, particularly for the discovery of retinal genes such as rhodopsin and peripherin. With the increasingly dense, high-resolution human transcript map, the positional candidate gene approach will soon become the predominant method of disease gene discovery.

4.2 Positional candidates for RP11

Although chromosome 19 has been described as a particularly gene rich chromosome, not many genes have yet been mapped to chromosome 19q13.4, which contains the RP11 locus (Larsen *et al.*, 1992; Schuler *et al.*, 1996). Of the few genes localised to the RP11 region, those encoding interleukin-11 (*IL11*), the slow skeletal isoform of troponin T (*TNNT1*) and the two isoforms of synaptotagmin (*SYT3* and *SYT5*) were excluded through the refinement of the RP11 locus, which is now confined between the markers D19S572 and D19S418 to a 4.1cM genetic interval (see fig 4.1).

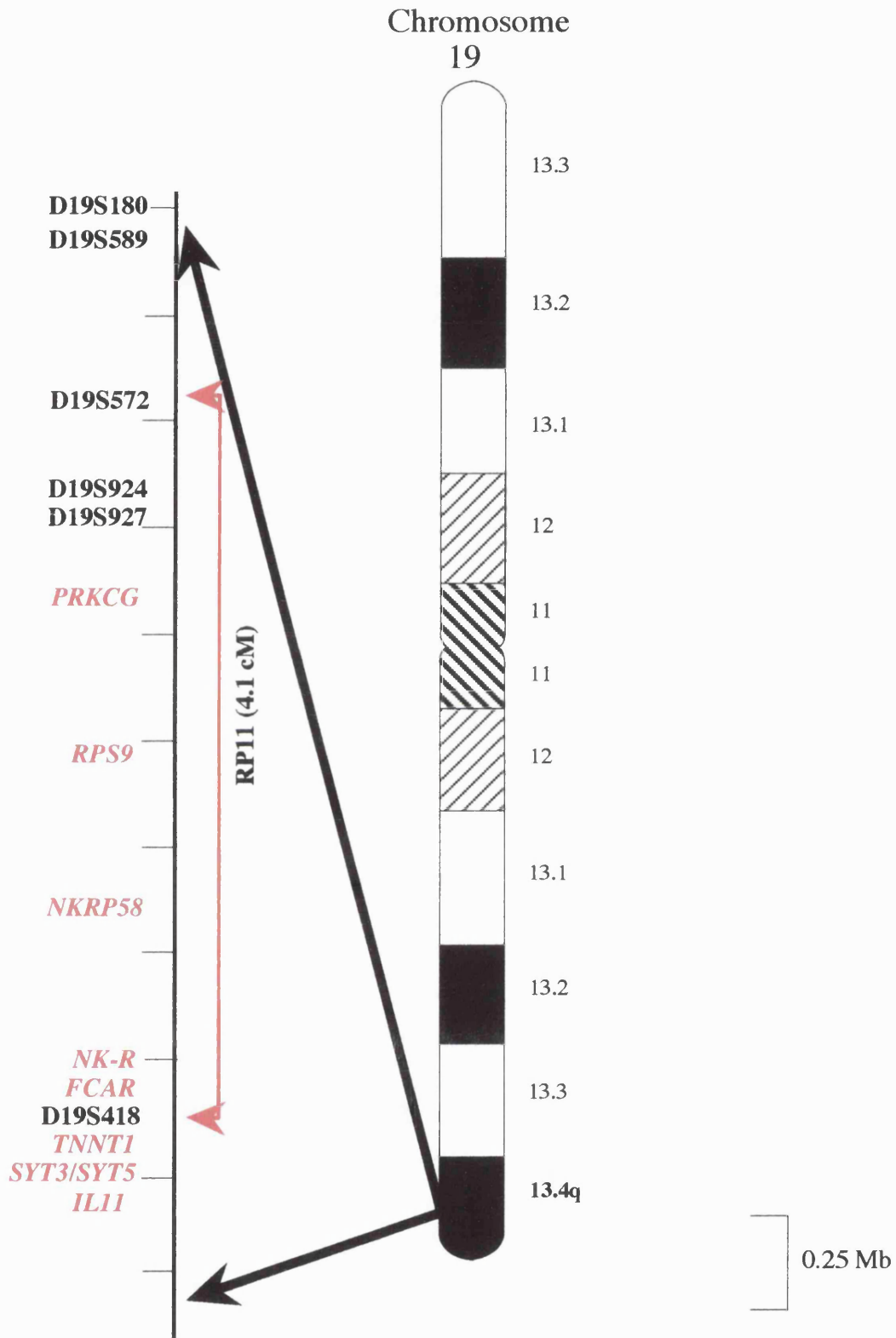


Figure 4.1

A sub-section of the physical map of chromosome 19q13.4, adapted from the LLNL integrated map of chromosome 19 (http://www-bio.llnl.gov/genome-bin/loadmap?m_region=mq2+arm) showing the candidate genes localised to the RP11 interval. Genes and microsatellite markers are indicated in red and black, respectively. The placement of interleukin 11 (*IL11*) in this map is based on its genetic localisation with respect to D19S418.

The genes that remained within the RP11 interval included those for the natural killer cell class 1 receptor family (*NKRP58* and *NK-R*), the receptor for Fc fragment of immunoglobulin A (*FCAR*), protein kinase C gamma (*PRKCG*) and ribosomal protein S9 (*RPS9*). When all the genes mapping to the RP11 interval were assessed as potential candidates for RP11, only *PRKCG* appeared to be a plausible candidate. This gene encodes an isoenzyme which is a member of a large, closely related family of proteins known as protein kinase C (PKC).

4.3 Protein kinase C (PKC)

Protein kinase C (PKC) constitutes a family of closely related isoenzymes that exhibit serine /threonine kinase activity. First characterised on the basis of their activation *in vitro* by Ca^{2+} , phospholipids and diacylglycerol (DAG), PKC isoenzymes are also activated by tumour promoting phorbol esters which have been used as pharmacological tools to investigate the numerous and diverse functions of PKC. Protein kinase C isoenzymes are key enzymes in membrane receptor signal transduction and play an important regulatory role in processes such as cellular proliferation, differentiation, pH homeostasis, tumourigenesis and gene induction (Nishizuka, 1988, 1992 and 1995). PKCs are also present in high concentrations in nervous tissues, and play an important part in the functioning of neuronal signalling by influencing the release of neurotransmitters, modulating ion channels and receptor function. To date, eleven PKC isoenzymes have been identified. These subspecies show subtly different enzymological properties, differential tissue expression and specific intracellular localisation (Nishizuka, 1988). The first and the best characterised are the conventional PKCs (cPKC), α , two alternative splice variants βI and βII , and γ . Ca^{2+} regulates all members in this class. The next well characterised are the novel PKCs (nPKC), δ , ϵ , η , θ , and μ , which are structurally similar to conventional PKCs, but do not contain the functional residues that mediate calcium binding. The third and the least understood group are the atypical PKCs, ζ and λ . The heterogeneous expression profiles of each of the different PKC isoenzymes in different tissues strongly suggest that each isozyme has a unique role. Even though PKCs have been extensively

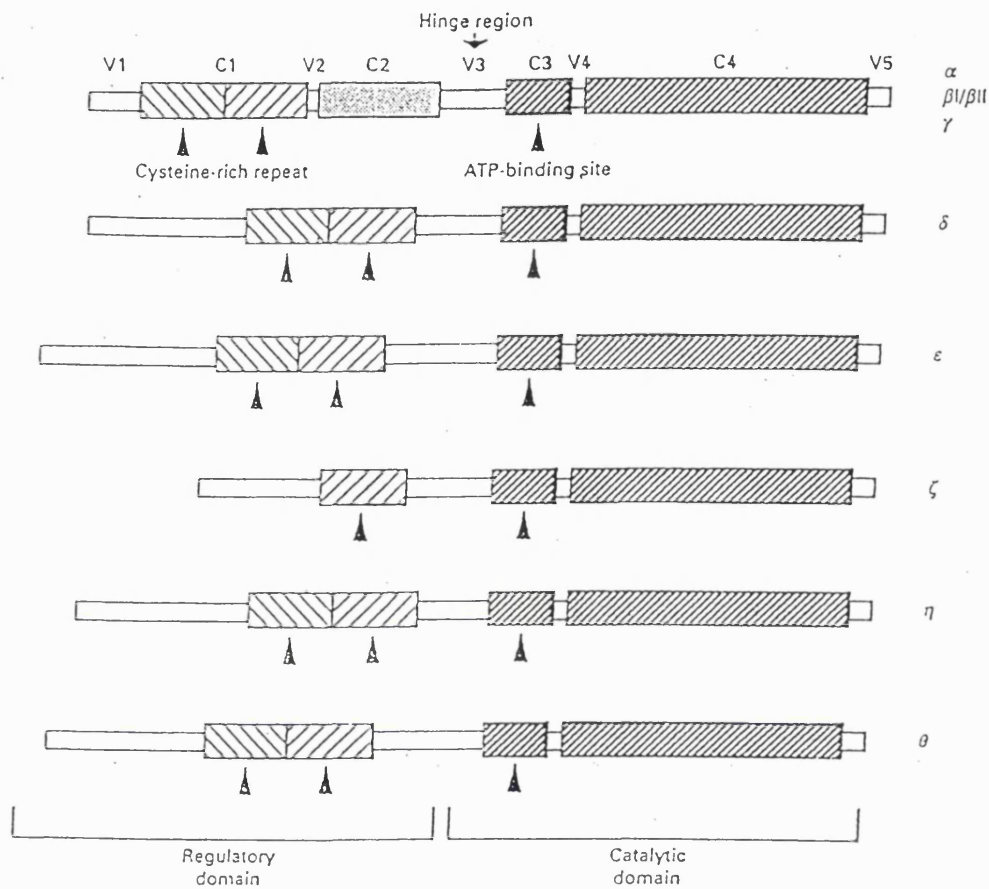


Figure 4.2 [Figure and legend from Hug and Sarre (1993)].

Domain structure of PKC isoenzymes.

All isoenzymes consist of constant (C) and variable (V) regions. The cysteine-rich repeats in the C1 region and the ATP binding site in the C3 region are indicated by arrowheads.

The arrow points to the hinge region in the V3 domain which separates the regulatory from the catalytic domain.

characterised biochemically, knowledge of the specific functions of each of the isoenzymes is still limited. Members of the PKC family are single subunit proteins. Conventional PKC isoenzymes consist of 4 constant (C1-C4) domains separated and flanked by 5 variable regions (V1-V5) (fig.4.2). The C-terminal regions C3-V5 have been defined as the catalytic domain, which is separated by the V3 or the hinge region from the N-terminal regulatory domain. The constant domains of the enzyme have the following functional motifs. The C1 domain contains an autoinhibitory pseudosubstrate sequence and a duplicated Cysteine- rich motif that forms a phorbol ester-binding site. The C2 and C3 domains contain a Ca^{2+} and an ATP binding motif respectively. The substrate-binding site and the phosphate transfer region are contained within the C4 domain.

4.3.1 Retinal protein kinase C

Immunological localisation studies in various vertebrate retinas have indicated the presence, so far, of eight of the PKC isoenzymes, each with a unique cellular distribution in this tissue (see table 4.1). Research work to date has assigned a number of roles for PKC in the retina. These include control of dopamine release, modulation of glutamate receptor responses and function, feedback control of enzymes involved in the process of inositol phosphate signalling and modulation of cyto-skeletal function, to name a few of these (Wood *et al.*, 1997). As mentioned earlier, the specific or the predominant function of each of the isoenzymes in the retina is not yet clear. Specific PKC isozyme activators, inhibitors or the use of antisense RNA techniques (to remove the expression of certain isoenzymes) will be required to determine each isozyme specific function. However, more relevant to this study is the regulatory role PKC has shown to have in the process of phototransduction.

4.3.2 Protein kinase C and Phototransduction

PKC appears to play a key role in the regulation of both vertebrate and invertebrate phototransduction. The specific function of PKC in the invertebrate visual cascade is better understood; hence it will be described first.

PKC Isozyme	Cellular localisation	Species of retina	References
α	ON-Bipolar Cells	Goldfish, Turtle, Rat, Rabbit	Suzuki and Kaneko, 1990; Cuenca <i>et al.</i> , 1990; Osborne <i>et al.</i> , 1992.
α	Cone Cells	Turtle	Cuenca <i>et al.</i> , 1990.
α	Rod Bipolar/Blue Cone Bipolar Cells	Human	Kolb <i>et al.</i> , 1993.
α	Photoreceptor Outer Segments	Rabbit, Bovine	Osborne <i>et al.</i> , 1992; Udovichenko <i>et al.</i> , 1993.
$(\alpha+\beta)^1$	Rod Bipolar Cells	Teleost Fish, Chick, Guinea Pig, Rabbit, Cat, Rat	Negishi <i>et al.</i> , 1988; Karschin and Wässle, 1990; Greferath <i>et al.</i> , 1990; Grünert and Martin, 1991; Wässle <i>et al.</i> , 1991; Yamashita and Wässle, 1991.
$(\alpha+\beta)^1$	Cone Photoreceptor (Inner/Outer segments)	Rat	Ohki <i>et al.</i> , 1994.
$(\alpha+\beta)^1$	Some Müller cell mitochondria	Carp	Fernández <i>et al.</i> , 1995
β^2	Ganglion/Amacrine Cells	Rabbit, Goldfish	Osborne <i>et al.</i> , 1992.
β^2	Bipolar Cells	Frog	Osborne <i>et al.</i> , 1992.
β^2	Cone Bipolar/Amacrine/Ganglion Cells	Human	Kolb <i>et al.</i> , 1993.
β^2	Photoreceptor Outer Segments	Bovine	Udovichenko <i>et al.</i> , 1993.
γ	Ganglion / Amacrine Cells	Rabbit, Human	Osborne <i>et al.</i> , 1992; Kolb <i>et al.</i> , 1993.
γ	Amacrine Cells	Goldfish	Osborne <i>et al.</i> , 1992.
γ	Ganglion Cells	Frog	Osborne <i>et al.</i> , 1992.
γ	Photoreceptor Outer Segment	Bovine	Udovichenko <i>et al.</i> , 1993.
δ	Müller Cells	Rat, Rabbit, Guinea Pig, Frog	Osborne <i>et al.</i> , 1994; Terasawa <i>et al.</i> , 1991.
δ	Ganglion Cells	Goldfish	Osborne <i>et al.</i> , 1994.
δ	Photoreceptor Inner Segments	Chick	Osborne <i>et al.</i> , 1994.
ϵ	Bipolar Cells (Terminals)	Rabbit, Rat, Guinea Pig, Goldfish	Osborne <i>et al.</i> , 1994.
ϵ	Amacrine Cells / Horizontal Cells	Rabbit	Koistinaho and Sagar, 1994.
ϵ	Photoreceptor Inner Segments	Chick	Osborne <i>et al.</i> , 1994.
ϵ	Photoreceptor Outer Segments	Bovine	Udovichenko <i>et al.</i> , 1993.
θ	Bipolar Cells	Rabbit, Rat, Guinea Pig	McCord <i>et al.</i> , 1996.
ζ	Cone Outer Segments /Müller Cells (microvilli)	Frog	Terasawa <i>et al.</i> , 1991.
ζ	Photoreceptor Inner Segments	Rabbit, Rat, Guinea Pig, Goldfish, Chick	Osborne <i>et al.</i> , 1994; Ghalayani <i>et al.</i> , 1994.
λ/ι	Horizontal Cells	Rabbit, Rat, Guinea Pig	McCord <i>et al.</i> , 1996.

¹Monoclonal antibody recognises both isoenzymes. ²No distinction made between β I and β II isoenzymes.

Table 4.1

Occurrences of PKC isoenzymes in retinas of different species as described in Wood *et al.* (1997).

4.3.2.1 Invertebrate phototransduction

Both vertebrate and invertebrate phototransduction systems are rhodopsin-based G protein coupled signalling cascades, beginning with the light activation of the visual receptor, rhodopsin, followed by the subsequent activation of a G-protein (Smith *et al.*, 1991). However, light activation of vertebrate photoreceptors leads to activation of a cGMP-phosphodiesterase effector and the generation of a hyperpolarising response (see sections 1.8 and 1.9). In contrast, activation of invertebrate photoreceptors, for example in *Drosophila*, leads to stimulation of phospholipase C (PLC) and the generation of a depolarising receptor potential. PLC catalyses the hydrolysis of phosphatidylinositol 4, 5 bisphosphate (PIP₂) to produce inositol triphosphate (IP₃) and diacylglycerol (DAG). While IP₃ acts as the second messenger and mobilises Ca²⁺, DAG functions in the feed back regulation through the activation of eye-PKC in *Drosophila*.

***Drosophila* eye-PKC:** The visual systems of both vertebrates and invertebrates modulate their sensitivity to light, in a process known as light adaptation, through the changes in intra photoreceptor Ca²⁺ concentration (Nakatani and Yau, 1988; Matthews *et al.*, 1988; Payne *et al.*, 1986). The influx of extracellular Ca²⁺, that follows light activation of *Drosophila* photoreceptors, is necessary for the regulation of activation and deactivation kinetics of the light activated conductance. An electrophysiological screen for *Drosophila* phototransduction mutants with defective deactivation kinetics demonstrated that photoreceptors of *inaC* mutants (Pak, 1979) are specifically defective in the calcium-dependent negative regulatory mechanisms (Ranganathan *et al.*, 1991). Molecular cloning of the *inaC* locus showed that it encodes a photoreceptor specific isoform of PKC known as eye-PKC (Smith *et al.*, 1991). Eye-PKC, activated by DAG and the influx of external Ca²⁺, mediates photoreceptor deactivation and rapid desensitisation by inactivating active intermediates through phosphorylation. The target(s) of eye-PKC phosphorylation have not yet been defined. However, the role of eye-PKC in invertebrate phototransduction has been established (figure 4.3). In contrast the role of PKC in vertebrate phototransduction is still under investigation and will be discussed in the following section.

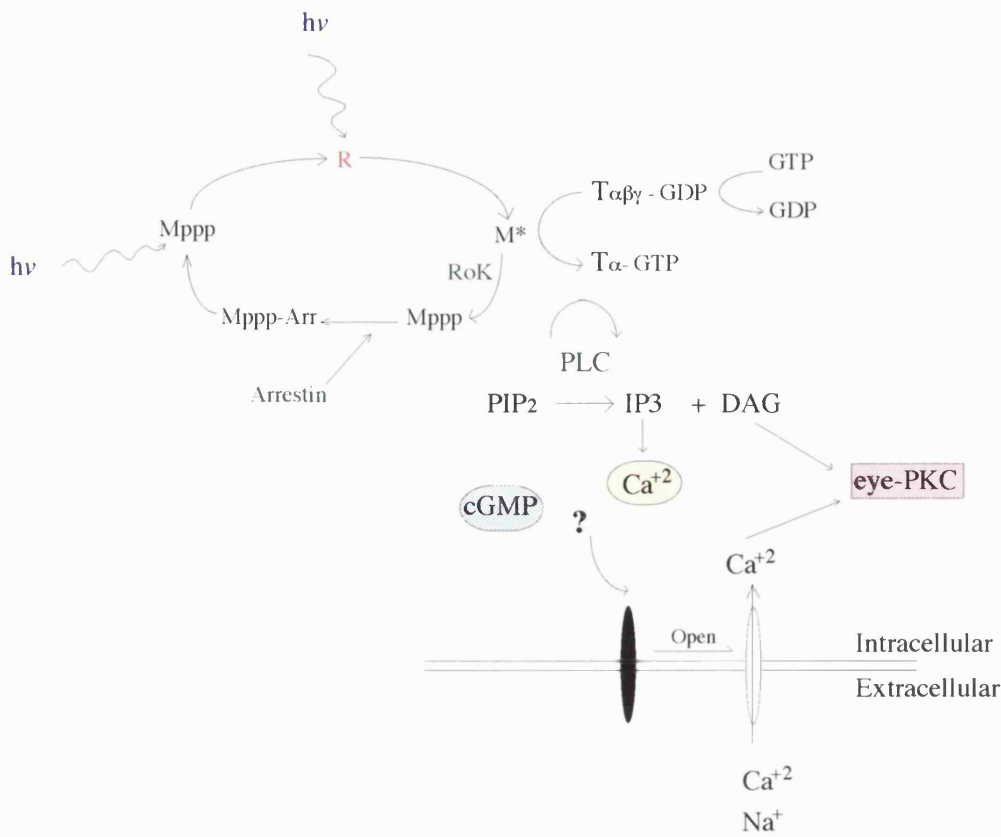


Figure 4.3.

Phototransduction in *Drosophila* photoreceptors.

Absorption of a photon of light causes a conformational change in the rhodopsin molecule (R) and activates its catalytic properties. Active metarhodopsin (M*) catalyses G protein activation. The G protein exchanges GDP for GTP and releases the inhibitory $\beta\gamma$ subunits. Active G protein catalyses the activation of phospholipase C (PLC). PLC hydrolyses PIP₂ into the intracellular messengers IP₃ and DAG. The IP₃ is released from the membrane and diffuses to receptors located on the subrhabdomeric cisternae (SRC). This binding is thought to release intracellular calcium from SRC, which appears to be involved in excitation. cGMP has also been implicated as a possible intracellular messenger mediating excitation. Extracellular sodium and calcium then enter the cell through the light-activated conductance and cause the depolarisation of the photoreceptor cells. DAG and the large influx of calcium is thought to modulate a photoreceptor cell-specific PKC (encoded by *inaC*) that regulates deactivation and desensitisation of the light response. Metarhodopsin is inactivated via phosphorylation by rhodopsin kinase (RoK) and arrestin binding. [Figure and legend adapted from Zuker, C. (1996)]

4.3.2.2 Regulation of vertebrate phototransduction and PKC

Phosphorylation plays an essential role in the regulation of G-protein-coupled receptors, where phosphorylation by both G-protein receptor kinases (GRK) and second messenger regulated kinases mediate receptor desensitisation (Lefkowitz, 1993.). In vertebrate phototransduction, it has been clearly established that a G-protein coupled receptor kinase, rhodopsin kinase, phosphorylates and desensitises light-activated rhodopsin. The primary phosphorylation sites of rhodopsin kinase *in vitro* have been identified as Ser³³⁸ and Ser³⁴³ of rhodopsin (Hargrave and McDowell, 1992). Phosphorylation on the carboxyl terminus decreases the interaction with transducin; an interaction that is effectively quenched when arrestin binds the polyphosphorylated carboxyl tail. Similar to other G protein coupled receptor kinases, rhodopsin kinase displays strict specificity for the active conformation of the receptor and only phosphorylates bleached rhodopsin.

Several reports over the past decade have also implicated a role for protein kinase C in the regulation of phototransduction. The presence of PKC in the rod outer segments (see section 4.3.1) and its phosphorylation of rhodopsin, transducin, arrestin and γ subunit of phosphodiesterase (PDE) have been reported (as reviewed by Udovichenko *et al.*, 1994; Yarfitz and Hurley, 1994; Newton and William, 1991, 1993). First, hyper-activation of PKC in the intact retinas, by phorbol ester treatment, was shown to alter the phosphorylation of rhodopsin in a light dependent manner. Then PKC was shown to phosphorylate rhodopsin *in vitro* and *in situ*, at different sites on the carboxyl terminus from that of rhodopsin kinase, and uncouple the receptor from transducin (Newton and William, 1993). Unlike rhodopsin kinase, this phosphorylation by PKC was independent of the activation state of rhodopsin, and was regulated by Ca^{2+} and lipids. Recently it was shown that PKC contributes to approximately half of the light dependent phosphorylation of rhodopsin in intact frog retinas (Udovichenko *et al.*, 1997). The PKC catalysed phosphorylation of rhodopsin is analogous to the ligand-independent phosphorylation of other G-protein coupled receptors, that is catalysed by second messenger-regulated kinases. Therefore as the evidence suggests, rhodopsin, similar to other G-protein coupled receptors, is phosphorylated by both a substrate-regulated kinase (rhodopsin kinase) and a second

messenger-regulated kinase (protein kinase C). Since PKC has equal affinity for both unbleached and bleached rhodopsin, it has been proposed that PKC may be involved in light adaptation at low light levels where most of the rhodopsin is unbleached, while rhodopsin kinase desensitises rhodopsin at higher light levels (Newton and William, 1993). It has been suggested that phosphorylation of rhodopsin by two differently regulated kinases may allow rhodopsin to respond to a much broader range of stimuli. The protein kinase C that catalyses the phosphorylation of rhodopsin in the rod outer segments is a member of the Ca^{2+} regulated protein kinase Cs. All three Ca^{2+} dependent isoforms α , β and γ have been found in the rod outer segments (Udovichenko *et al.*, 1993). In one study by Wolbring and Cook (1991), only the α isoform was purified from bovine rod outer segment. So far the exact identity of the PKC isozyme responsible for rhodopsin phosphorylation has not been determined. Yet it would appear that in both vertebrate and invertebrate phototransduction PKC plays a role in the desensitisation of the photoreceptor and termination of light response.

4.3.3 Protein kinase C gamma (*PRKCG*)

In the previous sections PKC was discussed in general without highlighting the γ isozyme in particular. The expression and the possible biological function(s) of this isozyme, as revealed by the few studies conducted on this isozyme, will now be discussed.

PKC γ isozyme is encoded by the protein kinase C gamma (*PRKCG*) gene, which is located on chromosome 19q13.4 within the RP11 interval. This isozyme has been reported to be expressed predominantly in the brain and spinal cord, (Saito *et al.*, 1988) but expression in the retina has also been reported (Udovichenko *et al.*, 1993). The high expression of PKC γ in the neuronal tissues has only implicated it in neuronal functions so far.

In the rat brain PKC γ is localised mainly in excitatory and inhibitory amino acid neurons such as cortical pyramidal cells, hippocampal pyramidal and granule cells, cerebellar Purkinje cells, and thalamic neurons (Hashimoto *et al.*, 1988; Hirata *et al.*, 1991). Due to the specific distribution of the γ isoform in the dendritic spines of Purkinje cells that form synaptic contacts with the parallel fibres, it has been

implicated in long term depression (LTD) of glutamatergic transmission in the dendritic spine of Purkinje cells. Because the γ isoform is localised in the postsynaptic side of the glutamatergic synapse related to LTP (Long Term Potentiation) or LTD, including for example hippocampal pyramidal cells and cerebellar Purkinje cells, this PKC isoform has been implicated in the postsynaptic regulation of glutamate receptors. Moreover, the presence of PKC γ in the presynaptic terminal of Purkinje cells has implicated it for the positive feedback control of GABA (γ amino butaric acid), the major inhibitory neurotransmitter of the CNS, release in the terminals of Purkinje cells. A null mutant mouse lacking PKC γ has also been created by gene-targeted homologous recombination (Abeliovich *et al.*, 1993a, Abeliovich *et al.*, 1993b). These null mutant mice display modest impairment in tests of learning and memory and slight ataxia but are otherwise normal in appearance (Abeliovich *et al.*, 1993a). Synaptic transmission is normal in the hippocampus, but long term potentiation is impaired (Abeliovich *et al.*, 1993b). These observations agree with the functions previously attributed to PKC γ purely on the basis of its localisation in the neural tissue.

4.4 *PRKCG*: a candidate for RP11?

The involvement of a Ca^{2+} dependent PKC isozyme in the regulation of phototransduction cascade in mammals (Newton and William, 1993) and the involvement of a PKC gene in retinal degeneration in *Drosophila* (Ranganathan *et al.*, 1991; Smith *et al.*, 1991) were supportive of *PRKCG* gene as a candidate for RP11. Even though expression of *PRKCG* in the rod outer segment has been controversial and its predominant expression is in the nervous tissue, the supportive evidence given above and the fortuitous localisation of *PRKCG* in the RP11 interval suggested that it is a gene worthy of a thorough investigation. The fact that *PRKCG* gene encodes an enzyme does not exempt it from being regarded as a candidate for autosomal dominant retinitis pigmentosa. Even though it is generally accepted that enzymes cause autosomal recessive rather than dominant diseases. This well-established dogma is not incontrovertible since a mutation in the β subunit of the enzyme phosphodiesterase has been associated with a dominant form of congenital stationary night blindness (Gal *et al.*, 1994). Some may also regard the lack of visual impairment in the PKC γ null mutant mouse as a reason to refute *PRKCG* as a candidate gene for

RP11. However considering that the phenotype of RP11 shows incomplete penetrance and the importance of genetic background for phenotypic expression of this disease (see section 3.1.3), it is obvious that the knockout mouse model does not mimic the genetic situation of RP11 and therefore cannot be expected to show the same phenotypic consequences. Moreover it can be argued that the PKC γ mouse is not a true model for an autosomal dominant disease caused by missense mutation, where the presence of the abnormal protein is likely to cause the disease rather than the absence of normal enzyme activity (gain of function mutation). Therefore despite this negative evidence the analysis of *PRKCG* gene was undertaken reassured by the fact that there are no other (known) genes that could be regarded as superior candidates for RP11 in the disease interval.

4.5 Genomic organisation of the *PRKCG* gene

In order to assess *PRKCG* as a potential candidate for the RP11 gene, the intron-exon structure of this gene was characterised. In this section the genomic organisation of the *PRKCG* gene, which consists of eighteen exons, will be presented along with a structural comparison with the *Drosophila* dPKC53E (brain) gene. The *Drosophila* dPKC53E (brain) gene was the only other PKC gene to be fully characterised prior to this study (Rosenthal *et al.*, 1987).

4.5.1 Materials and methods

4.5.1.1 The cDNA sequence of *PRKCG* and genomic clones of the *PRKCG* gene.

Two overlapping cDNA clones (phPKCgamma6, 2HVaK) and a genomic clone encompassing the 5' upstream region of the *PRKCG* gene (accession numbers M13977, Z15114 and X62533, respectively) were used for the analysis of the exon-intron boundaries. Starting with transcription initiation, the combined sequence of the three clones span 2574 nucleotides (fig 4.4). A polyadenylation signal was not identified in the cloned part of the 3'-untranslated region. The open reading frame extends from nucleotide 1147 through 3240, encoding a putative 697 amino acid, 78kDa protein.

Promoter region	CTCCCAGAGA	GGAGACAGAG	ACGGACTCCC	AAAGAGTCAA	GTTGCAGAGG	GAGGGGGAGA	60
	CAGAGTCCGC	CTGAGAGAGA	ATAGAGTTCA	CAGAGCCAGA	GAGACAAAGA	CAGGGAAAGA	120
	AACAAACTCA	GGGAGACAGA	GGCGGAAGAA	TTTATCAGAC	GTAGAGGGAG	AGACAGAGGG	180
	AGACTGATTC	ACAGTGAGGC	AGACCAAGAC	CCCGAGAAGA	ACCACGACAC	CAAGACCCAG	240
	AGATGGCGAC	AGATACAGAT	GGAGACACAC	AGCAGAGGAA	ATAGAGAAAT	AGAAAGAAAT	300
	AGAAACAGCA	TTTCTCGGAG	ACAGGGAAAA	GAAAAAGAT	CCACTGAGAG	GCAGAAACAG	360
	AGACACATCG	AGAAAGCTCG	CTCCAGGAAT	AGAGGGAGAG	GGACAGAGGT	CACAAGAAAG	420
	ACACACGCAG	GCAGAGAGCT	ACACAAATAA	TGGAGAGGCC	AGGGGAGGAA	ATAAAGACTC	480
	AGCCGGCATC	GGAGAAAAGT	AGAACCCTAG	CGCCCTGCCT	GTCCACTGCT	GGACCCCTAG	540
	CGTGGAGCAT	AAAGTTTGT	GAAGGAAGGA	GAGGGGCAGG	GTCAGACACA	GGACCCAGG	600
5'UTR sequence	GCGCCACAG	GACACACGAG	GCACCTAGTG	GGGGAGAAC	GCGGGCAGGA	TGACAGATTG	660
	CAGGTCGTTG	GGGGGAGCC	AGGCTCAGAG	GATGCCCTC	CCTCCAGCCA	GGCCCCGGGA	720
	GTGGGTGTGT	GCACGTGTGG	GGGCGGGCA	GGGAGGACAT	TTGTCCCGTG	TCTCCGGGAG	780
	GGGAGCGCCT	TTAAGCCGAA	ACCCGCCCTC	TCGGTCGTCC	TGCAACGCCT	CCCCCACCCG	840
	GGGCTCCAC	ATTTACAGCAG	GTGCCGGAGC	TGGAGCTCCC	ACC CCG CCG	CCCGTGCCTC	900
	CGGCTGCCGG	CGCCCTGCC	TTTGGCTCTT	CCTCCCACT	CGCCCGCTCC	CCCTGGCGGA	960
	GCCGGCGCGC	CCGGGTGCC	GTCCTCGCC	TGGCGCGCTC	CGCACCTGGA	GGTGCCTTGC	1020
	CCCTCTCCTG	CCCACCTCGG	AATTTCCTTG	TGGCTCCTTT	GATCCTTCGA	GTCTCCAGCT	1080
	CCCTCTCCTT	CCACCTGTTT	CCCCAAGAA	AGGCAGGATC	CTGGTCCCTG	CTACGTTTCT	1140
	GGGGC ATCG	CTGGTCTGGG	CCCCGGCGTA	GGCGATTGAG	AGGGGGGACC	CCGGCCCTTG	1200
Coding sequence	TTCTGCAGAA	AGGGGGCTCT	GAGGCAGAAG	GTGGTCCACG	AAGTCAAGAG	CCACAAGTTC	1260
	ACCGCTCGTG	TCTTCAAGCA	GCCCACCTTC	TGCAGCCACT	GCACCGACTT	CATCTGGGGT	1320
	ATCGGAAAGC	AGGGCCTGCA	ATGTCAAGT	TGCAGCTTTG	TGGTTCATCG	ACGATGCCAC	1380
	GAATTTGTGA	CCCTTCGAGTG	TCCAGGGCCT	GGGAAGGGCC	CCCAGACGGA	CGACCCCGG	1440
	AACAAACACA	AGTTCCGCCT	GCATAGCTAC	AGCAGCCCCA	CCTTCTGCGA	CCACTGTGGC	1500
	TCCTCCTCT	ACGGGCTTGT	GCACCAGGGC	ATGAAATGCT	CCTGCTGCGA	GATGAACGTG	1560
	CACCGCGCCT	GTGTGCGTAG	CGTGCCCTCT	CTGTGCGGTG	TGGACCACAC	CGAGCGCCGC	1620
	GGGCGCCTGC	AGTGGGAGAT	CCGGGCTCCC	ACAGCAGATG	AGATCCACGT	AACTGTTGGC	1680
	GAGGCCCGTA	ACCTAATTCC	TATGGACCCC	AATGGTCTCT	CTGATCCCTA	TGTGAAACTG	1740
	AAGCTCATCC	CAGACCCTCG	GAACCTGACG	AAACAGAAGA	CCCGAACGGT	GAAAGCCACG	1800
CTAAACCCTG	TGTGGAATGA	GACCTTTGTG	TTCAACCTGA	AGCCAGGGGA	TGTGGAGCGC	1860	
Partial 3' UTR sequence	CGGCTCAGCG	TGGAGGTGTG	GGACTGGGAC	CGGACCTCCC	GCAACGACTT	CATGGGGGCG	1920
	ATGTCCTTG	GCGTCTCGGA	GCTGCTCAAG	GCGCCCGTGG	ATGGCTGGTA	CAAGTTACTG	1980
	AACCAGGAGG	AGGGCGAGTA	TTACAATGTG	CCGGTGGCTG	ATGCTGACAA	CTGCAGCCTC	2040
	CTCCAGAAAGT	TTGAGGCTTG	TAACCTACCC	CTGGAATTGT	ATGAGCGGGT	GCGGATGGGC	2100
	CCCTCTCCT	CTCCCATCCC	CTCCCTTCC	CCTAGTCCCA	CCGACCCCAA	GCGCTGCTTC	2160
	TTCCGGGCGA	GTCACGAGC	CCTGCACATC	TCCGACTTCA	GCTTCTCAT	GGTCTAGGA	2220
	AAAGGCAGTT	TTGGGAAGGT	GATGCTGGCC	GAGCGCAGGG	GCTCTGATGA	GCTCTACGCC	2280
	ATCAAGATCT	TGAAAAAGGA	CGTGATCGTC	CAGGACGACG	ATGTGGACTG	CACGCTGGTG	2340
	GAGAAACGTG	TGCTGGCGCT	GGGGGGCCGG	GGTCTGGCG	GCCGGCCCCA	CTTCTCACC	2400
	CAGCTCCACT	CCACCTTCCA	GACCCCGGAC	CGCCTGTATT	TCGTGATGGA	GTACGTCACC	2460
GGGGGAGACT	TGATGTACCA	CATTCAACAG	CTGGGCAAGT	TTAAGGAGCC	CCATGCAGCG	2520	
TTCTACCGCG	CAGAAATCGC	TATCGGCCCT	TTCTTCCTTC	ACAATCAGGG	CATCATCTAC	2580	
AGGGACCTGA	AGTTGGACAA	TGTGATGCTG	GATGCTGAGG	GACACATCAA	GATCACTGAC	2640	
TTTGGCATGT	GTAAGGAGAA	CGTCTTCCCC	GGGACGACAA	CCCGCACCTT	CTGCGGGACC	2700	
CCGGACTACA	TAGCCCGGGA	GATCATTGCC	TACCAGCCCT	ATGGGAAGTC	TGTCGATTGG	2760	
TGGTCTTTG	GAGTCTGCT	GTATGAGATG	TTGGCAGGAC	AGCCTCCCTT	CGATGGGGAG	2820	
GACGAGAAGG	AGCTGTTTCA	GGCCATCATG	GAACAAACTG	TCACCTACCC	CAAGTCGCTT	2880	
TCCCGGGAAG	CCGTGGCCAT	CTGCAAGGGG	TTCTTGACCA	AGCACCCAGG	GAAGCGCCTG	2940	
GGCTCAGGGC	CTGATGGGGA	ACCTACCATC	CGTGACACATG	GCTTTTCCG	CTGGATTGAC	3000	
TGGGAGCGGC	TGGAACGATT	GGAGATCCCG	CCTCCTTTCA	GACCCCGCCC	GTGTGGCCGC	3060	
AGCGCGGAGA	ACTTTGACAA	GTCTTTCACG	CGGGCGGCGC	CAGCGCTGAC	CCCTCCAGAC	3120	
CGCTAGTCC	TGGCCAGCAT	CGACCAGGCC	GATTTCCAGG	GCTTACCTA	CGTGAACCCC	3180	
GACTTCGTGC	ACCCGATGC	CCGACGCCCC	ACCAGCCAG	TGCCCTGTCC	CGTCATG TAA	3240	
TCTCACCCGC	CGCCACTAGG	TGTCCCCAAC	GTCCCTCCG	CCGTGCCGGC	GGCAGCCCCA	3300	
CTTCACCCCC	AACTTACCA	CCCCCTGTCC	CATTCTAGAT	CCTGCACCCC	AGCATTCCAG	3360	
CTCTGCCCCC	GCGGGTTCTA	GACGCCCTC	CCAAGCGTTC	CTGGCCTTCT	GAACTCCATA	3420	
CAGCCTCTAC	AGCCGTCCCG	CGTTCAAGAC	TTGAGCG			3457	

Figure 4.4

The collated sequence of two overlapping cDNA clones (pPKC γ 6 and 2HV α K) and a genomic clone. The genomic clone encompasses the 5' upstream sequence of *PRKCG* gene. The 5' upstream sequence, the 5' UTR, the entire coding sequence and the partial 3' UTR region are indicated. The putative transcription initiation, translation initiation and translation termination sites are boxed.

The ATG start codon is included in exon 1, and the termination signal in exon 18 is followed by 217 bp of 3'-untranslated region.

A *PRKCG* specific DNA probe (724bp) was generated by PCR, using the phPKCgamma6 cDNA clone as a template, with the following primers: (forward/reverse, 5'-3') gaaaggcaggatcctgtgtcc/aacacaaaggtctcattccaca. A human cosmid genomic library was screened with this probe to isolate cosmids that contained the *PRKCG* gene (see section 2.12). However this endeavour was not very successful, and only resulted in the isolation of a single cosmid that contained the 5'end of the *PRKCG* gene. Later three cosmids (20019, 21280 and 25482) previously found to contain the *PRKCG* gene (Mohrenweiser *et al.*, 1995), were obtained from the Lawrence Livermore National Laboratories (LLNL) genome centre, U.S.A. These cosmids, obtained at a latter stage of the project, were predominantly used to confirm splice site junctions of *PRKCG*.

4.5.1.2 Elucidation of *PRKCG* genomic structure

Generally genomic clones of any gene are a prerequisite for characterisation of that particular gene. The strategy most often used starts with subcloning of these large insert clones into a vector that accommodates smaller inserts, then assembling an overlapping array of smaller clones which are then systematically sequenced to identify the exon-intron boundaries of that gene. However, as mentioned before the isolation of *PRKCG* genomic clones proved difficult, therefore a strategy had to be devised to elucidate the structure of *PRKCG* without having to use such genomic clones.

Since the entire coding sequence of *PRKCG* was available, first an idea was formulated to cover the entire cDNA sequence with a series of PCR primers to give overlapping fragments. Then use these primers on genomic DNA and sequence any large fragment that was produced presumably due to the presence of an intron, to identify the splice sites. This random strategy was abandoned in favour of a method, which had a higher success rate for 'hitting introns'. Existence of structural conservation between homologous genes is a well-documented fact (Lundin, 1993). Therefore the elucidation of *PRKCG* gene structure was subsequently based on a

method which assumed conservation of genomic organisation between a *Drosophila* PKC gene known as dPKC53E (brain) and *PRKCG*. This presumption was reinforced by the fact that although *Drosophila* has only ~10,000 genes in its genome, compared with ~100,000 genes in a mammalian cell, most human genes are duplications and elaborations of their insect equivalents (Lundin, 1993; Holland *et al.*, 1994; Miklos and Rubin., 1996). The method used, which was still a PCR based method, is described in detail in the following sections.

(a) *Drosophila* PKC genes (dPKC)

In *Drosophila*, 3 PKC genes have been identified (Schaeffer *et al.*, 1989). One maps to the third chromosome at position 98F and is expressed throughout the *Drosophila* central nervous system. The other two genes are located 50 kb apart at position 53E on the second chromosome. The 53E genes have been named dPKC53E (brain) and dPKC53E (eye), according to their cyto-genetic location and tissue expression. It is dPKC53E (eye) gene that encodes eye-PKC (see section 4.3.2.1). The 98F and the two 53E genes are homologous to the mammalian nPKC and cPKC genes respectively. Among the *Drosophila* PKC genes, dPKC53E (brain) is most similar to the conventional mammalian genes (PKC α , PKC β and PKC γ). The protein encoded by this gene is 10 amino acids longer at the amino terminus and 35 amino acids shorter at the carboxyl terminus than the mammalian PKCs, and unlike PKC γ , does not contain a V5 domain (Rosenthal *et al.*, 1987). The dPKC53E (brain) gene was the only PKC gene to be fully characterised prior to this study (Rosenthal *et al.*, 1987) and was of immense use for elucidation of *PRKCG* gene structure. The dPKC53E (brain) gene consists of 13 coding exons and at least one untranslated exon at its 5' end.

(b) Prediction of splice sites according to organisation of *Drosophila* dPKC53E gene

The main strategy employed to establish the genomic organisation of the *PRKCG* gene was based on the assumption of there being some degree of structural conservation between the *PRKCG* and *Drosophila* dPKC53E genes (Rosenthal *et al.*, 1987).

Figure 4.5

The alignment of amino acid sequences of *Drosophila* PKC(dPKC), human PKCs (hPKC α , hPKC β , hPKC γ), bovine PKC γ , rat PKC γ and mouse PKC γ genes showing the extent of homology that exist between these genes. Intron positions of the dPKC gene are marked by arrows.

dPKC	MSEGSNNNGD	PQQGAEGEA	VGENKMK SRL	RKGALKKKNV	FNVKDHCFIA	RFFKQPTFCS	HCKDFIWGFG	KQGFQCQVCS	YVVHKRCHEY	VTFICPGKDK	100
hPKC gamma	MAGLGPVGVD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
mPKC gamma	MAGLGPGGGD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
rPKC gamma	MAGLGPGGGD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
bPKC gamma	-----	-----	---RPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	75
hPKC alpha	MADVFPGNDS	T-----AS--	-QDVANRFA-	RKGALRQKNV	HEVKDHKFTA	RFFKQPTFCS	HCTDFIWGFG	KQGFQCQVCC	FVVHKRCHEY	VTFSCPGADK	91
hPKC beta	MADPAAGPPP	S-----EG--	-EESTVRFA-	RKGALRQKNV	HEVKNHKFTA	RFFKQPTFCS	HCTDFIWGFG	KQGFQCQVCC	FVVHKRCHEY	VTFSCPGADK	91
							↓	↓			
dPKC	GIDS DSPKTQ	HNFE PFTYAG	PTFCDHCGSL	LYGLVHQGLK	CSACDMNVHA	RCKENVPSLC	GCDHTERRGR	IYLEINV-KE	NLLTVQIKEG	RNLIPMDPNG	199
hPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTA	DEIHVTVGEA	RNLIPMDPNG	190
mPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHITVGEA	RNLIPMDPNG	190
rPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHITVGEA	RNLIPMDPNG	190
bPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHVTVGEA	RNLIPMDPNG	175
hPKC alpha	GPDTDDPRSK	HKFKIHTYGS	PTFCDHCGSL	LYGLVHQGMK	CDTCDMNVHK	QCVINVPSLC	GMDHTEKRGR	IYLKAEV-AD	EKLHVTVRDA	RNLIPMDPNG	190
hPKC beta	GPASDDPRSK	HKFKIHTYSS	PTFCDHCGSL	LYGLVHQGMK	CDTCDMNVHK	RCVMNVPSLC	GDHTERRGR	IYIQAHI-DR	DVLIIVLVRDA	RNLVPMDPNG	190
							↓				
dPKC	LSDPYVKVKL	IPDDKDQSKK	KTRTIKACLN	PVWNETLTYD	LKPEDKDRRI	LIEVWDWDR	SRNDFMGALS	FGISEIKNP	TNGWFKLLTQ	DEGEYINVPC	299
hPKC gamma	LSDPYVKLKL	IPDPRNLTKQ	KTRTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDR	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
mPKC gamma	LSDPYVKLKL	IPDPRNLTKQ	KTKTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDR	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
rPKC gamma	LSDPYVKLKL	IPDPRNLTKQ	KTKTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDR	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
bPKC gamma	LSDPYVKLKL	IPDPRNLTKQ	KTRTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDR	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	275
hPKC alpha	LSDPYVKLKL	IPDPKNESKQ	KTKTIRSTLN	PQWNESFTFK	LKPSDKDRRL	SVEIWDWDR	TRNDFMGSL	FGVSELMKMP	ASGWYKLLNQ	EEGEYINVPI	290
hPKC beta	LSDPYVKLKL	IPDPKSESQ	KTKTIKCSLN	PEWNETFRFQ	LKESDKDRRL	SVEIWDWDLT	SRNDFMGSL	FGISELQKAS	VDGWFKLLSQ	EEGEYFNVPV	290
								↓			
dPKC	ADDEQ-DLL-	KL-----	---KQKPSQK	-KPMVMRS	NTHT-----S	SKKDMIRATD	FNFIKVLGKG	SFGKVL LAER	KGSEELYAIK	ILKKDVIIQD	380
hPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPS	PTDPKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIVQD	390
mPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIVQD	390
rPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIVQD	390
bPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRTGPS	SSPIPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIVQD	375
hPKC alpha	PEGDE-----	--EGNMLRQ	KFEKAKL GPA	GNKVISPS	--RK-QP--S	NNLDRVKLT	FNFLMVLGKG	SFGKVMLADR	KGTEELYAIK	ILKKDVIIQD	378
hPKC beta	PPEGS-----	--EANEELRQ	KFERAKISQ	TKVPEEKT	TVSKFDN--N	GNRDRMKLT	FNFLMVLGKG	SFGKVMLSER	KGTEELYAVK	ILKKDVIIQD	381

First, to see if such a presumption is valid the homology that exists at the protein level between dPKC, PKC γ , α and β was investigated by aligning the amino acid sequences of dPKC, human, mouse, rat and bovine PKC γ , human PKC α and PKC β (Fig 4.5). This revealed a considerable degree of homology and splice sites in the dPKC gene could be correlated to positions in the amino acid sequence of human PKC γ .

Next PCR primers had to be designed from PKC γ cDNA sequence to flank these 12 putative splice sites. As these primers needed to be PKC γ specific, it was important to consider the extent of homology that existed at the DNA level between each of the different cPKC isoenzymes. To ensure these primers amplified *PRKCG* DNA and not DNA of other homologous members of the PKC gene family, the primers needed to include as many gamma specific nucleotides as possible. Therefore, the sequence homology between PKC α , β and γ was first tested by aligning the cDNA sequence of these three isozyme species using the software package Gene Works™ and PCR primer pairs were designed accordingly. Figure 4.6 depicts the homology of the sequence and the 11 primer pairs designed to capture introns at positions predicted according to dPKC gene structure. Since genomic clones of the *PRKCG* gene were not available at this time, these PCR primers pairs were used on genomic DNA instead.

4.5.2 Results

4.5.2.1 Identification of splice sites conserved between *Drosophila* dPKC53E gene and *PRKCG*

The primers that were described in the previous section (and depicted in figure 4.6) were used for PCR amplification on genomic DNA from normal subjects. Clear single PCR products were obtained for 10 of the 11 primer pairs designed, and all these fragments were larger than expected indicating the presence of introns (see fig 4.7). The primer pair that failed to amplify flanked the third putative splice site (F2/R2).

Figure 4.6

The alignment of the cDNA sequence of PKC α , β and γ for the design of PKC γ specific primers. The vertical black arrows indicate putative splice sites according to *Drosophila* PKC gene. The 11 PKC gamma specific primer pairs (e.g. F1 and R1) designed to amplify the putative introns are shown in blue.



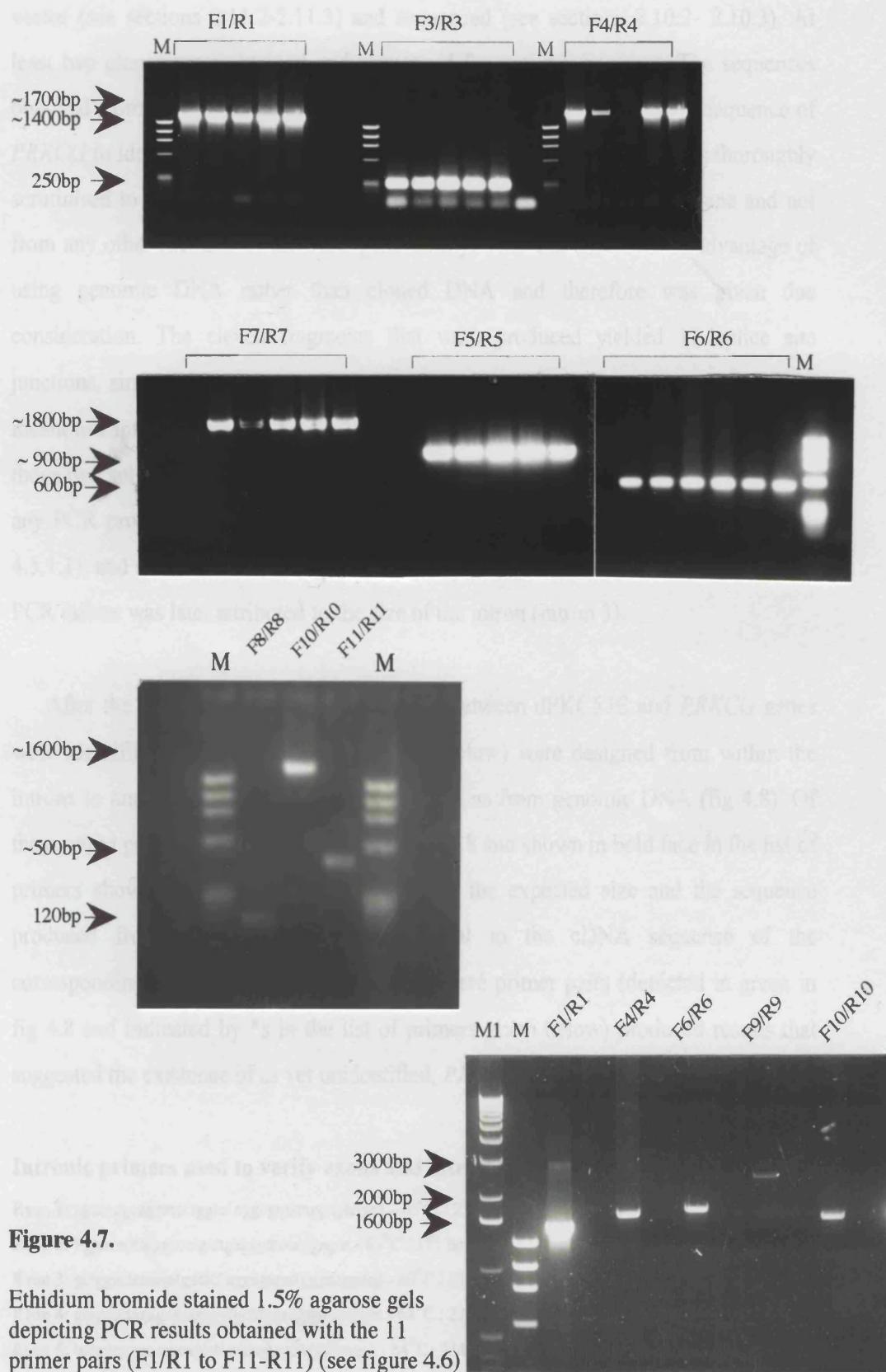


Figure 4.7.

Ethidium bromide stained 1.5% agarose gels depicting PCR results obtained with the 11 primer pairs (F1/R1 to F11-R11) (see figure 4.6) on a genomic DNA template. The sizes of PCR fragments greater than 500 bp were estimated by eye using the markers, M- ϕ X174DNA marker and M1- λ 1kb ladder.

The 'large' PCR fragments were then cloned individually in to an appropriate vector (see sections 2.11.2-2.11.3) and sequenced (see sections 2.10.2- 2.10.3). At least two clones were obtained and sequenced for a given fragment. The sequences obtained from these cloned fragments were then compared to the cDNA sequence of *PRKCG* to identify intron-exon boundaries. At this stage the sequence was thoroughly scrutinised to ensure that the sequence was obtained from the *PRKCG* gene and not from any other member of the PKC gene family. This was the main disadvantage of using genomic DNA rather than cloned DNA and therefore was given due consideration. The eleven fragments that were produced yielded 12 splice site junctions, since the fragment produced by the primer pair F1/R1, which was initially meant to capture two putative *PRKCG* introns, did indeed lead to the identification of those two splice sites (see figure 4.6). The primer pair F2/R2, which failed to produce any PCR product, was later used to direct-sequence the cosmid 20019 (see section 4.5.1.1), and another conserved splice site junction was identified. The reason for this PCR failure was later attributed to the size of the intron (intron 3).

After the splice sites that were conserved between dPKC53E and *PRKCG* genes were identified, eleven primer pairs (shown below) were designed from within the introns to amplify *PRKCG* coding sequence/exons from genomic DNA (fig 4.8). Of these, eight primer pairs (depicted in red in fig 4.8 and shown in bold face in the list of primers shown below) produced fragments of the expected size and the sequence produced from each fragment was identical to the cDNA sequence of the corresponding *PRKCG* exon. The remaining three primer pairs (depicted in green in fig 4.8 and indicated by *s in the list of primers given below) produced results that suggested the existence of as yet unidentified, *PRKCG* specific introns.

Intronic primers used to verify exons and intron-exon boundaries (F/R 5'-3'):

Exon 1: agaaggcaggatcctggtc/ cggcgtgataggagtctgca - 65°C / 274 bp

Exon 2: ttgacacctgggccctg/ctgagggtcccaggagcc - 65°C / 171 bp

Exon 3: gctggactaatccatgcctc/ aggagaaattgggacggacg - 60°C / 213 bp

Exon 4: gctgacctagagagcaaggc/gctttggaagggccctggca - 64°C / 221 bp

Exon 5: tgaggctctaccgcagctt/ acaagtgccttgggtcagcc - 64°C / 234 bp

Exon 6: ctctaaccctcacactctt/ tctgtcagctgtcattgct - 60°C / 234 bp

* Exon 7: gccatgagctcggctctgca/ aaccagaaatctgacctccc - failed PCR

Exon 8: aggtcctgtaccactgggtt/ atccaacgcagatgtccag - 65°C / 318 bp

Exon 9: gtagatggatcccgcctcta/ acgtcagaaggctcaggct - 58°C / 205 bp

* Exon 10: agccactgacctctgacgt/ taaggatctcaaagcgtg- 56°C / ~ 650 bp

* Exon 11: gcacttaacgtggtagcg/ agtgacttcaggaatgggag - 60°C / ~ 1800 bp

Exon 12: atgtacctgtccgcact/ accagggtttgtgcctgg - 55°C / 232 bp

Exon 13: ctggagctgcttaacttcc/ acgtggggacacctagtg - 64°C / 262 bp

4.5.2.2 Identification of *PRKCG* specific splice sites

Even though the characterisation of *PRKCG* gene was initiated with a strategy that assumed a high degree of structural conservation between dPKC gene and *PRKCG*, structural features unique to *PRKCG* were also expected. Therefore the results obtained from the first set of intronic primers more than complied with what was anticipated. These results not only support the assumption made about structural conservation between the dPKC and *PRKCG* genes, but three intronic primer sets gave different results to indicate that there still were introns, specific to the *PRKCG* gene, to be defined. Of the three primer pairs, two produced PCR fragments larger than expected and one pair failed to produce a fragment (see section 4.5.2.1). First, the two large PCR fragments were cloned and sequenced. This led to the identification of two *PRKCG* specific splice sites and as before intronic primers were designed to flank these two newly defined exons. These primers yielded products of the expected size and the sequence obtained from these fragments was also identical to the corresponding *PRKCG* cDNA sequence.

The primer pair, which failed to produce any PCR fragment, flanked 406 bp of *PRKCG* cDNA (Fig 4.8). Characterising this portion of the gene proved to be difficult. By virtue of failed PCRs, it was assumed that this coding region was interrupted by more than one intron. The cosmid clone 20019, fortunately available by this time, was subcloned to characterise this part of the *PRKCG* gene. Cosmid DNA was digested with *Bam*HI and subcloned into *Bam*HI digested pUC18 vector. Mini libraries were plated and replica filters were hybridised with the 724bp DNA probe generated from the partial cDNA clone of *PRKCG* (phPKCgamma6). The technical details of this procedure are described in Materials and Methods section 2.13. The sequence data obtained from positive clones were aligned with *PRKCG* cDNA using GeneWorks™ to identify splice site junctions.

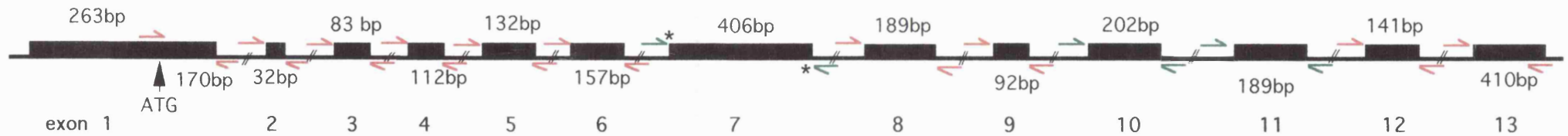


Figure 4.8

Schematic diagram of the structure of *PRKCG* gene, predicted according to the organisation of the dPKC gene, showing the primers designed to verify the identified exons. The primer pairs in red gave PCR products of the expected size whereas the primer pairs shown in green either failed to amplify [indicated by (*)] or produced PCR fragments which were larger than the expected size.

Three other *PRKCG* specific splice sites were identified by this method making the total number of *PRKCG* specific splice sites five.

4.5.2.3 Confirmation of splice junctions using *PRKCG* genomic clones

With the exception of the last three splice sites mentioned above, all exon-intron boundaries of *PRKCG* were identified by a PCR based method, where genomic DNA was used as a template. This situation was not ideal, as there was a risk of amplifying fragments from other members of the PKC gene family. However due to the quality of the primers used, the intron sequences identified by this method proved to be from the *PRKCG* gene. The intronic primers designed to amplify a given exon not only produced fragments of the expected size but also the sequence of these fragments matched the *PRKCG* cDNA sequence. Even though this was the case splice sites were confirmed once again by using the cosmid clone, 20019 (and later a PAC clone). The PCR primers previously used on genomic DNA to capture *PRKCG* introns were used on the cosmid DNA, to ensure that both templates produced identical sized fragments. Moreover, the intronic primers were also used on cosmid DNA and the fragments produced (putative exons) were sequenced. At this point it became clear that the cosmid 20019 only contained exons 1 to 14 of *PRKCG*, and we were able to confirm all of these intron-exon boundaries. The cosmid clone was also directly sequenced, using primers that flanked the putative human introns according to organisation of the *Drosophila* dPKC53E gene, to confirm splice sites identified by other means. A PAC clone (275-e-9), isolated at a later stage (see chapter 5 section 5.5.2 and figure 5.7), was found to contain exons 3 to 3'-UTR of *PRKCG* gene and using this clone, the rest of the *PRKCG* gene structure was confirmed. Table 4.2 lists all the splice sites, along with the method employed for the identification of each.

Intron	Intron Size(bp)	Identifying Method	Primers designed for PCR and/or sequencing (5'-3')	T _m for PCR (°C)
1 and 2	500 and 1000	B (F1/R1)	F-tggtccacgaagtcaagagc / R-tcgatgaaccacaaagctgca	62°C
3	-	B/D (F2/R2)	F-cttcgagtgccaggcgctgg / R-gcacaagcccgtagaggagg	
4	135	B (F3/R3)	F-ccctcctctacgggcttgctg / R-gatctccagctgcaggcgccc	68°C
5	1600	B (F4/R4)	F-gggcgctgcagctggagatc / R-ggaattagggtacgggctcg	62°C
6	690	B (F5/R5)	F-ggagggcccgtaacctaa / R-acatcccctggcttcaggt	62°C
7	360	C	-	
8	285	C	-	
9	1300	C	-	
10	370	B (F6/R6)	F-acttcagcttctcatggttc / R-gagctgggtgaggaagtgg	62°C
11	1600	B (F7/R7)	F-aaagctgtgctggcgctggg / R-ctcctaaactgcccagctg	66°C
12	94	B (F8/R8)	F-cgtgatggagtacgtcacc / R-agcgatttctgcccgctag	60°C
13	210	A	F-agccactgaccttctgacgt / R- taaggatctcaaagcgtg	60°C
14	2400	B (F9/R9)	F-ctgaagctggacaatgtgatg / R-atcgacagactcccattag	60°C
15	1400	A	F- gcacttaacgtgggtagcg / R- agtgacttcaggaatgggag	60°C
16	1400	B (F10/R10)	F-cctaccccgaagctgttcc / R-gatctccaatggtccagccg	64°C
17	270	B (F11/R11)	F-cggctggaacgattggagatc / R-ttcatgacgggcacaggcac	66°C

Note. Sizes of introns >500 bp were estimated from PCR products and therefore are not exact values (see figure 4.7).

Table 4.2

Methods used to elucidate the structure of *PRKCG* organisation.

- A- Sequencing large PCR fragments C- By subcloning cosmid 20019
 B- Based on dPKC53E organisation D- Direct sequencing cosmid 20019

4.5.3 Discussion

4.5.3.1 Genomic organisation of the *PRKCG* gene

The intron-exon boundaries of the *PRKCG* gene, and the comparison with the boundaries of the *dPKC* gene are shown in tables 4.3 and 4.4, respectively. Of the 17 splice junctions identified, 11 are conserved between the *Drosophila* and human *PRKCG* genes, the rest being unique to *PRKCG*. All splice sites display the consensus sequences (C/T) AG-exon-GT (Breathnach *et al.*, 1978). The conserved splice sites are at cDNA positions 170, 202, 397, 529, 686, 1092, 1281, 1373, 1575, 1764 and 1905 of *PRKCG*, where position 1 of the cDNA is defined as the first base of the start codon (Table 4.3). The intron phases of the conserved exon-intron boundaries are also identical (Table 4.4). The third splice site of *PRKCG* was found three nucleotides away from the predicted site based on the *dPKC* sequence (Fig 4.9). The close proximity of these two corresponding introns of *dPKC* and *PRKCG* implies slippage of this intron position. The splice sites not found in *dPKC* are located at positions 821, 909, 939, 1436 and 1656 of the *PRKCG* cDNA. The first three “novel” sites divide the largest exon of the *dPKC* gene (exon 8) into four exons in *PRKCG*, namely exons 7, 8, 9 and 10. The other two novel sites divide exon 11 and 12 of *dPKC* to form exons 13, 14, 15 and 16 of *PRKCG* (see figure 4.10). The human *PRKCG* gene, therefore, consists of 18 coding exons whereas *dPKC* has 13 coding exons and at least one untranslated exon at its 5' end (Rosenthal *et al.*, 1987). In comparison, no such untranslated exons are found at the 5' end of *PRKCG*. All introns have been sized except intron three, which could not be amplified by PCR; therefore it is possible that this intron is considerably large. Exon sizes range from 32bp to 433bp and the known intron sizes range from 94 to <2500bp (Table 4.2). Thus, the *PRKCG* gene should span at least 16 kb of genomic DNA. The genomic sequence of *PRKCG*, generated in this study is depicted in figure 4.11.

Exon/size bp	cDNA position	Splice Acceptor	Splice Donor
1	5' UTR- 170	5'UTR ATG GCT GGT	C TTC ATC TG gtgagg gaa
2 32	171-202	ctctccag G GGT ATC GG	AA TGT CAA G gtaagag ct
3 83	203-285	catctatag TC TGC AGC T	CAG ACG GAC gtgagt gctc
4 112	286-397	tgccccag GAC CCC CGG	AA TGC TCC T gtgagt gacc
5 132	398-529	cccctccag GC TGC GAG A	AC GTA ACT G gtgagg cccc
6 157	530-686	ccccgttag TT GGC GAG G	C TTT GTG TT gtgagt ctgg
7 135	687-821	cttctgcag C AAC CTG AA	G GAT GGC TG gtgagg agca
8 88	822-909	tgtatgtcag G TAC AAG TT	AAG TTT GAG gtacc cagaa
9 30	910-939	tttctccag GCT TGT AAC	TTG TAT GAG gtgagt agaa
10 153	940-1092	ctttccacag CGG GTG CGG	TTT GGG AAG gttg gattcc
11 189	1093-1281	gtgcgcatag GTG ATG CTG	CAG ACC CCG gtaagg atgg
12 92	1282-1373	ttctccgag GAC CGC CTG	C CAT GCA GC gtgagt ctcg
13 63	1374-1436	cgctctccag G TTC TAC GC	C ATC TAC AG gtgagc agcc
14 139	1437-1575	catcgtccag G GAC CTG AA	GCC CCG GAG gtaacc ccaa
15 81	1576-1656	cctccccag ATC ATT GCC	GCA GGA CAG gtaagg gaag
16 108	1657-1764	actttgatag CCT CCC TTC	TGC AAG GGG gtgagag cccc
17 141	1765-1905	tttccacag TTC CTG ACC	CCC CGC CCG gtcagt cacc
18	1906-3'UTR	tctccacag TGT GGC CGC	GTC ATG TAA -3'UTR

Note. Intronic sequence is indicated in lowercase letters, exonic sequence in uppercase letters. Position 1 in cDNA is defined as the first base of the start codon.

Table 4.3.

Intron-exon boundary sequences of the human *PRKCG* Gene.

Figure 4.9

Amino acid sequence alignment between dPKC and mammalian PKCs showing the conserved splice site junctions and those specific for *PRKCG*. Intron positions are marked by arrows, while conserved splice sites are indicated by black arrows, *PRKCG* specific splice sites are indicated by red arrows. For the third splice site junction two positions are shown, one predicted according to dPKC structure (marked by a •) and the other as found in reality within the *PRKCG* gene (shown in red). The close proximity of these two splice sites imply intron slippage.

dPKC	MSEGSNNNGD	PQQQGAEGEA	VGENKMSRL	RKGALKKKNV	FNVKDHCFIA	RFFKQPTFCS	HCKDFIWGFG	KQGFQCQVCS	YVVHKRCHEY	VTFICPGKDK	100
hPKC gamma	MAGLGPVVD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
mPKC gamma	MAGLGPGGGD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
rPKC gamma	MAGLGPGGGD	S-----EG--	-GPRPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	90
bPKC gamma	-----	-----	---RPL-FC-	RKGALRQKVV	HEVKSHKFTA	RFFKQPTFCS	HCTDFIWGIG	KQGLQCQVCS	FVVHRRCHEF	VTFECPGAGK	75
hPKC alpha	MADVFPNGDS	T-----AS--	-QDVANRFA-	RKGALRQKNV	HEVKDHKFA	RFFKQPTFCS	HCTDFIWGFG	KQGFQCQVCC	FVVHRRCHEF	VTFSCPGADK	91
hPKC beta	MADPAAGPPP	S-----EG--	-EESTVRFA-	RKGALRQKNV	HEVKNHKFTA	RFFKQPTFCS	HCTDFIWGFG	KQGFQCQVCC	FVVHRRCHEF	VTFSCPGADK	91
		•					↓	↓			
dPKC	GIDS DSPKTQ	HNFEFPTYAG	PTFCDHCGSL	LYGIYHQGLK	CSACDMNVHA	RCKENVPSLC	GCDHTERRGR	IYLEINV-KE	NLLTVQIKEG	RNLIPMDPNG	199
hPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTA	DEIHVTGGEA	RNLIPMDPNG	190
mPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHITVGEA	RNLIPMDPNG	190
rPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHITVGEA	RNLIPMDPNG	190
bPKC gamma	GPQTDDPRNK	HKFRLHSYSS	PTFCDHCGSL	LYGLVHQGMK	CSCCEMNVHR	RCVRSVPSLC	GVDHTERRGR	LQLEIRAPTS	DEIHVTGGEA	RNLIPMDPNG	175
hPKC alpha	GPDTDDPRSK	HKFKIHTYGS	PTFCDHCGSL	LYGLIHQGMK	CDTCDMNVHK	QCVINVPSLC	GMDHTEKRRG	IYLAKEV-AD	EKLHVTVRDA	RNLIPMDPNG	190
hPKC beta	GPASDDPRSK	HKFKIHTYSS	PTFCDHCGSL	LYGLIHQGMK	CDTCMMNVHK	RCVMNVPSLC	GDHTERRGR	IYIQAHI-DR	DVLIVLVRDA	RNLVPMDPNG	190
				↓							
dPKC	LSDPYVKVKL	IPDDKDQSKK	KTRTIKACLN	PVWNETLTYD	LKPEDKDRRI	LIEVWDWDRD	SRNDFMGALS	FGISEIKNP	TNGWFKLLTQ	DEGEYINVPC	299
hPKC gamma	LSDPYVKLKL	IPDPRNLTQK	KTRTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDRD	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
mPKC gamma	LSDPYVKLKL	IPDPRNLTQK	KTKTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDRD	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
rPKC gamma	LSDPYVKLKL	IPDPRNLTQK	KTKTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDRD	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	290
bPKC gamma	LSDPYVKLKL	IPDPRNLTQK	KTRTVKATLN	PVWNETFVFN	LKPGDVERRL	SVEVWDWDRD	SRNDFMGAMS	FGVSELLKAP	VDGWYKLLNQ	EEGEYINVVP	275
hPKC alpha	LSDPYVKLKL	IPDPKNESKQ	KTKTIRSTLN	PQWNESFTFK	LKPSDKDRRL	SVEIWDWDRD	TRNDFMGSL	FGVSELMKMP	ASGWYKLLNQ	EEGEYINVPI	290
hPKC beta	LSDPYVKLKL	IPDPKSESQK	KTKTIKCSLN	PEWNETFRFQ	LKESDKDRRL	SVEIWDWDLT	SRNDFMGSL	FGISELQKAS	VDGWFKLLSQ	EEGEYINVPI	290
				↓							
dPKC	ADDEQ-DLL-	KL- ↓ -----	---KQKPSQK	-KPMVMRSDT	NHTH-----S	SKKDMIRATD	FNFIKVLGKG	SFGKVLLAER	KGSEELYAIK	ILKKDVIQD	380
hPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPSPS	PTDPKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIQD	390
mPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIQD	390
rPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRMGPS	SSPIPSPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIQD	390
bPKC gamma	ADADNCSLLQ	KFEACNYPLE	LYERVRTGPS	SSPIPSPSPS	PTDSKRCFFG	ASPGRLHISD	FSFLMVLGKG	SFGKVMLAER	RGSDELYAIK	ILKKDVIQD	375
hPKC alpha	PEGDE-----	--EGNMELRQ	KFEKAKLGPA	GNKVISPSED	--RK-QP--S	NNLDRVKLTD	FNFLMVLGKG	SFGKVMLADR	KGTEELYAIK	ILKKDVIQD	378
hPKC beta	PPEG-----	--EANEELRQ	KFERAKISQK	TKVPEEKTTN	TVSKFDN--N	GNRDRMKLTD	FNFLMVLGKG	SFGKVMLSER	KGTDELYAVK	ILKKDVIQD	381

Gene	Intron/Intron phase		5' splice junction	3' splice junction
dPKC	2	II	Phe, Ile, Tr- Intron	Intron- p, Gly, Phe
<i>PRKCG</i>	1		Phe, Ile, Tr- Intron	Intron- p, Gly, Ile
dPKC	3	I	Cys, Gln, V- Intron	Intron- al, Cys, Ser
<i>PRKCG</i>	2		Cys, Gln, V- Intron	Intron- al, Cys, Ser
dPKC	4	0	Ile, Asp, Ser- Intron	Intron- Asp, Ser, Pro
<i>PRKCG</i>	3		Gln, Thr, Asp- Intron	Intron- Asp, Pro, Arg
dPKC	5	I	Cys, Ser, A- Intron	Intron- la, Cys, Asp
<i>PRKCG</i>	4		Cys, Ser, C- Intron	Intron- ys, Cys, Glu
dPKC	6	I	Val, Gln, I- Intron	Intron- le, Lys, Glu
<i>PRKCG</i>	5		Val, Thr, V- Intron	Intron- al, Gly, Glu
dPKC	7	II	Leu, Thr, Ty- Intron	Intron- r, Asp, Leu
<i>PRKCG</i>	6		Phe, Val, Ph- Intron	Intron- e, Asn, Leu
dPKC	-	-	-	-
<i>PRKCG</i>	7	II	Asp, Gly, Tr- Intron	Intron- p, Tyr, Lys
dPKC	-	-	-	-
<i>PRKCG</i>	8	0	Lys, Phe, Glu- Intron	Intron- Ala, Cys, Asn
dPKC	-	-	-	-
<i>PRKCG</i>	9	0	Leu, Tyr, Glu- Intron	Intron- Arg, Val, Arg
dPKC	8	0	Phe, Gly, Lys- Intron	Intron- Val, Leu, Leu
<i>PRKCG</i>	10		Phe, Gly, Lys- Intron	Intron- Val, Met, Leu
dPKC	9	0	Gln, Thr, Met- Intron	Intron- Asp, Arg, Leu
<i>PRKCG</i>	11		Gln, Thr, Pro- Intron	Intron- Asp, Arg, Leu
dPKC	10	II	Val, Ala, Va- Intron	Intron- l, Phe, Tyr
<i>PRKCG</i>	12		His, Ala, Al- Intron	Intron- a, Phe, Tyr
dPKC	-	-	-	-
<i>PRKCG</i>	13	II	Ile, Tyr, Ar- Intron	Intron- g, Asp, Leu
dPKC	11	0	Ala, Pro, Glu- Intron	Intron- Ile, Ile, Leu
<i>PRKCG</i>	14		Ala, Pro, Glu- Intron	Intron- Ile, Ile, Ala
dPKC	-	-	-	-
<i>PRKCG</i>	15	0	Ala, Gly, Gln- Intron	Intron- Pro, Pro, Phe
dPKC	12	0	Cys, Lys, Gly- Intron	Intron- Phe, Leu, Thr
<i>PRKCG</i>	16		Cys, Lys, Gly- Intron	Intron- Phe, Leu, Thr
dPKC	13	0	Pro, Lys, Ile- Intron	Intron- Lys, His, Arg
<i>PRKCG</i>	17		Pro, Arg, Pro- Intron	Intron- Cys, Gly, Arg

Note. Intron 3 of *PRKCG* occurs 3 nucleotides away from the corresponding intronic position of dPKC. The positions of introns between codons are indicated by phase 0, interruption after the first nucleotide by phase I, and interruption after the second nucleotide by phase II.

Table 4.4.

Intron-exon boundary comparison between dPKC and *PRKCG* genes.

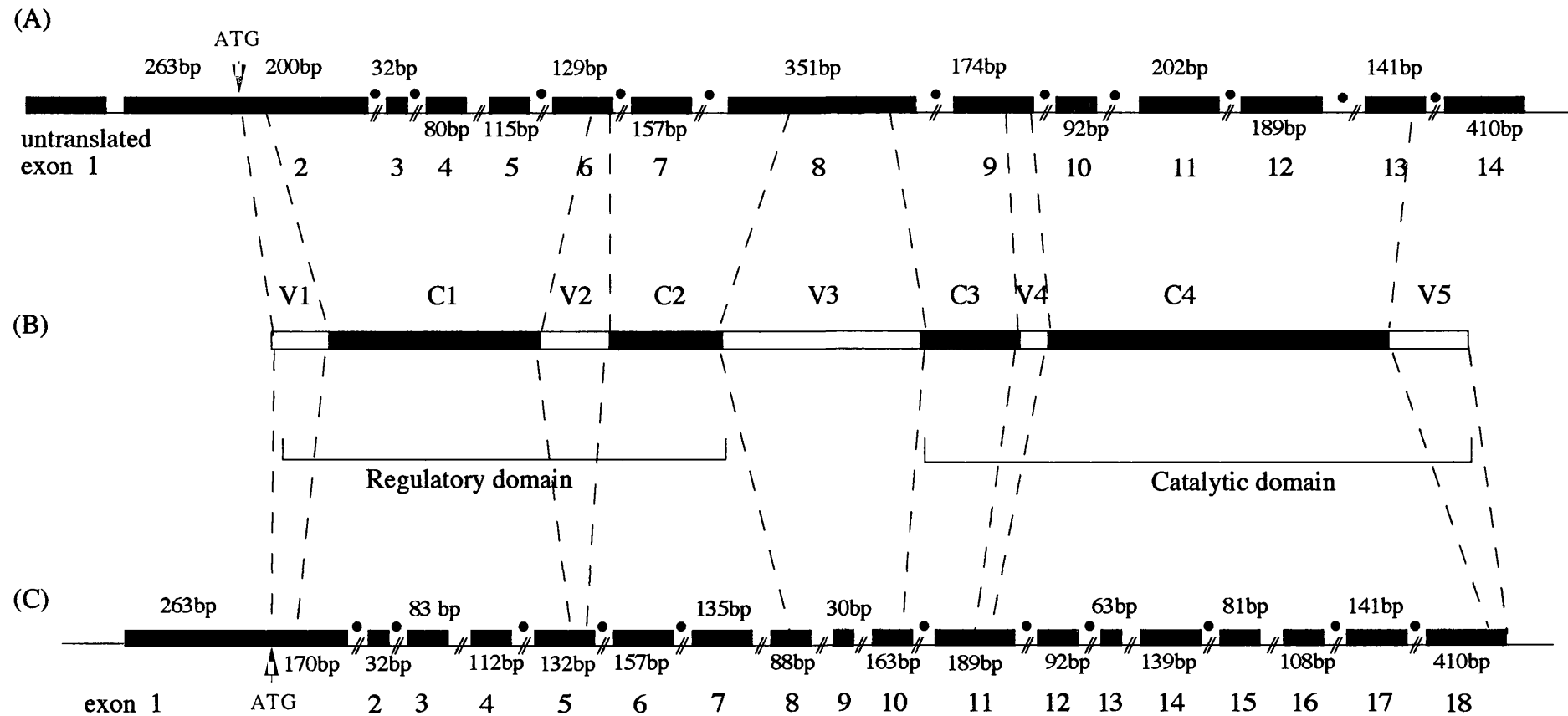


Figure 4.10

Schematic diagram showing the structure of the dPKC gene (A), the domain structure of a cPKC isoenzyme (B) and the structure of *PRKCG* gene (C). For both the genes the sizes of exons are shown and the conserved splice sites are indicated by a •. Dashed lines show the correspondance between the genomic structure (of both dPKC and *PRKCG*) and the domain structure of the isoenzyme.

4.5.3.2 Comparison of structure between *PRKCG* and *Drosophila* dPKC53E genes

The genomic structure of the *PRKCG* was related to the functional topology of the protein (Fig. 4.10). The pseudosubstrate site of the C1 domain is included within exon 1. The cysteine-rich region of C1 domain is formed from a perfect repeat of the motif Cys-X₂-Cys-X₁₃-Cys-X₂-Cys-X₇-Cys-X₇-Cys at amino acids 49-85 and 114-150 (Rosenthal *et al.*, 1987). This Cys-rich region is located within exons 1-5, where the first repeat is included within exons 1-3 and the second within exons 4-5. The distribution of the two Cys-rich motifs within the exons is exactly as described for the *Drosophila* PKC gene (Rosenthal *et al.*, 1987). The two repeats are separated by intron 3 and each repeat is split by an intron at the same position to give the structure Cys-X₂-Cys-X₁₃-Cys-X-intron-X-Cys-X₇-Cys-X₇-Cys.

The variable domains V1, V2, V4 and V5 are each contained within a single exon, namely exons 1, 5, 11 and 18 respectively. The only exception is the V3 region, which is included within exons 8-10. In contrast, the entire V3 coding region of the *Drosophila* PKC protein is contained within exon 8 of the dPKC gene. V3, the largest of the variable regions forms the hinge region, which separates the regulatory and the catalytic halves of the protein. If the ancestral PKC gene, like the cGMP dependent kinases, was derived from a fusion of two genes, providing regulatory and catalytic domains (Parker *et al.*, 1986), intron 8 of the *PRKCG* gene can be regarded as the boundary between these domains. In such an event, intron 8 may also represent an evolutionary footprint of the recombination event that brought the two genes together (fig 4.10). Since dPKC has no equivalent intron it could be argued that during the course of its evolution dPKC lost the introns equivalent to numbers 7, 8 and 9 of *PRKCG*, allowing the expression of the V3 domain from one exon. Intron loss through evolution has been postulated for the opsin genes of *Drosophila* (Carulli *et al.*, 1992). However, the possibility that *PRKCG* gained these introns after the two genes diverged cannot be discounted. In both the dPKC and *PRKCG* genes, an intron splits the ATP binding site (C3 domain). In *PRKCG*, the splice site at cDNA position 1436 enables the entire phosphate transfer region of the C4 domain to be included within exon 14. In contrast there are no obvious structural implications associated with the other *PRKCG* specific splice site at cDNA position 1656. As a result of

previous work on the characterisation of the promoters of human PKC beta and rat PKC gamma genes, it is known that the first two splice site junctions of those genes also correspond exactly with two splice junctions in the dPKC gene (Chen *et al.*, 1990; Mahajna *et al.*, 1995). This added with the data presented here highlights the high degree of structural conservation of PKC genes over an evolutionary period of approximately 400 million years. The conserved splice junctions in the both *PRKCG* and dPKC genes correspond to the conserved C1-C4 domains of the protein (Fig 4.10), which suggest an evolutionary conserved protein core that is not susceptible to major change. The PKC gene family probably arose from a single ancestral gene by gene duplication, and the conserved position of introns indicates that the ancestral gene contained introns. The absence of some introns in the dPKC gene implies that human *PRKCG* is closer to the ancestral PKC gene than dPKC. The position of these “lost” introns in the gene also strengthens the argument that the ancestral PKC gene was derived from the fusion of two genes. Overall the data generated in this study support the ‘intron early’ theory of gene evolution, which states that introns existed from the beginning of evolution where their role was to enhance recombination between exons in order to facilitate the first creation of genes i.e. shuffling exons to make novel genes (Gilbert, 1978; Gilbert, 1987). Subsequent evolutionary processes such as recombination within introns, the sliding and drift of introns to change the peptide sequence around the splicing site, and the loss of introns to make more complex exons, to be shuffled in turn are postulated to create diversity in related genes.

The work on the genomic organisation of the human *PRKCG* gene has given us some insight into the structural evolution of PKC genes. Comparative structural analysis of more PKC genes, ideally a member from every class of the PKC gene family, would further our understanding of the evolution of these genes. However in relation to this study, the genomic organisation of *PRKCG* gene enabled us to design intronic primers for a thorough mutation screen of this gene as a candidate for RP11.

Figure 4.11.

The genomic sequence of *PRKCG* gene. Exonic sequence of the gene are boxed and written in uppercase whereas the intronic sequence is in lower case. Sequence encoding the variable regions (V1-V5) of the protein are highlighted in yellow and the sequence encoding the constant (C1-C4) domains are highlighted in blue. The primers used to amplify the 5'upstream region (in red) and exons of *PRKCG* (in black) are also shown.

```

CTCCCAGAGA GGAGACAGAG ACGGACTCCC AAAGAGTCAA GTTGCAGAGG
                                     UF5
GAGGGGGAGA CAGAGTCCGC CTGAGAGAGA ATAGAGTTCA CAGAGCCAGA
GAGACAAAGA CAGGGAAAGA AACAAACTCA GGGAGACAGA GGCGGAAGAA
TTTATCAGAC GTAGAGGGAG AGACAGAGGG AGACTGATTC ACAGTGAGGC
AGACCAAGAC CCCGAGAAGA ACCACGACAC CAAGACCCAG AGATGGCGAC
AGATACAGAT GGAGACACAC AGCAGAGGAA ATAGAGAAAT AGAAAGAAAT
                                     UF4
AGAAACAGCA TTTCTCGGAG ACAGGGAAAA GAAAACAGAT CCACTGAGAG
UR5
GCAGAAACAG AGACACATCG AGAAAGCTCG CTCCAGGAAT AGAGGGAGAG
GGACAGAGGT CACAAGAAAG ACACACGCAG GCAGAGAGCT ACACAAATAA
TGGAGAGGCC AGGGGAGGAA ATAAAGACTC AGCCGGCATC GGAGAAAGTG
                                     UF3
                                     UR4
AGAACCCTAG CGCCCTGCCT GTCCACTGCT GGACCCCTAG CGTGGAGCAT
AAAGTTTGTT GAAGGAAGGA GAGGGGCAGG GTCAGACACA GGACCCAGG
GCGCCACAG GACACACGAG GCACCTAGTG GGGGAGGAAC GCGGGCAGGA
                                     UF2
TGACAGATTG CAGGGTGGTG GGGGGGAGCC AGGCTCAGAG GATGCCCTC
                                     UR3
CCTCCAGCCA GGCCCCCGGA GTGGGTGTGT GCACGTGTGG GGGCGGGCA
GGGAGGACAT TTGTCCCGTG TCTCCGGGAG GGGAGCGCCT TTAAGCCGAA
ACCGCCCTC TCGGTCGTCC TGCAACGCCT CCCCACCCG GGGCTCCCAC
                                     UF1
                                     UR2
ATTTCAGCAG GTGCCGAGC TGGAGCTCCC ACCGCCGCGG CCCGTGCCTC
CGGCTGCCGG CGCCCTGCC TTTGGCTCTT CCTCCCCACT CGCCCGCTCC
CCCTGGCGGA GCCGGCGCGC CCGGGGTGCC GCTCCCTGCC TGGCGCGCTC
```


CGCACCTGGA GGTGCCTTGC CCCTCTCCTG CCCACCTCGG AATTTCCCTG

TGGCTCCTTT GATCCTTCGA GTCTCCAGCT CCTCTCCCTT CCACCTGTTT

CCCCAAGAA AGGCAGGATC CTGGTCCCTG CTACGTTTCT GGGGCCATGG
UR1 ← IR1 →

CTGGTCTGGG CCCCGGCGTA GCGATTTCAG AGGGGGGACC CCGGCCCTG

VI
DOMAIN

TTCTGCAGAA AGGGGGCTCT GAGGCAGAAG GTGGTCCACG AAGTCAAGAG

EXON 1

CCACAAGTTC ACCGCTCGCT TCTTCAAGCA GCCCACCTTC TGCAGCCACT

GCACCGACTT CATCTG gtg agggaagggg gctgggggac tgggggacga

C1
DOMAIN

ggggactagg ggtgcagact cctatcacgc cgaccctgt ggaaggaaga

IR1 ←

aggagggggc tgtagtcccg actcccaggt tctaggatgg ccagggaa

// tggcc ggggcttggc cacctgggcc ctgcggaag aaggtcacan

IF2 →

ancgatgcg cctgtggctc ncagagttgg ggtccaggt acccctttct

gactgacct aggatccctg actcttcag GGGTATCGGA AAGCAGGGCC

EXON 2

TGCAATGTCA AG gtaagag ctggggaccg gggctcctgg gaccctcagg

IR2 ←

agggtggagg ctggggcccc aacagctgag gctgcttgac acacgtgttc

tctggtcccc agagagggcg gggg //

c cctctcttc tggttttctc tgtgtccgag

ttccgctctc tctttccaat tttctgtctg ctggggtctccc gctggact

IF3 →

aatccatgcctc cgtctgtgtc tctatgattt tcatctatag TCTGCAGCT

EXON 3

TTGTGGTTCA TCGACGATGC CACGAATTTG TGACCTTCGA GTGTCCAGGC

GCTGGGAAGG GCCCCCAGAC GGAC gtgag tgctcggaca cctggttctc

ctcctcgggc cgtgcccccg cctcaccctc ctgggctcc gtcccaattt

IR3 ←

ctcctgctat ttttatggct gggagggg // ac acagcaggtg

ctcaatgatt attggtacat agagtгааag agatggagcc tcaggctgac

ctagagagca → IF4 aaaagataaa agggcccctc ccctgggggtt

ttaggaccct cccaacgccc cctaagccag tcttctctgc ccccag GAC

CCCCGGAACA AACACAAGTT CCGCCTGCAT AGCTACAGCA GCCCCACCTT

EXON 4

CTGCGACCAC TGTGGCTCCC TCCTCTACGG GCTTGTGCAC CAGGGCATGA

AATGCTCCT gtgagtgacc tgggccttgc cagggccctt ccaagcgcc
IR4 ←

cggctctgggt tccgggaaat gcccgggatg ggggtgggggg tggagtcttg

gcttggggcg gggcctgagg tgctaccgc agctttcccc tccag GCTG
IF5 →

CGAGATGAAC GTGCACCGGC GCTGTGTGCG TAGCGTGCCC TCTCTGTGCG

EXON 5

GTGTGGACCA CACCGAGCGC CGCGGGCGCC TGCAGCTGGA GATCCGGGCT

V2
DOMAIN

CCCACAGCAG ATGAGATCCA CGTAACTG g tgaggccccg cccctcgcc

tggccccgcc ccctcccaa gtgtgaggcg gggctgacce aaggcacttg
IR5 ←

tcgtggcca gccctacccc aaagatgggg ccacgcctct ttctatggtc

acgccacac tcttgacccc accccaaagg ccgagcacac ccagccata

// acctgccac cagccccagc taattttttt

tttttttttg tatttttagt ggnagacgg ggttcacca ttcacaggat

ggtcccgatc tctgacctt gtgatccgcc cgccttggcc tcccaaagtg

ctggggatta caggcatgag ccgccgtgcc tggccaagct tggactctt

gattgctgac tggaggagge tgggagcccc ttctggatc tctaaccgt

cacactcttc → IF6 ctcactcccc gtttag TTG GCGAGGCCCG TAACCTAATT

C2
DOMAIN

CCTATGGACC CCAATGGTCT CTCTGATCCC TATGTGAAAC TGAAGTCAT

EXON 6

CCCAGACCCT CGGAACCTGA CGAAACAGAA GACCCGAACG GTGAAAGCCA

CGCTAAACCC TGTGTGGAAT GAGACCTTTG TGTT gtgag tctgggggtgc

aggaaggca atgacagctg acagagaatg aggg // aaggga

IR6 ←

gagaagagct ctctaggttt acttcaggcc ccaaagccct agctggagag
agagcccggc tgggaaggca gaggtcggag accgacaaag caggagagga
gncccagctg gctgggtttg cccccacctc cagcaccaag gatggggaac
cgaggggagc catgagctcg gctctgcacc ^{IF7} ccaccacccc accttctctg

ag CAACCTG AAGCCAGGGG ATGTGGAGCG CCGGCTCAGC GTGGAGGTGT
GGGACTGGGA CCGGACCTCC CGCAACGACT TCATGGGGGC CATGTCCTTT
GGCGTCTCGG AGCTGCTCAA GGCGCCCGTG GATGGCTG g tgaggagcag

EXON 7

ggctggggcc tggggatgga gcgcaatatt accatctcca tctgtgtgtg
^{IR7} gtctctctcc tccaggccac tgtccttccc tctgectccc agcatgcgcg
cacacacaca cacacacaca cacacgcaca cacacgcaca cacacacccc
tctctctcta ttcttctett cttctcccct cccctttctc cctctccctc
tctttttatc tcaactcttc tctcttccat ctctgtgtcc gtctctctgt
gtctctttcc tcccttcaa tgtctttgcc tctcccatgg gtgccccatc ^{IF8}

cccgctgccc gctctggtc tccgtctgta tgtcag GTA CAAGTTACTG

EXON 8

AACCAGGAGG AGGGCGAGTA TTACAATGTG CCGGTGGCTG ATGCTGACAA

V3
DOMAIN

CTGCAGCCTC CTCCAGAAGT TTGAGgtacc cagaaccctg gcttctcaa
gggagcccag cccagcctcc cacggttcag acctggcctt tccttccacc
cctgagtgcc cgctggctct ggggactaca gttcccagaa gaccctagga
ctccctcctc tgctcttcta ggggactcga gcccagggt ctgatgggaa
ttatagttcc tatctatcgc catggcttga ^{IF9} gggactaggg gccaccagc

ccctgttcta gggcgatccc ctgcattctt tgggaccctg actctctctt
tcttttctcc cag GCTTGT AACTACCCCC TGAATTGTA TGAG gtgag

EXON 9

tagaaccagg gcggtgaatg gaggcagttt ttgcctactt ctctgatttc
^{IR9}

ttattcctcc tctgacttct gtcttcaatt cccacacat gagttgagca

cacatttggtg ctaggcctgt // ggngct gtgntcttg

gggagcattt ccttatcggc **IF10** tgtgtaaggt ctaactgcnt ctggctcttt

V3
DOMAIN

ctttctcctt tccacag CG GGTGCGGATG GGCCCTCTT CCTCTCCCAT

EXON 10

CCCCTCCCCT TCCCCTAGTC CCACCGACCC CAAGCGCTGC TTCTTCGGGG

CGAGTCCAGG ACGCCTGCAC ATCTCCGACT TCAGCTTCCT CATGGTTCTA

C3
DOMAIN

GGAAAAGGCA GTTTTGGGAA G gttggatt cctgggggttc tgggggaaag

ggaggatgtc tgtgggaagg tcagatttct ggttcttagg gaggaagtgg

IR10 ←

gggtgggaag agactgggt cctgtgcac ttcaaatatg gttaggttgg

gccgttcagg ttcttgaga ggagaggtt acagatgtgg acactctcct

tgaggggacg ggcggcaagt cagggctgtc agtcccttaa gagatggagg

aa // ttaggg agggggcagg tctgtacca ctgggtccc **IF11** →

aacatggact ggccttttg gaactgtgag catag GTGA TGCTGGCCGA

EXON 11

GCGCAGGGGC TCTGATGAGC TCTACGCCAT CAAGATCTTG AAAAAGGACG

TGATCGTCCA GGACGACGAT GTGGACTGCA CGCTGGTGGG GAAACGTGTG

V4
DOMAIN

CTGGCGTGG GGGGCCGGGG TCCTGGCGGC CGGCCCACT TCCTACCCA

GCTCCACTCC ACCTTCCAGA CCCC gta aggatggagg ggcggaggc

tgctctccgg gcctgcctt atccagttct ggacatctgc gttgggatc

IR11 ←

tgagtttagg gcgaggcaag agaacttt // tc aggaagaaa

ttctctact ctggtagat ggatcccgcc tctaagccca tgcacttctc

C4
DOMAIN

cgcag GACC GCCTGTATTT CGTGATGGAG TACGTCACCG GGGGAGACTT

EXON 12

GATGTACCAC ATTCAACAGC TGGGCAAGTT TAAGGAGCCC CATGCAGC

gtgagtctcg gccaacagag aatggctcggg gtggtggaag ggggcaggat

ccagccactg accttctgac gtccccaccc acccgtcct ccag GTTC
TACGCGCAG AAATCGCTAT CGGCCTCTTC TTCCTCACA ATCAGGGCAT
CATCTACAG gtgagcagcc ccaggaattt ccgtggagga aatcacgccc
ctggaaggg aagggatttg aatatgtggc tctagactgc tgaactcaac

EXON 13

acttottgca attcctgccc cacaccctg catcgtccag GGACCTGAA
GCTGGACAAT GTGATGCTGG ATGCTGAGGG ACACATCAAG ATCACTGACT
TTGGCATGTG TAAGGAGAAC GTCTTCCCGG GGACGACAAC CCGCACCTTC
TGCGGGACCC CGGACTACAT AGCCCCGGAG gtaacccc aacctgctg

EXON 14

ctctggtcac gctttgagat cccttagagg gtgtagctga tggccagta
ttcaccaggg gtgaaggcct gaccctcaga ccttgtcatg agttgtggcc
ttcttacaca // ggaagag ctttggtga
aagcacttaa cgtgggtagc gttcccagg gggtagggcc agaggggtcc
taggcttctt aaagaacgca tcatgattcc ctgccttcca cctcccctag

EXON 15

ATCATTGC CTACCAGCCC TATGGGAAGT CTGTCGATTG GTGGTCCTTT
GGAGTTCTGC TGTATGAGAT GTTGGCAGGA CAG gtaagg gaaggtgggg
agaagctggc ttggctaaaa gagacagaga ggggcacctg gatctcagga
ggagccagtt agaaaggagc ccagaaggtt gtgctcgaat agccc
// aaag gcagttgggc atgtccctga ctctctatcc cctcccacttt

EXON 16

gatag CCTC CCTTCGATGG GGAGGACGAG AAGGAGCTGT TTCAGGCCAT
CATGGAACAA ACTGTCACCT ACCCAAGTC GCTTCCCGG GAAGCCGTGG
CCATCTGCAA GGGG gtga gagccccctg actcccagct tctccaggct
cacaaccaca caccctattg ctgtctctgt gcctattaga aaaatgctcc
cattcctgaa gtcactttac ttccatctgt tggaaaagtt gatatgatgc

ataggttttg ttagaacaat gatttccagc cc // ttcc

tccatctgcc tgtctctgtg ggtctttctc tggatgtacc tgcctcgac

▶ IF17
tctgtctgtt tgtctgtctg tctctctctg tgtttccac ag TTCCTG

ACCAAGCACC CAGGGAAGCG CCTGGGCTCA GGGCCTGATG GGAACCTAC
CATCCGTGCA CATGGCTTTT TCCGCTGGAT TGA CTGGGAG CGGCTGGAAC
GATTGGAGAT CCCGCCTCCT TTCAGACCCC GCCCG gtca gtcaccctcc
IR17 ◀

EXON 17

aggcaacaaa aacctggtcc ctgaaggggt ggggttcccc tgggectcaa

tatacctcta tgtgggggtg gggttccctc tgcagagccc cccgccccca

acaaaaggag gtgcagacac catgaagcat gaatagagat tctgcaggag

acaggagatg agactggggt acacagaggg acacccgagg ag

// agagcctcgg agctgcttaa ctttccctcc cccagctctc
IF18A

ccacag TG TGGCCGCAGC GCGGAGAACT TTGACAAGTT CTCACGCGG
GCGGCGCCAG CGCTGACCCC TCCAGACCGC CTAGTCCTGG CCAGCATCGA
CCAGGCCGAT TTCCAGGGCT TCACCTACGT GAACCCCGAC TTCGTGCACC
CGGATGCCCG CAGCCCACC AGCCAGTGC CTGTGCCCGT CATGTAATCT
IF18B ▶

EXON 18

V5
DOMAIN

CACCCGCCGC CACTAGGTGT CCCCAACGTC CCCTCCGCCG TGCCGGCGGC
IR18A ◀

AGCCCCACTT CACCCCAAC TTCACCACC CCTGTCCCAT TCTAGATCCT

GCACCCAGC ATTCCAGCTC TGCCCCCGCG GGTTC TAGAC GCCCCTCCA

3' UTR

AGCGTTCCTG GCCTTCTGAA CTCCATACAG CCTCTACAGC CGTCCCGCGT
IR18B ◀

TCAAGACTTG AGCG

4.6 Mutation analysis of the *PRKCG* gene

The genomic structural analysis established that the *PRKCG* gene contains 18 exons spanning at least 16 kb of genomic DNA. In addition to this, the 5' promoter region of *PRKCG* extends 1146 bp upstream from the ATG translation initiation site (Mahajna *et al.*, 1995). In order to conduct a full mutation analysis of the *PRKCG* gene in families linked to RP11, intronic primers that were designed previously to cover the entire coding region were utilised. Moreover, primers were designed to cover about 1 kb of promoter region (fig. 4.11).

4.6.1 Methods and materials

4.6.1.1 Primers and mutation detection methods employed

All *PRKCG* exons, together with about 30-50 bp of flanking intronic sequence, were amplified from the five families linked to the RP11 locus at the time of screening. These were ADRP5, ADRP29, RP1907, a Japanese family and ADRP2. The panel of DNA for mutation analysis was composed of two symptomatic-affected haplotype carriers and two normal individuals from each RP11 linked family (altogether 20 DNA samples). PCR products, with sizes ranging from 168 - 318 bp, were first analysed on 2.5% agarose gels to investigate the quality of PCR amplification and also to detect possible insertion /deletion events. PCR products were then separated on MDE gel matrix for heteroduplex analysis (see section 2.5.4) and direct sequenced using ABI PRISM™ dye terminator cycle sequencing ready reaction kit with AmpliTaq® DNA polymerase FS, (Part number 402078). This kit was used for the direct sequencing of all *PRKCG* exons (see section 2.10.3). Amplimers for *PRKCG* exons are given below with their annealing temperature and size of PCR product. Exon 18, which also includes the available 3' UTR sequence, was amplified in two fragments as it was too large to be screened by heteroduplex analysis as a single fragment.

Intronic primers used to screen exons (F/R 5'-3'):

- Exon 1:** agaaaggcaggatcctggct/ cggcgtgataggagtctgca - 65°C / 274 bp
Exon 2: ttggacacctgggcccctg/ctgagggctcccaggagcc - 65°C / 171 bp
Exon 3: gctggactaatccatgcctc/ aggagaaattgggacggagc - 60°C / 213 bp
Exon 4: gctgacctagagagcaaggc/gctttggaaggccctggca - 64°C / 221 bp
Exon 5: tgaggctacctccgagctt/ acaagtgccttgggtcagcc - 64°C / 234 bp
Exon 6: ctctaaccgctcactctt/ tctgtcagctgtcattgct - 60°C / 234 bp
Exon 7: gccatgagctcggctctgca/ gtaatattgcgctccatccc - 65°C / 223 bp
Exon 8: tgcctctccatgggtgc/ aaggccagctctgaacctg - 60°C / 213 bp
Exon 9: ctatctatgcccattgct/ aactgcctccattcaacg - 58°C / 168 bp
Exon 10: gagcatttcctatcggctg/ aaccagaaatctgacctccc - 55°C / 280 bp
Exon 11: aggtcctgtaccactgggtt/ atcccaacgcagatgcccag - 65°C / 318 bp
Exon 12: gtgatggatcccgcctcta/ acgtcagaaggtcagtggtc - 58°C / 205 bp
Exon 13: agccactgacctctgacgt/ gtgtgagttcagcagtctag - 60°C / 196 bp
Exon 14: ctgactgctgaactcaaca/ taaggatctcaaagcgtg - 56°C / 242 bp
Exon 15: gcacttaacgtgggtagc/ tagccaagccagcttctcc - 56°C / 212 bp
Exon 16: gcatgtccctgactctctat/ agtgactcaggaatgggag - 60°C / 245 bp
Exon 17: atgtacctgtccggcact/ accaggtttttgtgctgg - 55°C / 232 bp
Exon 18A: ctcggagctgcttaacttcc/ acgtggggacacctagtg - 64°C / 262 bp
Exon 18B/ 3'UTR: actctgtcacccggatgc/ agtcttgaacggggacgg - 62°C / 270 bp

Primers used to screen 5' promoter region (F/R 5'-3'):

Five primer pairs were designed to screen about 1 kb of promoter region. The five fragments, called upstream fragment 1, 2, 3, 4 and 5 (UF1, 2, 3, 4 and 5), each had about 30 -50 bp of overlap with the adjacent fragment.

- UF 1-** ctcggtcgtcctgcaacgcc/ ggaccaggatcctgcctttc
UF 2- gatgacagattgcagggtgg/ cggcacctgctgaaatgtgg
UF 3- aagactcagccggcatcgg/ atcctctgacccctggctcc
UF 4- tggagacacacagcagagg/ tcactttctccgatccgg
UF 5- agacagagacggactccca/ cctgtctccgagaatgc

4.6 2 Results

4.6.2.1 Mutation in the family RP1907 and ADRP24 segregates with disease

On heteroduplex analysis, a subtle heteroduplex was detected in exon 18 (exon 18A) of *PRKCG* amplified from two affected individuals from the family RP1907. Direct sequencing of these PCR products revealed a heterozygous missense point mutation, a C to A transversion of the first nucleotide in codon 659 predicting an

Arginine to Serine substitution in the protein (Arg659Ser) (Fig 4.12a). This nucleotide alteration also eliminated an *AciI* restriction site (Fig 4.12b); a restriction enzyme digestion of the PCR products was therefore used to confirm the mutation. Using this assay the complete RP1907 family was typed, and we found perfect co-segregation of mutation with the disease haplotype (see fig 4.13).

Next exon 18, amplified from affected individuals from small RP families, sporadic RP cases and 250 unrelated normal controls of Caucasian origin, was analysed by heteroduplex analysis. This was to see if the codon 659 mutation was present in other RP cases, both familial and sporadic, and also to investigate the frequency of this mutation in the general population. The characteristic heteroduplex was detected in two affected members from the small family ADRP24 and two other sporadic RP cases, but it was not detected in any of the 250 normal controls. Direct sequencing and restriction enzyme assay confirmed the presence of the mutation in these individuals. The restriction enzyme assay performed on the complete ADRP24 pedigree also showed co-segregation of codon 659 mutation with the disease (see fig 4.13). ADRP24 had previously shown tentative linkage to RP11, but it was too small a family to give a lod score value above 3 which indicates definite linkage. The co-segregation of the disease with this highly informative mutation however proved linkage of ADRP24 to RP11 (see sections 3.3.2 and 3.3.3).

As the codon 659 mutation appeared to be a rare occurrence, the possibility of a founder event was investigated. Microsatellite markers that flanked *PRKCG* gene were typed in the complete pedigrees of RP1907 and ADRP24 and in the two sporadic cases. This revealed a shared haplotype that extended from the marker D19S921 to D19S605 (see section 3.3.3, fig 3.5 and fig 4.13) over 9 cM genetic distance, which confirmed a founder event. Therefore it appears that the mutation in codon 659 occurred as a single event, in a mutual ancestor of RP1907, ADRP24 and the two sporadic RP cases, rather than as three separate mutational events.



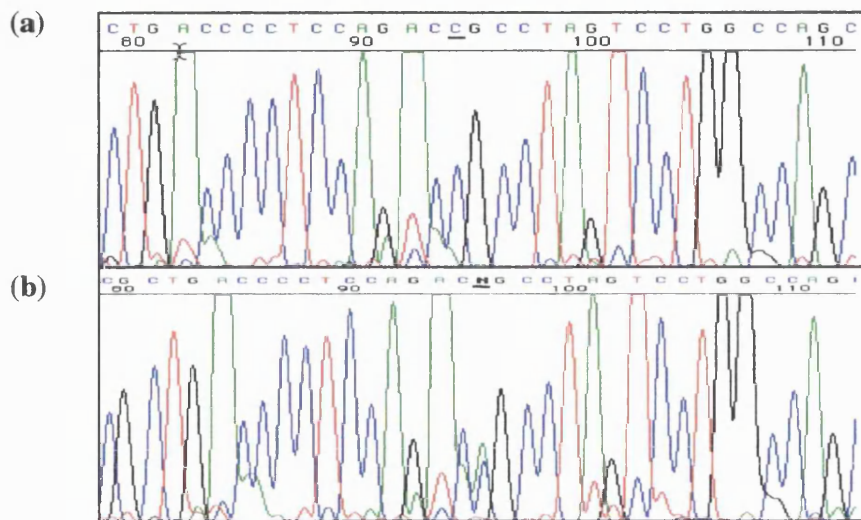


Figure 4.12 a.

Comparison of sequence data from a normal (a) and a heterozygous mutant (b) *PRKCG* exon 18 sequence. A C-A transversion (underlined bases) changes codon 659 from arginine (CGC) to serine(AGC).

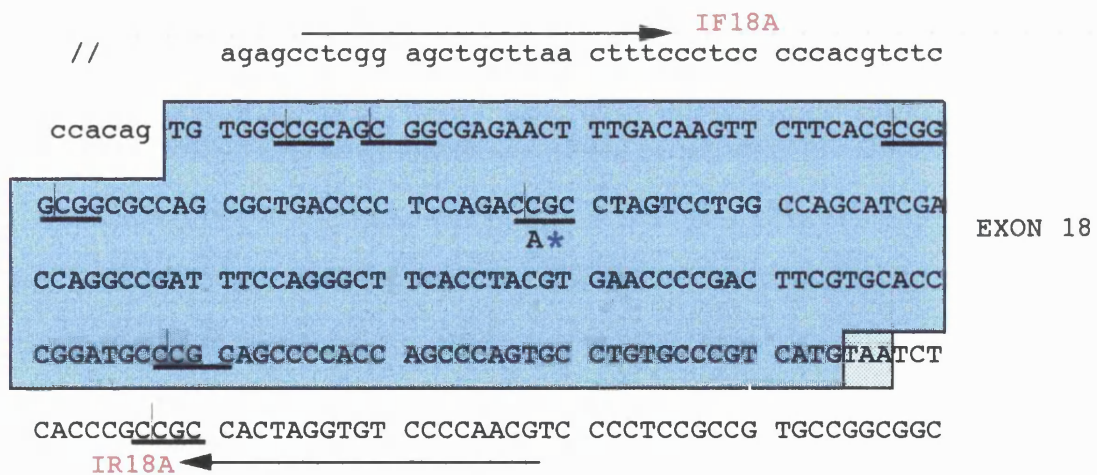
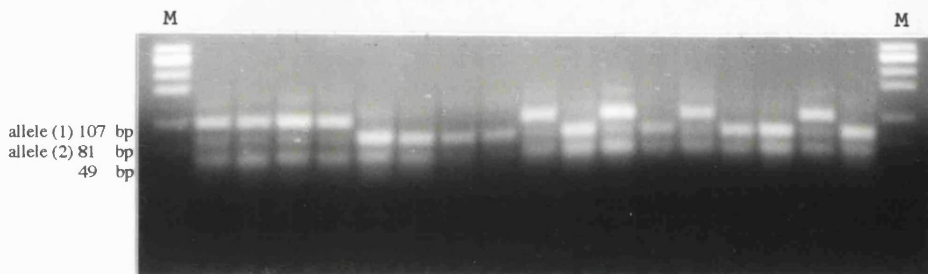
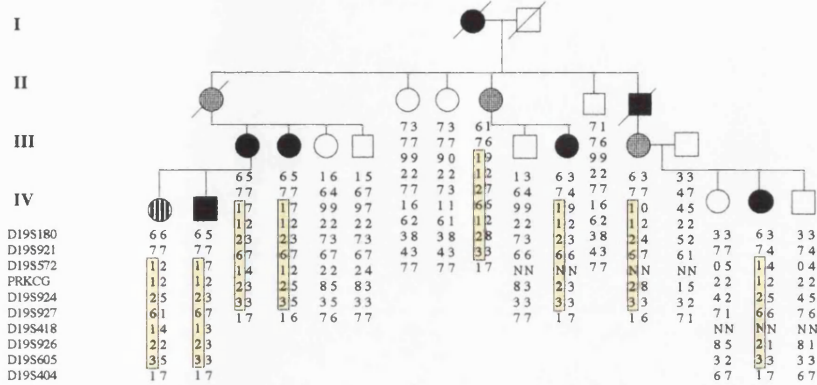


Figure 4.12 b

The sequence of exon 18 (262 bp) depicting the 7 *Aci* I restriction sites within the sequence (underlined 5' -3' C/CGC and 5'-3' G/CGG, *Aci* I has a non-palindromic recognition site). The C to A transversion of the first nucleotide in codon 659 (indicated by a *) eliminates the fifth restriction site producing a larger 107 bp fragment (allele 1/mutated allele) instead of the two fragments (allele 2/ normal allele), which are 26 bp and 81 bp in size.

(a) **RP1907**



(b) **ADRP24**

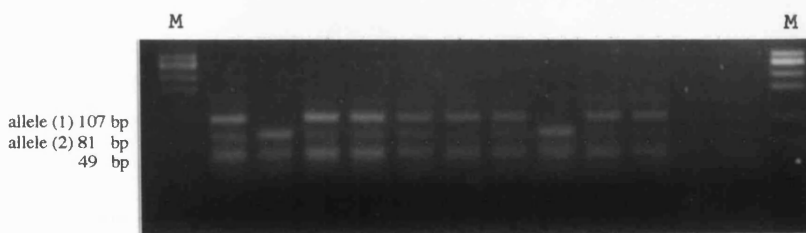
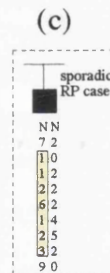
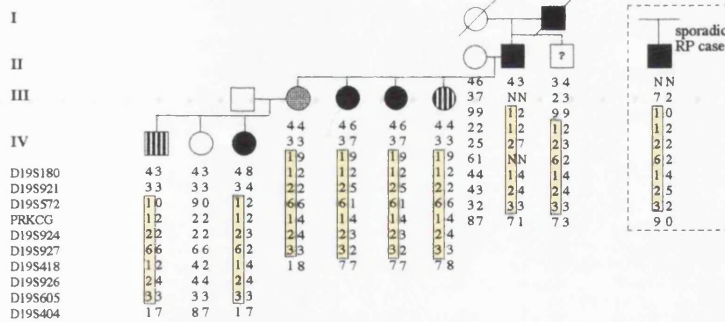


Figure 4.13

Pedigrees of (a) RP1907, (b) ADRP24, and (c) a sporadic RP case, showing the segregation of codon 659 mutation by way of an *Aci* I restriction assay on exon 18. The linked haplotype for 19q markers, including the *PRKCG* mutation, are shown below the pedigrees. Allele 1 for *PRKCG* represents the C-A mutation at codon 659, whereas allele 2 represents the normal sequence. The common haplotype in (a), (b), and (c), between markers D19S921 and D19S605 is boxed. Restriction digests of exon 18 PCR products, which demonstrate absence of the *Aci* I site in the mutated alleles, are shown, with fragment sizes of 107 bp (for the mutated allele), 81 bp (for the normal allele), and 49 bp. M- ϕ X174 DNA marker.

4.6.2.2 Mutation in the Japanese family

At the time the codon 659 mutation was detected, a second mutation in exon 18 was also detected in the Japanese family. While the codon 659 mutation produced a subtle heteroduplex band this second mutation gave a very clear heteroduplex band pattern. When exon 18 was screened in the complete Japanese family the mutation was also seen to co-segregate with the disease (see fig 4.14). Direct sequencing of PCR products identified the sequence change as a 'silent' third nucleotide C to T point mutation in codon 647 (Phe) of exon 18 (see fig 4.15). As mentioned before this mutation was not detected in any of the normal individuals of the Japanese family, nor was it found in any of the 250 normal unrelated controls or other RP cases screened subsequently. Unfortunately this mutation could not be tested on a panel of normal Japanese controls, therefore while this mutation appears to be absent (or very rare) among the Caucasian population the same cannot be said of its frequency in the Japanese population as yet. Codon 647 mutation, being silent, can be regarded as a rare polymorphism associated with the disease, similar to a rare allele of a microsatellite marker tightly linked with the disease. However the possibility of this mutation being pathogenic is discussed in section 4.7.2.

4.6.2.3 Polymorphisms in the *PRKCG* gene

After the discovery of the codon 659 mutation and its association with the disease phenotype in RP1907, all exons of *PRKCG* (inclusive of exon 18), amplified from affected and unaffected individuals from other families linked to RP11, were re-screened for mutations by direct sequencing and heteroduplex analysis. Yet we did not detect any other mutation in the coding sequence that also co-segregated with the disease phenotype. This was the case for families ADRP5, ADRP29 and ADRP2. However a few polymorphisms within the *PRKCG* gene were detected in both affected and unaffected family members, these are listed in table 4.5 and depicted in figure 4.16.

Japanese family

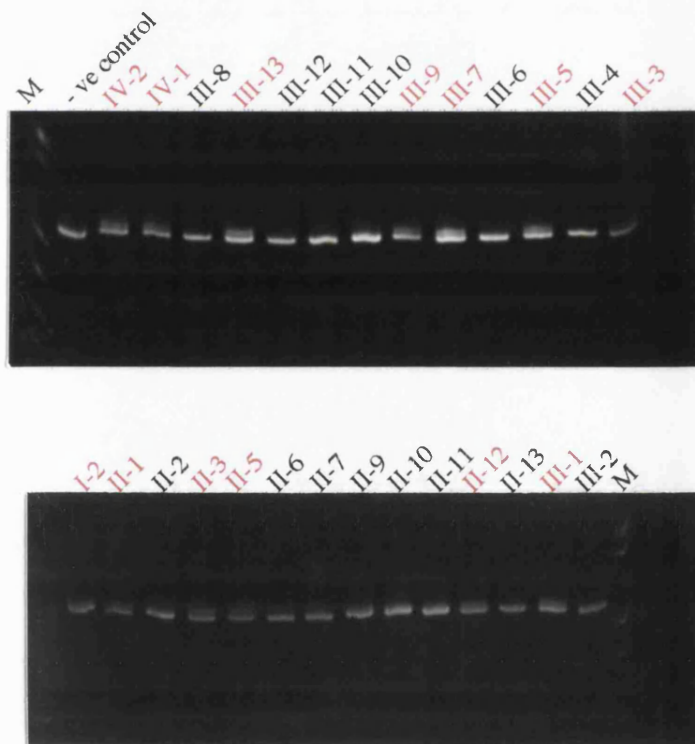
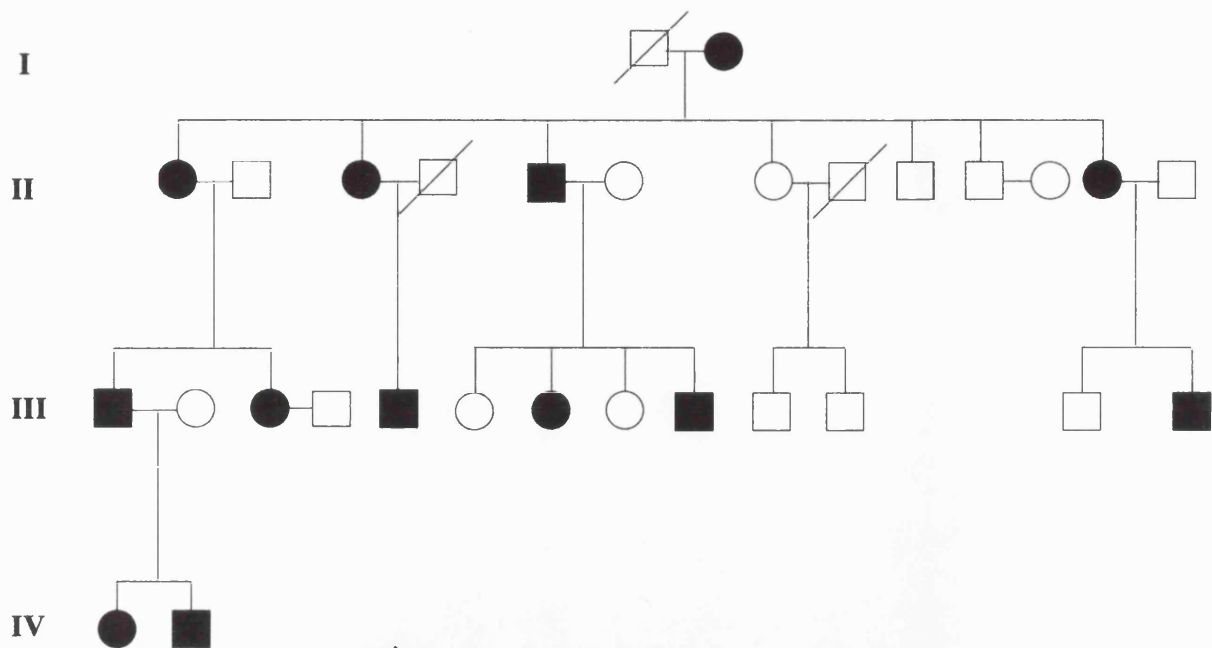


Figure 4.14

Ethidium bromide stained MDE gels depicting heteroduplex band patterns produced by codon 647 mutation of exon 18 in the entire Japanese family. Individuals showing heteroduplexes in exon 18 are indicated in red. The negative control is an exon 18 PCR product without any mutation. M- Φ X174 DNA marker.

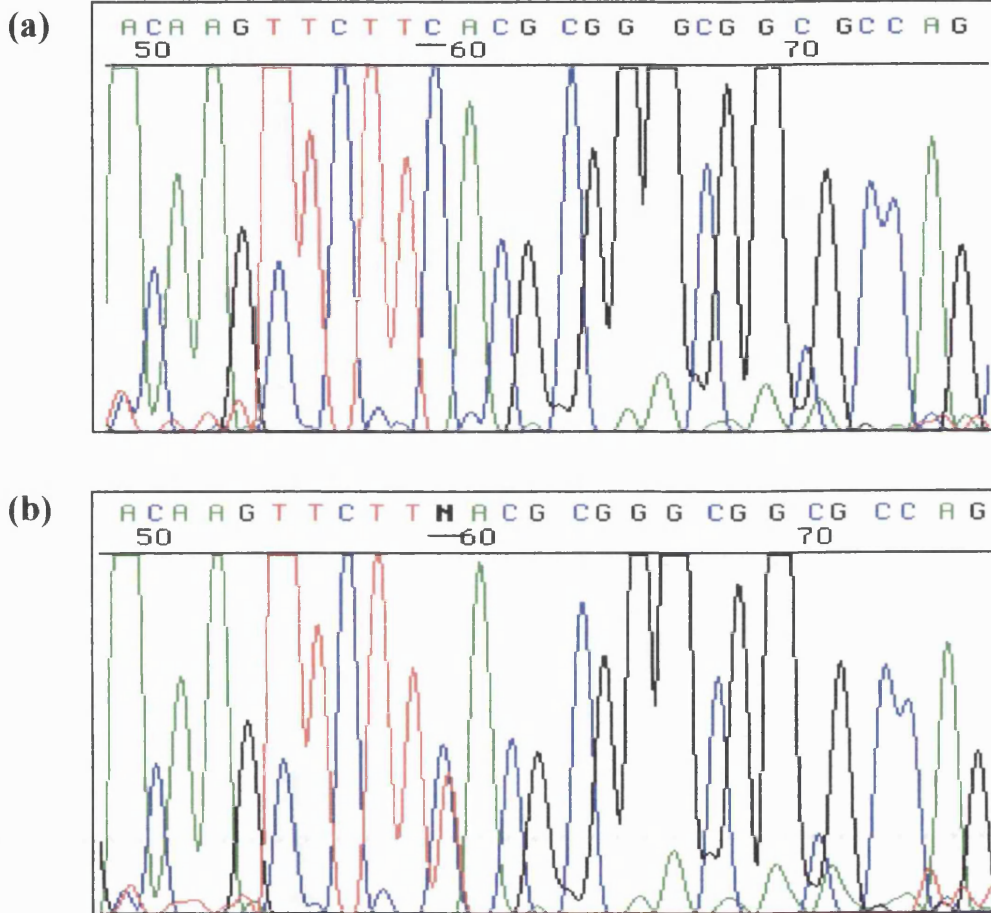
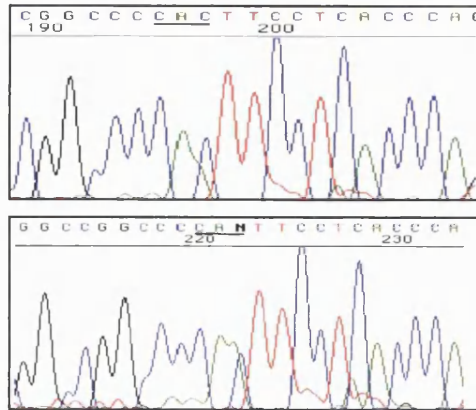


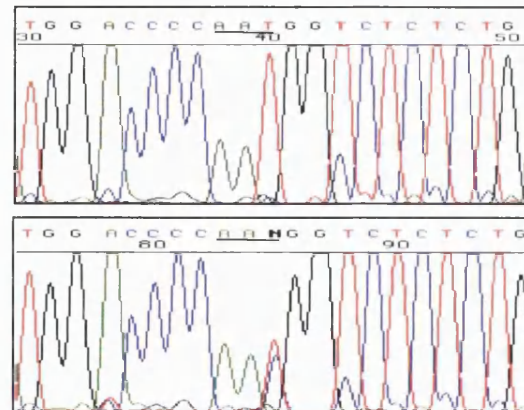
Figure 4.15 .

Comparison of sequence data from (a) a normal and (b) a heterozygous mutant *PRKCG* exon 18 sequence, from the Japanese family. The C to T change at codon 647 (underlined bases) does not cause an amino acid change.

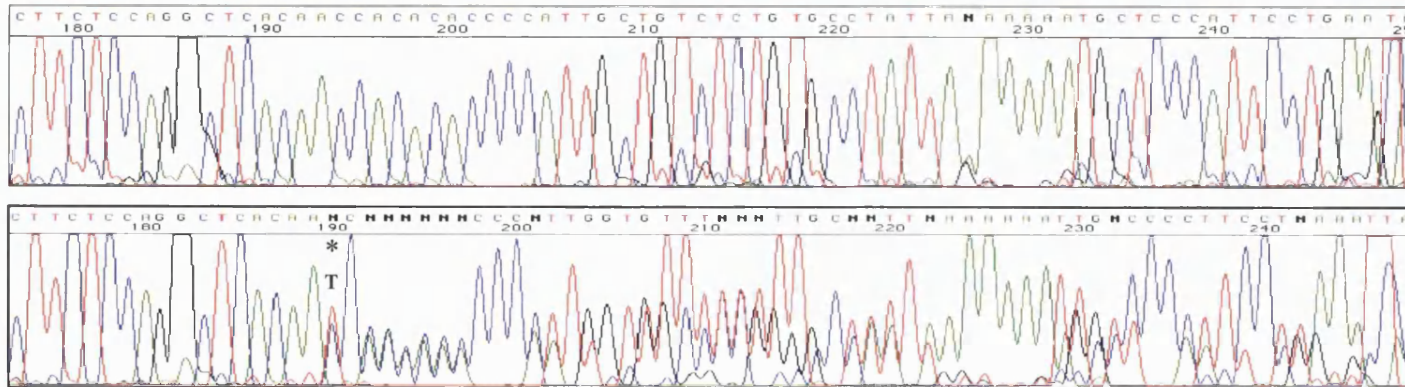
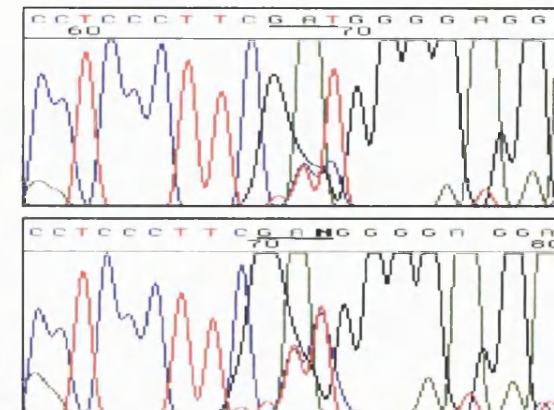
(a) Exon 11 (CAC-CAA)



(b) Exon 6 (AAT-AAC)



(c) Exon 16 (GAT-GAC)



(d) Exon 16
T bp insertion

Figure 4.16

Sequence depicting four DNA polymorphisms identified during the mutation analysis of the *PRKCG* gene. A homozygote (top panel) and a heterozygote (bottom panel) is shown for each polymorphism.

Exon	Codon	Sequence change	Mutation	Detection method Sequencing (S)/ Heterduplex Analysis (HA)
1	24	GCT → GCC	Silent (Ala)	S
6	189	AAT → AAC	Silent (Asn)	HA+S
11	415	CAC → CAA	His → Gln	S
14	499	TTT → TTC	Silent (Phe)	S
16	556	GAT → GAC	Silent (His)	S

Table 4.5

Polymorphisms identified during the mutation analysis of *PRKCG* gene.

In addition to the point mutations listed above a T nucleotide insertion in intron 16, 39 base pairs downstream of exon 16, was also detected in both affected and normal individuals. The frequency of these polymorphisms in the general population was not tested by analysis on normal controls.

4.6.2.4 Screening of the 5' untranslated region of the *PRKCG* gene

Five pairs of amplimers were designed to cover approximately 1 kb of 5' promoter region of *PRKCG* (see section 4.6.1.1). These were used on two affected and two normal members from each RP11 linked family to screen the promoter region of *PRKCG*. A heteroduplex was detected in the UF2 fragment amplified from two affected individuals from the family ADRP29. Direct sequencing of the PCR product revealed a point mutation at a potential sp1 site, which was a C to an A change in the sequence at the nucleotide position - 49 (+1 being the start of transcription). However, this mutation did not co-segregate with the disease in the family and it was also detected in normal controls. Another point mutation was detected at nucleotide position -300 which was an A to T change, as with the other mutation this was detected in both affected and normal individuals.

When the sequence data generated at the laboratory was compared with the *PRKCG* 5'upstream sequence originally reported by Mahajna *et al.* (1995), several discrepancies in the published sequence were identified. These sequence changes were not polymorphisms but actual sequence changes observed in all of the sequenced individuals. There were nucleotides missing at positions -53, -62, -82, -238, -258 and -292 (a C at -53, a G at -62, a C at -82, a G at -238, a C at -258 and a G at -292). Also there was an extra G at position -172 in the published sequence, which also reported a C at the nucleotide at position -135 instead of a G. Apart from these changes the sequence generated in our laboratory matched the published sequence.

4.6.2.5 Investigation of large rearrangements within the gene

To detect large rearrangements that involve several kilobases of DNA, such as deletions or insertions, southern blots of DNA digested with a range of restriction enzymes (i.e. *Bam*HI, *Msp*I, *Taq*I and *Hae*III) were prepared from RP11 linked families. Each family was represented by one affected and one normal member. The blot was hybridised individually, with the *PRKCG* cDNA probe (see section 4.5.1.1) and the purified exon 18 PCR product to detect possible mutations at both 5' and 3' UTR end of the *PRKCG* gene. However no significant rearrangements were detected. Moreover, to see if there were heterozygous intronic mutations such as deletions (or insertions) within introns at the primer-annealing site that may also extend in to the exon, long PCR was performed to amplify two adjacent exons including the intron in between. This was not very successful, as some of the relatively large introns (i.e. intron 5, 11 and 14) were not amenable for long PCR. Yet the introns that did amplify did not produce any anomalous PCR products, which would indicate the presence of such mutations.

4.6.2.6 Mutations in *Drosophila* eye-PKC gene (*inaC* mutants)

In *Drosophila*, the *inaC* (inactivation *no* after potential *C*) mutants with defective deactivation kinetics led to the discovery of the eye-PKC gene (see section 4.3.2.1). Three *inaC* mutants, described as *inaC*²⁰⁹, *US*³⁷⁹¹ and *inaC*²⁰⁷, have the following mutations in the eye-PKC gene (Smith *et al.*, 1991). The mutation in *inaC*²⁰⁹ is a replacement of Trp93 by a stop codon. In *US*³⁷⁹¹ Trp139 is replaced by a stop codon and in *inaC*²⁰⁷ the mutation leads to Val201 to Asp substitution. These mutations

correspond to codons 57, 103 and 165 of the *PRKCG* gene, respectively, and they were looked out for during the mutation screening procedure. However as described earlier no such mutations were identified during the analysis of *PRKCG*.

4.7 Discussion

4.7.1 The efficiency of mutation detection methods employed

In light of the conflicting results obtained in the mutation analysis of the *PRKCG* gene, the question that inevitably arises is “could we have missed mutations?” The *PRKCG* gene was screened by two methods simultaneously. While the heteroduplex analysis method is known to miss mutations, the direct sequencing method should detect all mutations. The direct sequencing method employed in this study was based on ABI technology, which is now used widely for mutation detection. The radioactive method of direct sequencing (see section 2.10.1.) was attempted initially, but was abandoned in favour of the ABI method as it was not successful for some exons in *PRKCG* that were particularly GC rich. Provided that normal controls are also sequenced, one could easily eliminate “false heterozygous mutations” or sequence anomalies sometimes encountered in the chosen method of sequencing. Also by including two affected and two normal individuals from a given family in the mutation analysis, the chance of detecting mutations was maximised. Moreover, it is highly unlikely that three independent mutations or a very common mutation were overlooked in three RP11 linked families in the mutation analysis. Therefore we can be quite certain of not missing any mutations in the coding region or in the 5' promoter region. However, we cannot be equally certain of not missing possible mutations within introns or in the still incomplete 3'UTR sequence of *PRKCG*. In answer to the latter possibility, cosmid clones that contain the 3' UTR of *PRKCG* gene (recently ascertained from the LLNL genome centre) are being sequenced. Once complete, the 3' UTR sequence will also be screened in all of the RP11 linked families.

4.7.2 The codon 647 mutation in exon 18 of *PRKCG*

The mutation at codon 647 is a C→T transition at position + 36 of exon 18. This mutation (TTC-647-TTT) does not cause an amino acid change and was therefore regarded as a silent mutation. Silent mutations that do not lead to an amino acid change are frequently found in coding sequences and are generally considered to be normal variants. However the fact that codon 647 mutation also segregates with the disease phenotype implies that it may be involved in causing RP in the Japanese family.

Involvement of silent mutations in disease causation has been reported. In two such reported instances, silent nucleotide substitutions created new donor splice sites whose exclusive utilisation led to stable mutant transcripts with in frame deletions in the growth hormone receptor and the fibroblast growth factor receptor II, which caused Laron dwarfism (Berg *et al.*, 1992) and Crouzon syndrome (Li *et al.*, 1995b) respectively. In Marfan syndrome a silent mutation in the fibrillin-1 gene (*FBNI*) induced skipping of an entire exon (Liu *et al.*, 1997). RT PCR is the most commonly used method to detect irregular transcripts produced by mutated copies of genes. But in the latter example of Marfan syndrome, the disease causing potential of the silent mutations via abnormal splicing was tested *in vitro* by using an exon-trapping expression vector. Abnormal splicing of the *PRKCG* transcript was not investigated in our study, the main reason being the unavailability of reagents for RT PCR such as fresh tissue/blood. However, the examples given above and the codon 647 silent mutation strongly favour an investigation in to such a possibility.

4.7.3 Is *PRKCG* the gene for RP11?

From the evidence presented so far, the main evidence that is in favour of *PRKCG* being the RP11 gene is the close genetic association between the RP11 phenotype and the Arg-659-Ser mutation in families RP1907 and AD24. Furthermore, the non-conservative replacement of a basic arginine with an uncharged serine at codon 659, a codon conserved in human, bovine and rat *PRKCG*, suggest that this mutation is unlikely to be a neutral polymorphism. These observations, in conjunction with the involvement of a PKC gene in retinal degeneration in *Drosophila* (Ranganathan *et al.*

1991; Smith *et al.*, 1991), suggests that Arg-659-Ser could be a possible cause of RP in these families.

If PKC γ is involved in rhodopsin phosphorylation, as described above, a mutation in this gene might be expected to have an effect on adaptation or recovery of photoreceptors in response to light flashes or to high or low light levels. Mutations in rhodopsin kinase and arrestin, both similarly involved in the restoration of resting potential in photoreceptors after light stimulus, have been shown to cause a rare recessive form of congenital stationary night blindness known as Oguchi disease (Fuchs *et al.*, 1995; Yamamoto *et al.*, 1997). As with Oguchi disease one might expect a mutation in *PRKCG*, an enzyme, to cause a recessive rather than a dominant phenotype. However a mutation which affected the ability of the photoreceptor to restore rhodopsin to an inactive state might lead to a background level of constitutive activation, as has been hypothesised for several dominant rhodopsin mutations which cause a dominant form of CSNB (congenital stationary night blindness) as well as RP (Rao *et al.*, 1994; Robinson *et al.* 1994). Furthermore, a mutation in the β -subunit of phosphodiesterase (PDE β), an enzyme was also found to cause a dominant form of CSNB (Gal *et al.*, 1994).

However, the entire *PRKCG* gene has been sequenced in four other RP11 linked families (ADRP5, ADRP29, ADRP2 and the Japanese family) and no mutations have been found. It is still possible though highly unlikely that another, perhaps common, mutation could have been missed or could lie in introns which have not yet been fully characterised. Alternatively, and assuming that *PRKCG* is the RP11 gene in these families, the absence of mutations in a proportion of RP11 families might imply micro-heterogeneity, a hypothesis proposed to explain the apparent lack of mutations in *RPGR* in over 50% of RP3-linked families (Meindl *et al.*, 1996).

Mice deficient in PKC γ have been created using gene knockout technology. These animals exhibit mild deficits in spatial and contextual learning but no mention is made, in the published description, of defective vision (Abeliovich *et al.*, 1993; Chen *et al.*, 1995b). Several explanations can be given for this observation; it is possible that these mice had visual defects that were not tested for. It could be also be argued,

given the incomplete penetrance for RP11 phenotype in some patients, that lack of visual defect in the knockout mice may be the result of the genetic background in the inbred mouse strain that protects against retinal degeneration. Alternatively it can be said that PKC γ knockout is not a true model for *PRKCG* missense mutation (gain of function mutations) or a suitable animal model for RP11, where the genetic background plays a major role in actual manifestation of the disease phenotype. A detailed histological examination of the *PRKCG* knockout phenotype on several different mouse backgrounds could provide further evidence as to whether *PRKCG* is the RP11 gene.

So far the evidence against *PRKCG* being the RP11 gene overrides the evidence for it. The lack of visual impairment in the PKC γ knockout mouse, the lack of expression of this isoenzyme in the retina and finally the absence of mutation in three RP11 linked families all act against *PRKCG* being the RP11 gene. Even though one can give valid reasons for each of these phenomenon as explained in the first part of this discussion. Therefore it remains possible that the Arg-659-Ser change is a rare polymorphism associated with the disease but is not itself causative. The same could be said for the codon 647 polymorphism. Even if this were so, these data suggest that RP11 is in linkage disequilibrium with the Arg-659-Ser *PRKCG* allele in these two families and two apparently unrelated patients. A search for other RP patients carrying this change may reveal a wider founder effect that could substantially narrow the locus for this common form of dominant RP.

4.7.4 Allelic effect, modifier genes and *PRKCG*

The characteristic partial penetrance phenotype of RP11 has been attributed to allelic effect (McGee *et al.*, 1997; Evans *et al.*, 1995). Sib pair analysis has shown with statistical significance that the allele of the normal parent determines the phenotype of the offspring (see section 3.1.3 for more detail). Therefore it has been suggested that a polymorphism, which on its own would not be pathogenic, in *trans* with the disease allele would cause the disease. This is the simplest explanation for a possibly more complex scenario, which may involve more than one allele combination in *trans* with the disease allele. It is also possible that the second polymorphism is not in the normal allele of the disease gene but in a second gene

close by! Therefore contribution of a modifier gene may also need to be considered for disease pathogenesis of RP11.

So where does this fit in with mutation analysis of *PRKCG* gene? The affected individuals chosen for mutation screening were all symptomatic patients. However this need not have been, for regardless of the actual phenotype of the individuals, if they were disease haplotype carriers the primary mutation would be present in those individuals. However to identify the genetic cause of the partial penetrance phenotype (provided this phenomenon is indeed due to an allelic variant of the same gene inherited from the normal parent), symptomatic and asymptomatic individuals need to be analysed separately to identify this secondary genetic lesion that sets the two phenotypic categories apart from each other. However if the partial penetrance phenotype is due to the influence of a nearby modifier locus, a similar analysis need to be carried out in the modifier gene. These are facts that need to be considered when mutation screening the RP11 gene in future. If *PRKCG* had indeed proved to be the RP11 gene a secondary mutation analysis would have been executed to test the 'allelic effect' hypothesis.

CHAPTER 5

YAC CONTIG OF THE RP11 INTERVAL AND ISOLATION OF NOVEL MICROSATELLITE MARKERS FOR FURTHER LOCUS REFINEMENT

5.1 Introduction

The physical mapping of a disease-linked region is initiated once the locus has been refined through genetic linkage analysis to a suitably small genetic interval more amenable to cloning strategies. It is generally accepted that the upper limit for this genetic distance be 1cM, which in molecular distance approximates a megabase (Mb). However in most situations such narrow refinement does not occur and physical maps have been constructed across disease intervals that exceed 1 Mb.

Based on recombination analysis in RP11 linked families the genetic interval of RP11 was confined to a genetic distance of 4.1 cM between the markers D19S572 and D19S418. Since this interval could not be genetically refined any further with the microsatellite markers available in the region, the physical mapping of the RP11 interval was initiated to provide a genomic resource from which new markers could be isolated. This endeavour was further supported by the fact that the actual physical distance of RP11 might be less than that suggested by its genetic distance. It is known that genetic distances at telomeric regions are most often exaggerated due to high recombination frequency. Indeed, in accordance with this fact the physical distance between D19S572 and D19S418 has been approximated at 1.5 Mb by the physical mapping group at Lawrence Livermore National Laboratories (LLNL) who constructed a metric physical map of chromosome 19 in its entirety (Ashworth *et al.*, 1995).

5.2 Physical mapping

Construction of a physical map of any given genomic region involves the assembly of genomic DNA clones in an overlapping array (described as a contig) such that the DNA reconstituted in the contig faithfully represents the genomic region of interest in linear order. The idea behind the physical cloning of a disease region is to have direct access to the genomic segments that harbours the disease gene. Traditionally, an in depth characterisation of that particular genomic region ensues after a contig has been established. This is accomplished either through mapping of pre-characterised genes to the region or by isolation of novel transcripts within the region utilising various techniques such as exon trapping/exon amplification (Duyk *et al.*, 1990; Buckler *et al.*, 1991) and cDNA selection (Parimoo *et al.*, 1991). Mutation analysis of genes isolated in this manner has on many occasions led to the identification of the disease gene. Therefore positional cloning approach assumes no functional information about the disease gene and locates a gene solely on the basis of its map position.

As mentioned earlier a physical map must include an ordered array of markers that faithfully represent their order and relative distances in the corresponding chromosomal DNA with flexibility to allow further refinement at the nucleotide level. To satisfy this criteria physical maps have mainly been constructed with overlapping Yeast Artificial Chromosomes (YACs) that provide long range-continuous coverage, and which are aligned by markers at appropriate distances.

5.2.1 Cloning in YACs

Yeast Artificial Chromosomes (YACs), first developed by Burke *et al* (1987), with their ability to accommodate 5-20 fold greater amount of DNA than cosmids promised early on that they could serve as the bridge between the 5-Mb resolution of genetic/cytogenetic maps and the 50 kb clone level of previous recombinant DNA technology. Increased resolution in physical mapping was also provided by the concomitant development of Pulse Field Gel Electrophoresis (PFGE), which allows modified electrophoretic apparatus to separate DNA molecules as large as 10 Mb enabling the intact yeast chromosomes to be separated and sized (Schwartz and Cantor, 1984). Yeast Artificial Chromosomes contain indispensable chromosomal elements, namely the telomeres, centromeres, an origin of replication for autonomous

replication and utilises the machinery of the host cell *Saccharomyces cerevisiae* to maintain these recombinant chromosomes. Of all the vectors developed so far YACs have the highest cloning capacity, with an insert capacity of up to 2Mb. The large insert capacity of YACs meant that large genomic regions could be cloned with a few YACs and that genes with genomic spans ranging to 100's of kilo bases could be isolated in one uninterrupted segment including their regulatory regions. YACs successfully introduced into mammalian cells were shown to encode normal enzymatic products (D' Urso *et al*, 1990; Huxley *et al*, 1991). Therefore YACs were shown to be useful for fundamental research into gene expression and regulation with potential clinical applications, as well as being useful vectors for constructing DNA contigs.

Even though YACs enabled large scale cloning of DNA of complex genomes that are unclonable in bacterial systems, there are a number of operational problems associated with the use of YACs (Kouprina *et al.*, 1994; Monaco and Larin, 1994). It is estimated that 10-60% of clones in existing libraries represent chimaeric DNA sequences, i.e. sequences from different regions of the genome cloned into a single YAC. Chimaeras may arise by co-ligation of DNA inserts *in vitro* prior to yeast transformation, or by recombination between two DNA molecules that were introduced into the same yeast cell. Such chimaeras can be detected by *in situ* hybridisation of the YAC to metaphase chromosomes. Unstable clones, with the tendency to delete internal regions from their inserts, are another problem with YACs. Deletions that vary in size from 20 to 260 kb can be generated both during the transformation process and during mitotic growth of transformants (Kouprina *et al.* 1994). Loss of whole YACs is also known to occur during mitotic growth. The frequency of deletion formation however can be reduced, if not eliminated, by use of a recombination deficient strain (Ling *et al*, 1993). Co-cloning events where independent YACs are cloned into the same cell, difficulties in purifying the YAC from the host yeast background, and poor DNA yield are some other limitations on the use of YACs. However, due to the general availability of YAC libraries (see section 5.3.1.1) and their insert capacity YACs are still being used to create maps that provide long range continuous coverage of numerous disease regions. Therefore YACs were also the initial choice of vector for the construction of the contig spanning the RP11 disease interval.

5.2.2 STS content mapping and chromosome walking for contig establishment

Construction of a contig first entails the identification of necessary clones followed by determination of degree of overlap between neighbouring clones, which harbour random end points (see section 1.4.2.2). Fingerprinting is a method often used to determine overlaps between YAC clones. It involves the restriction digestion of YAC clones followed by southern blot analysis using a repetitive sequence (such as human Alu repeat or L1 repeat), which allows related YAC clones to be identified by the presence of common bands (Bellane-Chatelote *et al.*, 1992). The fingerprint patterns of inter-Alu PCR products are also frequently used to determine the degree of overlap between neighbouring YACs (Nelson *et al.*, 1989). However the most favoured and reliable means of overlap detection is through PCR based, STS content mapping, which was also the approach utilised in this study in the construction of the YAC contig spanning the RP11 interval. The idea behind STS content mapping is simple, if two or more clones contain the same STS then they must overlap and the overlap must include the STS.

Whilst a physical map can easily be established of a region with a high density of markers or STSs, identification of overlaps between clones in a region with an insufficient number of STSs can be achieved through chromosome walking. The concept of chromosome walking relies on the isolation of a DNA fragment at or near an end of a cloned insert for use as an STS (or probe) to screen the library and identify more clones. Screening the YAC libraries with such STSs identifies neighbours with a bare minimum overlap and lead to a longer and more efficient walking procedure. Furthermore, the failure of a walking step with an end-clone is a quick indication of a gap in the collection of clones being screened. A number of methods have been developed to isolate YAC ends, these include the techniques of vectorette PCR (Riley *et al.*, 1990), inverse PCR (Silverman *et al.*, 1989), single primer extension PCR (Screaton *et al.*, 1993) and vector-Alu PCR (Nelson *et al.*, 1989, 1991). The latter method (vector-Alu PCR) was used in this study and is described in section 5.3.1.3.











5.2.3 The genomic region of chromosome 19q13.4

The human chromosome 19, at 60 megabases (Mb) constitutes slightly more than 2% of the human genome (Mayall *et al.*, 1984), and due to a relatively high GC content is described as a comparatively gene rich chromosome (Larsen *et al.*, 1992). Overall it is genetically well characterised with over 450 genes, cDNAs, STSs and polymorphic markers (Ashworth *et al.*, 1995) mapped to it. The recently published gene map (Schuler *et al.*, 1996) placed nearly 100 cDNAs in the region between the markers D19S246 and D19S218 which roughly encompass the 19q13.4 cytogenetic band, to which the gene for RP11 is also localised. In comparison to the 19q13.3 band relatively few genes have been characterised in the 19q13.4 region and most of the expressed sequences mapped to the region still remain as anonymous cDNAs.

Several interesting diseases have been mapped to the chromosome 19q13.3-13.4 region. Cone rod dystrophy (*CORD2*), another retinal disease locus was mapped to chromosome 19q13.3 in our laboratory and originally it was thought that RP11 and the CRD locus may be allelic despite considerable phenotypic variation between the two diseases. Subsequently the two diseases were genetically separated and recently the gene for CRD was identified as *CRX* (Freund *et al.*, 1997), an OTX-like homeobox gene that shows photoreceptor-specific expression and regulates photoreceptor differentiation (Furukawa *et al.*, 1997). The well-known Myotonic dystrophy (DM) causative *DMPK* gene is also located in the 19q13.3. The 19q13.4 region is also exceptionally rich in zinc finger genes (ZNF) and two main regions of tandemly clustered ZNF genes exist on this cytogenetic band (Ashworth *et al.*, 1995). One is located proximal to the marker D19S180 (and the RP11 locus) and the other cluster is located distal to RP11 within the sub telomeric region of 19q13.4 (see figure 5.1). The human homolog of a mouse imprinted gene, Peg 3, also map to the latter zinc finger rich region of 19q13.4. Therefore the region encompassing *PEG3* on 19q13.4 may indicate a new potentially imprinted domain (Kim *et al.*, 1997).

LEGEND

Cosmid names are listed vertically. Black indicates a FISH ordered clone where distance between clones has been measured. Green indicates that FISH order data was used to place the clone. Other clones are labeled in red. Diseases are shown in blue. Genes are labeled in red. Other markers are labeled in black.

-  CONTIG SEQUENCED OR IN SEQUENCING QUEUE
-  Restriction mapped contig
-  Estimated contig size
-  BAC, PAC, or P1 clone
-  YAC with known and concordant size
-  YAC with unknown or discordant size
-  Sequence Tagged Site (STS)
-  STS and/or Hybridization results
-  Polymorphic marker
-  Clonal coverage

6/15/98

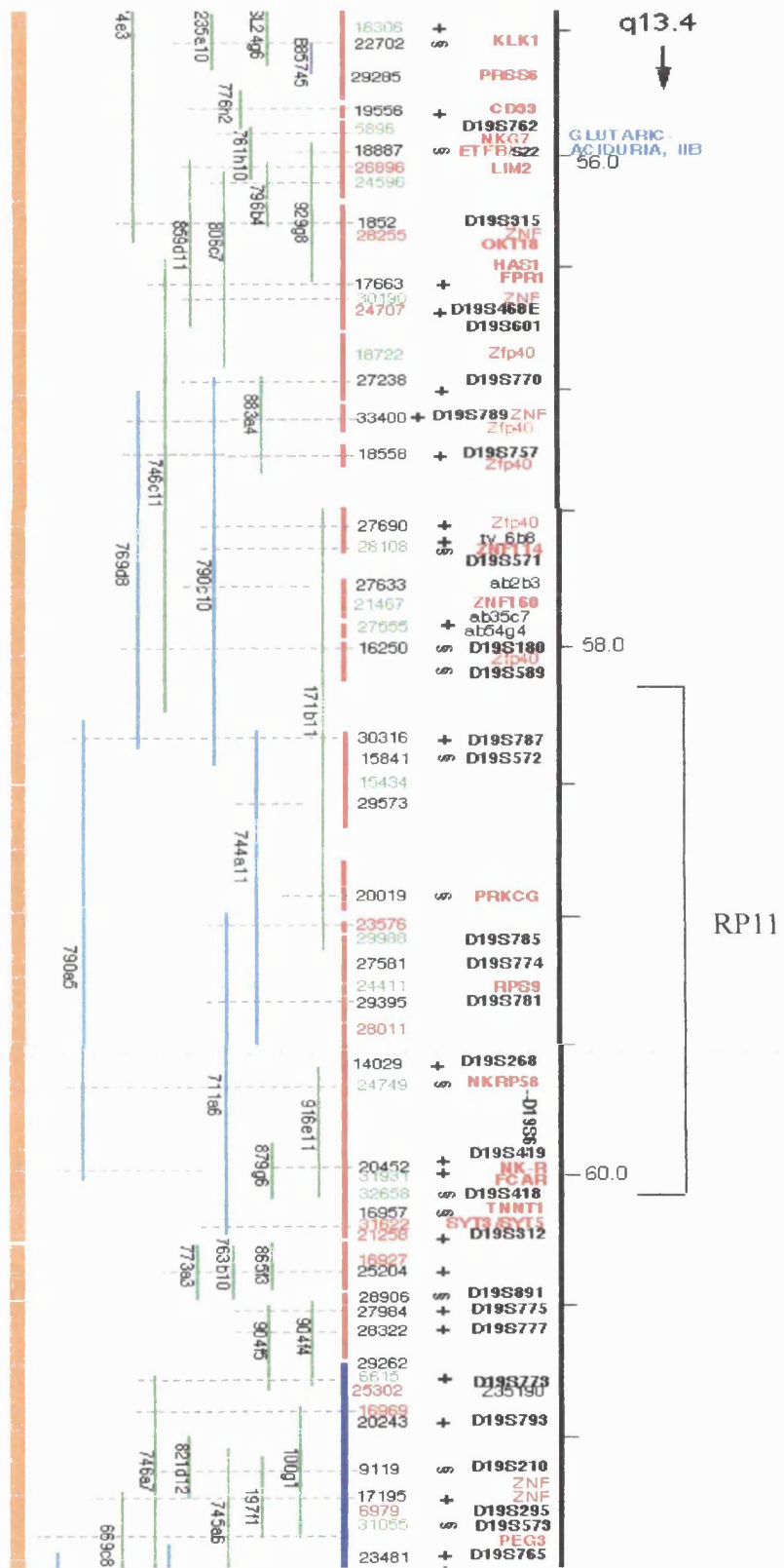


Figure 5.1

A portion of the metric physical map of human chromosome 19 constructed at the LLNL genome centre showing the RP11 genomic region and the existing YAC coverage. Figure legend is on opposing page.

Interestingly multiple genes encoding type I transmembrane molecules and belonging to the immunoglobulin superfamily have recently been localised distal to RP11 in 19q13.4 (Dupont *et al.*, 1997) (figure 5.1). These include a family of genes encoding killer cell inhibitory receptors (KIR) expressed on natural killer (NK) cells (also referred to as natural killer cell class 1 receptor family; the genes are called *NK-R* and *NKRP58*), immunoglobulin-like transcripts expressed on myeloid cells, the gp49 family of receptors expressed on mast cells and NK cells and the gene encoding the receptor for Fc fragment of IgA (*FCAR*). Additionally the genes for, the gamma isozyme of protein kinase C (*PRKCG*), the ribosomal protein S9 (*RPS9*), the slow skeletal isoform of troponin T (*TNNT1*), two isoforms of synaptotagmin (*SYT3* and *SYT5*) and the gene for interleukin-11 (*IL11*) have also been localised to this region.

Another interesting factor concerning 19q13.4 is that it belongs to one of the largest syntenically homologous blocks found between human and mouse; studies have established close evolutionary relationship between the proximal portion of mouse chromosome 7 and the long arm of human chromosome 19 (reviewed by Brilliant *et al.*, 1994). Despite the remarkable similarity between the two regions in both content and organisation of related genes, the studies have also indicated the presence of significant rearrangements in the relative orders of homologous mouse and human genes within this homologous block. The overall similarity is disrupted by four major rearrangements, which include transpositions of three blocks of genes and a relative inversion of sequences located near the 19q telomere (Stubbs *et al.*, 1996) (see figure 5.2). Three of the four rearrangements of mouse versus human 19q sequences involve segments that are located directly adjacent to each other in 19q13.3-q13.4. Interestingly the proximal border of the inverted human interval and the distal border of the adjacent segment containing genes that have been transposed, are both located near the RP11 region of 19q13.4 between *ZNF160* and *PRKCG* genes. The association of all these genomic segments (which were also coincidentally closer to each other) with unstable DNA sequence was given as a reason for these rearrangements (Stubbs *et al.*, 1996). This is also an important fact that needs to be considered during the physical mapping of this region.

MOUSE CHROMOSOME 7

HUMAN CHROMOSOME 19q

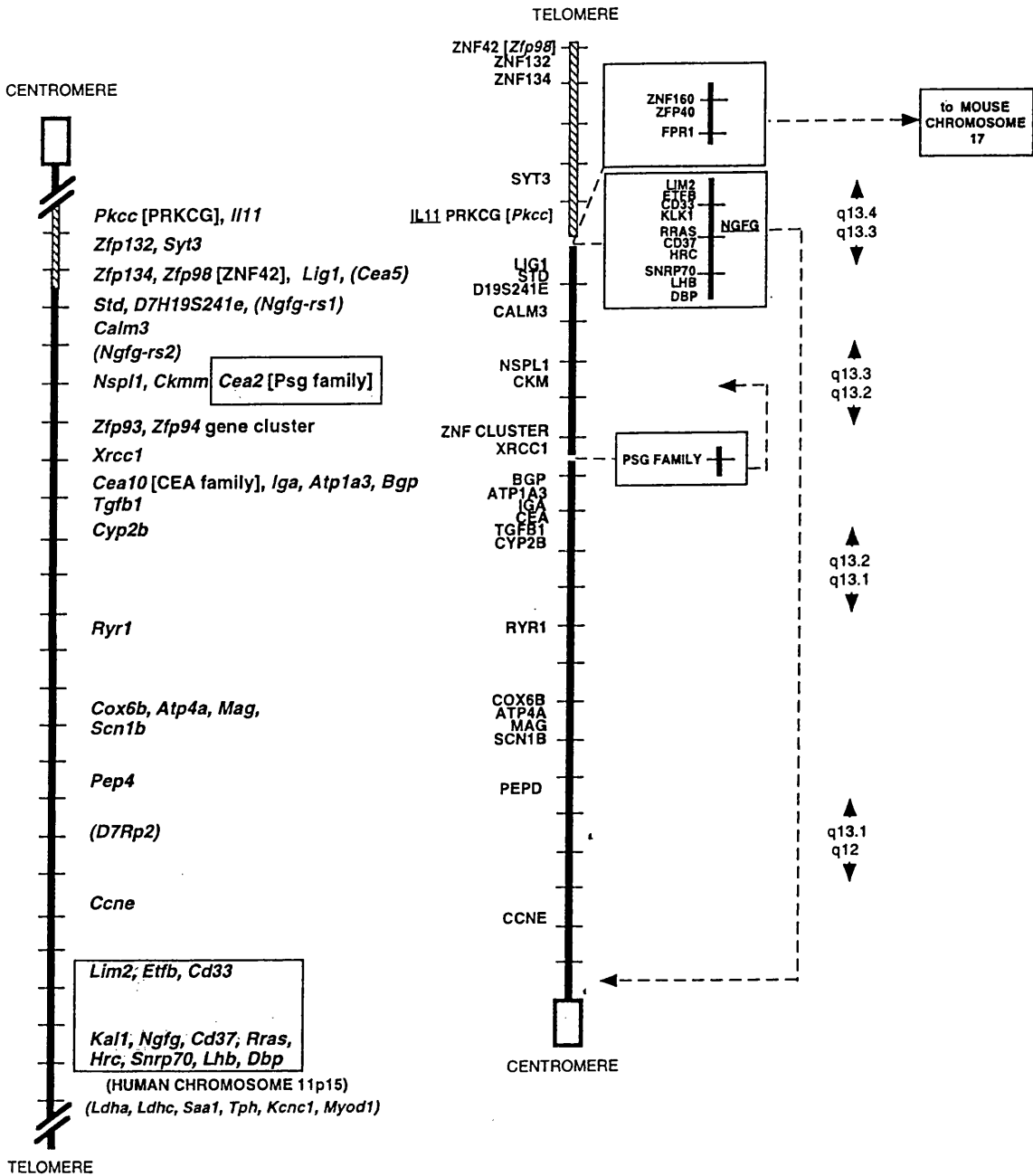


Figure 5.2 [figure and legend adapted from Stubbs *et al.* (1996)].

Comparative map of human chromosome 19q and related regions of the mouse genome. The map on the left represents the genetic map of proximal mouse chromosome 7. Markers listed together in a single line and separated by commas are indistinguishable from each other. Locus names presented in parentheses represent mouse genes or other markers without known human equivalents. Symbols in brackets indicate the name assigned to the homolog of the preceding mouse or human gene and are presented only where the two names differ significantly. The map on the right is extracted from the physical map of human chromosome 19 taken from Ashworth *et al.* (1995) (also see figure 5.1). Scales used are 1-cM increments in the mouse map and 1-Mb increments in the human map. The two maps are positioned to allow maximal alignment of related mouse and human markers. The hatched region on the scale of each map indicates a segment of the map that is inverted between the two species. Boxed regions indicate relative transpositions mentioned in the text, with a dashed arrow indicating the extent of the transposition. Gene symbols that are underlined represent sequences that have been assigned to a local region of human chromosome 19 by cytogenetic methods, but that have not yet been specifically ordered in the human physical map.

5.2.4 Physical maps of chromosome 19

Physical maps of chromosome 19 have been constructed by the Whitehead institute/MIT Centre for Genome Research (Hudson *et al.*, 1995) and the Lawrence Livermore National Laboratories (LLNL), USA (Ashworth *et al.*, 1995). The Whitehead map of chromosome 19 is composed entirely of YAC clones derived from the CEPH mega-YAC library. These mega YACs were isolated by PCR based screening with 548 STSs, which include Généthon microsatellite markers from the genetic map of chromosome 19 (Dib *et al.*, 1996), and STSs mapped and ordered by radiation hybrid (RH) mapping. The Whitehead map (as yet) does not provide complete coverage of chromosome 19 and is composed of several isolated, non-overlapping contigs distributed along the length of the chromosome.

The chromosome 19 physical map constructed at LLNL is a high resolution, cosmid- based map spanning the entire chromosome (Ashworth *et al.*, 1995). The foundation of this map consists of a set of 802 cosmid contigs assembled by automated restriction fragment analysis. More than 400 of the cosmid contigs have been mapped by fluorescence *in situ* hybridisation (FISH) to metaphase bands. From these, 236 cosmids have been further ordered along the chromosome by high resolution FISH mapping on sperm pronuclear targets (Brandriff *et al.*, 1991), a technique in which the distance between these reference cosmids is also determined. This ordered FISH map provided a metric map, upon which to anchor cosmid contigs. The average distance between these cosmid reference points is 230 kb, with a range of 50-700 kb. The metric map also contains 'islands' of large insert clones such as YACs, BACs, PACs and P1 clones, that span gaps between ordered cosmid contigs (The physical map shown in figure 5.1 is a diagrammatic representation of the LLNL metric map of the region surrounding the RP11 locus). The most interesting feature of this map, and also of immense use to gene hunters working on specific regions of chromosome 19, is the mapping of genetic markers and genes to the ordered cosmid contigs that is still on going work at LLNL. The physical map spanning the RP11 interval that is the focus of this study had its basis in these two physical maps of chromosome 19.

5.2.5 Aims of study

Earlier in this study the genetic refinement of the RP11 locus and the analysis and exclusion of a positional candidate gene was described. Following the strategy outlined for identifying a disease gene in positional cloning, we next embarked upon the physical characterisation of the RP11 region and the construction of a YAC contig across the RP11 interval with the ultimate aim of identifying / isolating retinally expressed transcripts localised to the RP11 region.

Although large-scale physical maps of chromosome 19 exist, they all have relatively poor YAC coverage in the RP11 region of chromosome 19. For example in the LLNL physical map, the entire RP11 region is covered in 6 CEPH mega-YAC clones (see fig 5.1), all of which are chimaeric according to the CEPH genome database. Even though these mega-YACs provide long range coverage, being chimaeric they are not at all suitable for the isolation of region specific genes, which was the ultimate purpose of our study. Therefore while part of the effort was to obtain all YACs that map to the RP11 by scrutiny of available genome databases including the chimaeric mega-YACs mapped by other groups, the main focus was on the construction of a much deeper physical map of the RP11 region with smaller non-chimaeric YACs isolated from several YAC libraries. If satisfactory coverage was not achieved with these YAC clones, PAC clones were also to be isolated from a PAC library to consolidate the contig across the RP11 interval.

The physical map was constructed in collaboration with LLNL from whom we obtained cosmids localised to the RP11 region (see section 5.2.4). The next objective in this study was to use these cosmids, especially those mapped to the distal region of RP11, to isolate novel microsatellite markers (and STSs) and increase the density of STSs in this region. There is a scarcity of STS/polymorphic markers in the distal region of RP11, particularly within the 3cM genetic interval between D19S927 and D19S418 where the distal recombination was observed in the family ADRP2 (see section 3.3.2). These markers once isolated and characterised were to serve a dual purpose. If polymorphic they were to be analysed in ADRP2 to explore further possibilities of locus refinement at the distal end of the RP11 region. Secondly these

markers were to be used as STSs for the isolation of more clones to increase the depth of the YAC contig.

5.3 Isolation and characterisation of YACs in the RP11 region of chromosome 19q13.4

The YAC clones that map to the RP11 interval were obtained in two ways; they were obtained either from the available physical maps of the region or by PCR screening YAC libraries with primers for the STSs located within the RP11 interval. The overlaps between YAC clones were detected through STS content mapping; where gaps were present attempts were made to bridge them through chromosome walking, for which ends of non-chimaeric YACs were isolated and used to isolate further YACs. The specific methods employed to achieve these targets are described in the following section.

5.3.1 Materials and methods

5.3.1.1 YAC libraries

YACs used in this study were all obtained through screening of YAC libraries provided by the UK HGMP Resource Centre, which has in its possession the three human YAC libraries (table 5.1).

5.3.1.2 Isolation of YACs by PCR screening of YAC libraries

All of the individual YAC clones in the ICI and ICRF libraries are arranged in 96 well plates (each with 8 rows and 12 columns), and the libraries are screened by PCR in a two step pooling procedure. The ICI library is divided in to 40 primary pools with each pool containing DNA of 9 plates of clones (864 clones each). The ICRF YAC library has 27 primary pools of YAC clones and each primary pool has DNA of 8 plates of YAC clones (768 clones each). Once the primary pool has been identified in the primary screening of 40 (ICI) or 27 (ICRF) pools, a secondary screening is executed to identify the plate from which the clone is derived and the plate coordinates (positive row and column). Therefore, 28 (ICRF) 29 (ICI) smaller pools, made up of 8 (ICRF) or 9 (ICI) plate pools, 8 pooled rows (the corresponding rows from all 8/9 plates) and 12 pooled

YAC library	Derivative Human Cell line	Average Insert size (kb)	No of clones	Reference
ICI	Lymphoblastoid 48XXXX-cell line	350	34,500	Anand <i>et al.</i> , 1990
ICRF¹ 164 plates (labelled 4X1-4X164)	Lymphoblastoid 48XXXX-cell line	600		Larin <i>et al.</i> , 1991
ICRF² 24 plates (labelled 4Y1-4Y24)	Male lymphoblastoid 49 XYYYYY-cell line	400-500		Larin <i>et al.</i> , 1991
ICRF³ 26 plates (labelled HD1-HD26)	Lymphoblastoid, Huntingdon's disease 46XX-cell line	600	Total of 20,500	Larin <i>et al.</i> , 1991
CEPH		920	35,600	Chumakov <i>et al.</i> , 1992

Note: ^{1,2,3} ICRF library is composed of these three sub libraries

Table 5.1.

The human YACs libraries available from the HGMP resource centre.

columns (the corresponding columns from all 8/9 plates) are tested for a given positive primary pool. The PCR should give at least 3 positives: the positive plate, the positive row and the positive column i.e. plate number and plate co-ordinates.

All YAC library pools were received as agarose plugs in 5 mM EDTA and the plugs were prepared for PCR as described in section 2.1.4. The YAC libraries were screened with a variety of microsatellite markers, STS and ESTs from the RP11 interval following standard PCR protocol described in section 2.4. The sequence of primers for the microsatellite markers were obtained from genome databases and manufactured commercially. Once a YAC positive for a given STS was identified through the screening procedure as described earlier the YAC was ordered from the HGMP Resource Centre.

5.3.1.3 Isolation of YAC insert end sequences for chromosome walking by Alu-Vector arm PCR

Alu PCR has been adapted for use in end recovery by including Alu primers (Ale 1 and Ale 3) and vector specific primers (RA and LA) (Nelson *et al.*, 1991; Ragoussis *et al.*, 1991). RA and LA represent pYAC4 vector right arm and left arm primer sequences, respectively. The end fragment is amplified in cases where an Alu repeat is sufficiently close to the end of the YAC. To detect the Alu-Vector amplification product(s) the following combination of primers are used in a PCR reaction that contain ~ 250 ng of YAC DNA solution (see section 2.1.3).

Ale 1 + RA (5'- ata tag gcg cca gca acc gca cct gtg gc - 3')

Ale 1 + LA (5'- cac ccg ttc teg gag cac tgt ccg acc gc - 3')

Ale 3 + RA

Ale 3 + LA

Ale 1 only (5'- gcc tcc caa agt gct ggg att aca -3')

Ale 3 only (5'- cca t/ctg cac tcc agc ctg gg -3')

All PCR reactions were carried out at an annealing temperature of 60°C and products were visualised by agarose gel electrophoresis. Unique products present within the Ale-Vector arm samples, which represent YAC insert termini, were directly sequenced using internal pYAC4 vector arm primers (see section 2.10.3).

PYAC4 LA internal primer sequence: (5'- gtt ggt tta agg cgc aag -3')

PYAC4 RA internal primer sequence: (5'- gtc gaa cgc ccg atc tca a -3')

5.3.2 Results

5.3.2.1 Verification of the integrity of mega-YACs obtained from pre existing contigs.

The CEPH mega-YACs that map to the RP11 region were identified from pre-existing physical maps of chromosome 19 and the YAC clones were obtained from the HGMP Resource Centre. In total 11 mega-YACs were obtained and their total yeast DNA was extracted. These were then used to test for the presence of STSs the YAC inserts were reported to contain and also the 10 STSs localised to the RP11 region, which included the microsatellite markers D19S571, D19S589, D19S180, D19S921, D19S572, D19S924, D19S927, D19S418 D19S926 and exons of the gene *PRKCG*.

Figure 5.3 depicts the results of the STS content mapping of these mega-YACs. As can be seen few of the mega-YACs (i.e. 916-e-11, 711-a-6, 790-a-5 and 955-g-11) proved either to be rearranged or deleted. The extent of deletions in YACs can easily be determined if the region being tested is saturated with STSs of known order. It was therefore difficult to determine the integrity of YACs such as 790-a-5 and 955-g-11 that contained inserts extending towards the distal half of the RP11 interval, which as mentioned earlier is scarce of markers. Deletions in these mega-YACs were first revealed with the testing for *PRKCG* exons, which showed that these YACs only contained a portion of this gene (i.e. exons 1-10). As expected the deletions in YACs 790-a-5 and 955-g-11 became more obvious after they were assayed for novel STSs localised to the distal portion of the RP11 interval (see figure 5.22). YACs 916-e-11 and 711-a-6, which according to the physical map of LLNL were integral YACs of the RP11 interval, both proved to be either deleted or rearranged as certain STSs were absent in these YACs in an intermittent manner. The genetic order of these 'missing' STSs, which were mostly microsatellite markers, had been established through genetic recombination and was therefore reliable; hence the gaps in these YACs could not have been due to erroneous order of markers. Due to these rearrangements the two YACs were excluded from the contig. This discrepancy between the LLNL data and

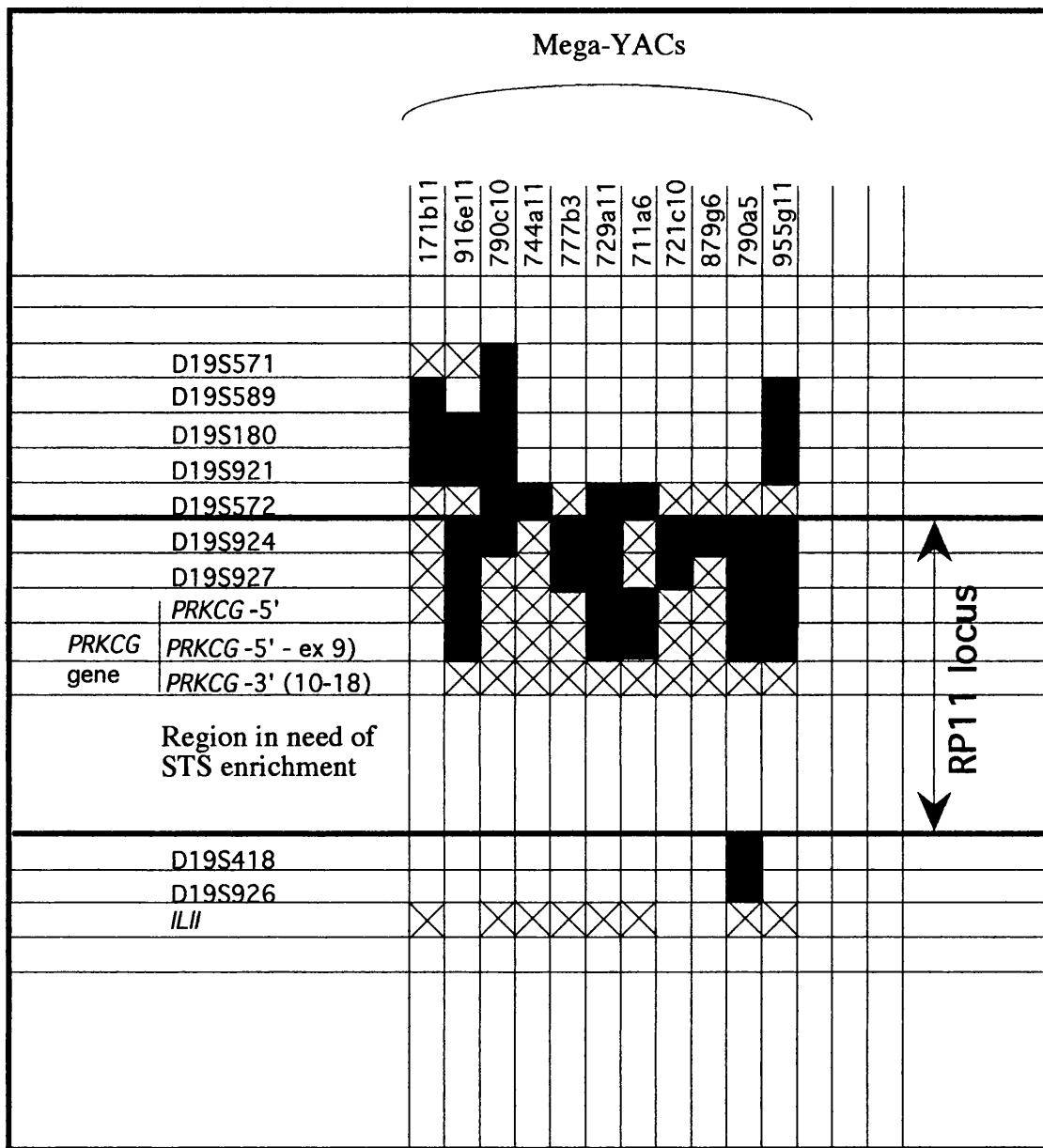


Figure 5.3

The STS content map of the CEPH mega YACs. Solid boxes denote the presence of the STS on PCR analysis and the crosses the absence of the STS. STSs from genes i.e. *PRKCG* and Interleukin 11 (*IL11*) are shown in italic.

our data could partly be due to the different screening methods applied in the two laboratories. In LLNL YACs were isolated through the hybridisation of radiolabelled cosmids to gridded YAC clones, and the YACs anchored to the RP11 interval were isolated through < 4 separate cosmid hybridisations to each YAC (fig 5.1). Unlike the more systematic approach of STS content mapping utilised in this study this hybridisation technique is more likely to overlook deletions and rearrangements. Most of the mega YACs, especially those that contained a significant number of STSs from the RP11 region, were also sized by pulsed field gel electrophoresis to see if the size of the YAC isolated (in the laboratory) was different from its stated size in the databases. If different it would be a further indication of YAC rearrangements (i.e. deletions) while undergoing mitotic growth. Most of the YAC clones were smaller than the expected size. The most likely possibility is that these YACs originally contained unstable sequences, which were lost during growth in yeast producing a stable deletion derivative. YAC 955-g-11 can be cited as a classic example for the aforementioned events; it was considerably smaller than its expected size and also contained two YAC species in one cell instead of just one. Presence of two YACs in one cell can lead to rearrangements during growth and may have resulted in YAC 955-g-11 being rearranged with considerable sequence loss, alternatively the two small YAC species could be stable deletion derivatives of a larger original YAC. Majority of the mega-YAC clones localised to the RP11 region revealed such deficiencies, which may explain the discrepancy observed between our YAC data and that of LLNL.

The STS content of the mega YACs 790-c-10, 171-b-11, 729-a-11, 790-a-5 and 955-g-11 suggest that they provide reasonable if not complete coverage of the RP11 region. However, due to their many deficiencies the mega YACs were mainly used to provide a framework map of the region and to test for the presence of Expressed Sequence Tags (ESTs) and other STSs localised to the RP11 region. Table 5.2 lists all the features of the mega YACs including the physical map from which they were derived, their size, both expected and observed (on PFGE analysis in the laboratory), and known FISH data and

YAC ID Located map	YAC size kb		FISH data	STS Hits in Chromosomes ²	No of bands on PFGE ¹
	Expected	Observed			
955-g-11 (MIT)	1620	500, 550	–	19, 2, 5 , (3, 8, 10, 12, 13, 20, X)	2
790-a-5 (LLNL)	1770	~ 1000	–	19, 18 , (11, 17)	1
711-a-6 (LLNL)	1290	~ 550	–	19 , (3, 9)	1
171-b-11 (LLNL)	No size data	450	–	19	1
744-a-11 (LLNL)	950	825-945	–	19, 4 , (12, X)	1
777-b-3 (MIT)	220, 400	200, 220	–	19 , (5, 7)	2
721-c-10 (MIT)	350	–	–	19, 17, 4 , (6, 10, 13, 21)	–
879-g-6 (LLNL)	870	650	–	19, 20 , (1, 17, X)	1
916-e-11 (LLNL)	330	–	–	(2, 7, 10, 12, 18, 19)	–
790-c-10 (LLNL)	1650	1650	–	19 , (18, 16)	1
729-a-11 (MIT)	760	500	–	19 , (1, 17)	1

¹ Presence of more than 1 YAC indicated by two (or more) bands on PFGE gel.

² Unambiguous hits are highlighted in bold, (ambiguous hits are not highlighted)

Table 5.2.

Features of CEPH mega YACs obtained from chromosome 19 maps.

STS content. This information was obtained from electronic genome databases maintained by Whitehead Institute/MIT Centre for Genome Research (<http://www-genome.wi.mit.edu/cgi-bin/contig/STS-by-chrom>) and CEPH- Généthon (<http://www.cephb.fr/bio/ceph-genethon-map.html>)

5.3.2.2 Isolation of novel YACs by PCR screening of ICI and ICRF YAC libraries

Upon further analysis, the mega YACs mapped to the RP11 region proved to be either deleted or rearranged and of little use for isolating region specific transcripts. Therefore in order to isolate relatively small and non-chimaeric YACs, the ICI and ICRF libraries (see table 5.1) were also screened with all of the aforesaid STSs mapping to the RP11 region by PCR. For each screening experiment, a negative control (water only) and positive (human genomic DNA) control was included. In total 18 YACs were isolated with the 10 STSs as some of these STSs yielded several YACs on library screening. The result of PCR screening of the ICIYAC library for the isolation of YAC 6GB8 associated with marker D19S926 is presented in figure 5.4. To determine overlaps between YACs and their STS content, all YAC clones were tested for the presence of all the other STSs in addition to the STS responsible for their isolation. The overall results are depicted in figure 5.5.

To isolate YACs that contained the *PRKCG* gene both the ICI and ICRF libraries were screened with exon 7 (to represent the 5'end) and exon 18 (to represent the 3'end) of this gene. YACs (HD12D9 and 4X5F1) that contained exon 7 were also tested for all other *PRKCG* exons, upon which it showed that these YACs only contained the 5' portion of this gene (i.e. exons 1-10). Only YAC 4X42H10 was positive with exon 18, further analysis of this YAC showed that it contained exons 1-13 and exon 18 but not exons 13-17, indicating that this YAC was rearranged or deleted. Therefore none of the isolated YACs appeared to contain the complete *PRKCG* gene. Similar results were obtained with the mega-YACs, which implied that this region was prone to deletions or rearrangements. The deletions in these smaller YACs first brought to light with the testing of *PRKCG* exons in YACs became more obvious with the testing of newly isolated STSs in the RP11 interval (see section 5.5.1).

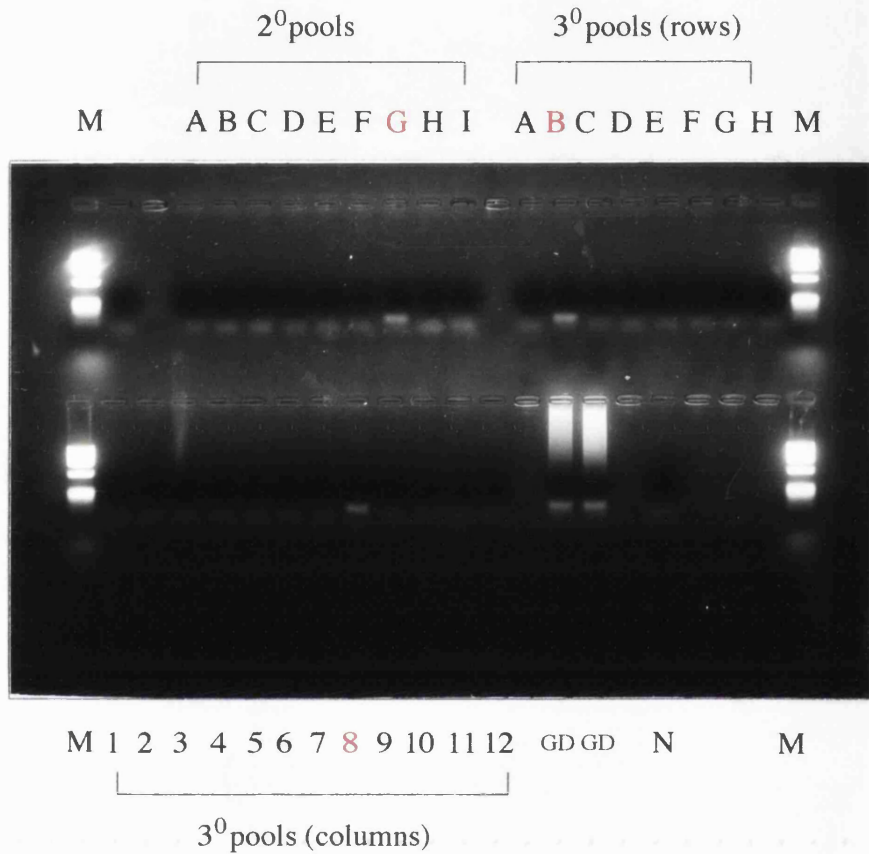


Figure 5.4

Ethidium bromide stained 2% agarose gel depicting PCR products obtained by screening of ICI YAC library with marker D19S926. Positive pools are indicated in red; since ICI primary pool 6 was found positive the identity of the isolated YAC clone is 6GB8.

- M - Φ x174 DNA size marker.
- N - negative control (water only).
- GD - genomic DNA

As an example, the PCR result obtained when exon 8 was tested on a selection of YACs from the RP11 region is presented in figure 5.6. Figure 5.7 shows the distribution of *PRKCG* exons within the various genomic clones (YACs, PACs and cosmids) that contain portions of the *PRKCG* gene. As can be seen satisfactory clonal coverage across the *PRKCG* gene is only provided by the cosmid clones and the PAC clone, which was isolated at a later stage (see section 5.5.2).

All the YACs were also sized by pulsed field gel electrophoresis (see table 5.3). An example of an autoradiograph of a PFGE gel is shown in fig. 5.8. A few of the YACs showed multiple bands on PFGE gel analysis, i.e. YAC HD12D9. There can be trivial reasons for the observation of multiple species, such as mixture of several clones, or independent YACs within the same cell. However, in a pure clone, the presence of multiple species is an indication of a structurally unstable YAC. To come to a definite conclusion about the instability of a YAC clone, several independent colonies of that YAC need to be analysed by PFGE to see if the YAC continues to show multiple species. At this stage it was not necessary to conduct such a test on these YACs, as they were not being used for any purpose other than for testing their STS content. However it is important to investigate the structural stability of the YAC HD12D9, which contains all of the critical STSs of the RP11 interval and was therefore one of the most important YACs of the contig.

The contig assembled with ICI and ICRF YACs also had two regions, where there was no overlap between the YACs. These two gaps were between the markers D19S921 and D19S572, and between the gene *PRKCG* and the marker D19S418. In order to bridge these gaps, it was necessary to isolate end sequences from YAC clones anchored on either side of the gaps and produce STSs to conduct a bi-directional chromosome walk. However prior to embarking on a chromosome walk, chosen YACs were investigated for chimaerism by fluorescent in situ hybridisation.

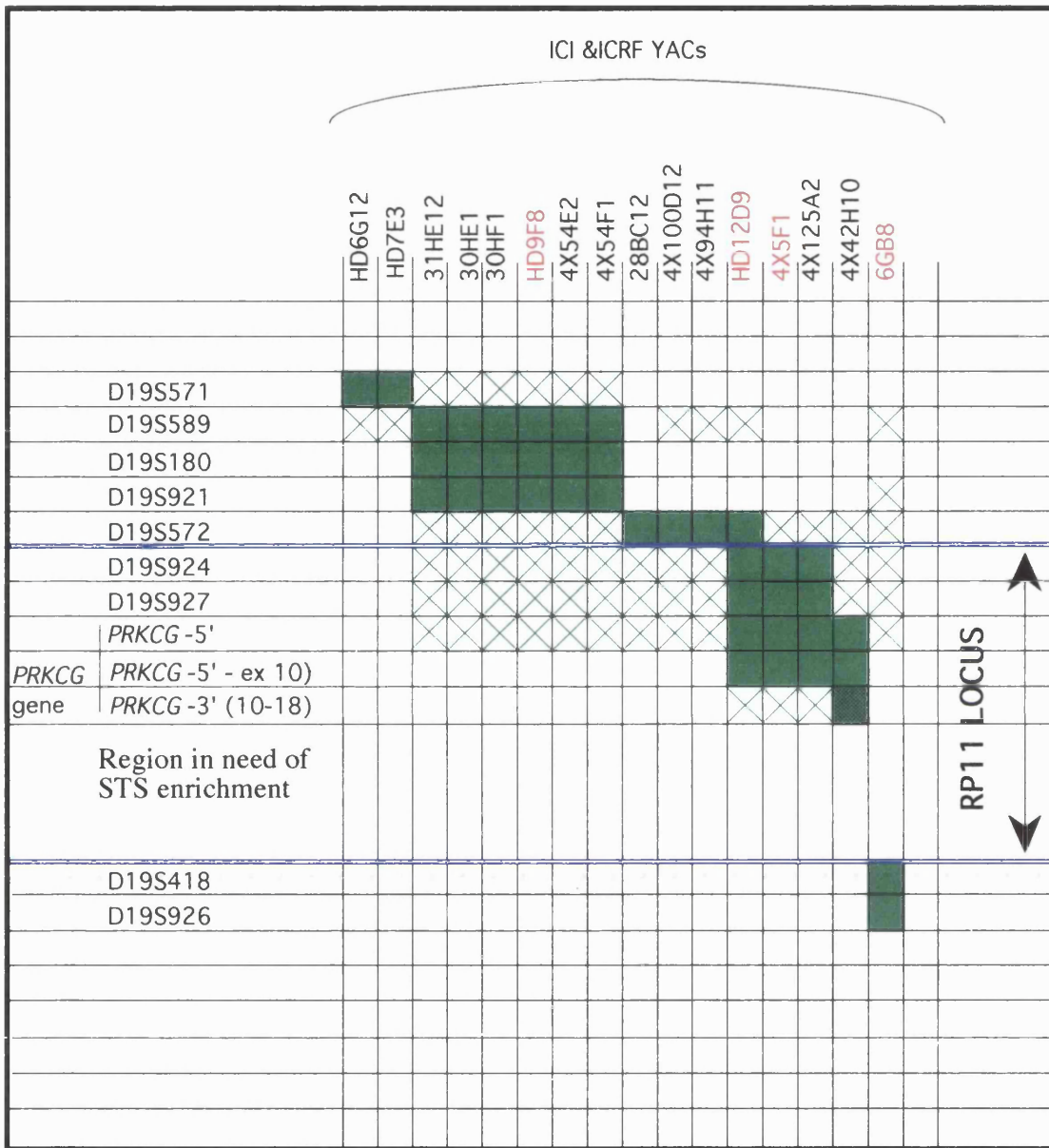


Figure 5.5

ST content mapping of the ICI and ICRF YACs. YACs written in red are critical YACs of the contig as they proved to be non-chimaeric on FISH analysis (section 5.3.2.3). YAC 4X42H10 is missing certain 3' exons of *PRKCG* which is indicated by a dotted filled box. There are two gaps in the YAC contig; one between D19S921 and D19S572 and the other between *PRKCG* and D19S418.

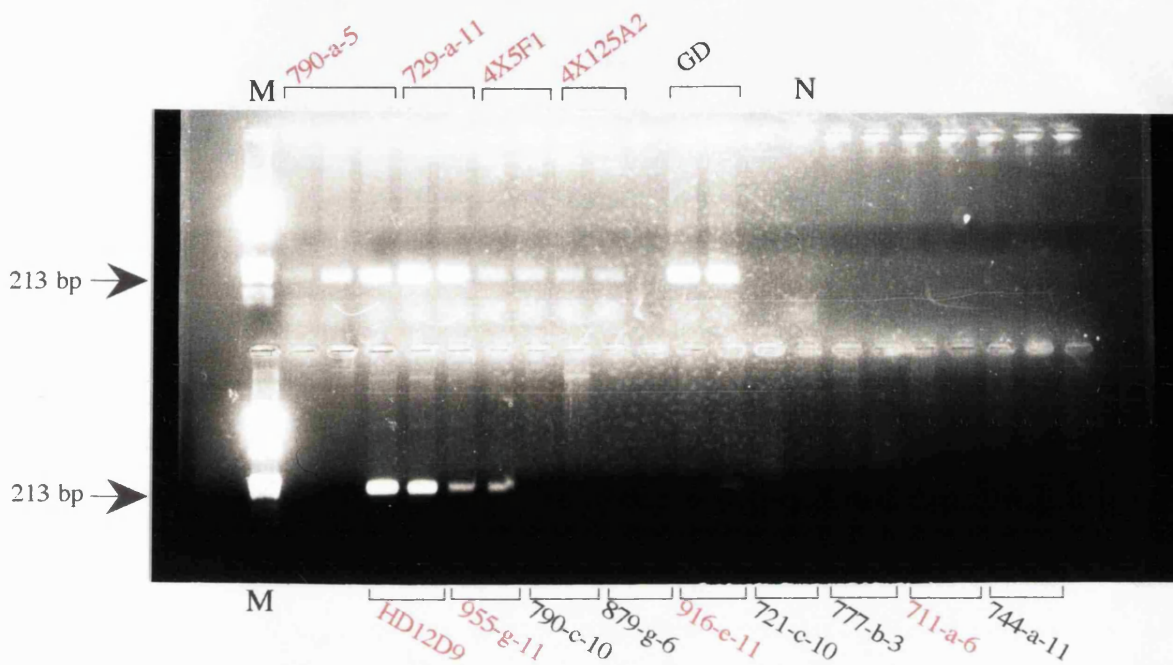


Figure 5.6

Ethidium bromide stained 2% agarose gel depicting the STS mapping results obtained with *PRKCG* exon 8 in a selection of YACs from the RP11 YAC panel. YACs showing a positive result are indicated in red. M denotes ϕ x174 DNA size marker, GD- the positive control (human genomic DNA) and N the negative control (water only).

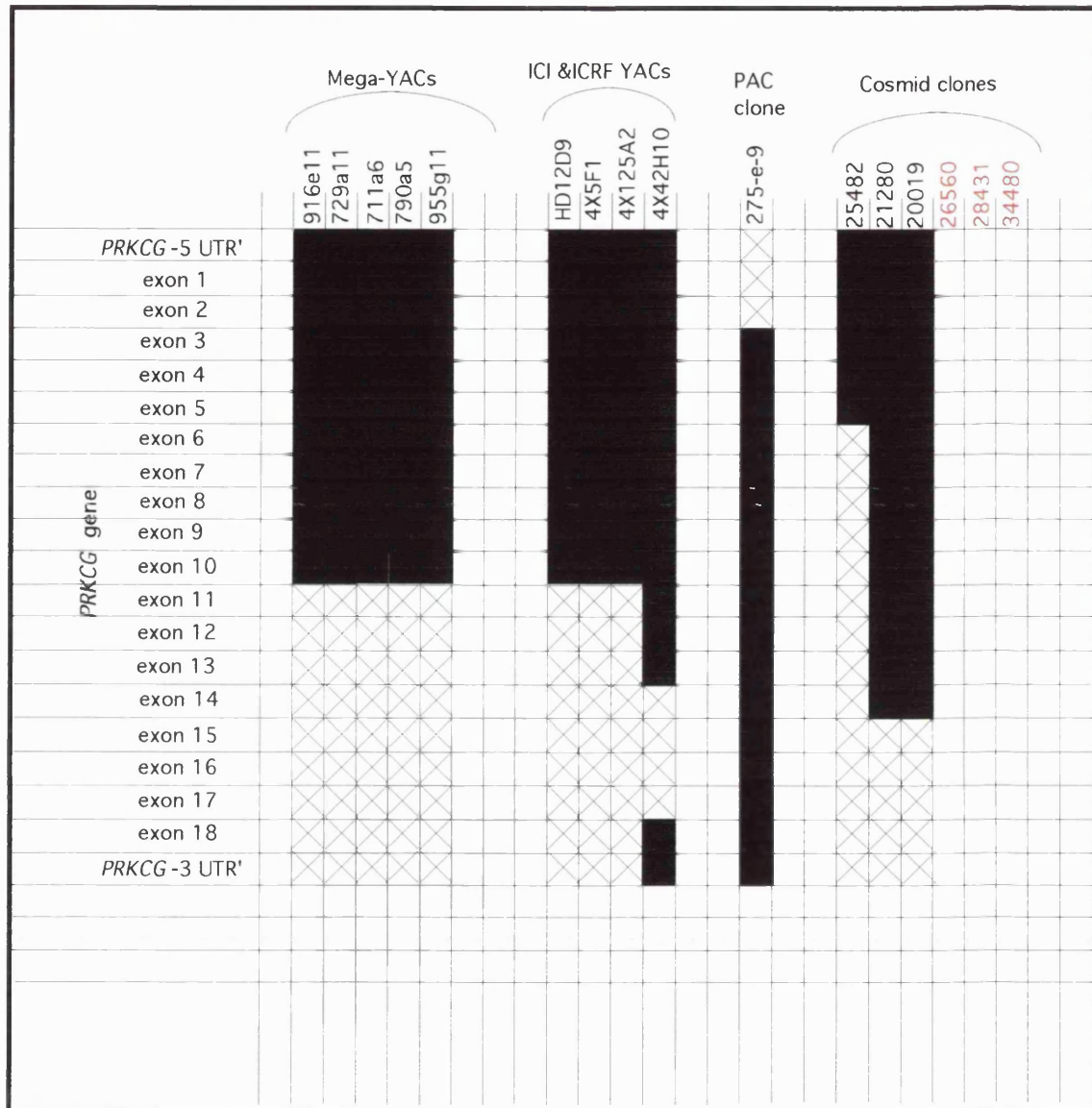


Figure 5.7

The distribution of *PRKCG* exons in YACs, PACs and cosmid clones localised to the RP11 region. Solid boxes indicate the presence of an exon whereas crosses denote a negative result on PCR. As can be seen the mega-YAC clones only contain the 5' portion of the gene. YAC clone 4X42H10 on further analysis proved to be deleted for certain exons of *PRKCG*. According to LLNL mapping data cosmid clones in red should contain the 3' portion of the gene; but they have yet to be tested.

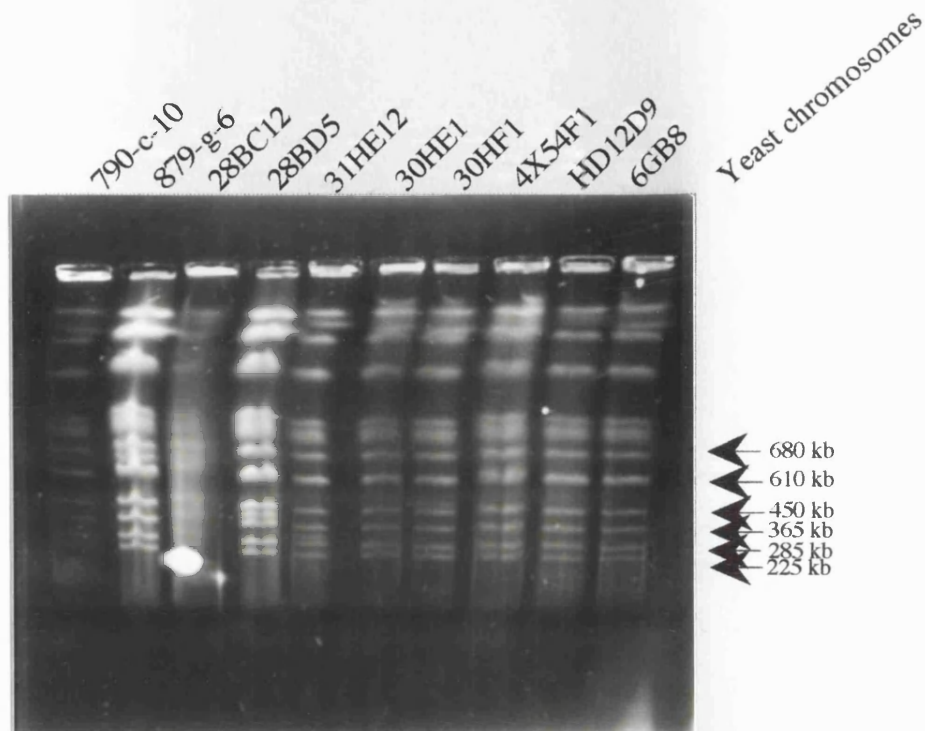


Figure 5.8a

Ethidium bromide stained 1% agarose gel depicting YACs subjected to PFGE. YAC 879-g-6 is visible between chromosomal bands 680 kb and 610 kb while YAC 28BD5 is visible between chromosomal band 450 kb and 365 kb. YAC 31HE12 is visible between 285 kb and 365 kb. The yeast colony harbouring YAC HD12D9 also retains an additional YAC, both YACs being less than 225 kb in size.

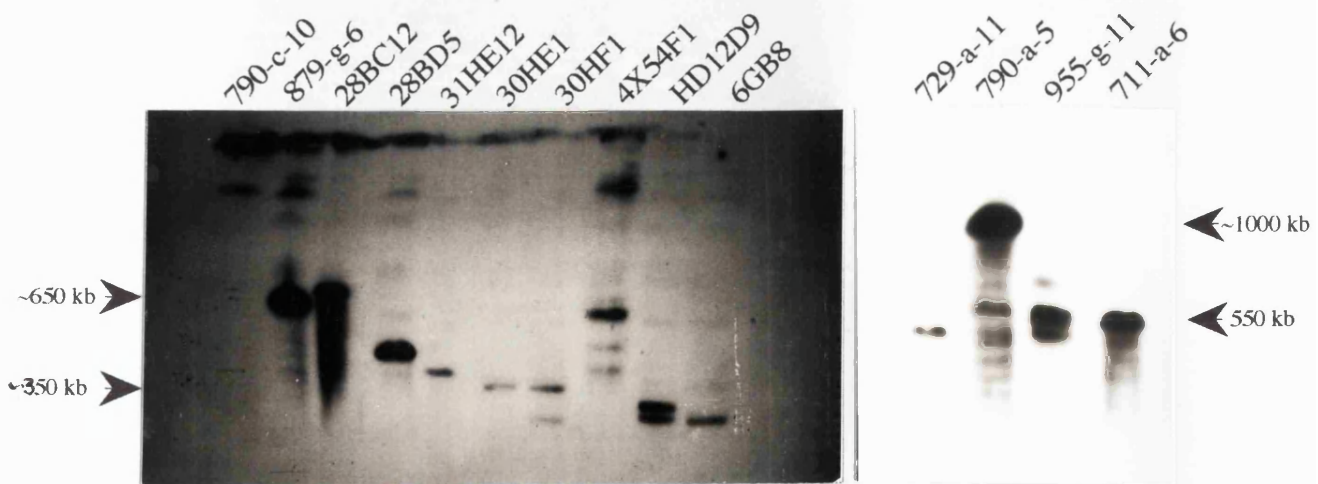


Figure 5.8b

Autoradiographs of the PFGE gel presented in figure 5.8a and of a PFGE gel that fractionated the mega YACs 729-a-11, 790-a-5, 955-g-11 and 711-a-6, following southern blotting and hybridisation with human Cot1 DNA. While the yeast chromosomal background appears clearly for YAC 790-a-5, the yeast colony harbouring YAC 955-g-11 seems to retain two YACs that are 500 kb and 550 kb in size.

5.3.2.3 Investigation of chimaerism in critical YACs spanning the RP11 interval by FISH

Chimaeric YACs contain fragments that are derived from different regions of the genome and a proportion of clones in most YAC libraries is chimaeric. One method used to detect chimaerism is fluorescent *in situ* hybridisation (FISH) of YAC clones to metaphase chromosome spreads. The chimera is detected by the observation of a positive signal at two distinct chromosomal locations. The YACs that mapped within the critical region were FISH mapped (by courtesy of Ms M. Fox at University College London (UCL)), to investigate possible chimaerism prior to end sequence isolation (see fig 5.9). In total 10 YACs were FISH mapped but only 4 YACs proved to contain sequences exclusively from chromosome 19q13.4. These were YACs HD9F8, HD12D9, 4X5F1 and 6GB8. Four of the YACs did not appear to contain any sequences from 19q13.4, even though they were positive for STSs from the RP11 region. This observation can be explained if within the concatemeric/chimaeric clone, the DNA fragment from 19q13.4 was small compared to the other co-ligated fragments, from other regions of the genome. This essentially small probe from 19q13.4 would then give rise to a signal too weak to be detected. These YACs were excluded from subsequent studies conducted on the RP11 YAC contig. Two other YACs, (4X94H11 and 4X100D12) also hybridised to chromosome 19q13.4, but signals were also detected from other chromosomes implicating them as chimaeric clones that should be used with caution.

5.3.2.4 Closure of gaps by chromosome walking

In total 8 YACs were used to isolate end sequences, which included the 4 non-chimaeric clones, the 2 chimaeric YAC clones (4X94H11 and 4X100D12) and 2 other YAC clones not tested by FISH (28BC12 and 28BD5). All of these YACs, chosen for end sequence isolation, were located on either side of the two existing gaps in the contig. The YAC HD12D9 was used even though according to the PFGE result, there were essentially 2 YAC clones (hence 4 YAC termini), present within one cell. The FISH result showed that YAC DNA of HD12D9 map exclusively to 19q13.4.



(a) ICRF YAC HD12D9 (non-chimaeric)



(b) ICRF YAC 4X94H11 (chimaeric)

Figure 5.9

Two YAC Fluorescent *in situ* hybridisation (FISH) results to metaphase chromosomes (from normal 46, XY male). (a) depicts the FISH result of non-chimaeric YAC HD12D9 and (b) depicts the FISH result of chimaeric YAC clone 4X94H11 showing signals on chromosomes Xp and 19q13.4.

YAC ID and Library	Size (kb)	FISH data	No of bands on PFGE	Usage
28BC12 – ICI	610-680	–	1	Included
28BD5 – ICI	365-450	–	1	Included
30HE1 – ICI	<285	–	1	Included
30HF1 – ICI	<285, <200	–	2	Included
31HE12 – ICI	285 –365	–	1	Included
6GB8 – ICI	<225	19q13.4	1	Included
4X54F1 – ICRF	565-610	–	Multiple bands	Discarded
4X54E2 – ICRF	–	1q41, 17p12	–	Discarded
HD9F8 – ICRF	265-365	19q13.4	1	Included
4X94H11 – ICRF	<225	19q13.4, Xp11.2	1	Used with care
4X100D12 – ICRF	785-825	19q13.4, 4p, 4q, 13q, 17q	1	Used with care
HD12D9 – ICRF	<225, <225	19q13.4	2	Used with care
4X108G12 – ICRF	360	Mainly repeats	1	Discarded
4X5F1 – ICRF	<225	19q13.4	1	Included
4X125A2 – ICRF	500, 650	Xq21-23, 13q22-31	2	Discarded
4X42H10 – ICRF	600	2p13,17q11,17p11,5q14-15, 13q	1	Discarded

Table 5.3.

Characteristics of the ICI and ICRF YACs. YACs shown in bold face are critical and valuable YACs for being non-chimaeric. Instabilities are suggested by the presence of multiple species in one clone. Sizes are indicated where known. As the smallest yeast chromosome was 225 kb in size, YACs smaller in size could not be accurately sized and are denoted as being <225 kb.

Since it is highly unlikely that two separate YAC species, co-cloned in to one cell, should both map to the same chromosome location, the smaller of the two YACs is more likely to be the stable deletion derivative of the larger original YAC, HD12D9.

The specific method used to isolate end sequences was Alu-vector arm PCR (described in section 5.3.1.3). When the Alu-PCR reactions were analysed by gel electrophoresis, unique products were observed within the Alu-vector arm samples, for all the YAC clones used for end sequence isolation. Indicating the fortuitous location of Alu repeat sequences close enough to the vector junction in the correct orientation in all of these YAC clones for the production of unique products. Figure 5.10 depicts PCR results of YAC insert-vector junction products obtained for YACs 4X94H11, 6GB8 and 28BD5. The unique products that represent YAC insert termini were directly sequenced using internal vector arm primers. Unfortunately some of the unique PCR products of certain YAC clones, within which the Alu sequences were located too close to the vector-insert junctions, were too small to enable the design of useful STSs. This was the case for both left and right arm products of YACs 4X5F1 and 28BD5 and the left arm product of the YAC 4X100D12. Prior to designing STS primers, the sequence obtained from YAC termini were screened for identities with entries within the genome database using the BLAST/FASTA computer programs (Pearson and Lipman, 1988; Altschul *et al.*, 1990). While this screening did not result in any interesting 'hits' with known unique sequences, some of the end sequences were discovered as being particularly rich in Alu repeat sequences or other repetitive elements, and could not be used for generation of STSs. For these reasons outlined above, even though YAC end sequence was obtained for nearly all of the YACs, useful STSs could only be generated from a few of these YAC termini (see table 5.4 and figure 5.11 which presents sequence obtained from YACs HD12D9, 6GB8 and 28BC12).

The STSs thus generated were first tested on their derivative YACs and then on all of the YACs from the RP11 region (CEPH mega YACs, ICI and ICRF YACs). All of the STSs mapped back to their derivative YACs. The STS, 23L, derived from the YAC 28BC12, was located on YAC HD9F8 also known to be non-chimaeric through FISH analysis and positive with the STSs D19S180, D19S589 and D19S921. Therefore the gap between D19S921 and D19S572 was bridged by these two non

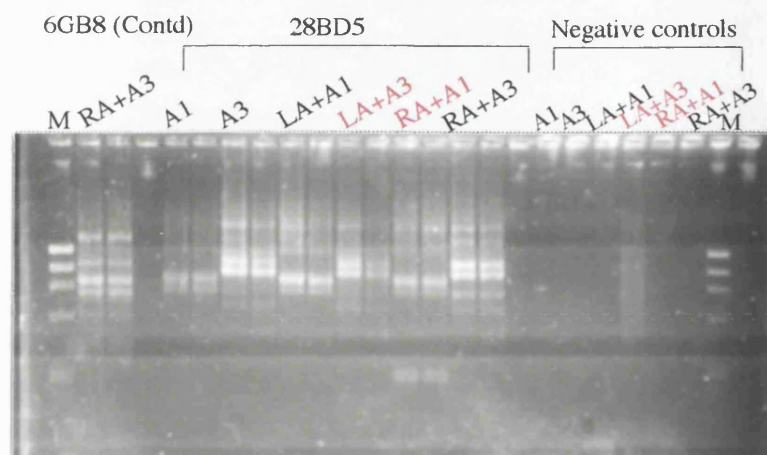
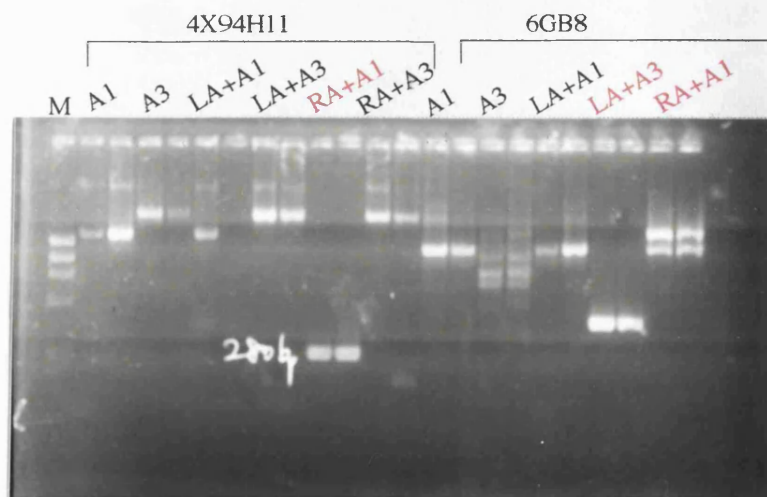


Figure 5.10

Ethidium bromide stained 2% agarose gels depicting the results of Alu-vector PCR performed on three of the YACs within the RP11 contig. For each of the YACs, Alu-PCR reactions were performed in duplicates. PCR reactions that show unique Alu-vector products are indicated in red.

M- ϕ x174 DNA size marker.

Ale1 and Ale3- human Alu repeat specific primers.

LA and RA- pYAC4 vector arm primers.

```

CTCCCGGGG CGAGTCGAAC GCCCGATCTC AAGATTACGG AATTCCTAG
AGGGTTTGGT AAAACAGATT GCTGGGTCCC ACTGCCAGAT TTTCTAACTC
TATAAATCTG GGGTGGGACC TGANAATCGT TATGTGTAAA AAGCTTCTGG
GTGACATGGG TGCTGCCAGT CCAGGGGAGC ACACCTGGCT TTCAGCTAGG
GTCCCTTCTC TACCTTTTCC TGCCTCGGCC CCCCAGGGAG GGGGCCCAAN
ACTAAGCTGC TGTCACANNT TCCACACTTG TCCCTCTACA TGGACGGCTG
CCCTTTCTAA TGCATTCTCT ACTCAGCNGG GCNCAGGGAG CCAATTTAAA
AATAAATAAA CCAGGTCATG TNTCCCCTAT GCTTGTA AAC CTTTAAACT

```

5.11 a

YAC end sequence data generated from YAC HD12D9 showing the forward and reverse primers of STS **9R**.

```

AAATCTCCCG GGGGNGAGTC GAACGCCCGA TCTCAAGATT ACGGAATTCT
AGGCCAGTCT GGACAAAATA GCAAGACCCT GGCTCTACAA AAAATAAAAA
ATTAGCCAGG CGTGGTGGAG TGCACCTGTA NACCCAGCTN NTCGGGAGGC
TGAGGCNTGA AAATCACTTG AACGCAGGAG ACAAATGCTG CATTGAGCTG
AGATAGCGCC ANNGCNCTCC AGCCTGGGTG ACGGAGTTAG ACTGTCTCAA

```

5.11 b

YAC end sequence generated from YAC 6GB8 showing the forward and reverse primers of STS **11R**

```

AATTTAATTT ATCACTACGG GAATTCTCTC AGGAGGAGTG GAAATGCCTG
GACCCTGCTC AGAGGACTCT ATACAGGGGA CGTGATGCTG GAGAATTATA
GGAACCTGGT CTCCCTGGGT GAGGATAACT TCCCTCCAGA AGTGGGGATG
TGCCCTTG TG TATCTTTGTA TTTTCTCTTG TTTT TAGATA CAGTGTCTTG
CTCTGTCACC CAGGCTGGAG TGCTGGGG

```

5.11 c

YAC end sequence data generated from YAC 28BC12 showing the forward and reverse primers of STS **23L**.

Figure 5.11

YAC end sequence data generated from Alu-vector PCR method. Only the sequence of useful STSs are shown. PCR primers designed to amplify STSs are underlined in red with orientation indicated by arrows. The pYAC4 vector sequence is highlighted in blue.

chimaeric YACs. 23L was also located on mega YACs such as 790-c-10, 744-a-11 and 171-b-11 that provided coverage at the proximal end of RP11 region. These results are depicted in figure 5.12. The STS 9R derived from the YAC HD12D9 was also located on other mega YACs and ICI and ICRF YACs, however unlike 23L it was difficult to distinguish the origin of 9R by these 'STS hits'. It appeared to map to the middle of HD12D9 rather than either of its termini. This observation implied that 9R was derived from the deletion derivative of YAC HD12D9, the smaller of the two YACs seen on PFGE of HD12D9, rather than from the larger intact YAC. The other STS from HD12D9, 9L, failed to produce a single product on PCR, this may have been due to the fact that the STS 9L was derived from a stretch of sequence rich in repetitive sequence. Due to its dubious structure, no further effort was made to isolate STSs from HD12D9. However as it was non-chimaeric and appeared to contain a considerable number of STSs from the RP11 critical region, although unstable it was retained within the RP11 contig.

The STSs 7R and 8R derived from YACs 4X100D12 and 4X94H11, respectively, were only located on the derivative YAC and not in any of the other overlapping YACs. This implied that these ends might be chimaeric and were therefore abandoned. The STSs 11R and 11L derived from the YAC 6GB8 mapped back to 6GB8. The STS 11R was also located on YACs 790-a-5 and 955-g-11, implying that it was derived from the proximal end of 6GB8. The localisation of the novel STSs generated from YAC termini and the gap closure achieved is depicted in figure 5.13. The STS 11R was used to screen the ICI and ICRF YAC libraries, and two YACs 4X108H11 and 4X111H11 were isolated. Therefore, of all the STS generated from end sequence of YACs only 11R could be used to isolate further YACs. However it was not possible to investigate overlaps between these newly isolated YACs and YAC clones located proximal to them due to the deficiency of STSs in the region between the gene *PRKCG* and D19S418. Therefore the chromosome walking strategy to close the gap between D19S927 and D19S418 was temporarily abandoned until more region specific STSs were generated. Cosmids that have been FISH mapped to the RP11 interval by our collaborators at the Lawrence Livermore National Laboratories (LLNL), were used to generate these novel region specific STSs.

YAC ID	La b ID	Ends Isolated	STS Designed	STS Primers (5'→3')	Size (bp)	Tm °C
HD12D9	9	LA, RA	RA (9R)	F- aagcttctgggtgacatgg / R- atgtagagggacaagtgtg	150	58°C
			LA (9L)	F- agaagggatagaggcctgga / R- ccaacttgctaaccgagag	170	62°C
HD9F8	4	LA, RA	-	-		
4X5F1	27	LA, RA	-	-		
6GB8	11	LA, RA	RA (11R)	F- gccagtctggacaaaatagc / R- ctatctcagctcaatgcagc	155	60°C
			LA (11L)	F- acccttctttatggctgc / R- tcaacaagagtgaaactcca	145	56°C
4X100D12	7	LA, RA	RA (7R)	F-aacctggggaacacgggtgg / R- aggagtgcagtggcacgct	170	56°C
4X94H11	8	RA only	RA (8R)	F- ctctactactcgatgtcc / R- gacagagtgagactccgt	120	56°C
28BC12	23	LA, RA	LA (23L)	F-ctcaggaggagtggaaatgc / R- gatacacaagggcacatccc	137	62°C
28BD5	20	LA, RA	-	-		

Table 5.4.

Details of YAC end recovery. Despite recovering YAC ends from all the YACs useful STSs could be designed from the YACs shown bold face.

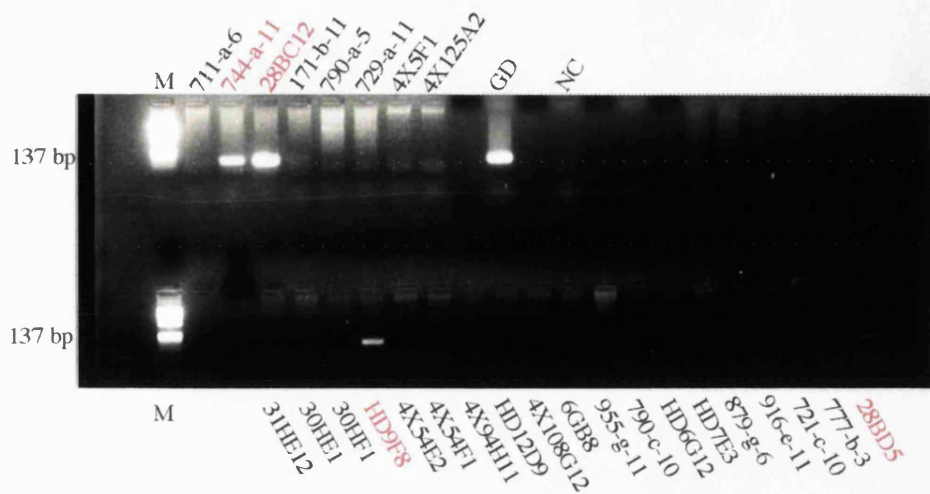


Figure 5.12a

Ethidium bromide stained 2.5% agarose gel presenting results of PCR analysis of YAC insert end sequence STS 23L (derived from YAC 28BC12) in a selection of YACs from the RP11 YAC panel. YACs presenting a positive result are indicated in red. M denotes the ϕ x174 DNA size marker, GD the genomic DNA/ positive control and NC the negative control (water only).

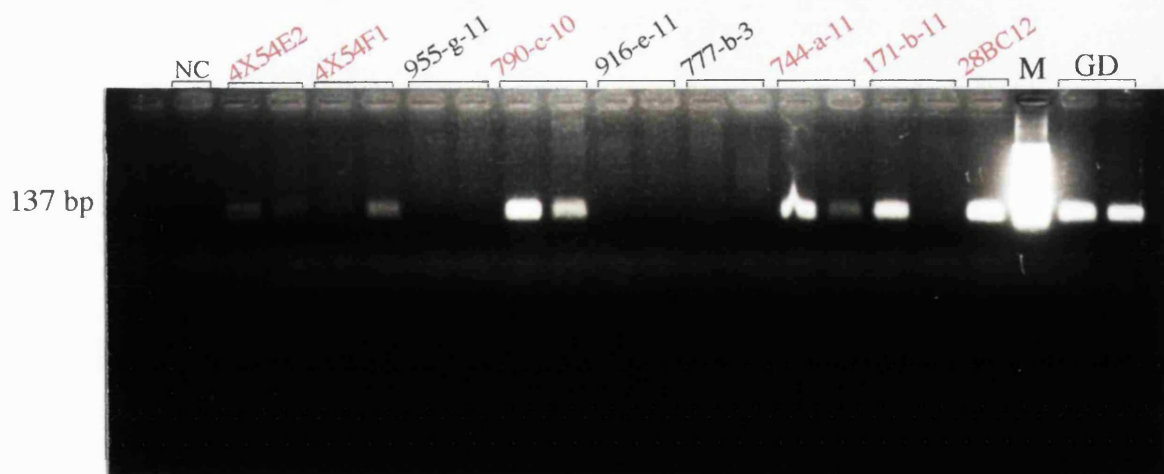


Figure 5.12b

Ethidium bromide stained 2.5% agarose gel presenting more STS content mapping results of STS 23L. Three YACs (i.e. 4X54E2, 4X54F1 and 790-c-10) that did not show positive results in the earlier analysis gave positive results in the PCR assay shown above.

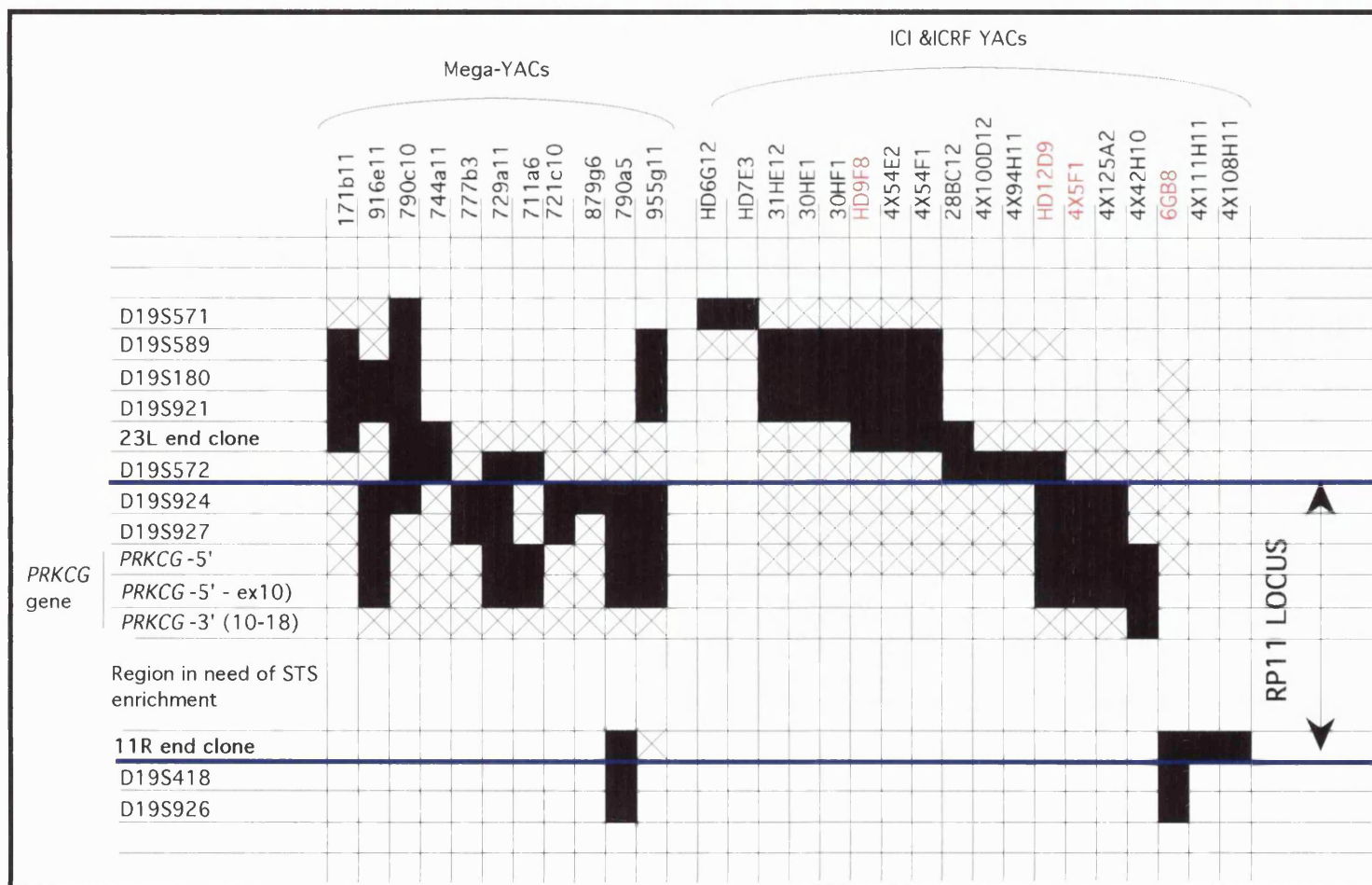


Figure 5.13

Mapping of two novel STSs generated from YAC termini in all the YACs located in the RP11 interval showing the gap closure achieved between D19S921 and D19S572 with STS 23L.

5.4 Isolation, characterisation and genetic mapping of microsatellites from cosmids in the RP11 region

Due to high STS density between the markers D19S180 and D19S927 complete YAC coverage was achieved with relative ease at the proximal half of the RP11 interval. In contrast the region between D19S927 and D19S418, which was particularly deficient of STS proved difficult to bridge by YACs. It became apparent that isolation of novel microsatellite markers within this interval could have a dual advantage. If the endeavour was successful not only would it increase the STS density and facilitate the isolation of further YACs, also if the microsatellite markers proved to be polymorphic, there was an opportunity for further locus refinement from the distal end through analysis of these novel markers in ADRP2. As described in chapter 3, the recombination event in the family ADRP2 (that led to locus refinement at the distal end) had occurred within the region targeted for STS enrichment between *PRKCG* and D19S418.

Fortuitously the LLNL had localised several cosmid contigs between *PRKCG* and D19S418. Each cosmid contig was assembled by *EcoRI* restriction mapping to determine the overlaps between cosmid clones, and the order and the distance between selected cosmid clones (usually representing a contig) was determined by fluorescence *in situ* hybridisation in sperm pronuclei. These selected cosmids are referred to as reference cosmids and their FISH distance estimates provide the metric scale for the LLNL physical map (see section 5.2.4). Cosmid clones from contigs thus assembled were utilised for the isolation of microsatellite markers. An adaptation of the LLNL physical map of the RP11 region of chromosome 19 at the onset of this project, depicting the cosmid contigs utilised in this study is presented in figure 5.14.

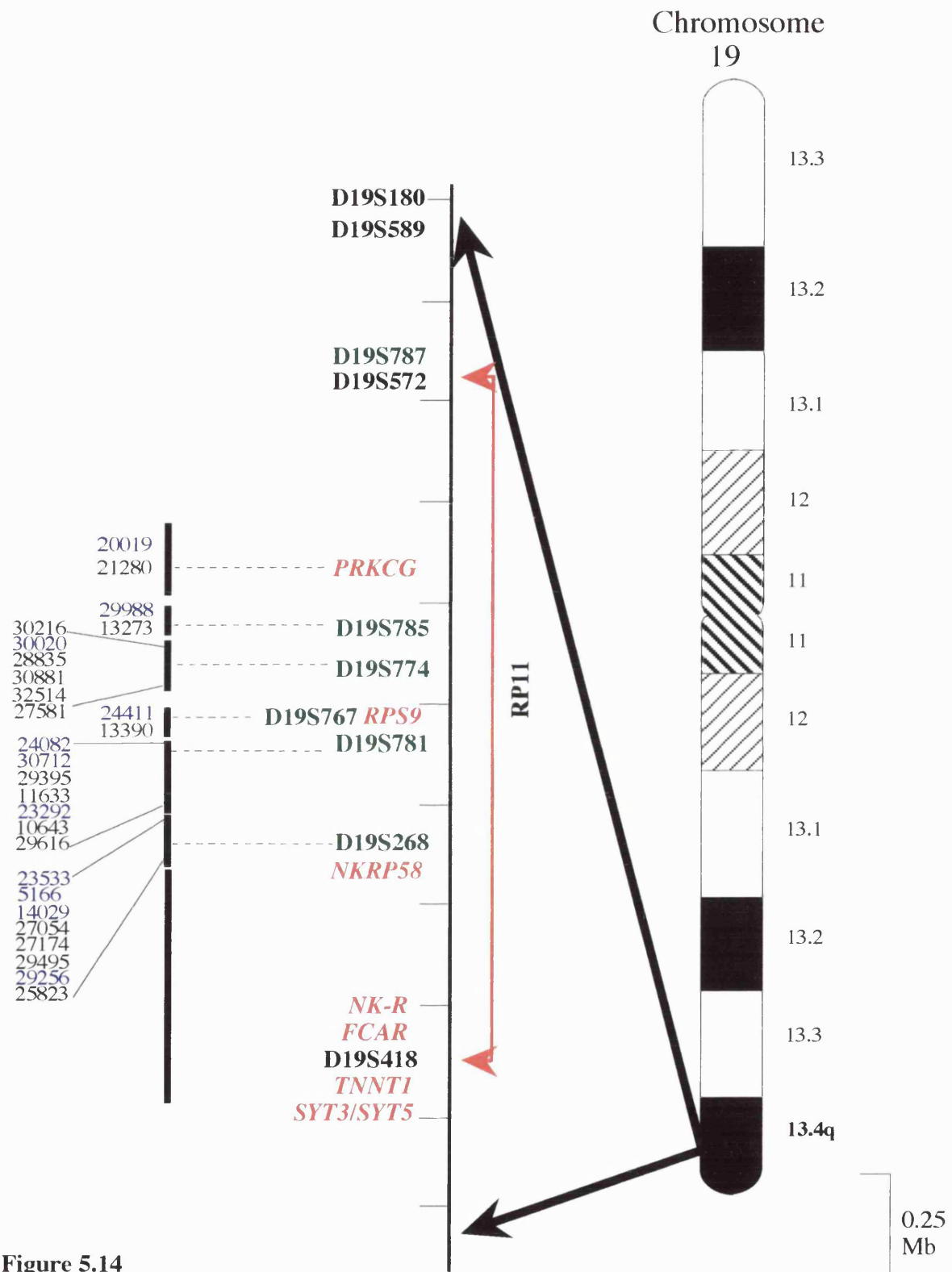


Figure 5.14

A sub-section of the physical map of chromosome 19q13.4, adapted from the LLNL integrated map of chromosome 19 (http://www-bio.llnl.gov/genome-bin/loadmap?m_region=mq2+arm). The 27 cosmids analysed in this study are presented to the left of the figure; the cosmids in blue were used for the isolation of (CA)_n repeats. The black bars represent *EcoRI* restriction mapped cosmid contigs. Genes are indicated in red, cosmid contigs are indicated in green and microsatellite markers are in bold face.

5.4.1 Results

5.4.1.1 Obtaining cosmids from the proximal end of RP11

The aforementioned cosmid contigs assembled by LLNL have each been given a GDB accession number. Henceforth each of the cosmid contigs utilised in this study would be identified by this number e.g. D19S268. The cosmid contigs used for the isolation of microsatellite markers were located in the following order within the RP11 interval: *PRKCG*, D19S785, D19S774, D19S767, D19S781 and D19S268, with *PRKCG* lying most centromeric (fig 5.14). Constituent cosmids of these contigs and their *EcoRI* restriction maps were obtained from LLNL, USA. Analysis of the restriction maps revealed that most contigs contained redundant cosmid clones with strong overlap to a minimal path clone. To avoid redundancies only cosmid clones with minimal overlap to each other as determined by *EcoRI* restriction mapping were selected for our study. Table 5.5 gives details of cosmid clones of each contig chosen for the isolation of markers. All cosmids were cloned in the LAWRIST4 cosmid vector and received as 'stabs' in agar. Selected cosmids were grown and DNA was isolated from single colonies as described in section 2.1.2. DNA concentrations were estimated by agarose gel electrophoresis.

Cosmid contig	Cosmids from the minimal path of the contig
<i>PRKCG</i>	20019 ¹ , 21280
D19S785	29988 ² , 13273
D19S774	30216, 30020, 28835, 30881, 32514, 27581
D19S767	13390, <i>24411</i>
D19S781	24082, 30712, 29395 , 11633, 23292, 10643, 29616
D19S268	23533, 5166, 14029 , 27054, 27174, 29495, 29256, 25823

¹ Cosmids given in bold face belong to the 236 reference cosmids set of the LLNL physical map (see section 5.2.4).

² Cosmids in italic have been ordered relative to the 236 clones in the reference set, but distance relative to the reference clones are unknown.

Table 5.5.

Identities of cosmid clones chosen for isolation of (CA)_n markers.

5.4.1.2 Screening cosmids for the presence of (CA)_n repeat sequences

The 27 cosmid clones selected were first screened for presence of (CA)_n repeat sequences. To do so 500ng of DNA from each cosmid was digested overnight with *EcoRV* enzyme, fractionated by gel electrophoresis and subjected to southern blot analysis. The cosmid filters were hybridised with [γ ³²P]-dATP end labelled G₄(GT)₁₃ probe at 65⁰C overnight. Auto-radiographic exposure of the washed filters identified the following cosmids to be positive with this probe:

Cosmid contig	Cosmids positive with G ₄ (GT) ₁₃ probe
<i>PRKCG</i>	20019, 21280
D19S785	29988, 13273
D19S774	30020
D19S767	24411
D19S781	24082, 30712, 11633, 23292
D19S268	23533, 5166, 14029, 29256

Hybridisation with the G₄(GT)₁₃ probe also indicated that a few cosmids, which belonged to one contig and overlapped with each other, were positive for the same repeat sequence. This being the case with cosmid pairs (**20019**, 21280), (**29988**, 13273) and (**23292**, 11633) from respective contigs *PRKCG*, D19S785 and D19S781, only one cosmid (shown in bold face) of each cosmid pair was used for the isolation of the repeat sequence. Therefore from the initial 27 cosmids chosen only 11 cosmids were used for further analysis. In order to identify the enzyme/enzymes producing the smallest (CA)_n positive clonable fragments, the selected cosmids were digested with following restriction enzymes, *Bam*HI, *Sau*3AI, *Msp*I and *Hae*III and fractionated by gel electrophoresis. Following southern blot analysis the resulting filters were hybridised with the same G₄(GT)₁₃ probe.

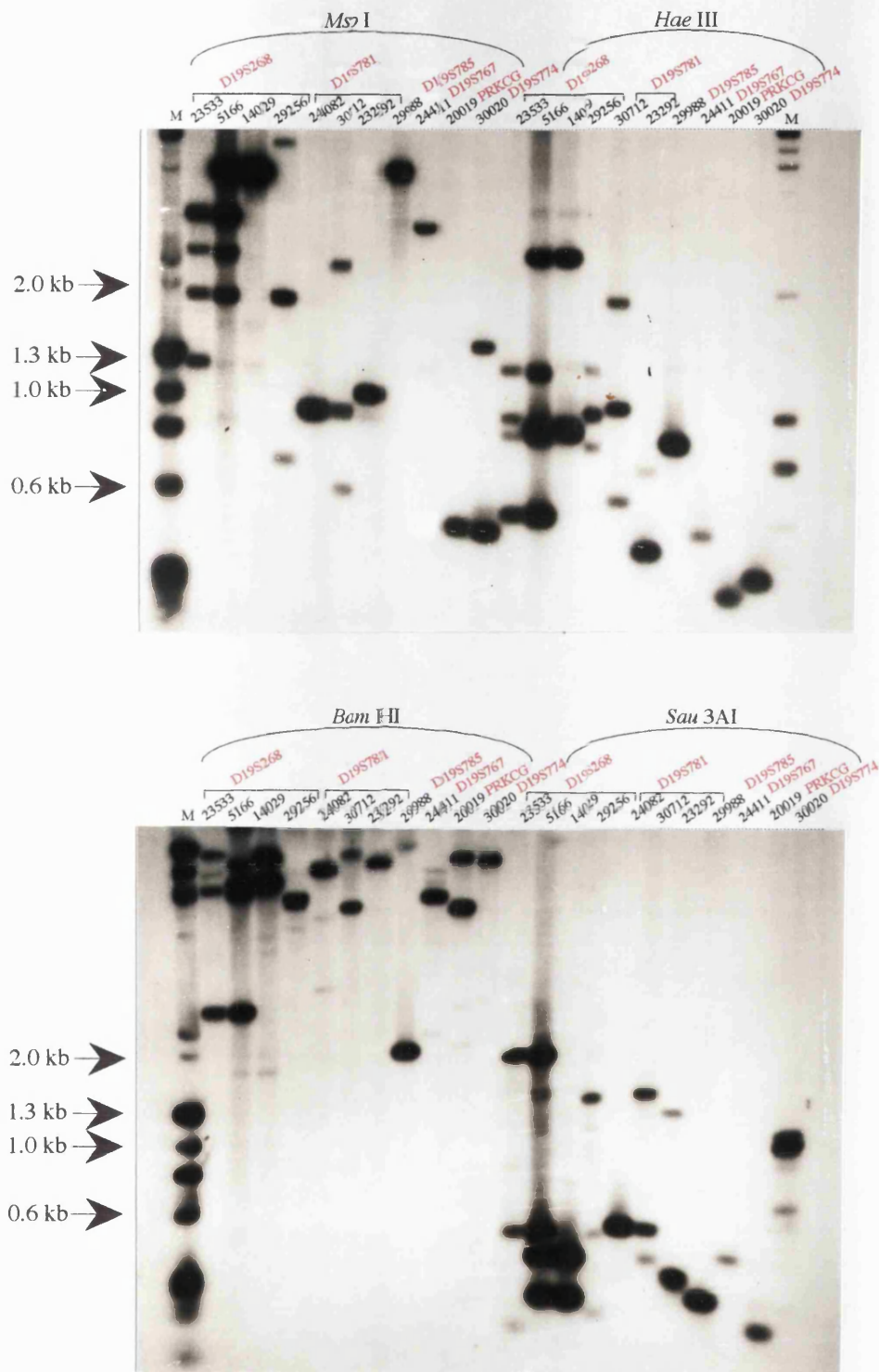


Figure 5.15

Autoradiographs of the 11 cosmids belonging to 'contigs' D19S268, D19S781, D19S785, PRKCG, D19S774 and D19S767 digested to completion with *Msp* I, *Hae* III, *Sau* 3AI and *Bam* HI followed by hybridisation to G₄(GT)₁₃ probe. Overall comparison reveals *Sau* 3AI and *Hae* III enzymes to produce the smallest (CA)_n positive fragments for purpose of cloning. M denotes combined λ Hind III and ϕ X 174 size marker.

The auto-radiographic analysis revealed certain cosmid clones to contain more than one (CA)_n repeat, which they appear to share with adjoining cosmid(s). For example cosmid 5166 of contig D19S268, appeared to contain 4 separate (CA)_n repeats or at least 2 (CA)_n repeats, interrupted by enzyme restriction sites, and cosmids 23533 and 14029, which overlapped 5166, each appeared to contain one (or two) of the two (or four) separate repeats contained within 5166. Similarly cosmid 30712 (of D19S781) appeared to contain two (CA)_n repeats, one of which was also the only (CA)_n repeat to be present in cosmid 24082. Cosmids 20019, 29988, 30020, 24411, 24082, 23292 and 29256 with single strong hybridisation signals implied the presence of single repeats. Auto-radiographs of all the chosen 11 cosmids are presented in figure 5.15. Size markers radiolabelled by random priming method, were also hybridised to the filters subsequent to hybridisation with the G₄(GT)₁₃ probe to prevent cross-hybridisation by their simultaneous addition. Even though plasmid vectors are capable of accommodating inserts up to 10-15 kb, smaller inserts are cloned with greater efficiency and can also be easily sequenced. Since *Sau3AI* and *HaeIII* provided fragments that were less than 1.5 kb they were selected as the enzymes of choice for creating cosmid mini libraries.

5.4.1.3 Construction of cosmid mini libraries

Mini libraries were constructed for each of the 11 aforesaid cosmids by digesting with the enzyme that produced the smallest clonable fragment, which was either *Sau3AI* or *HaeIII*, and cloning in to the appropriately cut vector. To increase the efficiency of capturing all putative repeats present within cosmids, mini libraries were constructed with both *Sau3AI* and *HaeIII* for some cosmids and those cosmids that harboured more than one repeat (i.e. 5166) were sub cloned in two different vector systems such as the pBluescript[®] II KS phagemid vector (pBS) and the commercially available *Bam*HI digested pUC18 vector in Ready-To-Go ligation kit (see table 5.6).

More detailed description of this sub cloning procedure can be found in the Materials and Methods chapter under section 2.13. After the mini library of each cosmid was constructed, plated out and incubated overnight at 37⁰C, recombinant colonies were identified as white colonies due to the absence of functional β-galactosidase activity. No colonies were observed in the negative controls.

Cosmid contig	Cosmid clone	Digestive enzyme	Plasmid vector
<i>PRKCG</i>	20019	<i>Hae</i> III <i>Sau</i> 3AI	<i>Eco</i> RV digested pBluescript [®] II KS <i>Bam</i> HI digested pBluescript [®] II KS
D19S785	29988	<i>Sau</i> 3AI	<i>Bam</i> HI digested pBluescript [®] II KS <i>Bam</i> HI digested pUC18
D19S774	30020	<i>Hae</i> III	<i>Eco</i> RV digested pBluescript [®] II KS
D19S767	24411	<i>Hae</i> III	<i>Eco</i> RV digested pBluescript [®] II KS
D19S781	24082 30712 23292	<i>Sau</i> 3AI <i>Hae</i> III <i>Sau</i> 3AI <i>Hae</i> III	<i>Bam</i> HI digested pBluescript [®] II KS <i>Eco</i> RV digested pBluescript [®] II KS <i>Bam</i> HI digested pBluescript [®] II KS <i>Eco</i> RV digested pBluescript [®] II KS
D19S268	23533 5166 14029 29256	<i>Hae</i> III <i>Sau</i> 3AI <i>Sau</i> 3AI <i>Sau</i> 3AI	<i>Eco</i> RV digested pBluescript [®] II KS <i>Bam</i> HI digested pBluescript [®] II KS <i>Bam</i> HI digested pUC18 <i>Bam</i> HI digested pBluescript [®] II KS <i>Bam</i> HI digested pUC18 <i>Bam</i> HI digested pBluescript [®] II KS

Table 5.6.

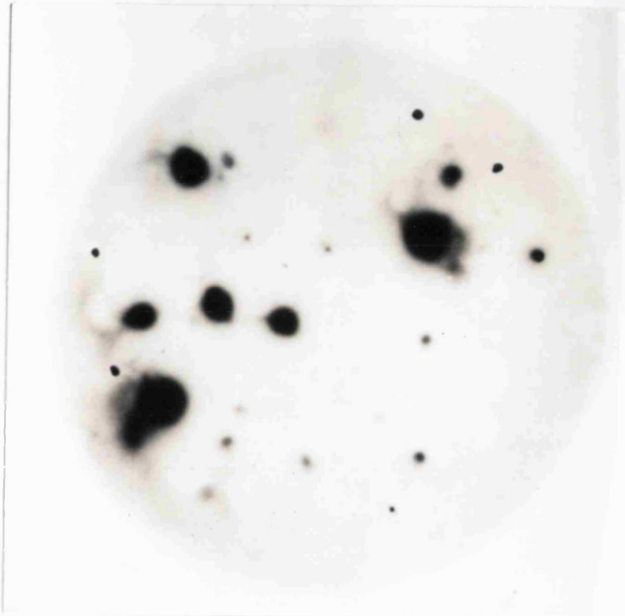
Details of mini library construction of cosmid clones used for isolation of (CA)_n repeats.

5.4.1.4 Isolation of (CA)_n positive subclones and sequencing

Two replica filters were made from the master mini library plate of each cosmid and processed as described in section 2.12b. These filters were then hybridised under stringent conditions with the G₄(GT)₁₃ probe. Authentic positive signals were identified as those that were present on both of the X-ray films of the replica filters. An example of a set of replica filters is shown in figure 5.16. With the exception of a few master plates where there were no positive colonies present, approximately 2-5 positive colonies were present on all other master plates and these were isolated as described in section 2.12.c. The selected positive colonies were inoculated in 10 ml of LB-broth supplemented with ampicillin. After overnight growth the DNA (of positive clones) was isolated using the alkaline lysis mini prep method. ~ 1 µg of recombinant DNA was double digested with an appropriate enzyme pair to separate the cloned insert from vector arms. Hence recombinant pBS and pUC18 plasmids were digested with *Xba*I and *Hind*III and *Kpn*I and *Hind*III, respectively. The DNA digests were gel fractionated, southern blotted and the filters re-probed with G₄(GT)₁₃ probe to confirm the positive nature of the selected clones. An example for positive clones isolated from cosmids 14029, 24411, 30020, 29988, 24082, 30712, 23292 and 20019 is shown in figure 5.17.

All the positive clones thus isolated were directly sequenced in a ABI 373a DNA sequencer, recombinant pBS clones were sequenced with T3 and T7 primer pairs whilst pUC18 clones were sequenced with forward and reverse M13 primers. Both primer sets are situated adjacent to the vector cloning sites. This led to the identification of most of the (CA)_n repeats whose presence was suggested from the initial analysis of the 11 cosmids. Unfortunately not all of the promised (CA)_n repeats could be isolated from every cosmid despite the effort made to ensure the isolation of all possible (CA)_n tracts. This was particularly the case for cosmids belonging to the contig D19S268. (CA)_n repeats harboured within cosmids 23533 and 29256 could not be isolated, furthermore only two (CA)_n repeats could be isolated from cosmid 5166, which on initial analysis appeared to harbour four (CA)_n repeats (see figure 5.15). One of the (CA)_n repeats isolated from cosmid 5166 was also the only (CA)_n to be isolated from its neighbouring cosmid 14029.

Replica filter 1



Replica filter 2

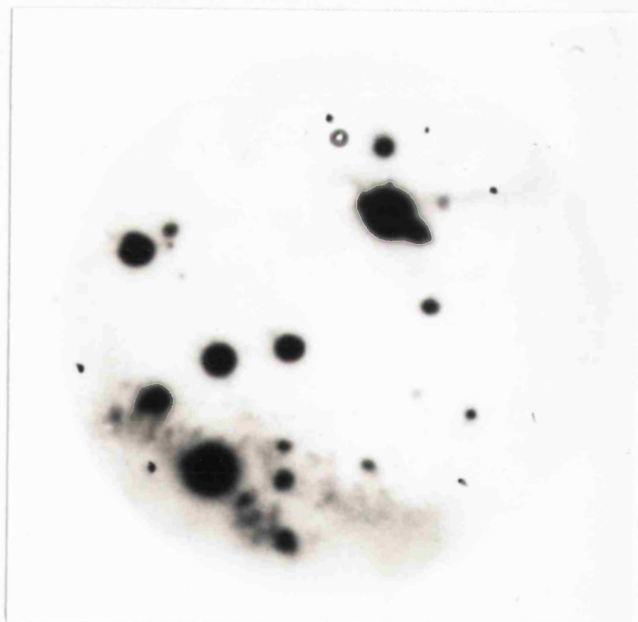


Figure 5.16

G4(GT)₁₃ probed replica filters of the cosmid 20019 pBluescript II KS mini library. Colonies corresponding to the replicate signals were selected from the plated libraries.

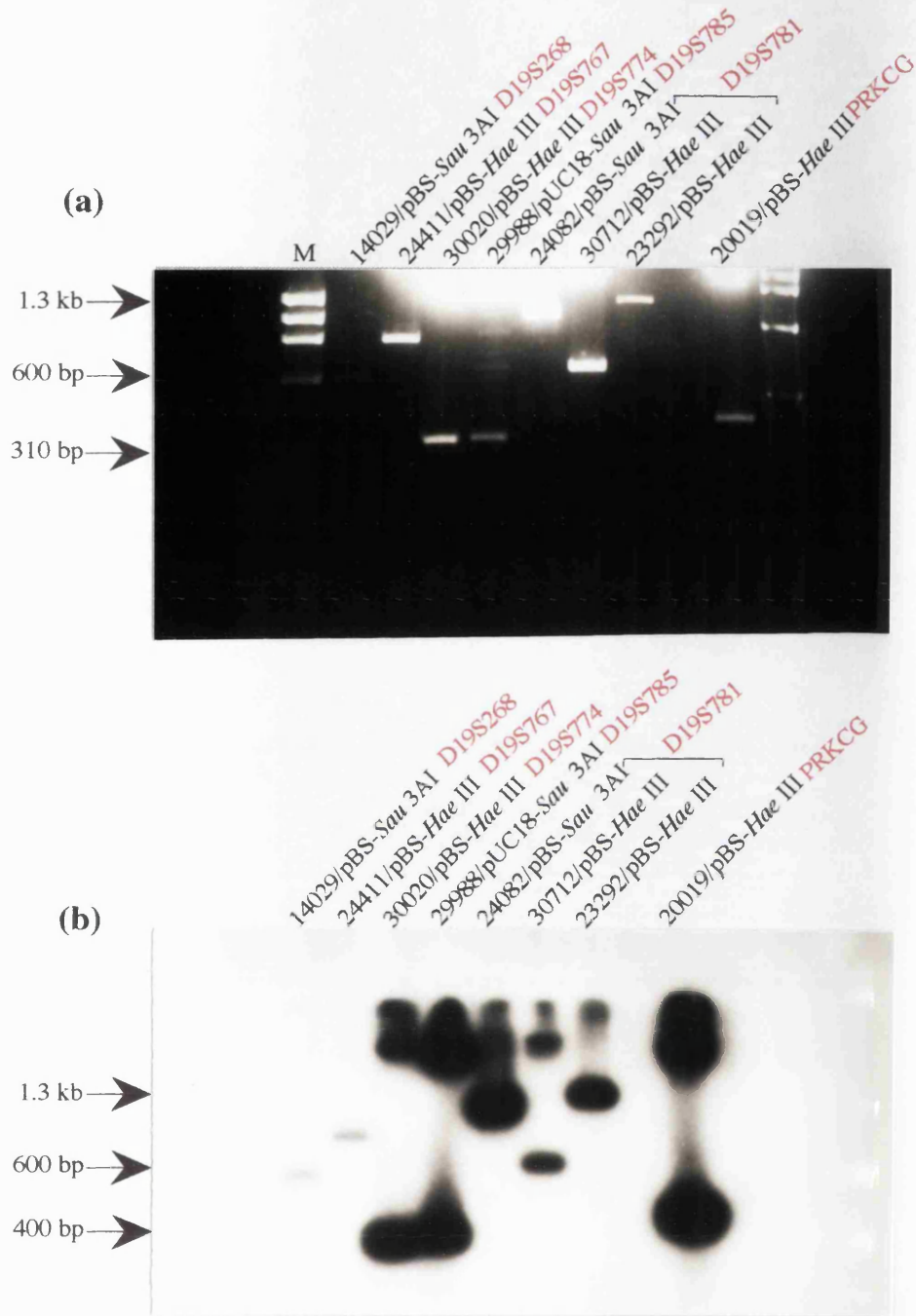


Figure 5.17

(a) Ethidium bromide stained 2.5% gel and (b) its corresponding autoradiograph following hybridisation with G4(GT)₁₃ probe, showing the fractionation of the restriction enzyme digested positive plasmid subclones isolated for each of the cosmids utilised for (CA)_n isolation. A single representative subclone is shown for a given cosmid. The derivative cosmid, the digestive enzyme used in the subcloning, and the cloning vector are indicated for all the plasmid subclones. *Xba*I + *Hind*III enzymes were used to excise the inserts from the pBS subclones while *Hind*III + *Kpn*I enzymes were used for the pUC18 subclones. M denotes the ϕ x174 size marker.

After the repeat motif from each subclone/cosmid was identified they were classified as perfect (no interruptions in the run of dinucleotide repeat), imperfect (one or more interruptions in the run of repeats), or compound (a run of perfect or imperfect CA or GT repeat adjacent to a run of another simple sequence repeat) as described by Weber (1990). Equal numbers of perfect and imperfect repeat sequences and two relatively infrequent compound repeat sequences were identified in this study. Of the two, the repeat sequence isolated from cosmid 23292 was an uninterrupted perfect compound repeat (D19S781.5N). The microsatellites identified in cosmids 29988, 30020 and 24411 all were uninterrupted perfect (CA)_n repeats (see table 5.7).

The sequence obtained from all subclones was also analysed for cleavage sites of the restriction enzyme used for the construction of cosmid mini libraries in order to investigate the possibility of the cloned insert being composed of concatemeric fragments. It was important to analyse the sequence in this manner since during the shot gun cloning procedure precautions were not taken to prevent co-ligation of non-contiguous fragments derived from restriction digests of cosmids. The sequence was also screened for identities against entries within the GenBank database, using the BLAST/FASTA computer programs (Pearson and Lipman, 1988; Altschul *et al.*, 1990) and any homology to Alu or LINE 1 repetitive sequences was given particular attention. Such preliminary analysis of sequence prior to designing primers flanking the (CA)_n repeats was conducted to prevent difficulties with subsequent PCR reactions. In the process of such sequence analysis, one (CA)_n positive subclone of cosmid 20019 (also known to contain the *PRKCG* gene) was found to contain sequence identical to that of *PRKCG* exon 8, and subsequent sequence comparisons with the already-known genomic sequence of *PRKCG* positioned the repeat motif within intron 7 of *PRKCG* (see chapter 4). Prior to this study the only intragenic marker associated with the *PRKCG* gene was a RFLP marker (Johnson *et al.*, 1988).

Once the sequence flanking each repeat was analysed to ensure that it was derived from a single fragment and that the DNA was non-repetitive, oligonucleotide primer pairs were designed to amplify each (CA)_n repeat so as to yield unique products ranging in size from 100-350 bp (see table 5.7). It was not possible to design both primers to encompass the (CA)_n repeat in cosmid clones, 24411 and 24082, since the

available sequence was only suitable to design one primer. These primers were used to directly sequence the cosmids 30020 and 24082 in order to obtain enough sequence to design the other primer. However, sequence could only be obtained thus for cosmid 24082, and another primer was designed accordingly. In contrast to 24082 all attempts made at directly sequencing the cosmid 30020 failed. Therefore a primer was designed from the already available repetitive sequence flanking the $(CA)_n$ motif in 30020. Unfortunately this primer pair also failed to produce a product, subsequently the $(CA)_n$ in cosmid 30020 (D19S774.2N) was considered as an ineffectual microsatellite for being embedded among highly repetitive sequence and was not further characterised. Later a primer pair ([5'→3'] F- agtcccgccagccaactgat/ R- cctgtaatcccagctcgga) was designed from the sequence downstream of $(CA)_n$ motif in cosmid 30020 to produce a PCR fragment of 200 bp. This STS identified as D19S774N/STS contributed towards STS enrichment in the region between *PRKCG* and the marker D19S418. The primer pairs flanking the other eight novel microsatellites were tested on their derivative cosmids and genomic DNA and they all produced the expected size product. The characteristics of all identified $(CA)_n$ motifs and their primer sequences designed for their amplification are presented in table 5.7 and examples of repeat containing sequence data identified from cosmids 30020 and 20019 are presented in figure 5.18. Next all novel repeats were tested for their heterozygosity and usefulness as microsatellites. From henceforth all the microsatellites would be referred to by the laboratory identity number given to them (see table 5.7).

Locus/ Cosmid contig	Cosmid ID	Repeat unit	Repeat category	ID of repeat ¹	PCR primer sequence (5'-3')	Ann. Temp (°C)	Size (bp)	P/S
<i>PRKCG</i>	20019	(CA) ₁₂ CG(CA) ₄ CG(CA) ₅	Imperfect	PRKCG-IM ²	(F) ggagaggcaaacattgga (R) ggatggagcgcaatattacc	60°C	272	P
D19S785	29988	(CA) ₂₃	Perfect	D19S785.1N	(F) aggtcatcctcagacaagt (R) tcagctcttcacagcttgc	60°C	185	P
D19S774	30020	(CA) ₂₇	Perfect	D19S774.2N	Primers could not be designed			
D19S767	24411	(CA) ₆	Perfect	D19S767.2N	(F) ttctgaccatcttacctggc (R) cgtcagaagacttacacgtg	60°C	245	S
D19S781	24082	(CA) ₁₅ (TA) ₄ (CA) ₄ TA (CA) ₄	Imperfect	D19S781.1N	(F) ctgaactggagttgatacga (R) cataatgaacgtaagagtaatc	58°C	150	P
	30712	(GT) ₃ AT(GT) ₇ (AT) ₁₃	Compound (imperfect)	D19S781.2N	(F) tctgtcccacacagaagtgg (R) ggctgtagtggatgatgc	60°C	310	P
	23292	(GT) ₁₆ (AT) ₂₂	Compound (perfect)	D19S781.5N	(F) tccatggatgacgggttggg (R) tatgacacctgtaacaccata	62°C	270	P
D19S268	5166	(CA) ₃ AA(CA) ₂ AA(CA) ₃	Imperfect	D19S268.21N	(F) actcgtagacttgtaagtga (R) tcactctctgttagagtc	60°C	185	S
		(CA) ₇ AA(CA) ₆	Imperfect	D19S268.22N/ D19S268.31N	(F) atcactagcatttatatacc (R) tccttgaagaaattgtcca	52°C	150	S
	14029	(CA) ₇ AA(CA) ₆	Imperfect	D19S268.31N	(F) atcactagcatttatatacc (R) tccttgaagaaattgtcca	52°C	150	S

¹ The ID for the (CA)_n repeats isolated from a given cosmid was derived thus, the first part eg D19S268 was the ID of the cosmid contig, this is followed the lab number given to the cosmid from which the repeat was derived (D19S268.2) this is then followed by the number of the repeat within that cosmid (D19S268.21 or D19S268.22) and finally N for novel.

² IM stands for Intragenic Marker

Table 5.7.

Presentation of all polymorphic markers (P) and STSs (S) isolated and characterised from cosmids located in the distal RP11 region.

5.4.1.5 Determining the heterozygosity of novel microsatellites

The informativeness of the 8 STSs as genetic markers was determined by genotyping a cohort of 20-30 unrelated 'normal' Caucasian individuals for the identification of polymorphisms. In this analysis the following STSs, PRKCG.IM, D19S785.1N, D19S781.2N, D19S781.5N and D19S781.1N proved to be polymorphic, whereas the STSs D19S767.2N, D19S268.21N and D19S268.31N did not. The non polymorphic nature of these STSs was not surprising considering that the (CA)_n motif in marker D19S767.2N was only composed of 6 CA units and that the respective repeat motifs of D19S268.21N and D19S268.31N, which were (CA)₃AA(CA)₂AA(CA)₃ and (CA)₇AA(CA)₆, were both short and interrupted.

The polymorphic STSs, D19S785.1N and D19S781.5N both contained uninterrupted perfect repeat sequences. The STSs PRKCG.IM, D19S781.1N and D19S781.2N all contained interrupted hence imperfect repeat sequences but still proved to be polymorphic. PRKCG.IM and D19S781.2N, with the respective interrupted repeats (CA)₁₂CG(CA)₄CG(CA)₅ and (GT)₃AT(GT)₇(AT)₁₃, had 6 and 7 different alleles segregating in the population (figure 5.19) whereas D19S781.1N with a highly interrupted CA)₁₅(TA)₄(CA)₄TA(CA)₄ motif only presented 2 alleles. With 8 different alleles the STS D19S785.1N proved to be the most polymorphic genetic marker of all. The marker D19S781.5N, though polymorphic proved to be a difficult microsatellite marker in practice, having many 'stutter bands' associated with its alleles (data not shown), which was attributed to "slippage" during PCR (Litt and Luty, 1989; Petersen *et al.*, 1990). This led to difficulties in scoring of alleles, which were of similar size and in close proximity in the polyacrylamide gel, therefore this repeat was not characterised any further. Hence 4 useful polymorphic markers were isolated from the 11 cosmids used in this study. The characteristics of all polymorphic markers identified have been presented previously (table 5.7). The estimated allele frequencies of all polymorphic markers with their corresponding observed heterozygosity values and GenBank accession numbers are presented in table 5.8.

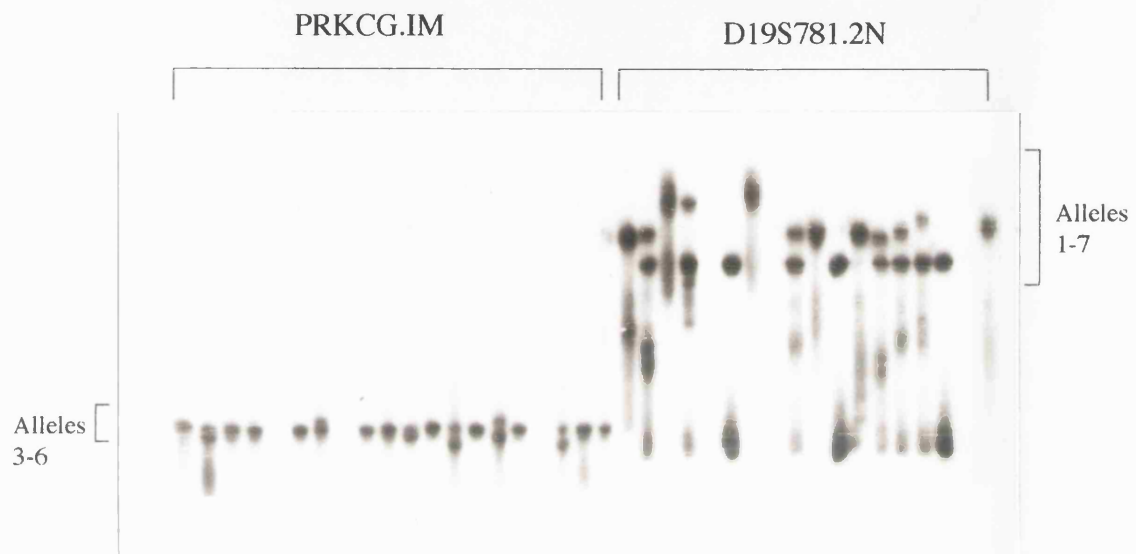


Figure 5.19

Autoradiograph depicting the polymorphic nature of novel microsatellites PRKCG.IM and D19S781.2N. Even though PRKCG.IM showed 6 alleles only 4 of those alleles are depicted above, alleles 1 and 2 which are relatively rare are not shown. D19S781.2N presents with 7 alleles. For both markers only a portion of the individuals analysed in total are shown above. In total 20-30 individuals were analysed to determine the heterozygosity values of each of the novel markers.

Cosmid ID	Repeat unit	Lab ID	PCR Prod Size (bp)	Allele No:	(Allele no. and frequency)	Obs Het.	GDB Acc No:
20019	(CA) ₁₂ CG(CA) ₄ CG(CA) ₅	PRKCG.IM	272	6	(1) 0.02 (2) 0.02 (3) 0.10 (4) 0.70 (5) 0.10 (6) 0.06	0.45	AF030445
29988	(CA) ₂₃	D19S785.1N	185	8	(1) 0.06 (2) 0.06 (3) 0.06 (4) 0.06 (5) 0.32 (6) 0.27 (7) 0.11 (8) 0.06	0.72	AF069628
24082	CA ₁₅ (TA) ₄ (CA) ₄ TA (CA) ₄	D19S781.1N	150	2	(1) 0.44 (2) 0.56	0.57	AF069629
30712	(GT) ₃ AT(GT) ₇ (AT) ₁₃	D19S781.2N	310	7	(1) 0.06 (2) 0.09 (3) 0.03 (4) 0.03 (5) 0.35 (6) 0.06 (7) 0.38	0.65	AF069627

Table 5.8

The characteristics of all polymorphic markers identified from cosmids in the RP11 region chromosome 19q13.4.

Obs Het – Observed heterozygosity values

(Obs Het = No of heterozygotes/ Total no of individuals typed)

5.4.1.6 Verification of genetic location of novel microsatellites by haplotype analysis in the RP11 linked families

Most of the novel microsatellite markers characterised in this study were isolated from cosmids whose physical position was established by Fluorescence *in situ* hybridisation (FISH) to highly extended chromatin from sperm pronuclear interphase. Use of chromatin targets of increasing resolution permits high resolution cosmid ordering as well as generation of estimates of genomic distance separating cosmid pairs in the range of less than 50 kb to 1.2 Mb (Brandriff *et al.*, 1991, 1994). Cosmids that were ordered thus were listed in table 5.5. Therefore the position of the genetic markers could be established from the physical position of their derivative cosmids, especially since the robustness of the technique used to establish the order of the cosmids renders the order information reliable. However it is still necessary to confirm their chromosomal position genetically by analysing these novel markers in pedigrees with informative recombination events. The novel markers PRKCG.IM, D19S785.1N, D19S781.1N and D19S781.2N were genotyped in families, ADRP11, ADRP2, RP1907 and ADRP24, which had individuals (both normal and affected) with recombination events between certain key genetic markers in the RP11 interval that would help confirm the genetic position of the novel markers. For example the normal Individual III.2 and the affected individual III.9 of ADRP11 each had a recombination event between the markers D19S927 and D19S418 in the haplotype inherited from their non-carrier parent. Individual III.9 also had a proximal recombination event between the marker D19S921 and D19S572 in the affected haplotype inherited from the carrier parent. In ADRP2 the affected individual II.8 had a distal recombination event between the gene *PRKCG* and D19S418 in her affected haplotype (see section 3.3.2).

Fortunately all the novel markers were informative in at least one of the families mentioned earlier. The segregation of alleles of the novel markers in the above mentioned recombinant individuals placed the novel markers between D19S927 and D19S418 in the descending order, PRKCG.IM, D19S785.1N, D19S781.1N and D19S781.2N with PRKCG.IM lying most centromeric. This order was also in agreement with the map position of the cosmids from which the novel markers were

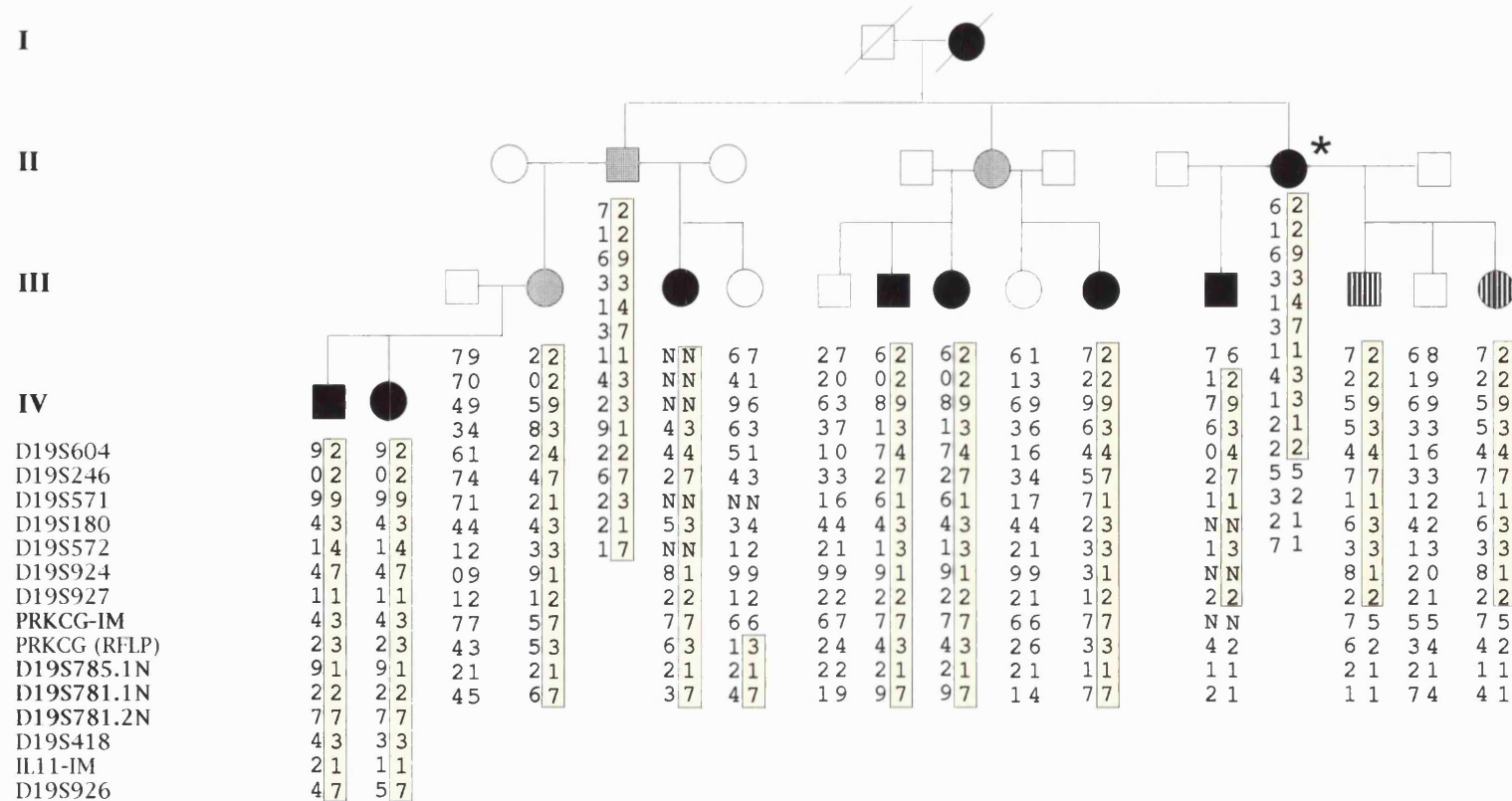
isolated. Haplotypes constructed using these novel markers in ADRP2, ADRP11, RP1907 and ADRP24 are presented in fig. 5.20a, b, c and d, respectively.

5.4.1.7 Genetic analysis of novel microsatellites in ADRP2 for locus refinement

Haplotype data of ADRP2 had revealed a recombination event between the disease and the marker D19S418 in the affected individual II-8, which changed the telomeric flanking marker of the RP11 interval from D19S926 to D19S418 (see section 3.3.2). The genotyping of the microsatellite marker D19S927 and the RFLP marker associated with the *PRKCG* gene, both of which are localised centromeric to D19S418, in ADRP2 further revealed that the recombination event in II-8 had occurred telomeric to the *PRKCG* gene. Since all the novel markers were placed between *PRKCG* and D19S418 there was an opportunity for further locus refinement from the distal end through analysis of novel markers in ADRP2. In this analysis markers PRKCG.IM, D19S785.1N and D19S781.2N proved informative in ADRP2 except for marker D19S781.1N. While PRKCG.IM and D19S785.1N did not recombine with the disease phenotype in ADRP2, D19S781.2N showed recombination with disease in the affected individual II-8 refining the disease locus telomerically and confining the disease interval between D19S572 and D19S781.2N. Haplotypes constructed using these novel markers in ADRP2 are presented in fig. 5.20a.

According to the metric map of LLNL the physical distance between D19S572 and D19S781.2N is ~ 1 Mb. Considering that the physical distance between D19S572 and the previous distal flanking marker D19S418 is ~1.7 Mb a substantial refinement of the locus was achieved with D19S781.2N, thus reducing the region across which a YAC contig has to be established. It also excluded several positional candidate genes that were located at the distal end of the RP11 region such as interleukin-11 (*IL11*), the receptor for Fc fragment of IgA (*FCAR*), natural killer cell class 1 receptor gene family (*NK-R* and *NKRP58*), the slow skeletal isoform of troponin T (*TNNT1*) and isoforms of synaptotagmin (*SYT3* and *SYT5*).

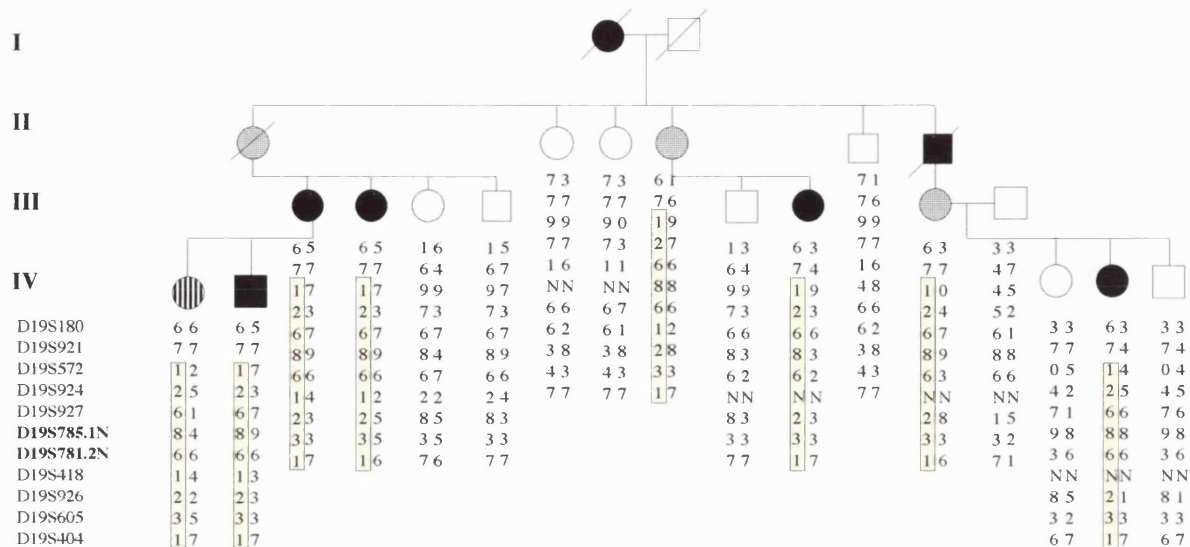
ADRP2



5. 20 a

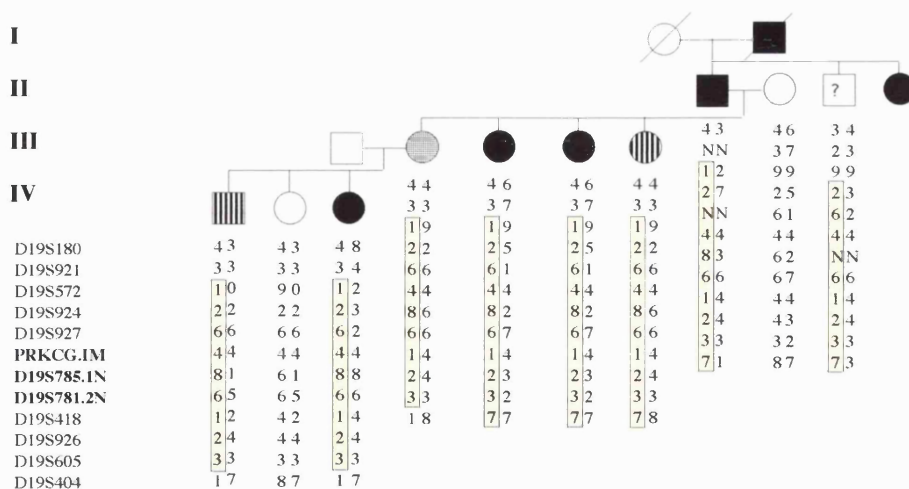
Analysis of PRKCG-IM, D19S785.1N, D19S781.1N and D19S781.2N in ADRP2. Allele segregation in recombinant individuals III-4 and II-8 map all the novel markers centromeric of D19S418. (*) indicates individual II-8 that shows the recombination with D19S781.2N.

RP1907



5.20c

ADRP24



5.20d

Figure 5.20

Analysis of novel markers in pedigrees ADRP2, ADRP11, RP1907 and ADRP24 showing the localisation of the novel markers relative to other markers within the RP11 region. The affected haplotype in each of the four families is boxed in yellow.

5.5 Integration of novel STSs in to the YAC contig

The polymorphic markers and non-polymorphic STSs generated in this study considerably increased the STS density between the markers D19S927 and D19S418, where previously there was a paucity of markers. The novel STSs provided PCR typeable markers, which can be used to isolate further YACs and PACs to consolidate the already existing YAC contig across the RP11 interval.

5.5.1 STS content mapping with the novel markers

Prior to isolating novel YACs by screening the ICI and ICRF YAC libraries with the new STSs, they were first tested on the existing YACs of the contig including the mega-YACs. The markers D19S781.1N, D19S781.2N, D19S781.5N, D19S767.2N, D19S774/STS, D19S268.21N and D19S268.22N were present in YACs 4X5F1 and 4X125A2. (The STS content mapping result obtained for D19S767.2N is presented in figure 5.21). The presence of novel STSs in the critical YACs further verified the genetic localisation and order of these STSs within the RP11 interval. However the marker D19S785.1N that was localised between *PRKCG* and D19S781.1 was not present in YACs 4X5F1 and 4X125A2. With the exception of YAC 955-g-11 this marker was also absent in all other mega-YACs anchored to the distal half of RP11 region. This was not surprising since the 3' end of the gene *PRKCG* was also absent from these YACs (see section 5.3.2.2). Therefore it appears that the majority of YACs localised to RP11 are deleted for the region that include the 3' end of *PRKCG* gene and the marker D19S785.1N. It is possible that this genomic portion gives rise to instability in YAC clones and was deleted during successive propagation of the YAC clone.

The analysis of novel STSs in the mega YAC clones 790-a-5 and 955-g-11 laid bare the highly rearranged nature of these YACs and also defined the boundaries of deletions within these YACs. In YAC 790-a-5 the region bounded by the 3'end of *PRKCG* (exon 10) and D19S781.2N was deleted and in YAC 955-g-11 the deleted region was bounded by the 3'end of *PRKCG* (exon 10) and D19S785.1N. However YAC 955-g-11 also harboured other proximal and distal deletions, which indicated it to be a highly rearranged YAC as suggested in the earlier PFGE analysis

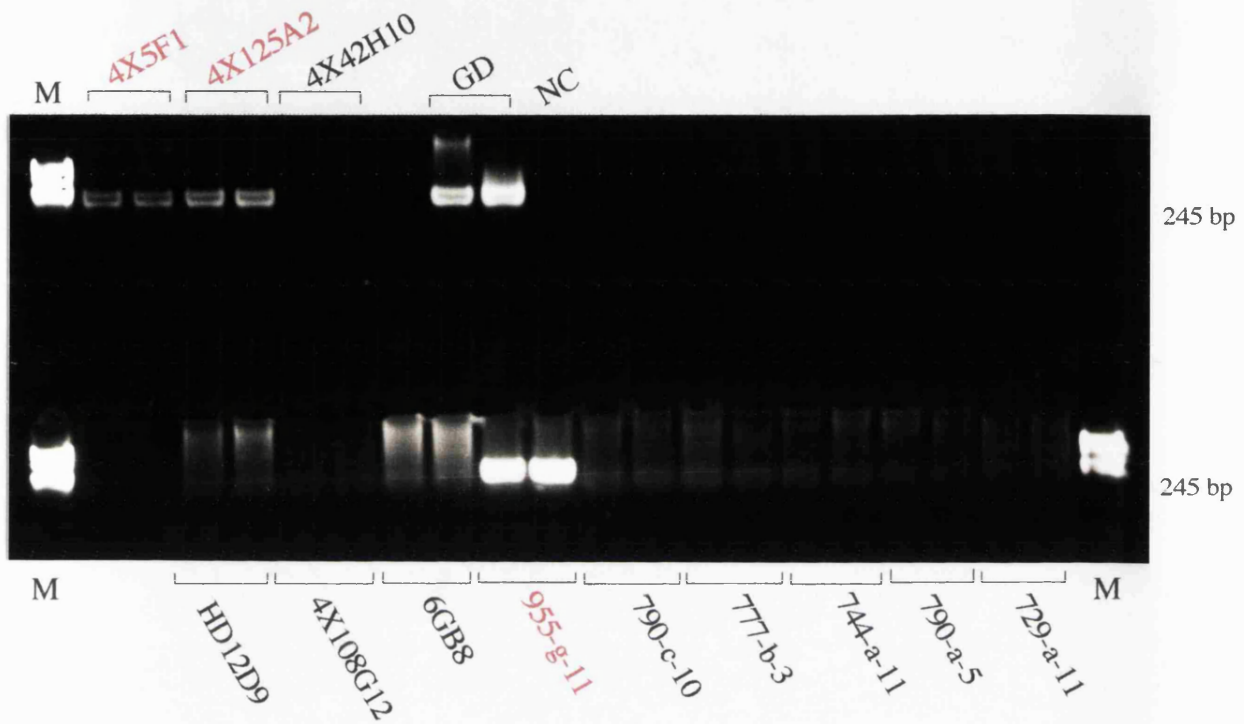


Figure 5.21

Ethidium bromide stained 2% gel depicting the STS content mapping results obtained with the novel STS, D19S767.2N, in a selection of YACs from the RP11 contig. For each YAC the STS PCR was performed in duplicates. M denotes the ϕ X174 size marker, GD the genomic DNA control and NC the negative (water only) control.

The most up to date YAC (and PAC) contig of the RP11 interval is depicted in figure 5.22. In order to isolate YACs to cover the gap between *PRKCG* and D19S785.1N, ICI and ICRF YAC libraries were screened with STSs localised in the vicinity of this gap, which were exons from 3' end of *PRKCG* gene, D19S785.1N, D19S774N/STS and D19S767.2N. Unfortunately this screening was not successful in isolating further novel YAC clones and only identified already known YAC clones, adding weight to the notion that this region either represented an unstable genomic region not well maintained within YACs or a region unclonable in YACs. Therefore focus was next placed on isolating PAC clones with the novel STSs.

5.5.2 Isolation of PAC clones with the new markers to establish a deeper contig across the RP11 interval

Similar to YAC libraries, for facility of use the PAC library (Ioannou and de Jong, 1996) constructed in the vector, pCYPAC2N was also organised in a hierarchical pool system with primary, secondary and tertiary pools for PCR screening. However instead of being dispatched as DNA in agarose plugs (as with the YAC pools) the PAC pools were dispatched as live cultures. (For expediency the screening of the PAC libraries with the novel STSs was done by a colleague within the group). For each screening experiment a negative control (water only) and positive human genomic DNA control for comparison of human and PAC product sizes was included.

In total 10 PAC clones were isolated with the novel STSs localised to the RP11 region (see fig. 5.22). Once isolated the PAC clones were also assayed for markers in the vicinity of the original STS with which they were isolated. Certain PAC clones were found to contain more than one STS. The distribution of STSs within these 'multiple hit' PAC clones also helped determine the orientation of previously genetically indistinguishable markers. For example PAC 53-k-12 defined the order of D19S924 and D19S927 and placed D19S924 centromeric to D19S927. Between the overlapping PAC clones 105-m-21, 282-j-5, 130-l-3 and 15-j-7 the STSs D19S785.1N, D19S774N/STS, D19S767.2N, D19S781.1N, D19S781.2N and D19S781.5N were oriented as written with D19S785.1N lying most centromeric. This order was also in agreement with the LLNL physical map. The PAC library screening also led to the isolation of a PAC clone (275-e-9) that contained the genomic sequence of *PRKCG* gene from exons 3 - 3' UTR, which by far was the

furthermost extending genomic clone of *PRKCG* gene to be isolated (see figure 5.7). STS content mapping of PACs also revealed a single gap that had yet to be bridged between the 3' end of *PRKCG* gene and D19S785.1N for completion of the RP11 contig. End clones of PACs 105-m-21 (that contains D19S785.1N and D19S774N/STS), 275-e-9 (that contains *PRKCG*) can be used in a chromosome walk to bridge this gap between *PRKCG* and D19S785.1N and constitutes future work.

5.6 Genes and ESTs located in the RP11 region

Only two pre-characterised genes (*PRKCG* and *RPS9*) are localised within the RP11 critical interval confined by the markers D19S572 and D19S781.2N. In an effort to place yet more genes within the RP11 interval and thereby facilitate the identification of the RP11 gene, information on partial cDNA sequences or Expressed Sequence Tags (ESTs) of genes that were placed on 19q13.4 by the gene map consortium (Schuler *et al.*, 1996) was accessed through the electronic data base at <http://www.ncbi.nlm.nih.gov/SCIENCE96/>. Scrutiny of this EST database revealed that at least 100 ESTs have been mapped to the 19q13.4 region. However the localisation of these ESTs was not precise and needed to be tested on a physical contig for a more accurate localisation (see 1.5.2). With the YAC/PAC contig as yet incomplete this work was essentially a preliminary exercise. As yet only 8 ESTs have been assayed in the RP11 contig, which were chosen on the basis of their proximity to markers either flanking or within the RP11 interval and expression in the brain or retina (table 5.9). ESTs derived from brain cDNA libraries are worthy of investigation in light of the fact that retina is part of the central nervous system and derived from the neural ectoderm. Following the completion of the contig many more ESTs localised to 19q13.4 region will be ascertained for assaying within the YAC/PAC contig.

Of the 8 ESTs assayed only 3 ESTs (i.e. WI-15638, WI-17997 and SHGC-13495) could be localised within the existing physical contig (fig 5.22). Due to the lapse in the clonal coverage of the RP11 interval (see section 5.5.2) those ESTs that do not map within the existing contig cannot be definitively excluded from the region as yet. They will only be regarded as excluded if found to be negative in the completed YAC/PAC contig. Therefore as mentioned earlier the presented results are essentially preliminary. Nevertheless these ESTs are also STSs and through their localisation can

still provide information on the integrity of the constituent clones within the contig and are useful for that purpose. The following deductions were made on the localisation of the 4 ESTs on the basis of their presence/absence in the YAC/PAC clones of the contig. EST SHGC-13495, which is found only in 790-a-5 and 6GB8, appears to map outside of the RP11 critical interval and was excluded. EST WI-15638 is present in YACs, 955-g-11, 790-a-5, 4X5F1 and 4X125A2 and was placed between D19S268.3N and D19S418. This placement was later proved to be correct when the available DNA sequence of the 4 ESTs were screened for identities against entries within GenBank, whereby EST WI-15638 showed high homology to a gene belonging to the immunoglobulin superfamily. Since it is known that multiple genes that are part of the immunoglobulin superfamily are localised between D19S268 and D19S418 on 19q13.4 (see section 5.2.3, and figure 5.1), EST WI-15638 is also most likely to be located within this gene cluster. According to this localisation EST WI-15638 is excluded from the RP11 critical interval. Finally EST WI-17997, which was present within YACs 729-a-11, 711-a-6 and HD12D9 was tentatively placed between D19S572 and D19S924. This EST requires further testing in the completed contig to confirm its localisation, which if correct would make it the only EST so far to be located within the RP11 interval. Identity screening of WI-17997 did not produce any interesting 'hits' with known genes, unlike the case of WI-15638. The results of the EST mapping is depicted in figure 5.22 and table 5.9 lists the details of the 8 EST clones mapped so far within contig. The work presented above constitutes initiation of future work required for this research and is discussed in more detail in section 5.7.5.

EST clone ID	CDNA Library	Primer sequences 5'-3'	Tm (°C)	Size (bp)	Expression Information
WI- 15638	Infant Brain	F- ttgcttaattaaaaatgtagga R- tggaccaatgaacccc	58	128	Brain, Breast, Liver, Prostate, Uterus
WI- 17997	Lung	F- ataccacagagaatggtgggatg R- cactcagcctgccagagat	55	126	Lung, Aorta, Prostate, Breast, Foreskin, Colon, Germ cell, Uterus
SHGC 30929	Heart Retina	F- ggggtgccaaaccagttg R- aaggcctcaatgacatccac	62	126	Heart, Retina
SHGC-30985	Retina Melanocyte	F- catatcatagccagatctacaacc R- atccagaataatctcccaatgc	56	125	Eye, Foreskin, Tonsil
SHGC-13495	Placenta	F- catgagaattgataggaaggtgg R- cagagcccagccagagag	66	135	Adrenal gland, Breast, Colon, Heart, Kidney, Lung, Prostate, Tonsil, Uterus, Whole embryo.
SHGC-8081	Fetal brain	F- tggacagtgagacatcttccc R- aaaaaactggcctctggtgt	60	165	Adipose, Adrenal gland, Blood, Bone, Brain, Breast, Colon, Foreskin, Germ Cell, Heart, Kidney, Muscle, Ovary, Parathyroid, Prostate, Skin, Spleen, Testis, Uterus
WI-15464	Brain Heart Liver Retina	F- aggccttattggggaaacgt R- acacgctctgagcaggg	57	114	Ubiquitous expression including retina (whole embryo)
A002145	Brain Lung, Retina Liver	F- ctcttctccacatctcag R- ctgtatcatctagacgttata	54	159	Brain, Lung, Retina Liver

Table 5.9.

Characteristics of the ESTs assayed within the RP11 contig, ESTs that have been mapped within the present contig are in bold type.

LEGEND

The identity of YAC clones are listed horizontally while the STSs tested on the clones are listed vertically. YAC clones that are non-chimaeric are labelled in red.

■ STS is present on PCR analysis (YAC)

▣ STS is present on PCR analysis (PAC)

⊠ STS is absent on PCR analysis

Green labelling indicates an EST

Blue labelling indicates a STS generated from an end clone of a YAC

Red labelling indicates a gene

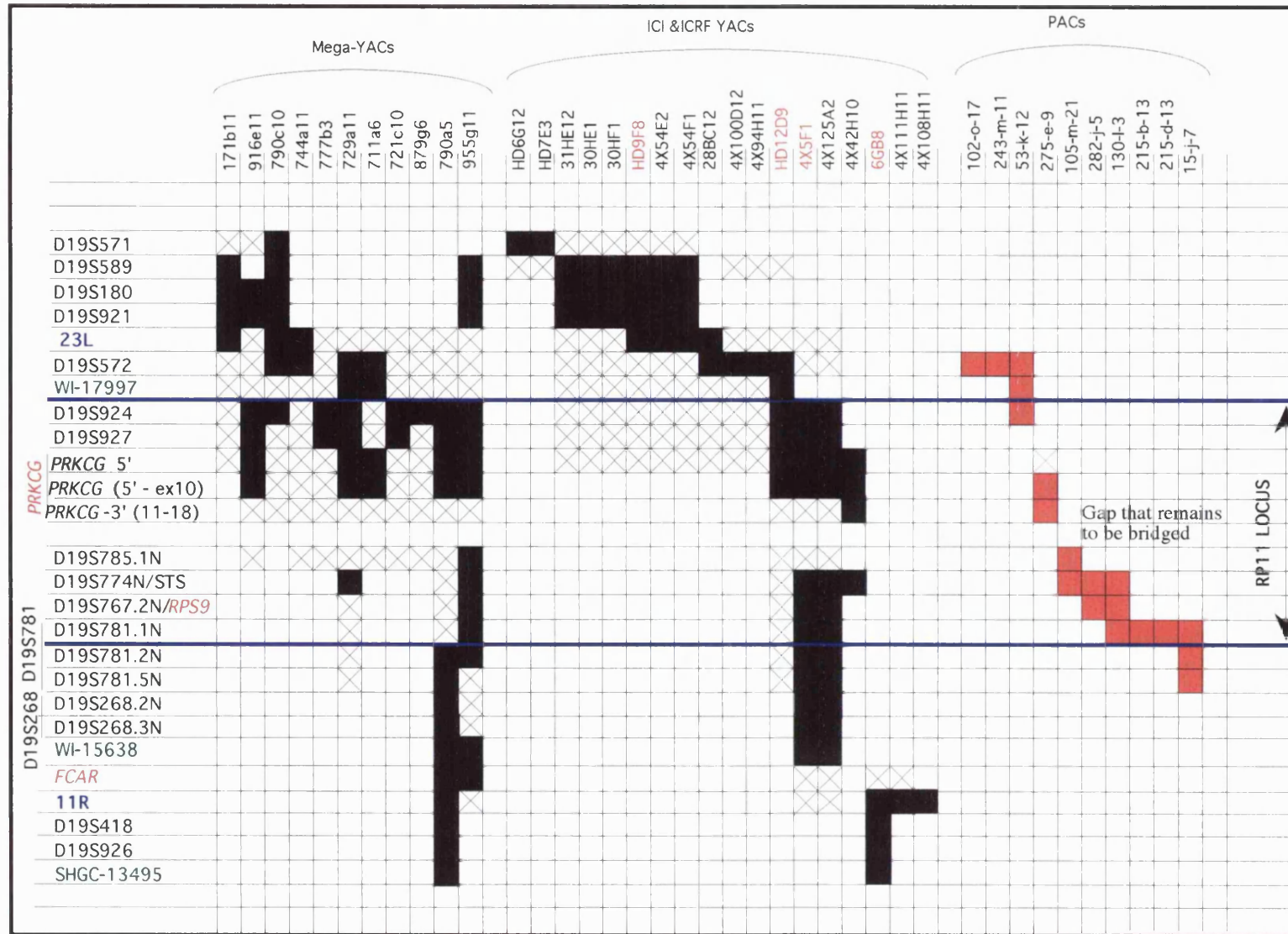


Figure 5.22

The most up to date YAC and (PAC) contig across the RP11 interval. The more detailed legend is on the opposing page.

5.7 Discussion

5.7.1 The establishment of a YAC contig in the RP11 and verification of integrity

Using a selection of STSs obtained either through the chromosome 19 map of CEPH/Genethon (1996) or developed in house with cosmids localised to the RP11 region a YAC and PAC contig was assembled spanning the RP11 interval. The proximal and distal regions of this contig were reinforced with numerous overlapping YACs and PACs indicating a stable region, whereas the middle region near the *PRKCG* gene is incomplete and will require further characterisation. The mega-YAC clones that are anchored to the RP11 interval obtained from pre existing genome maps all proved to be rearranged, deleted or chimaeric on close inspection. Therefore these YACs were of little use apart from providing a crude framework map of the region. The smaller ICI and ICRF YACs provided more reliable coverage of the region and the non-chimaeric YAC clones HD12D9 and 4X5F1 both of which are approximately 200 kb in size contain a majority of STSs from the refined RP11 interval. In fact if it were not for the deletion in 4X5F1, which include the 3' end of *PRKCG* gene and the marker D19S785.1N, the DNA content of these two YAC clones would have completely represented the genomic region of RP11. However, given the deletion in YAC 4X5F1 and the possible unstable nature of YAC HD12D9 (see section 5.3.2.4 and 5.7.3), even though both YACs were non chimaeric, one could not help but harbour reservations about the integrity of the DNA contained within these two YACs. Therefore the coverage provided by YAC clones of the RP11 interval is not as reliable as one would hope for and is a great incentive to consolidate the contig with more stable PAC clones.

Nevertheless according to STS content mapping of all the genomic clones (YACs and PACs) that span the RP11 interval clonal coverage has been achieved of the RP11 interval except for the gap between *PRKCG* and D19S785.1N. In order to bridge this gap STSs need to be designed from the proximal terminal of PAC clone 105-m-21 (that contains D19S785.1N) and the distal terminal of clone 275-e-9 (that contains *PRKCG*), to be utilised in a chromosomal walk to isolate further YAC and PAC clones. This region also contains several cosmid contigs assembled by LLNL, which

however do not overlap one another. If necessary further STSs can be generated from cosmids that were not utilised for (CA)_n isolation from the above mentioned cosmid contigs (see section 5.4.1.1) and these STSs could also be used for further YAC, PAC isolation.

The deletion between *PRKCG* 3'end and D19S785.1N, which most of the YAC clones with inserts extending distally (including the smaller non-chimaeric ICI and ICRF YAC clones) had in common, raises interesting possibilities. The deletion implies that this region is highly unstable, which is not entirely a novel idea as chromosome 19 has been reported to harbour regions of instability more frequently towards the telomere, which may promote chromosomal rearrangements. Stubbs and coworkers (1996) who studied the mouse syntenic region to human chromosome 19q13.3-13.4 identified several rearrangements within the proximal portion of mouse chromosome 7, which included a local inversion and 3 small transpositions of gene rich segments (see section 5.2.3 and figure 5.2). They suggested that instead of being a coincidental occurrence these rearrangements were due to unstable sequences located in the telomeric regions of chromosome 19. The numerous repetitive sequences encountered within 19q13.3-13.4 (Martin-Gallardo *et al.*, 1992) may give rise to this instability suggested by Stubbs and coworkers (1996). These unstable repetitive sequences include the *Alu* sequences, which are found at a higher frequency on chromosome 19 than its predicted occurrence of 1 repeat in every 4 kb of genomic DNA (Korenberg and Rykowski, 1988). It has been estimated that 1.4 *Alu* repeats per kilobase of sequence is present in the 19q13.3 region alone (Martin-Gallardo *et al.*, 1992).

The 19q13.4 region is also rich in gene families, apart from the tandemly clustered Zn finger containing genes for which 19q13.4 is particularly renown, two clusters of gene families have also been identified within the distal portion of the RP11 interval itself. One cluster contains numerous genes encoding proteins belonging to the immunoglobulin superfamily and the other contain two isoforms of synaptogamin (*SYT3* and *SYT5*). It is postulated that clustered gene families have evolved by tandem gene duplication as a result of unequal cross over events or unequal sister chromatid exchange mediated by short interspersed repeats such as *Alu* sequences. The original mispairing of chromatids prior to recombination being facilitated by the high degree

of sequence homology between the non-allelic *Alu* repeats. Therefore it is feasible to suggest that the numerous *Alu* repeats encountered in the 19q13.3-13.4 region were responsible for the initial tandem gene duplications during the evolution of the above mentioned clustered multigene families. If one considers the fact that the same genetic mechanisms as well as producing DNA duplication also produce reciprocal deletions, then it is conceivable that the deletions evident in nearly all of the RP11 YACs resulted due to unequal cross over events (mediated by repetitive elements found in 19q13.4) during the mitotic growth of YACs in yeast.

In recent years a number of microsatellite repeats have also been shown to be associated with *Alu* elements (Beckmann and Weber, 1992). These repeat elements have been postulated to contribute to the genesis of microsatellites due to the presence of middle A-rich and 3' oligo (dA)-rich regions in these elements (Arcot *et al.*, 1995). It therefore seems conceivable that certain cosmids from the 19q13.4 region, which incidentally is rich in *Alu* repeats, should contain up to 2-3 different (CA)_n tracts (see section 5.4.1.4). Therefore in summary it can be postulated that the high incidence of repetitive sequences in the 19q13.3-13.4 region are responsible for the evolution of the clustered gene families, numerous microsatellite repeats and finally the deletions found commonly in YAC containing inserts of this region.

5.7.2 Isolation of novel microsatellites and refinement of the RP11 locus

Using 11 cosmid clones FISH mapped to the RP11 interval four informative microsatellite markers, D19S781.1N, D19S781.2, PRKCG.IM and D19S785.1 were isolated and characterised. With the use of pedigrees ADRP2 and ADRP11 these polymorphic markers were genetically placed within the RP11 interval whilst by assaying for these novel markers within the already existing YAC contig by PCR the physical location of these markers was also confirmed. Markers D19S781.2N and D19S785.1N that show moderate to high polymorphic levels, respectively, are extremely useful, being localised to the 2.8 cM genetic interval between D19S927 and D19S418, which prior to this study was completely absent of (CA)_n markers. The marker PRKCG.IM also localised within the aforementioned interval is located within intron 7 of *PRKCG* gene and show moderate heterozygosity with 6 different alleles segregating in the population. This marker is of immense use as it provides an easily assayed intragenic marker, which obviates the use of the previous *PRKCG*-RFLP

marker (Johnson *et al.*, 1988) that required laborious and time consuming blotting techniques. Along with the 5 non-polymorphic STSs generated in this study the polymorphic markers provide PCR typeable genetic markers that serve as anchor points on the physical map of RP11 interval and have considerably increased the density of markers in this region. The analysis of these novel markers in ADRP2 led to the refinement of the RP11 interval, defining the proximal and distal boundaries of the critical interval as D19S572 and D19S781.2N, respectively. According to LLNL physical mapping data the distance between these two markers approximate 1 Mb, which is a significant refinement of the region when one compares it to the previous RP11 interval of 1.7 Mb between D19S572 and D19S418.

5.7.3 The size of the RP11 interval

According to the LLNL physical map the FISH estimated distance between D19S572 and D19S781.2N is 1 Mb. This distance is significantly greater than the value obtained by adding the total length of the two overlapping YAC clones HD12D9 and 4X5F1, the sum of which only approximates 400 kb. If the size of the RP11 interval estimated by LLNL is accurate and the YAC HD12D9 is an intact YAC without any deletions, the deletion within YAC 4X5F1 should at least be 600 kb. Yet this YAC is only deleted for the region harbouring *PRKCG* and D19S785.1N, the size of which according to LLNL physical map is no greater than 250 kb in size. If this was the case together with the total length of the two overlapping YAC clones HD12D9 and 4X5F1 the size of the RP11 interval should not exceed 650 kb.

This discrepancy between the LLNL FISH estimated distance and the distance as estimated by YAC sizes question the integrity of YAC HD12D9. The genomic interval between D19S572 and *PRKCG* as estimated by the size of YAC HD12D9 that spans this region is 200 kb, which is an underestimation according to LLNL who have estimated it to be 500 kb. Unless the distance between D19S572 and *PRKCG* was overestimated by LLNL, YAC HD12D9 must harbour a deletion, which has not been detected by the STS content mapping to date. PFGE analysis of YAC HD12D9 revealed the presence of two YAC clones (see section 5.3.2.4). It was surmised from the FISH analysis of this YAC, which indicated that this YAC only contained sequences from 19q13.4, that the of smaller of the two YACs was the stable deletion derivative of the larger YAC, which itself was only 200 kb in size. If YAC HD12D9

is unstable, it is possible that the larger YAC clone (200-kb) taken as the unstable but intact YAC clone could also harbour undetectable deletions. Homologous recombination could also have taken place between the original unstable YAC and the deletion derivative such that either YAC resulting from this process would not possess DNA in their original form as a consequence of some sequence loss. Homologous recombination is believed to be responsible for some internal deletions that are known to occur within YACs. For reasons outlined above the size of YAC HD12D9 would not be a reliable basis for inference of distance between genetic markers. However all of the above arguments are purely speculative and STS content mapping has not revealed the suspected deletions in YAC HD12D9, therefore the size discrepancy could simply be a miscalculation in LLNL data.

However due to the aforementioned discrepancies in size estimations between key genetic markers, which as yet cannot be clarified with the existing clonal coverage of the RP11 interval, a precise estimation of the size of the RP11 interval cannot be given as yet. Once continuous coverage was achieved between *PRKCG* and D19S781.5N with PAC clones, which unlike YAC clones are less likely to harbour deletions, insert sizes of those PAC clones that provide a minimal path of the region together with the YAC 4X5F1 may give a more accurate estimate of the size of the RP11 interval. Other approaches that may aid clarification of the precise interval size include the establishment of a genomic PFGE map of the region by restriction mapping of genomic DNA with rare cutting enzymes, followed by successive hybridisation with an ordered array of cosmid, PAC or YAC probes or preferably single copy probes isolated from such clones.

5.7.4 Use of PACs in preference to YACs

The ideal cloning system would produce clones that contain a perfect representation of the original DNA. Close analysis of YAC clones isolated from the RP11 genomic region revealed that most of them were associated with problems that could be encountered during the cloning procedure. While the larger mega-YAC clones were either deleted or chimaeric, some of the smaller ICI and ICRF YAC clones also were deleted or showed evidence of being unstable. Unfortunately these YACs were also the critical YACs that spanned the RP11 interval. Therefore PAC

clones were chosen to consolidate the contig in regions where the representation in YAC clones was either weak or suspect.

Bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs) have been developed as alternative vectors to YACs (Shizuya *et al.*, 1992; Ioannou *et al.*, 1994). Both BAC and PAC possess features from the bacteriophage P1, which has a capacity for DNA fragments as large as 100 kb with an ability to produce a high copy number of clones with reduced chimaerism (Sternberg 1990; Pierce *et al.*, 1992). The PAC vector in particular can handle inserts in the 100-300 kb range and as yet no chimaeras or clone instability have been detected. For these reasons PACs were chosen to consolidate the contig spanning the RP11 interval and to date 10 PAC clones have been isolated and have been analysed by STS content mapping. This PAC contig, which has minimum redundancies, has two gaps one between D19S924 and *PRKCG* and the other between *PRKCG* and D19S785.1N. The PAC library screening with D19S927 did not yield any result, therefore further PAC clones are unlikely to be isolated from the available PAC library to bridge this gap in the contig. However this region is represented in the YAC clones HD12D9 and 4X5F1 and the coverage of the second gap has therefore been given first priority. A chromosome walking strategy initiated with STSs generated from insert termini of the flanking PAC clones *105-m-21* and *275-e-9* can bridge the second gap in the PAC contig between *PRKCG* and D19S785.1N. Many methods have been developed for rapid isolation and sequencing of insert termini from large insert clones (see section 5.2.2). In addition to this, a method that has been termed 'DOP-Vector PCR' (Wu *et al.*, 1996) has been specifically derived to isolate end sequences from PAC clones efficiently. The basic concept of this method is to employ a partially degenerate oligonucleotide primer (DOP) in combination with a vector primer. DOP-Vector PCR will be employed for the isolation of PAC insert termini, which is necessary for closing the existing gap in the contig across RP11 interval.

5.7.5 Future endeavours

The current status of the YAC/PAC contig across the RP11 interval is as depicted in figure 5.22. The consolidation of the physical map of the RP11 interval by isolation of more reliable genomic clones, less prone to chimaerisms or rearrangements, and the chromosome walking strategy to cover existing gaps in the

contig constitute the immediate future work and have been discussed previously. DNA can be easily purified from PAC clones and they are just as suitable for detailed molecular analysis as cosmids. Therefore once the dubious YAC clones are replaced by the more reliable PAC clones it will not be necessary to characterise the contig any further by building a high resolution cosmid map of the region, which has been the method of choice in the detailed analysis of the RP3 genomic region that led to the identification of the *RPGR* gene (Meindel *et al.*, 1996).

Once complete coverage has been achieved the next goal is to map more genes within the RP11 interval to facilitate the identification of the RP11 gene. Even though there are numerous approaches and protocols available for rapid transcript isolation, with the current progress made in the human genome mapping community, it may not be necessary to resort to any of these protocols for the identification of the disease gene. As described in section 5.6 at least 100 uncharacterised ESTs have been mapped to the 19q13.4 region by the gene map consortium (Schuler *et al.*, 1996). To date only 8 ESTs of these 100 or so ESTs have been assayed in the RP11 contig and only 3 have been localised within the existing physical contig, none of which are expressed in the retina or promising as a possible candidate for RP. However several of the ESTs that were not found within any of the YACs or PACs of the present contig do show expression in the retina, such as EST SHGC-30929 and 30985. Therefore once the PAC contig is complete more ESTs would be assayed including those that showed a negative result in the previous assay to see if any retinally or brain-expressed ESTs would localise to the RP11 interval. If expression profiles of the ESTs that map within the RP11 interval were incomplete, northern blot analysis would also be done to obtain the necessary information. Further characterisation of ESTs would involve the determination of exon intron structure of its derivative gene after the complete cDNA was obtained through sequencing of all available partial cDNA clones of the gene. To obtain the complete cDNA sequence it might be necessary to screen a retinal cDNA library or a library made from the tissue in which the gene is abundantly expressed or perform RACE experiments. RACE (rapid amplification of cDNA ends)-PCR, which is a form of RT-PCR that amplifies sequences between a single previously characterised region in the mRNA and either the 5' or the 3' end, is a popular method of obtaining full-length cDNA sequence from low abundance mRNAs that may prove difficult to obtain by conventional cloning

(Frohman *et al.*, 1988). Once the complete cDNA sequence is available the PAC clones that contain the gene would be used to determine the genomic organisation of the gene, which would be obtained in a manner similar to that described for *PRKCG* gene (see chapter 4). After the gene has been characterised it will be subjected to mutation analysis.

In the event of EST mapping effort failing to identify the RP11 gene, it may be necessary to identify retinal specific transcripts or genes highly expressed in the retina. This can be executed through the application of direct selection method utilising the smaller non-chimaeric YAC clones, HD12D9, 4X5F1 and/or the PAC clones on a retinal cDNA library (Monaco, 1994). It is not worthwhile to attempt to search for transcripts in a seemingly gene rich region (according to the gene map) using the technology of exon trapping (Buckler *et al.*, 1991), the detection of CpG islands associated with genes (Bird, 1986), island rescue PCR (Valdes *et al.*, 1994) or other non-specific transcript isolation methods that are more likely to isolate house keeping genes. Instead, selective transcript identification methods such as cDNA selection (Lovett *et al.*, 1991; Parimoo *et al.*, 1991) or direct selection on a retinal cDNA library are more likely to result in the identification of far worthier candidate genes for RP11. Moreover, a retinal cDNA library that has not been subtracted for house keeping genes should be utilised in case that the RP11 gene is not retina specific.

CHAPTER 6

GENERAL DISCUSSION AND FUTURE PROSPECTS

6.1 Overview of the work presented

The adRP locus on chromosome 19q13.4 (RP11) was found by a genome wide linkage analysis on a large British pedigree in our laboratory and since its discovery great progress has been made towards identifying the disease gene. At the start of this project RP11 was confined by markers D19S180 and D19S926 to a genetic interval of approximately 11cM and 3 adRP families had shown linkage to RP11 (ADRP5, ADRP29 and RP1907). Subsequent research identified two additional RP11 families (ADRP2 and ADRP11) and haplotype analysis in these two families confined the disease region to a 5.8 cM interval between D19S921 and D19S418. In order to further refine the locus novel microsatellite markers were isolated from the distal end of the RP interval using cosmid clones FISH mapped to the region between D19S927 and D19S418, and analysed in the pedigree ADRP2, which presented the distal flanking cross over with D19S418. In this endeavour a total of four microsatellite markers and five useful STSs were generated. Fortunately one of the novel markers, D19S781.2N, recombined with the disease phenotype in ADRP2 and superseded D19S418 as the distal flanking marker. At this stage McGee *et al.* (1997) linked three additional families of North American origin to 19q13.4 and subsequent haplotype analysis revealed D19S572 to be the novel proximal flanking marker of the RP11 disease interval. Therefore through the combined genetic data generated both in our laboratory and elsewhere, the RP11 locus was confined between the markers D19S572 (proximal) and D19S781.2N (distal). According to the physical mapping data generated at Lawrence Livermore National Laboratories (LLNL, U.S.A) of chromosome 19, the physical distance between these two markers is 1 Mb, a physical interval relatively easy to traverse by a chromosome walk across the RP11 interval.

However construction of a YAC contig was initiated prior to this refinement of the RP11 interval, consequently YAC coverage is not limited to just the absolute RP11 interval between D19S572 and D19S781.2N but extends approximately 0.5 Mb in both directions and include YAC clones from three different YAC libraries. Interestingly, STS content mapping, which was facilitated by the novel STSs generated, revealed deletions in several YAC clones that were meant to provide coverage across the RP11 interval, thus indicating a region of instability or a region unclonable in YACs. Therefore PAC clones have replaced YACs as the cloning vector of choice and the current on going work of the project involves the construction of a complete PAC contig across the RP11 interval. As yet complete coverage has not been achieved, and there is a gap that still remains to be closed in the DNA contig composed of both YAC and PAC clones.

The gene for protein kinase gamma (*PRKCG*) is localised in the refined RP11 interval and was considered as a positional candidate gene worthy of a thorough investigation. Therefore the genomic organisation of this gene was determined and exon-intron structure was characterised. *PRKCG* gene was found to consist of 18 exons and a genomic span of at least 16 kb. Following its characterisation this gene was subjected to a thorough mutation analysis. The exons, 5' promoter region, and the 3' UTR of this gene were screened in 6 families linked to RP11 locus (ADRP5, ADRP29, ADRP2, RP1907, ADRP24 and a Japanese family). In this analysis a mutation was identified in two adRP families (RP1907 and ADRP24) and two sporadic RP cases, which were subsequently identified as being founded upon a common ancestor and therefore a single mutation event rather than several independent events. This mutation resulted in the non-conservative replacement of a basic arginine with an uncharged serine at codon 659, a codon conserved in human, bovine and rat *PRKCG*. Moreover, this C to A transversion of the first nucleotide in codon 659 was not found in over 500 normal chromosomes indicating it to be a rare event. However the absence of potentially pathogenic mutations in the other four RP11 families screened along with RP1907 and ADRP24 indicated *PRKCG* as not being the sole RP11 gene. The pedigree ADRP11, which was not screened with the other RP11 linked pedigrees mentioned earlier, is now being screened along with another adRP pedigree of Russian origin recently linked to 19q, bringing to a total 8 the number of adRP pedigrees analysed for mutations in *PRKCG*. As yet no mutations

of *PRKCG* have been identified in either of these two families reinforcing the exclusion of *PRKCG* gene further. Currently the 3'UTR of *PRKCG*, the known sequence of which does not as yet extend to the polyadenylation signal, is being sequenced and once complete will also be screened in all of the RP11 families.

Therefore at the end of this research the RP11 disease interval has been refined to a ~1 Mb physical distance, which is traversed by a YAC/PAC contig that is near complete. A candidate gene has been characterised and potentially excluded. The immediate research work involves the completion of the contig followed by mapping of retinally or brain expressed ESTs, already placed in the 19q13.4 region by the EST mapping consortium (Schuler *et al.*, 1996), to the RP11 critical interval to facilitate the identification of the RP11 gene.

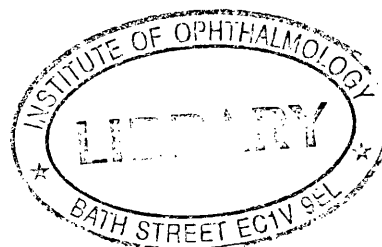
6.2 Allelic effect Vs Modifier gene in RP11: implications for future research

There is statistically significant evidence to suggest that the partial penetrance phenotype characteristic of RP11 is either due to an allelic effect or a mutation in a modifier gene at close proximity to the RP11 gene on 19q13.4 (see chapter 3) (McGee *et al.*, 1997). Modifier genes are genes capable of modifying the manifestations of a mutant gene without having an obvious effect on a normal individual (Gruneberg, 1963) and such genes constitute the 'genetic background' in which the mutant gene finds itself (Romeo and McKusick, 1994). The actions of modifier genes in human disease have been encountered before. For example from studies carried out in a *cftr* mouse model the cystic fibrosis (CF) disease severity in mice has been shown to be modulated by a major modifier gene locus that map near the centromere of mouse chromosome 7 (Rozmahel *et al.*, 1996).

The dominant inheritance pattern of RP11 should make the identification of the primary genetic mutation rather straight forward. If the partial penetrance phenotype is due to only an allelic effect then symptomatic and asymptomatic individuals need to be analysed separately to identify the secondary genetic lesion (within the same gene) that sets the two phenotypic categories apart from each other. This secondary mutation is also likely to be a DNA polymorphism of a reasonable frequency in the normal population. However if the partial penetrance is due to a second mutation in a nearby modifier gene, the identification of this mutation could prove to be very

difficult but is also likely to be facilitated by the identity of the primary gene and its function. Since it is possible that the functions of these two genes are related or even complementary to each other. It is also possible that the two genes are part of one family and therefore functionally related to each other. This is an interesting possibility in light of the fact that 19q13.4 genomic region is particularly rich in gene families (see section 5.2.3 and figure 5.1). Nevertheless to enable the execution of a logical mutation analysis some functional knowledge of the genes in the region is essential. Otherwise all genes in the region would have to be screened for mutations in a panel of asymptomatic and symptomatic RP11 patients to identify the distinguishing mutation(s). This possibility of having to identify not just one but two genes, further accentuate the need to characterise the RP11 interval in even more detail and saturate this region with genes of known function.

Once all the mutations in the RP11 gene have been identified, a series of detailed and comprehensive biochemical analysis of both the wild type and the mutated protein will be instigated, following their expression in an appropriate expression system, in order to understand the functional abnormalities and the underlying pathogenic mechanism associated with each of the different mutations. Such a study will address many fundamental problems associated with retinal degenerations such as whether it is the lack of protein function that leads to degeneration or the loss of protein structure. If the partial penetrance phenotype in RP11 is due to interaction of two gene products, the yeast two-hybrid system can also be utilised to confirm the interaction between these two proteins. The expression of the mutated copies of the gene in heterozygote (+/-) mice will also help us to understand the functional abnormalities associated with the gene mutation that lead to retinal degeneration. If the partial penetrance phenotype is influenced by genetic loci other than the one linked to the primary mutation (i.e. if the genetic background also has some influence) then this can be investigated by breeding the transgenic mice carrying mutations in the RP11 gene into different genetic backgrounds. Similar experiments have been carried out to map genetic modifiers of embryonic lethality in TGF β 1 knockout mice (Bonyadi *et al.*, 1997).



The research that would follow RP11 gene identification would be of utmost interest. Since the identification of the secondary genetic factors responsible for the 'all or nothing' disease phenotype of RP11 could benefit future development of therapies for RP caused by mutations at this locus or may be for RP in general. Hopefully studies into interactions between gene products and their level of involvement in health and disease would provide insight into future RP therapies.

6.3 Progress made in RP genetics and its future

The molecular genetics of retinitis pigmentosa described earlier (chapter 1) gives one some idea of the phenomenal progress that has been made in the field of retinal genetics. Linkage analysis has implicated several loci for RP and through the success of candidate gene approach and mutation screening of good positional candidates several causative genes have also been identified (especially genes responsible for arRP). Most of the RP loci identified by linkage analysis are currently at the 'rate limiting' stage of positional cloning, if there are no attractive candidate genes in the region the researchers have resorted to isolating novel genes. The fact that these endeavours are worthwhile has been proved by some success stories such as the cloning of the RP3 gene (Meindell *et al.*, 1996). Hopefully with the rapid development of the human transcript map most of these genes will also be cloned in the near future. However lack of additional families confirming the original linkage assignments that are predominantly based on single families imply a situation where each RP gene only accounts for a few percent of the cases. Therefore it is likely that many more RP causative genes still remain to be identified indicating a state of genetic heterogeneity for RP similar to that observed in *Drosophila*, where mutations at over 70 loci can cause photoreceptor degeneration or dysfunction (Pak, 1991).

With the identification of genes remarkable allelic heterogeneity among retinal diseases has also been revealed, which in the near future is likely to bring about a complete overhaul of the clinical classification system. Discovery of genes such as *RLBP1* (encoding CRALBP) and *RPE65*, whose expression is in the RPE and Müller cells rather than in the photoreceptor cells, has dismissed the idea that disease causative genes must be expressed in the photoreceptor cells. However with the exception of few genes that have well defined functions (i.e. rhodopsin, peripherin and β PDE) the exact pathogenic mechanisms associated with mutations in a majority

of these newly identified genes or how they bring about photoreceptor death are still unknown. For example, genes such as *RPGR*, *RPE65*, *ABCR* and *TULP1* have as yet unknown or poorly defined functions. Further functional studies of these proteins are required to enable conjectures about pathogenic mechanisms to be drawn, once the role of these proteins in the retina is fully understood they will also provide new insights in to the functioning of the normal retina. The elucidation of pathogenic mechanisms will no doubt be facilitated by future transgenic studies involving these genes just as mouse transgenic studies of rhodopsin, peripherin, β -PDE have unravelled the possible pathogenic mechanisms and a potential cause of photoreceptor death via the apoptosis mechanism.

6.4 Therapeutics in RP research

The ultimate aim of understanding the cause of disease is to provide treatment that can cure or retard the degenerative process of the disease. With RP even though great strides have been made in understanding the molecular cause of disease, the treatment as yet has not progressed beyond identifying pre-symptomatic individuals and providing genetic counselling. However avenues of cure that can be achieved through both non-genetic and genetic intervention are always being sought. As for therapies of non-genetic nature, the use of growth factors, cytokines and neurotrophins that reside in the microenvironment of the photoreceptor cells have been promoted for therapeutic use. For example a great deal of study has been focused upon the basic fibroblast growth factor (β FGF) which has been shown to delay photoreceptor degeneration in the RCS rat and also to protect photoreceptor cells from damaging effects of constant light (Faktorovich *et al.*, 1990). However none of the products have reached the market as therapeutic agents as yet and will only do so once they have been intensely examined for possible toxicity, potential side effects, dosage, half-life and efficient means of administration.

Eye is seen as an attractive target for gene therapy due to ease of accessibility and well-defined anatomy. The ideal end stage of gene therapy protocols would be the ability for the transgene to be functional within its specific cellular environment, in this case within the photoreceptors, while being appropriately regulated by cell specific promoters and regulatory systems. For rescue of recessive mutations the transfer of a normal copy of the gene should theoretically be sufficient to alleviate the

condition. However for it to be successful the gene has to be targeted to the correct cells and preferably integrate into the host genome to ensure long lasting expression driven by the natural promoter and enhancer elements within the cell thus avoiding the need for repeated delivery of the transgene. In the case of dominant disease, gene therapy is more complicated, especially if the gene mutation has a dominant negative effect and the disease is not due to haploinsufficiency. In these situations it has been suggested that a similar mutant gene be transferred in its reverse orientation into the affected cells so as to allow RNA-RNA hybrid arrest to occur, thus removing the mutant protein from expression (anti-sense RNA therapy) (Pepose and Leib, 1994).

Currently ocular gene therapy research is occupied with the testing of various viral derived vector systems for their ability to transfer the therapeutic gene into photoreceptor cells and various means of delivering the vector to the tissue in animal models. Means of administration include intravitreal injections and sub-retinal injections but other radical approaches for administration have also been suggested. The viral derived vectors being tested are the adenovirus (AV), adeno associated virus (AAV) and herpes simplex type 1 (HSV1) based vectors. So far none of these viral vectors have managed to transduce sufficient numbers of photoreceptor cells or produce stable long lasting expression. It has been suggested that modified retroviral vectors which have low immunogenicity and the potential for integration into non-dividing neural cells would be more effective as a vector (Wright, 1997b). Non-viral methods (i.e. liposomes and receptor-mediated endocytosis) for introducing genes into the retina have also been tested. These approaches are attractive as they circumvent some of the safety concerns raised by viral vectors and the risk of an immune response to the introduced vector proteins. However the heterogeneous nature of RP complicates the application of gene therapy. As yet only a few retinal genes responsible for causing retinal degeneration have as yet had the research input for them to be able to partake in gene therapy experiments. These include rhodopsin (*RHO*), peripherin/*RDS* and the cGMP phosphodiesterase (*PDEB*) genes, for which there are a plethora of mutations with known functional consequences and animal models for testing the feasibility of somatic gene therapy. Therefore even though greatly promising as a therapeutic alternative to retinal degeneration, retinal gene therapy is still far from reaching its clinical trial stage and requires more basic research to ensure its safety and efficacy. Also given the large number of mutations

that cause RP, strategies of gene therapy aimed at correcting each individual mutation may be an overwhelming task and may not even be practical.

The final therapeutic solution may lie with the fact that irrespective of the primary genetic lesion photoreceptor cell death is finally mediated by apoptosis. Strategies aimed at introducing a gene that would interfere with the cells' ability to carry out apoptosis is a more practical approach since it could be applicable to multiple mutations. Over expression of *bcl-2*, which prevent apoptosis, in mice with retinal degeneration due to 3 different molecular causes have been already been shown to result in increased photoreceptor survival (Chen *et al.*, 1996). However the therapeutic utility of blocking cell death remains unclear as the question remain whether cells rescued from apoptosis can function. In an effort to answer this question Davidson and Steller (1998) expressed the baculoviral cell survival protein p35 (which antagonises cell death by blocking cysteine proteases known as caspases) in *Drosophila* carrying retinal degeneration mutations in either *ninaE* (encoding rhodopsin 1) or *rdgC* (encoding a putative rhodopsin phosphatase). In both cases retinal degeneration was prevented and behavioural tests on the *ninaE* mutant flies showed that visual function was also preserved. These results provide a strong rationale for further exploration of anti apoptotic strategies in the treatment of RP, especially for rhodopsin mediated forms of RP given the rescue of *ninaE* phenotype in flies, which is equivalent to rhodopsin mutations in humans. It is hoped that by the time gene therapy is ready to be applied in humans as a therapeutic alternative in a more global scale that it would only be used in a strict moral and ethical context.

REFERENCES

- Abeliovich, A., Paylor, R., Chen, C., Kim, J. J., Wehner, J. M., and Tonegawa, S. (1993a) PKC γ mutant mice exhibit mild deficits in spatial and contextual learning. *Cell* **75**, 1263-1271.
- Abeliovich, A., Chen, C., Goda, Y., Silva, A.J., Stevens, C, F. and Tonegawa, S. (1993b) Modified hippocampal long-term potentiation in PKC γ mutant mice. *Cell* **75**, 1253-1262.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B. and Moreno R.F. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**, 1651-1656.
- Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrad, B., Baird, L., Stauffer, D., Peiffer, A., *et al.* (1997a) A photoreceptor cell-specific ATP binding transporter protein gene (ABCR) is mutated in recessive Stargardts macular dystrophy. *Nat. Genet.* **15**, 236-246.
- Allikmets, R., Shroyer, N.F., Singh, N., Seddon, J.K., Lewis, R.A., Bernstein, P., Peiffer, A., Zabriskie, N., Li, Y., Hutchinson, A., *et al.* (1997b) A mutation in the Stargardts disease gene (*ABCR*) in age related macular degeneration. *Science* **277**, 1805-1807.
- Al- Maghtheh, M., Inglehearn, C. F., Keen, T. J., Evans, K., Moore, A.T., Jay, M., Bird, A. C., and Bhattacharya, S. S. (1994) Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. *Hum Mol. Genet.* **3**, 351-354.
- Al- Maghtheh, M., Vithana, E. N., Tartelin, E., Evans, K., Moore, A.T., Jay, M., Inglehearn, C. F. and Bhattacharya, S. S. (1996) Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11), and association with a unique bimodal expressivity phenotype. *Am. J. Hum. Genet.* **59**, 864-871.
- Al- Maghtheh, M., Vithana, E. N., Inglehearn, C. F., Moore, A.T., Bird, A.C. and Bhattacharya, S. S. (1998) Segregation of a *PRKCG* mutation in two RP11 families. *Am. J. Hum. Genet.* **62**, 1248-1252.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Anand, R., Riley, J.H., Butler, R., Smith, J.C. and Markham, A.F. (1990) A 3.5 genome equivalent multi-access YAC library: construction, characterisation, screening and storage. *Nucleic Acids Research* **18**, 1951-1956.

- Anderson, D.H. and Fisher, S.K. (1976) The photoreceptor of diurnal squirrels: outer segment disc shedding and protein renewal. *J. Ultrastruct. Res.* **55**, 119. (Bok paper 35)
- Arcot, S.S., Wang, Z., Weber, J.L., Deininger, P.L. and Batzer, M.A. (1995) Alu repeats: a source for the genesis of primate microsatellites. *Genomics* **29**, 136-144.
- Ashworth, L.K. and 13 others (1995) An integrated metric physical map of human chromosome 19. *Nat. Genet.* **11**, 422-427.
- Attwood, J. and Bryant, S. (1988) A computer program to make analysis with LIPED and LINKAGE easier and less prone to errors. *Ann. Hum. Genet.* **52**, 259.
- Azarian, S.M. and Travis, G.H. (1997) The photoreceptor rim protein is an ABC transporter encoded by the gene for recessive Stargardt's disease (ABCR). *FEBS Lett* **409**, 247-252.
- Baer, W., Devlin, M. J. and Applebury, M.L. (1979) Isolation and characterisation of cGMP phosphodiesterase from bovine rod outer segments. *J. Biol. Chem.* **245**, 11669-11677.
- Baer, W., Morita, E.A., Swanson, R. J. and Applebury, M.L. (1982) Characterisation of bovine rod outer segment G-protein. *J. Biol. Chem.* **257**, 6452-6460.
- Ballabio, A. (1993) The rise and fall of positional cloning. *Nat. Genet.* **3**, 277-279.
- Balciuniene, J., Johanssen, K., Sandgren, O., Wachmeister, L., Holmgren, G. and Forsman, G. (1995) A gene for autosomal dominant progressive cone dystrophy (CORD5) maps to chromosome 17p12-p13. *Genomics* **30**, 281-286.
- Banerjee, P., Kleyn, P.W., Knowles, J.A., Lewis, C.A., Ross, B.M., Parano, E., Kovats, S.G., Lee, J.J., Penchaszadeh, G.K., Ott, J., Jacobson, S.G. and Gilliam, T.C. (1998) TULP1 mutation in two extended Dominican kindreds with autosomal recessive retinitis pigmentosa. *Nat. Genet.* **18**, 177-179.
- Banfi, S., Borsani, G., Ross, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O. and Ballabio, A. (1996) Identification of human cDNAs homologous to *Drosophila* mutant genes through EST database searching. *Nat. Genet.* **13**, 167-174.
- Bardien, S., Ebenezer, N., Greenberg, J., Inglehearn, C.F., Bartmann, L., Goliath, R., Beighton, P. and others (1995) An eighth locus for autosomal dominant retinitis pigmentosa is linked to chromosome 17q. *Hum. Mol. Genet.* **4**, 1459-1462.
- Bascom, R.A., Manara, S., Collins, L., Molday, R.S., Kalnins, V.I. and McInnes, R.R. (1992) Cloning of the cDNA for a novel photoreceptor membrane protein (rom-1) identifies a disk rim protein family implicated in human retinopathies. *Neuron* **8**, 1171-1184.

- Bassett, D.E., Boguski, M.S., Spence, F., Reeves, R., Kim, S., Weaver, T. and Hieter, P. (1997) Genome cross-referencing and XREFdb: Implications for the identification and analysis of genes mutated in human disease. *Nat. Genet.* **15**, 339-344.
- Bauer, P.H., Muller, S., Puzicha, M., Pippig, S., Obermaier, B., Helmreich, E.J. and Lohse, M.J. (1992) Phosducin is a protein kinase A-regulated G-protein regulator. *Nature* **358**, 73-76.
- Bayes, M., Goldaracena, B., Martinez-Mir, A., Iragui-Madoz, M.I., Solans, T., Chivelet, P., Bussaglia, E., Ramos-Arroyo, M.A., Baiget, M., Vilageliu, L., Balcells, S., Gonzalez-Duarte, R. and Grinberg, D. (1998) A new autosomal recessive retinitis pigmentosa locus maps on chromosome 2q31-q33. *J. Med. Genet.* **35**, 141-145.
- Baylor, D.A., Nunn, B.J. and Schnapf, J.L. (1984) The photocurrent, noise and spectral sensitivity of rods of the monkey *Macaca fascicularis*. *J. Physiol.* **357**, 575-607.
- Baylor, D. (1996) How photons start vision. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 560-565.
- Bech-Hansen, N.T., Moore, B.J., Pearce, W.G. (1992) Mapping of locus for X linked congenital stationary night blindness (CSNB1) proximal to DXS7. *Genomics* **12**, 409-411.
- Beckmann, J. S. and Weber, J. L. (1992) Survey of human and rat microsatellites. *Genomics* **12**, 627-631.
- Bellanné-Chantelot, C. and 21 others (1992) Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70**, 1059-1068.
- Bellingham, J., Gregory-Evans, K. and Gregory-Evans, C. (1997) A polymorphic dinucleotide repeat in the 5' flanking region of the human interleukin 11 (*IL11*) gene. *Immunogenetics* (in press).
- Bender, W., Sperer, P. and Hogness, D.S. (1983) Chromosome walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the *bithorax* complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168**, 17-33.
- Berg, M.A., Guevara-Aguirre, J., Rosenbloom, A.L., Rosenfield, R.G. and Francke, U. (1992) Mutation creating a new splice site in the growth hormone receptor genes of 37 Ecuadorean patients with Laron syndrome. *Hum. Mut.* **1**, 24-34.
- Bhattacharya, S.S., Wright, A.F., Clayton, J.F., Price, W.H., Phillips, C.L., McKeown, C.M., Bird, C.M., Pearson, R.L., Southern, E.M. and Evans, H.J. (1984) Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe LI.28. *Nature* **309**: 253-255.
- Bird, A.C. (1988) Clinical investigations of retinitis pigmentosa. *Br. J. Ophthalmol.* **59**, 177-199.

- Bird, A.C. (1995) Retinal photoreceptor dystrophies. *Am. J. Ophthalmol.* **119**, 543-562.
- Bird, A.P. (1986) CpG-islands and the function of DNA methylation. *Nature* **321**, 209-213.
- Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids. Res.* **7**, 1513-1523.
- Blanton, S.H., Heckenlively, J.R., Cottingham, A.W., Friedman, J., Sadler, L.A., Wagner, M., Friedman, L.H. and Daiger, S.P. (1991) Linkage mapping of autosomal dominant retinitis pigmentosa (RP1) to the pericentric region of human chromosome 8. *Genomics* **11**, 857-869.
- Blattner, F and 16 others (1996) The *E. coli* Genome project. *Microb. Comp. Genome* **1**, 357.
- Boehnke, M., Lange, K. and Cox, D.R. (1991). Statistical methods for multipoint radiation hybrid mapping. *Am. J. Hum. Genet.* **49**: 1174-1188.
- Boguski, M.S. and Schuler, G.D. (1995) ESTablishing the human transcript map. *Nat. Genet.* **10**, 369-371.
- Bok, D. (1985) Retinal photoreceptor-pigment epithelium interactions. *Invest. Ophthalmol. Vis. Sci.* **26**, 1659-1694.
- Bonyadi, M., Rusholme, S.A, Cousins, F.M., Su, H.C., Biron, C.A., Farrall, M. and Akhurst, R.J. (1997) Mapping of a major genetic modifier of embryonic lethality in TGF beta 1 knockout mice. *Nat. Genet.* **15**, 207-211.
- Botstein, D., White, R.L., Skolnick, M. and Davies, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314-331
- Boughman, J.A., Confally, P.M. and Nance, W.E. (1980) Population genetic studies of retinitis pigmentosa. *Am. J. Hum. Genet.* **32**, 223-235.
- Bourne, H. R., Sanders, D. A. and McCormic, F. (1990) The GTPase superfamily: a conserved switch for diverse cell function. *Nature* **348**, 125-132.
- Bowers, B. J., Parham, C. L., Sikela, J. M. and Wehner, J. M. (1993). Isolation and sequence of a mouse brain cDNA coding for protein kinase C-gamma isozyme. *Gene* **123**: 263-265.
- Bowes, C., Li, T., Danciger, M., Baxter, L.C., Applebury, M.L. and Farber, D.B. (1990) Retinal degeneration in the rd mouse is caused by a defect in the β subunit of rod cGMP-phosphodiesterase. *Nature* **347**, 677-680.
- Brandriff, B.F., Gordon, L.A. and Trask, B.J. (1991) A new system for high-resolution DNA sequence mapping in interphase pronuclei. *Genomics* **10**, 75-82.

- Brandriff, B.F., Gordon, L.A., Fertitta, A., Olsen, A.S., Christensen, M., Ashworth, L.K., Nelson, D.O., Carrano, A.V. and Mohrenweiser, H.W. (1994) Human chromosome 19p: A fluorescent in situ hybridisation map with genomic distance estimates for 79 intervals spanning 20 Mb. *Genomics* **23**, 582-591.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* **75**, 4953-4957.
- Brilliant, M.H., Williams, R.W., Conti, C.J., Angel, J.M., Oakey, R.J and Holdener, B.C. (1994) Mouse chromosome 7. *Mamm. Genome*. **5** Spec No, S104-S123.
- Buckler, A.J., Chang, D.D., Graw, S.L., Brook, D., Haber, D.A., Sharp, P.A. and Houseman, D.E. (1991) Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**, 4005-4009.
- Buetow, K. H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Duyk, G.M., Sheffield, V.C., Wang, Z. and Murray, J.C. (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. *Nat. Genet. ics* **6**, 391-393.
- Bundy, S. and Crews, S.J. (1984) A study of retinitis pigmentosa in the city of Birmingham. II. Clinical and genetic heterogeneity. *J. Med. Genet.* **21**, 421-428.
- Bunker, C.H., Berson, E.L., Bromley, W.C., Hayes, R.P. and Roderick, T.H. (1984) Prevalence of retinitis pigmentosa in Maine. *Am. J. Ophthalmol.* **97**, 357-365.
- Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-813.
- Burmeister, M., Monaco, A.P., Gillard, E.F., Van Ommen, G.B., Affara, N.A., Ferguson-Smith, M.A., Kunkel, L.M. and Lehrach, H. (1988) A 10-megabase physical map of human Xp21, including the Duchenne muscular dystrophy gene. *Genomics* **2**, 189-202.
- Cachon-Gonzales, M.B., Delhanty, J.D., Burn, J., Tsioupra, K., Davis, M.B., Attwood, J. and Chapman, P. (1991) Linkage analysis in adenomatous polyposis coli: the use of four closely linked DNA probes in 20 UK families (see comments). *J. Med. Genet.* **28**, 681-685.
- Carle, G.F. and Olson, M.V. (1984) Separation of chromosomal DNA molecules from yeast by orthogonal-field-alteration gel electrophoresis. *Nuc. Acids. Res.* **12**, 5647-5664.
- Carmi, R., Rokhlina, T., Kwitek-Black, A., Elbedour, K., Nishimura, D., Stone, E.M. and Sheffield, V.C (1995) Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. *Hum Mol. Genet.* **4**, 9-13.
- Carulli, J. P. and Hartl, D. L. (1992) Variable rates of evolution among *Drosophila* opsin genes. *Genetics* **132**, 193-204.

- Chang, G.Q., Hao, Y. and Wong, F. (1993) Apoptosis: final common pathway of photoreceptor death in rd, rds and rhodopsin mutant mice. *Neuron* **11**, 595-605.
- Chen, C.K., Inglese, J., Lefkowitz, R.J. and Hurley, J.B. (1995a) Ca (2+)-dependent interaction of recoverin with rhodopsin kinase. *J. Biol. Chem.* **270**, 18060-18066.
- Chen, C., Kano, M., Abeliovich, A., Chen, L., Bao, S., Kim, J. J., Hashimoto, K. (1995b) Impaired motor co-ordination correlates with persistent multiple climbing fibre innervation in PKC γ mutant mice. *Cell* **83**,1233-1242.
- Chen, J., Flannery, J.G., LaVail, M.M., Steinberg, R.H., Xu, J. and Simon, M.I. (1996) bcl-2 overexpression reduces apoptotic photoreceptor cell death in three different retinal degenerations. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7042-7047.
- Chen, K. H., Widen, S., Wilson, S. G., and Huang, K. P. (1990) Characterisation of the 5'flanking region of the rat protein kinase C γ gene. *J. Biol. Chem.* **265**, 19961-19965.
- Chen, T. Y., Peng, Y. W., Dhallan, R. S., Ahamed, B., Reed, R.R. and Yau, K.W. (1993) A new subunit of the cyclic nucleotide-gated cation channel in retinal rods. *Nature* **362**, 764-767.
- Chumakov, I. and 35 others. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380-387.
- Chumakov, I. and 60 others. (1995) A YAC contig map of the human genome. *Nature* **377**, 175-297S
- Coffey, A.J., Roberts, R.G., Green, E.D., Cole, C.G., Butler, R., Anand, R., Giannelli, F. and Bently, D.R. (1992) Construction of a 2.6 Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. *Genomics* **12**, 474-484.
- Cohen, A.I. (1983) Some cytological and initial biochemical observations on photoreceptors in retina of rds mice. *Invest. Ophthamol. Vis. Sci.* **24**, 9677-9688.
- Coleman, D.L. and Eicher, E.M. (1990) Fat (fat) and tubby (tub): two autosomal recessive mutations causing obesity syndrome in the mouse. *J. Hered.* **81**, 424-427.
- Colley, N., Cassill, J., Baker, E. and Zuker, C. (1995) Defective intracellular transport is the molecular basis of rhodopsin-dependent retinal degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3070-3074.
- Collins, F.S. (1995) Positional cloning moves from perditional to traditional. *Nat. Genet. ics* **4**, 347-350.
- Collins, R.J., Harmon, B.V., Gobe, G.C. and Kerr, J.F. (1992) Internucleosomal DNA cleavage should not be the sole criterion for identifying apoptosis. *Int. J. Radiat. Biol.* **61**, 451-453.

- Connell, G.J. and Molday, R.S. (1990) Molecular cloning, primary structure, and orientation of the vertebrate photoreceptor cell protein peripherin in the rod outer segment disc membrane. *Biochemistry*. **29**, 4691-4698.
- Cook, N. J., Molday, L. L., Reid, D., Kaupp, U.B. and Molday, R. S. (1989) The cGMP-gated channel of bovine rod photoreceptors is localised exclusively in the plasma membrane. *J. Biol. Chem.* **264**, 6996-6999.
- Coulson, A., Sulston, J., Brenner, S. and Karn, L. (1986) Toward a physical map of the genome of nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **83**, 7821-7825.
- Coussens, L., Parker, P. J., Rhee, L., Yang-Feng, T. L., Chen, E., Waterfield, M. D., Francke, U. and Ullrich, A. (1986) Multiple distinct forms of bovine and human protein kinase C suggest diversity in cellular signalling pathways. *Science* **233**, 859-866.
- Cox, D.R., Burmeister, M., Price, E.R., Kim, S. and Myers, R.M. (1990) Radiation hybrid mapping: a somatic cell genetics method for constructing high resolution maps of mammalian chromosomes. *Science* **250**, 245-250.
- Craig, J. M. and Bickmore, W.A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.* **7**, 376-381.
- Cremers, F.P., van de Pol, D.J., van Driel, M., den Hollander, A.I., van Haren, F.J., Knoers, N.V., Tijmes, N., Bergen, A.A., Rohrschneider, K., Blankenagel, A., Pinckers, A.J., Deutman, A.F. and Hoyng, C.B. (1998) Autosomal recessive retinitis pigmentosa and cone-rod dystrophy caused by splice site mutations in the Stargardt's disease gene ABCR. *Hum. Mol. Genet.* **3**, 355-362.
- Crouch, R.K., Goletz, P., Yu, S. and Redmond, T.M. (1997) A possible role for RPE65 in retinoid processing. *Invest. Ophthalmol. Vis. Sci.* **38**, 1422.
- Cuenca, N., Fernandez, E. and Kolb, H. (1990) Distribution of immunoreactivity to protein kinase C in the turtle retina. *Brain Res.* **532**, 278-287.
- Dacey D.M. and Lee, B.B. (1994) The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature* **367**, 731-735.
- Daiger, S.P., Sullivan, L.S. and Rodriguez, J.A. (1995) Correlation of phenotype with genotype in inherited retinal degeneration. *Behav. Brain Sci.* **18**, 452-467.
- Danciger, M., Bowes, C., Kozak, C., LaVail, M.M. and Farber, D.B. (1990) Fine mapping of a putative *rd* cDNA and its co-segregation with *rd* expression. *Invest. Ophthalmol. Vis. Sci.* **31**, 1427-1432.
- Davidson, F.F. and Stellar, H. (1998) Blocking apoptosis prevents blindness in *Drosophila* retinal degeneration mutants. *Nature* **391**, 587-591.

- Defoe, D.M. and Bok, D. (1983) Rhodopsin chromophore exchanges among opsin molecules in the dark. *Invest. Ophthalmol. Vis. Sci.* **24**, 1211-1226.
- DeRouck, A., de Bie, S. and Kayembe, D. (1986) Statistical evaluation of visual functions in dominant and recessive autosomal pigmentary retinopathy. *Doc. Ophthalmol.* **62**, 265-280.
- Deterre, P., Bigay, J., Forquet, F., Robert, M. and Chabre, M. (1988) cGMP phosphodiesterase of retinal rods is regulated by two inhibitory subunits. *Proc. Natl. Acad. Sci. USA* **85**, 2424-2428.
- Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-154.
- Dizhoor, A.M., ray, S., Kumar, S., Niemi, G., Spencer, M., Brolley, D., Walsh, K.A., Philipov, P.P., Hurley, J.B. and Stryer, L. (1991) Recoverin: a calcium sensitive activator of retinal rod guanylyl cyclase. *Science* **251**, 915-918.
- Dizhoor, A.M., Lowe, D.G., Olshevskaya, E.V., Laura, R.P. and Hurley, J.B. (1994) The human photoreceptor membrane guanylyl cyclase, retGC, is present in outer segments and is regulated by calcium and a soluble activator. *Neuron* **12**, 1345-1352.
- Dizhoor, A.M., Olshevskaya, E.V., Henzel, W.J., Wong, S.C., Stults, J. T., Ankoudinova, I. and Hurley, J.B. (1995) Cloning, sequencing, and expression of a 24-kDa Ca(2+)-binding protein activating photoreceptor guanylyl cyclase. *J. Biol. Chem.* **270**, 25200-25206.
- Dotd, G., Braveman, N., Valle, D. and Gould, S.J. (1996) From expressed sequence tags to peroxisome biogenesis disorder genes. *Proc. NY. Acad. Sci.* **804**, 516-523.
- Donis-Keller, H and 32 others (1987) A genetic linkage map of the human genome. *Cell* **51**, 319-337.
- Dowdy, S.F., Scanlon, D.J., Fasching, C.L., Casey, G. and Stanbridge, E.J. (1990) Irradiation microcell-mediated chromosome transfer (XMMCT): the generation of specific chromosomal arm deletions. *Genes Chromosome Cancer* **2**, 318-327
- Dowling, J.E. (1960) Chemistry of visual adaptation in the rat. *Exp. Eye Res.* **3**, 348-356.
- Downing, J.P. and Zimmerman, A.L. (1995) *J. Physiol.* **486**, 533-546.
- Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A. and Berson, E.L. (1990) A point mutation of rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364-366.

- Dryja, T.P., Berson, E.L., Rao, V.R. and Oprian, D.D. (1993) Heterozygous missense mutation in the rhodopsin gene as a cause of congenital stationary night blindness. *Nat. Genet.* **4**, 280-283.
- Dryja, T.P., Finn, J.T., Peng, Y-W., McGee, T.L., Berson, E.L. and Yau, K-W. (1995) Mutations in the gene encoding the α subunit of the rod cGMP-gated channel in autosomal recessive retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10177-10181.
- Dryja, T.P., Hahn, L.B., Reboul, T. and Arnaud, B. (1996) Missense mutation in the gene encoding the α subunit of rod transducin in the Nougaret form of congenital stationary night blindness. *Nat. Genet.* **13**, 358-360.
- Dupont, B., Selvakumar, A. and Steffens, U (1997) The killer cell inhibitory receptor genomic region on human chromosome 19q13.4. *Tissue Antigens* **49** (6), 557-563.
- D'Urso, M., Zucchi, I., Ciccodicola, A., Palmieri, G., Abidi, F.E. and Schlessinger, D. (1990) Human glucose 6-phosphate dehydrogenase gene carried on a yeast artificial chromosome encodes active enzyme in monkey cells. *Genomics* **7**, 531-534.
- Duyk, G.M., Kim, S.W., Myers, R.M. and Cox, D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. U.S.A* **87**, 8995-8999.
- Evans, G.A. and Lewis, K.A. (1989) Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Natl. Acad. Sci. USA* **86**, 5030-5034.
- Evans, K., Fryer, A., Inglehearn, C. F., Duvall-Young, J., Whittaker, J., Gregory, C.Y., Ebenezer, N., Hunt, D. and Bhattacharya, S.S. (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence of segregation distortion. *Nat. Genet.* **6**, 210-213.
- Evans, K., Al-Maghteh, M., Fitzke, F., Moore, A.T., Jay, M., Inglehearn, C. F., Arden, G.B. and Bird, A.C. (1995) Bimodal expressivity in dominant retinitis pigmentosa genetically linked to chromosome 19q. *Br. J. Ophthalmol.* **79**, 841-846.
- Evans, K., Duvall-Young, J., Fitzke, F. W., Arden, G. B., Bhattacharya, S. S. and Bird, A.C. (1995) Chromosome 19q cone-rod retinal dystrophy: ocular phenotype. *Arch. Ophthalmol.* **113**, 195-1102
- Fain, G.L. and Lisman, J.E. (1993) Photoreceptor degeneration in vitamin A deprivation and retinitis pigmentosa: The equivalent light hypothesis. *Exp. Eye Res.* **57**, 335-340.
- Faktorovich, E.G., Steinberg, R.H., Yasumura, D., Matthes, M.T. and LaVail, M.M. (1990) Photoreceptor degeneration in inherited retinal dystrophy delayed by basic fibroblast growth factor. *Nature* **347**, 83-86.

- Farber, D.B., Danciger, J.S. and Aguirre, G. (1992) The beta subunit of cyclic GMP phosphodiesterase mRNA is deficient in canine rod-cone dysplasia 1. *Neuron* **9**, 349-356.
- Farber, D.B. (1995) From mice to men: the cyclic GMP phosphodiesterase gene in vision and disease. *Invest. Ophthalmol. Vis. Sci.* **36**, 263-275.
- Farrar, G.J., Kenna, P., Jordan, S.A., Kumar-Singh, R., Humphries, M.M., Sharp, E.M., Shiels, D.M. and Humphries, P. (1991) A three base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature* **354**, 478-480.
- Fernández, E., Cuenca, N., Garcia, M. and De Juan J. (1995) Two types of mitochondria are evidenced by protein kinase C Müller cells of the carp retina. *Neurosci. Letts.* **183** 202-205.
- Fesenko, E.E., Kolesnikov, S.S. and Lyubarsky, A.L. (1985) Induction by cGMP of cationic conductance in plasma membrane of retinal rod outer segments. *Nature* **313**, 310-313.
- Finckh, U., Xu, S., Kumaramanickavel, G., Schrmann, M., Mukkadan, J.K., Fernandez, S.T., John, S., Weber, J.L., Denton, M.J. and Gal, A. (1998) Homozygosity mapping of autosomal recessive retinitis pigmentosa locus (RP22) on chromosome 16p12.1-p12.3. *Genomics* **48**, 341-345.
- Fishman, G.A. (1978) Retinitis pigmentosa: Visual loss. *Arch. Ophthalmol.* **96**, 1185-1188.
- Fishman, G.A., Alexander, K.R. and Anderson, R.J. (1985) Autosomal dominant retinitis pigmentosa. A method of classification. *Arch. Ophthalmol.* **103**, 366-374.
- Franke, R.R., Sakmar, T.P., Graham, R.M. and Kohara, H.G. (1992) Structure and function in rhodopsin: Studies of the interaction between the rhodopsin in cytoplasmic domain and transducin. *J. Biol. Chem.* **267**, 14767-14774.
- Freund, C. L., and 16 others (1997) Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91**, 543-553.
- Freund, C. L., Wang, Q-L., Chen, S., Muskat, B.L., Wiles, C.D., Sheffield, V.C., Jacobson, S.G., McInnes, R.R., Zack, D.J and Stone, E.M. (1998) De novo mutations in the CRX homeobox gene associated with Leber congenital amaurosis. *Nat. Genet.* **18**, 311-312.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8998-9002.
- Fuchs, S., Nakazawa, M., Maw, M., Tamai, M., Oguchi, Y. and Gal, A. (1995) A homozygous 1-base pair deletion in the arrestin gene is a frequent cause of Oguchi disease in Japanese. *Nat. Genet.* **10**, 360-362. Fujisawa, N., Ogita, K., Saito, N., and

- Nishizuka, Y. (1992) Expression of protein kinase C subspecies in rat retina. *FEBS Letters* **3**, 409-412.
- Fulton, T.R., Bowcock, A.M., Smith, D.R., Daneshvar, L., Green, P., Cavalli-Sforza, L.L. and Donis-Keller, H. (1989) A 12 megabase restriction map at the cystic fibrosis locus. *Nucleic Acids Res.* **17**, 271-284.
- Fung, B. K-K. (1987) Transducin: Structure, function and role in phototransduction. *Prog. Retinal Res.* **6**, 151-177.
- Fung, B. K-K., Young, J.F., Yamane, H. K. and Griswol-Prenner, I. (1990) Subunit stoichiometry of retinal rod cGMP phosphodiesterase. *Biochemistry* **29**, 2657-2664.
- Furukawa, T., Morrow, E. M. and Cepko, C. L. (1997) *Crx*, a novel *otx*-like homeobox gene, shows photoreceptor-specific expression and regulates photoreceptor differentiation. *Cell* **91**, 531-541
- Gal, A., Orth, U., Baehr, W., Schwinger, E. and Rosenberg, T. (1994) Heterozygous missense mutation in the rod cGMP phosphodiesterase β -subunit gene in autosomal dominant stationary night blindness. *Nat. Genet.* **7**, 64-68.
- Gal, A., Apfelstedt-Sylla, E., Janecke, A. R. and Zenner, E. (1997) Rhodopsin mutations in inherited retinal dystrophies and dysfunctions. *Prog. Ret. Eye Res.* **16**, 51-79.
- Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H. and Bruns, G.A.P. (1990) Homozygous deletion in Wilms tumours of a zinc finger gene identified by chromosome jumping. *Nature* **343**, 774-778.
- Ghalayini, F., Koutz, C.A., Wetsel, W.C., Hannun, Y.A. and Anderson, R.E. (1994) Immunolocalisation of PKC ζ in rat photoreceptor inner segments. *Curr. Eye Res.* **13**, 145-150.
- Gilbert, W. (1978) Why genes in pieces? *Nature* **271**, 501.
- Gilbert, W. (1987). The exon theory of genes. *Cold Spring Harbour Symp. Quant. Biol.* **52**, 901-905.
- Gillett, G.T., McConville, C.M., Byrd, P.J., Stankovic, T., Taylor, A.M., Hunt, D.M., West, L.F., Fox, M.F., Povey, S. and Benham, F.J. (1993) Irradiation hybrids for human chromosome 11: characterisation and use for generating region specific markers in 11q14-q23. *Genomics* **15**, 332-341.
- Goffeau, A., Barrel, B.G., Bussey, R.W., Davies, W., Dujon, B., Feldmann, H., Gailbert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. *Science* **274**, 546-567.

- Gorczyca, W.A., Gray-Keller, M.P., Detwiler, P.B., Palczewski, K. (1994) Purification and physiological evaluation of a guanylate cyclase activating protein from retinal rods. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4014-4018
- Gorczyca, W.A., Polans, A.S., Surgecheva, I.G., Subbaraya, I., Baer, W. and Palczewski, K. (1995) Guanylyl cyclase activating protein: A calcium sensitive regulator of phototransduction. *J. Biol. Chem.* **270**, 22029-22036.
- Goss, S.J. and Harris, H. (1975) New method for mapping genes in human chromosomes. *Nature* **255**, 680-684.
- Gorodovikova, E.N. and Philippov, P.P. (1993) the presence of a calcium-sensitive p26-containing complex in bovine retina rod cells. *FEBS Lett.* **335**, 277-279.
- Gorodovikova, E.N., Gimelbrandt, A.A., Senin, I.I. and Philippov, P.P. (1994) Recoverin mediates the calcium effect upon rhodopsin phosphorylation and cGMP hydrolysis in bovine retina rod cells. *FEBS Lett.* **349**, 187-190.
- Gratzer, W. (1994) Silence speaks in spectrin. *Nature* **372**, 620-621.
- Gray-Keller, M.P., Polans, A.S., Palczewski, K. and Detweiler, P.B. (1993) The effect of recoverin-like calcium binding proteins on the photoresponse of retinal rods. *Neuron* **10**, 523-531.
- Green, E.D. and Olson, M.V. (1990) Chromosomal region of the cystic fibrosis gene in yeast artificial chromosomes: a model for human genome mapping. *Science* **250**, 94-98.
- Greenberg, J., Goliath, R., Beighton, P. and Ramesar, R. (1994) A new locus for autosomal dominant retinitis pigmentosa on the short arm of chromosome 17. *Hum. Mol. Genet.* **3**, 915-918.
- Greferath, U., Grünert, U. and Wässle, H. (1990) Rod bipolar cells in the mammalian retina show protein kinase C-like immunoreactivity. *J. Comp. Neurology* **301**, 433-442.
- Gregory, C.Y., Evans, K., Whittaker, J.L., Fryer, A., Weissenbach, J. and Bhattacharya, S.S. (1994) Refinement of the cone-rod retinal dystrophy locus on chromosome 19q. *Am. J. Hum. Genet.* **55**, 1061-1063.
- Gruneberg, H. (1963). The pathology of development. 5-Blackwell scientific, Oxford.
- Grünert, U. and Martin, P.R. (1991) Rod bipolar cells in the macaque monkey retina: immunoreactivity and connectivity. *J. Neurosci.* **11**, 2742-2758.
- Gu, S., Thompson, D.A., Srikumari, S., Lorenz, B., Finckh, U., Nicoletti, A., Murthy, K.R., Rathmann, M., Kumaramanickavel, G., Denton, M.J. and Gal, A. (1997) Mutations in RPE65 cause autosomal recessive childhood onset severe retinal dystrophy. *Nat. Genet.* **17**, 194-197.

- Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994) The 1993-94 Génethon human genetic linkage map. *Nat. Genet.* **7**, 246-249.
- Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vegar-Czarny, V., Spillett, D., Muselet, D., Prud'Homme, J., Dib, C., Auffray, C., Morissette, J., Weissenbach, J. and Goodfellow, P.N (1996) A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**, 339-346.
- Hagstrom, S.A., North, M.A., Nishina, P.L., Berson, E.L. and Dryja, T.P. (1998) Recessive mutations in the gene encoding the tubby-like protein TULP1 in patients with retinitis pigmentosa. *Nat. Genet.* **18**, 174-176.
- Hahn, L.B., Berson, E.L. and Dryja, T.P. (1994) Evaluation of the gene encoding the gamma subunit of rod phosphodiesterase in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **35**, 1077-1082.
- Hamel, C.P., Tsilou, E., Pfeiffer, B.A., Hooks, J.J., Detrick, B. and Redmond, T.M. (1993) Molecular cloning and expression of *RPE65*, a novel retinal pigment epithelium-specific microsomal protein that is post-transcriptionally regulated *in vitro*. *J. Biol. Chem.* **268**, 15751-15757.
- Hardcastle, A.J., David-Gray, Z.K., Jay, M., Bird, A.C. and Bhattacharya, S.S. (1997) Localisation of CSNBX (CSNB4) between the retinitis pigmentosa loci RP2 and RP3 on proximal Xp. *Invest. Ophthalmol. Vis. Sci.* **38**, 2750-2755.
- Hargrave, P.A., Fong, S-L., McDowell, J.H., Mas, M.T., Curtis, D.R., Wang, J.K., Juszczak, E. and Smith, D.P. (1980) The partial primary structure of bovine rhodopsin and its topography in the retinal rod cell disc membrane. *Neurochem. Int.* **1**, 231-244.
- Hargrave, P. A. and McDowell, J.H. (1992) Rhodopsin and phototransduction: a model system for G protein linked receptors. *FASEB J.* **6**, 2323-2331.
- Hashimoto, T., Ase, K., Sawamura, S. and others. (1988) Postnatal development of a brain specific subspecies of protein kinase C in rat. *J. Neurosci.* **8**, 1678-1683
- Heckenlively, J.R. (1988) 'Retinitis Pigmentosa'. (Lippincott, Philadelphia, PA).
- Heckenlively, J.R. and Foxmann, S.G. (1988) Congenital and early-onset forms of retinitis pigmentosa. In 'Retinitis Pigmentosa'. (ed. Heckenlively, J.R.) 107-149 (Lippincott, Philadelphia, PA).
- Heckenlively, J.R., Rodriguez, J.A. and Daiger, S.P. (1991) Autosomal dominant sectorial retinitis pigmentosa: Two families with transversion mutation in codon 23 of rhodopsin. *Arch. Ophthalmol.* **109**, 84-91.
- Herrmann, B.G., Labeit, S., Poustka, A., King, T.R. and Lehrach, H. (1990) Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-622.

- Herrmann, K., Meindel, A., Apfelstedt-Sylla, E. et al. (1996) RPGR mutation analysis in patients with retinitis pigmentosa and congenital stationary night blindness. *Am. J. Hum. Genet.* **59**, 1518.
- Hieter, P., Bassett, D.E. and Valle, D. (1996) The yeast genome- a common currency. *Nat. Genet.* **13**, 253-255.
- Hirata, M., Saito, N., Kono, M. and Tanaka, C. (1991) Differential expression of the β I and β II PKC subspecies in the post natal developing rat brain; an immunocytochemical study. *Dev. Brain Res.* **62**, 229-238
- Hoheisel, J.D., Maier, E., Mott, E., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F. and Lehrach, H. (1993) High resolution cosmid and P1 maps spanning the 14 Mb genome of fission yeast *S. pombe*. *Cell* **73**, 109-120.
- Hogan, M.J., Wood, I. and Steinberg, R.H. (1974) Phagocytosis by pigment epithelium of human retinal cones. *Nature* **252**, 305.
- Holland, P.W.H., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Development Suppl*, pp125-133
- Hsu, Y-T. and Molday, R. (1993) Modulation of the cGMP-gated channel of rod photoreceptor cells by calmodulin. *Nature* **361**, 76-79.
- Huang, P.C., Gaitan, A.E., Hao, Y., Petters, R.M. and Wong, F. (1993) Cellular interactions implicated in the mechanism of photoreceptor degeneration in transgenic mice expressing a mutant rhodopsin gene. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8484-8488.
- Huang, S.H., Huang, X., Pittler, S.J., Oliveira, L., Berson, E.L. and Dryja, T.P. (1995) A mutation in the gene encoding the α -subunit of rod cGMP phosphodiesterase (*PDEA*) in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **36** (suppl): S825 (abstract no. 3815)
- Hudson, T. and 50 others. (1995) An STS-based map of the Human Genome. *Science* **270**, 1945-1954.
- Hug, H., and Sarre, T. F. (1993) Protein Kinase C isoenzymes: divergence in signal transduction? *Biochem. J.* **291**, 329-343
- Humphries, M.M., Rancourt, D., Farrar, G.J., Kenna, P., Hazel, M., Bush, R.A., Sieving, P.A., Shiels, D.M., McNally, N., Creighton, P., Erven, A., Boros, A., Gulya, K., Capocchi, M.R. and Humphries, P. (1997) Retinopathy induced in mice by targeted disruption of the rhodopsin gene. *Nat. Genet.* **15**, 216-219.
- Hurley, J. B., Dizhoor, A.M., Ray, S. and Stryer, L. (1993) Recoverin's role: conclusion withdrawn. *Science* **260**, 740.

- Huxley, C., Hagino, Y., Schlessinger, D. and Olson, M.V. (1991) The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* **9**, 742-750.
- Illing, M., Colville, C.A., William, A.J. and Molday, R.S. (1994) Sequencing cloning and characterisation of a third subunit of the cyclic nucleotide gated channel complex of rod outer segments. *Invest. Ophthalm. Vis. Sci.* **35**, 1474 (no.1022).
- Inglehearn, C.F., Keen, T.J., Bashir, R., Jay, M., Fitzke, F., Bird, A.C., Crombie, A. and Bhattacharya, S.S. (1992) A completed screen for mutations of the rhodopsin gene in a panel of patients with autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* **1**, 41-45.
- Inglehearn, C.F., Carter, S.A., Keen, T.J., Lindsay, J., Stephenson, A.M., Bashir, R., Al-Magthteh, M., Moore, A.T., Jay, M., Bird, A.C. and Bhattacharya, S.S. (1993). A new locus for autosomal dominant retinitis pigmentosa on chromosome 7p. *Nat. Genet.* **4**, 51-53.
- Inglehearn, C.F., Tartttelin, E.E., Plant, C., Peacock, R.E., Al-Magthteh, M., Vithana, E., Bird, A.C. and Bhattacharya, S.S. (1998) A linkage survey of 20 dominant retinitis pigmentosa families: frequencies of the nine known loci and evidence for further heterogeneity. *J. Med. Genet.* **35**, 1-5.
- Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and de Jong, P.J. (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet. ics* **6**, 84-89.
- Ioannou, P.A. and de Jong, P.J. (1996) Construction of Bacterial Artificial chromosome Libraries using the modified P1 (PAC) System. In Current Protocols in Human Genetics. (eds. Dracolpoli *et al.*), Unit 5.15. John Wiley and Sons, NY.
- Jay, M. (1982) Figure and fantasies: the frequencies of the different genetic forms of retinitis pigmentosa. *Birth Defects* **18**,167-173.
- Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985) Hyprvariable 'minisatellite' regions in human DNA. *Nature* **314**, 67-73.
- Jordan, S.A., Farrar, G.J., Kumar-Singhe, R., Kenna, P., Humphries, M.M., Allamand, V., Sharp, E.M. and Humphries, P. (1992) Autosomal dominant retinitis pigmentosa (adRP; RP6): Cosegration of RP6 and the peripherin-RDS locus in a late-onset family of Irish origin. *Am. J. Hum. Genet.* **50**, 634-639.
- Jordan, S.A., Farrar, G.J., Kenna, P., Humphries, M.M., Sheils, D., Kumar-Singhe, R., Sharp, E.M., Benitez, J., Carmen, A. and Humphries, P. (1993) Localisation of an autosomal dominant retinitis pigmentosa gene to 7q. *Nat. Genet.* **4**, 54-58.
- Kajiwara, K., Hahn, L.B., Mukai, S., Travis, G.H., Berson, E.L. and Dryja, T.P. (1991) Mutations in the human retinal degeneration slow gene in autosomal dominant retinitis pigmentosa. *Nature* **354**, 480-483.

- Kajiwara, K., Berson, E.L. and Dryja, T.P. (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-1608.
- Kaplan, J., Bonneau, D., Frezal, J., Munnich, A. and Dufier, J-L. (1990) Clinical and genetic heterogeneity in retinitis pigmentosa. *Hum. Genet.* **85**, 635-642.
- Kärschin, A. and Wässle, H. (1990) Voltage- and transmitter gated currents in isolated rod bipolar cells of the rat retina. *J. neurophysiol.* **63**, 860-876.
- Kawamura, S. and Murakami, M. (1991) Calcium-dependent regulation of cyclic GMP phosphodiesterase by a protein from frog retinal rods. *Nature* **349**, 420-423.
- Kawamura, S., Hisatomi, O., Kayada, S., Tokunaga, F. and Kuo, C.H. (1993) Recoverin has S-modulin activity in frog rods. *J. Biol. Chem.* **268**, 14579-14582.
- Kelsell, R.E., Evans, K., Gregory, C.Y., Moore, A.T., Bird, A.C. and Hunt, D.M. (1997) Localisation of a gene for dominant cone-rod dystrophy (CORD6) to chromosome 17p. *Hum Mol. Genet.* **6**, 597-600.
- Kelsell, R.E., Gregory-Evans, K., Payne, A.M., Perrault, I., Kaplan, J., Yang, R.B., Garbers, D.L., Bird, A. C. Moore, A.T. and Hunt, D.M. (1998) Mutations in the retinal guanylate cyclase (RETGC-1) gene in dominant cone-rod dystrophy. *Hum Mol. Genet.* **7**, 1179-1184.
- Khan, A.S. Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K. and Sikela, J.M. (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nat. Genet. ics* **2**, 180-185.
- Kim, R.Y., Fitzke, F.W., Moore, A.T., Jay, M., Inglehearn, C., Arden, G.B., Bhattacharya, S.S et al (1995) Autosomal dominant retinitis pigmentosa mapping to chromosome 7p exhibits variable expression. *Br. J. Ophthalmol* **79**, 23-27.
- Kim, J., Ashworth, L., Branscomb, E. and Stubbs, L (1997) The human homolog of mouse imprinted gene, Peg 3, maps to a zinc finger gene-rich region of human chromosome 19q13.4. *Genome Res.* **7**, 532-540
- Klenchin, V.A., Calvert, P.D. and Bownds, M.D. (1995) Inhibition of rhodopsin kinase by recoverin. Further evidence for a negative feedback system in phototransduction. *J. Biol. Chem.* **270**, 16147-16152.
- Kleyn, P.W. and 24 others (1996) Identification and characterisation of the mouse obesity gene tubby: a member of a novel gene family. *Cell* **85**, 281-290.
- Knopf, J. L., Lee, M. H., Sultzman, L. A., Kriz, R. W., Loomis, C.R., Hewick, R. M., and Bell, R. M. (1986) Cloning and expression of multiple protein kinase C cDNAs. *Cell* **46**, 491-502
- Knowles, J.A., Shugarts, Y.Y., Banerjee, P., Gilliam, T.C., Lewis, C.A., Jacobson, S.G. and Ott, J. (1994) Identification of a locus, distinct from RDS-peripherin, for autosomal recessive retinitis pigmentosa on chromosome 6p. *Hum. Mol. Genet.* **3**, 1401-1403.

- Koch, K.W. Stryer, L. (1988) Highly co-operative feedback control of retinal rod guanylate cyclase by calcium ions. *Nature* **334**, 64-66.
- Kohara, Y., Akiyama, K. and Isono, K. (1987) The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495-508.
- Koistinaho, J. and Sagar, S.M. (1994) Localisation of protein kinase C subspecies in the rabbit retina. *Neurosci. Letts* **177**, 15-18.
- Kolb, H., Zhang, L. and Dekorver, L. (1993) Differential staining of neurons in the human retina with antibodies to protein kinase C isoenzymes. *Vis. Neurosci.* **10**, 341-351.
- Korenberg, J.R. and Rylowski, M.C. (1988) Human genome organisation: Alu, lines and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391-400.
- Kosambi, D.D. (1944) The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172-175.
- Kouprina, N., Eldarov, M., Moyzis, R., Resnick, M. and Larionov, V. (1994) A model system to assess the integrity of mammalian YACs during transformation and propagation in yeast. *Genomics* **21**, 7-17.
- Koutalos, Y. and Yau, K.W. (1996) Regulation of sensitivity in vertebrate rod photoreceptors by calcium. *Trends Neurosci.* **19**, 73-81
- Krill, A.E., Deutman, A.F. and Fishman, M. (1973) The cone degenerations. *Doc. Ophthalmol.* **35**, 1-80.
- Kruglyak, L. (1997) The use of genetic map of biallelic markers in linkage studies. *Nat. Genet.* **17**, 21-24.
- Kumaramanickavel, G., Maw, M., Denton, M.J., John, S., Srikumari, C.R.S., Orth, U., Oehlmann, R. and Gal, A. (1994) Missense rhodopsin mutation in a family with recessive RP. *Nat. Genet.* **8**, 10-11.
- Kurada, P. and O'Tousa, J.E. (1995) retinal degeneration caused by dominant rhodopsin mutations in drosophila. *Neuron* **14**, 571-579.
- Kwitek-Black, A., Carmi, R., Duyk, G.M., Buetow, K.H., Elbedour, K., Parvari, R., Yandava, C.N., Stone, E.M. and Sheffield, V.C (1993) Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic heterogeneity. *Nat. Genet.* **5**, 392-396.
- Kylstra, J.A. and Aylsworth, A.S. (1993) Cone-rod retinal dystrophy in a patient with neurofibromatosis type 1. *Can. J. Ophthalmol.* **28**, 79-80.

- Lambrecht, H-G. and Koch, K-W. (1991) A 26 kd calcium binding protein from bovine rod outer segments as modulator of photoreceptor guanylate cyclase. *EMBO J.* **10**, 793-798.
- Lander, E.S. (1996) The new genomics: Global views of biology. *Science* **274**, 536-539.
- Lander, E.S. and Botstein, D. (1987) Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567-1570.
- Lander, E.S. and Green, P. (1987) Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. U.S.A* **84**, 2363
- Larin, Z., Monaco, A.P. and Lehrach, H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA* **88**, 4123-4127.
- Larsen, F., Gunderson, G., Lopez, R. and Prydz, H. (1992) CpG islands as gene markers for the human genome. *Genomics* **13**, 1095-1107.
- Lathrop, G.M. and Lalouel, J.M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36**, 460-465.
- Lee, R.H., Lieberman, B. and Lolley, R.N. (1990) Retinal accumulation of phosphodiesterase beta, gamma and transducin complexes in developing normal mice and in mice and dogs with inherited retinal degeneration. *Exp. Eye Res.* **51**, 325-333.
- Lefkowitz, R.J. (1993) G protein-coupled receptor kinases. *Cell* **74**, 409-412.
- Legouis, R. and 14 others (1991) The candidate gene for the X-linked Kallman syndrome encodes a protein related to adhesion molecules. *Cell* **67**, 423-435.
- Leppert, M., Baird, L., Anderson, K.L., Otterud, B., Lupske, J.R. and Lewis, R.A. (1994) Bardet-Biedl syndrome is linked to markers on chromosome 11 q and is genetically heterogeneous. *Nat. Genet.* **7**, 108-112.
- Li, T., Snyder, W.K., Olsson, J.E. and Dryja, T.P. (1996) Transgenic mice carrying the dominant rhodopsin mutation P347S: evidence for defective vectorial transport of rhodopsin to the outer segments. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14176-14181.
- Li, T., Franson, W.K., Gordon, J.W., Berson, E.L. and Dryja, T.P. (1995a) Constitutive activation of phototransduction by K296E opsin not a cause of photoreceptor degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3551-3555.
- Li, X., Park, W.J., Pyeritz, R.E. and Jabs, E.W. (1995b) Effect on splicing of a silent FGFR2 mutation in Crouzon syndrome. *Nat. Genet.* **9**, 232-233.
- Ling, L. L., Ma, N. S-F., Smith, D.R., Miller, D.D. and Moir D.T. (1993) Reduced occurrence of chimeric YACs in recombination deficient hosts. *Nucleic Acids Res.* **21**, 6045-6046

- Lisman, J. and Fain, G. (1995) Support for equivalent light hypothesis for RP. *Nature Med.* **1**, 1254-1255.
- Litt, M. and Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397-401.
- Liu, X., Seno, K., Nishizuka, Y., Hayashi, F., Yamazaki, A., Matsumoto, H., Wakabayashi, T. and Usukura, J. (1994) Ultrastructural localisation of retinal guanylate cyclase in human and monkey retinas. *Exp. Eye. Res.* **59**, 761-768.
- Liu, W., Qian, C. and Francke, U. (1997) Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nat. Genet.* **16**, 328-329.
- Lolley, R. N., Craft, C.M. and Lee, R.H. (1992) Photoreceptors of the retina and pinealocytes of the pineal gland share common components of signal transduction. *Neurochem. Res.* **17**, 81-89.
- Lovett, M. (1994) Fishing for complements: finding genes by direct selection. *Trends in Genetics* **10**, 352-357.
- Lovett, M., Kere, J. and Hinton, L.M. (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88**, 9628-9633.
- Lowe, D.G., Dizhoor, A.M., Liu, K., Gu, Q., Spencer, M., Laura, R., Lu, L. and Hurley, J.B. (1995) Cloning and expression of a second photoreceptor-specific membrane retina guanylyl cyclase (RetGC), RetGC-2. *Proc. Natl. Acad. Sci. U.S.A* **92**, 5535-5539.
- Ludin, L.G. (1993) Evolution of the vertebrate genomes as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16**, 1-19.
- Lyness, A.L., Ernst, W., Quinlan, M.P., Clover, G.M., Ardev, G.B., Carter, R.M., Bird, A.C. and Parker, J.A. (1985) A clinical, psychophysical and electroretinographic survey of patients with autosomal dominant retinitis pigmentosa. *Br. J. Ophthalmol.* **69**, 326-339.
- Matthews, R., Murphy, R.L.W., Fain, G.L. and Lamb, T.D. (1988) Photoreceptor light adaptation is mediated by cytoplasmic calcium concentration. *Nature* **334**, 67-69.
- Maier, E., Hoheisel, J.D., McCarthy, L., Mott, R., Grigoriev, A.V., Monaco, A.P., Larin, Z. and Lehrach, H. (1992) Complete coverage of the *Schizosacharomyces pombe* genome in yeast artificial chromosomes. *Nat. Genet. ics* **1**, 231-300.
- Mahajna, J., King, P., Parker, P., and Haley, J. (1995) Autoregulation of cloned human protein kinase C β and γ gene promoters in U937 Cells. *DNA and Cell Biol.* **14**, 213-222.

- Margulis, A., Goraczniak, R.M., Duda, T., Sharma, R.K. and Sitaramayya, A. (1993) Structural and biochemical identity of retinal rod outer segment membrane guanylate cyclase. *Biochem. Biophys. Res. Commun.* **194**, 855-861.
- Marlhens, F., Bareil, C., Griffon, C., Zrenner, E., Amalric, P., Eliaou, C., Liu, S-Y., Harris, E., Redmond, T.M., Arnaud, B., Claustres, M. and Hamel, C.P. (1997) Mutations in RPE65 cause Leber's congenital amaurosis. *Nat. Genet.* **17**, 139-140.
- Marra, M.A., Hillier, L. and Waterston, R.H. (1998) Expressed sequence tags-ESTablishing bridges between genomes. *Tren. in Genet.* **14**, 4-7.
- Martinez-Mir, A., Bayes, M., Vilageliu, L., Grinberg, D., Ayuso, C., del Rio, T., Garcia-Sandoval, B., Bussaglia, E., Baiget, M., Gonzalez-Duarte, R. and Balcells, S. (1997) A new locus for autosomal recessive retinitis pigmentosa (RP19) maps to 1p13-1p21. *Genomics* **40**, 142-146.
- Martinez-Mir, A., Paloma, E., Allikmets, R., Ayuso, C., del Rio, T., Vilageliu, L., Dean, M., Gonzalez-Duarte, R. and Balcells, S. (1998) Retinitis pigmentosa caused by a homozygous mutation in the Stargardt's disease gene ABCR. *Nat. Genet.* **18**, 11-12.
- Martin-Gallardo, A., McCombie, W.R., Gocayne, J.D., Fitzgerald, M.G., Wallace, S., Lee, B.M., Lamerdin, J., Trapp, S., Kelly, J.M. and Lui, L. (1992) Automated DNA sequencing and analysis of 106 kb from human chromosome 19q13.3. *Nat. Genet.* **1**, 34-39.
- Massof, R.W. and Finkelstein, D. (1981) Two forms of autosomal dominant primary retinitis pigmentosa. *Doc. Ophthalmol.* **51**, 289-346.
- Matthews, R., Murphy, R.L.W., Fain, G.L. and Lamb, T.D. (1988) Photoreceptor light adaptation is mediated by calcium concentration. *Nature* **334**, 67-69.
- Maw, M.A., Kennedy, B., Knight, A., Bridges, R., Roth, K.E., Mani, E.J., Mukkadan, J.K., Nancarrow, D., Crabb, J.W. and Denton, M.J. (1997) Mutation of the gene encoding cellular retinaldehyde-binding protein in autosomal recessive retinitis pigmentosa. *Nat. Genet.* **17**, 198-200.
- Mayall, B.H. et al. (1984) The DNA-based human karyotype. *Cytometry* **5**, 376-385
- McConkey, D.J., Hartzell, P., Nicotera, P. and Orrenius, S. (1989) Calcium-activated DNA fragmentation kills immature thymocytes. *FASEB J.* **3**, 1843-1849.
- McCord, R. J., Klein, A. and Osborne, N. N. (1996) *Neurochemical Res.* **21**, 259-266.
- McGee, T.L., Devoto, M., Ott, J., Berson, E.L. and Dryja, T.P. (1997b) Evidence that the penetrance of mutations at the RP11 locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous RP11 allele. *Am. J. Hum. Genet.* **61(5)**, 1059-1066.

- McGinnis, J.F., Austin, B., Bateman, B., Heinzman, C., Kojis, T., Sparkes, R.S., Bateman, J.B. and Lerious, V. (1995) Chromosomal assignment for the human gene for the cancer-associated retinopathy protein (recoverin) to chromosome 17p13.1. *J. Neuroscience res.* **40**, 165-168.
- McKinley, D., Wu, Q., Yang-Feng, T. and Yang, Y-C. (1992) Genomic sequence and chromosomal localisation of human interleukin 11 gene (*IL11*). *Genomics* **13**, 814-819
- McKusick, V.A. (ed.): Mendelian inheritance in man. Baltimore: John Hopkins University Press, 1992.
- McLaughlin, M.E., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1993) Recessive mutations in the gene encoding the β -subunit of rod phosphodiesterase in patients with retinitis pigmentosa. *Nat. Genet.* **4**, 130-134.
- McLaughlin, M.E., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1995) Mutation spectrum of the gene encoding the β -subunit of rod phosphodiesterase among patients with autosomal recessive retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3249-3253.
- McWilliam P., Farrar, G.J., Kenna, P., Bradly, D.G., Humphries, M.M., Sharp, E.M., McConnel, D.J., Lawler, M., Sheils, D., Ryans, C., Stevens, K., Daiger, S.P and Humphries, P. (1989) Autosomal dominant retinitis pigmentosa (ADRP): Localisation of an adRP gene to the long arm of chromosome 3. *Genomics* **5**, 619-622.
- Meindl, A., Dry, K., Herrmann, K., Manson, F., Ciccodicola, A., Edgar, A., Carvalho, M.R.S and others (1996) A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nat. Genet.* **13**, 35-42.
- Miklos, G.L.G and Rubin, G.M. (1996) The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**, 521-529.
- Milam, A. H., Dacey, D. M. and Dizhoor, A.M. (1993) Recoverin immunoreactivity in mammalian cone bipolar cells. *Visual Neurosci.* **10**, 1-12.
- Mizukami, T. and 11 others (1993) A 13 kb resolution cosmid map of the 14 Mb fission yeast genome by non random sequence tagged site mapping. *Cell* **73**, 121-132.
- Mohrenweiser, H. W., Tsujimoto, S., Tynan, K., Lamerdin, J. E. and Carrano, A.V. (1995) Unique sequence STSs for 21 cytogenetically mapped loci on human chromosome 19. *Cytogenet. Cell Genet.* **71**, 58-61
- Molday, L. L., Cook, N. J., Kaupp, U.B. and Molday, R. S. (1990) The cGMP-gated cation channel of bovine rod photoreceptor cells is associated with a 240 kDa protein exhibiting immunochemical cross reactivity with spectrin. *J. Biol. Chem.* **265**, 18690-18695.

- Monaco, A.F. (1994) Isolation of genes from cloned DNA. *Current Opinion in Genetics and Development* **4**, 360-365.
- Monaco, A.P. and Larin Z. (1994) YACs, BACs, and PACs, and MACs: artificial chromosomes as research tools. *Trends in Biotechnology* **12**, 280-286.
- Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertlson, C.J., Kurnit, D.M. and Kunkel, L.M. (1986) Isolation of candidate cDNAs portions of Duchenne Muscular Dystrophy gene. *Nature* **323**, 646-650.
- Moore, A.T. (1992) Cone and cone-rod dystrophies. *J. Med. Genet.* **29**, 289-290.
- Moore, A.T., Fitzke, F., Jay, M., Arden, G.B., Inglehearn, C.F., Keen, T.J., Bhattacharya, S.S. and Bird, A.C. (1993) Autosomal dominant retinitis pigmentosa with apparent incomplete penetrance: a clinical electrophysiological, psychophysical and molecular genetic study. *Br. J. Ophthalmol.* **77**, 473-479.
- Morrow, D.M., Tagle, D.A., Shilh, Y., Collins, F.S and Hieter, P. (1995) An *S. cerevisiae* homologue of the human gene mutated in ataxia telangiectasia, is functionally related to the yeast checkpoint gene *MEC1*. *Cell* **82**, 831-840.
- Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277-318.
- Murray, N. (1986) Phage lambda and molecular cloning. In *Lambda II* (ed. Hendrix, R., Roberts, J., Stahl, F. and Weisberg, R.), pp. 395-432. Cold Spring Harbour Laboratory, Cold Spring Harbour, NY.
- Mussarella, M.A., Weleber, R.G., Murphey, W.H., et al. (1989) Assignment of the gene for complete X-linked congenital stationary night blindness (CSNB1) to Xp11.3. *Genomics* **5**, 727-737.
- Nagao, S., Yamazaki, A. and Bitenski, K.M. (1987) Calmodulin and calmodulin binding proteins in amphibian rod outer segments. *Biochemistry* **26**, 1659-1665.
- Nakatani, K. and Yau, K.W. (1988) Calcium and light adaptation in retinal rods and cones. *Nature* **334**, 69-71.
- Nakazawa, M., Kikawa, E., Chida, Y. and Tamai, M. (1994) Asn244His mutation of the peripherin/RDS gene causing autosomal dominant cone-rod degeneration. *Hum Mol. Genet.* **3**, 1195-1196.
- Nakazawa, M., Kikawa, E., Chida, Y., Wada, Y., Shiono, T. and Tamai, M. (1996) Autosomal dominant cone-rod dystrophy associated with mutations in codon 244 (Asn244His) and codon 184 (Tyr184Ser) of the peripherin/RDS gene. *Arch. Ophthalmol.* **114**, 72-78.
- Negishi, K., Kato, S. and Teranishi, T. (1988) Dopamine cells and rod bipolar cells contain protein kinase C immunoreactivity in some vertebrate retinas. *Neurosci. Letts* **94**, 247-252.

- Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. and Caskey, T.C. (1989) Alu polymerase chain reaction : a method for rapid isolation of human-specific sequences from complex sources. *Proc. Natl. Acad. Sci. USA* **86**, 6686-6690.
- Nelson, D.L., Ballabio, A., Victoria, M.F., Pieretti, M., Bies, D., Gibbs, R.A., Maley, J.A., Chinault, A.C., Webster, T.D. and Caskey, C.T (1991) Alu-primed PCR for regional assignment of 110 YAC clones from the human X chromosome: Identification of clones associated with a disease locus. *Proc. Natl. Acad. Sci. USA* **88**, 6157-6161.
- Newton, A.C. and William, D.S. (1991) Involvement of protein kinase C in the phosphorylation of rhodopsin. *J Biol. Chem* **266**, 17725-17728.
- Newton, A.C. and William, D.S. (1993) Does protein kinase C play a role in rhodopsin desensitisation ? *Trends. Biochem. Sci.* **18**, 275-277.
- NIH/CEPH Collaborative Mapping Group (1992) A complete genetic linkage map of the human genome. *Science* **258**, 67-86.
- Nishizuka, N. (1988) The molecular heterogeneity of protein kinase C and its implications for cellular regulation. *Nature (london)* **334**, 661-665.
- Nishizuka, N. (1992) Intracellular signalling by hydrolysis of phospholipids and activation by protein kinase C. *Science* **258**, 607-614.
- Nishizuka, N. (1995) Protein Kinase C and lipid signalling for sustained cellular responses. *FASEB J.* **9**, 484-496.
- Noben-Trauth, K., Naggert, J.K., North, M.A. and Nishina, P.M. (1996) A candidate gene for the mouse mutation tubby. *Nature* **380**, 534-538.
- Noell, W.K., Walker, V.S., Kang, B.S. and Berman, S. (1966) Retinal damage by light in rats. *Invest. Ophthalmol.* **5**, 450-473.
- North, M.A., Naggert, J.K., Yan, Y., Noben-Trauth, K. and Nishina, P.M. (1997) Molecular characterisation of TUB, TULP1, and TULP2, members of the novel tubby gene family and their possible relation to ocular diseases. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3128-3133.
- Ohki, K., Yoshida, K., Imaki, J., Harada, T. and Matsuda, H. (1994) The existence of protein kinase C in cone photoreceptors in the rat retina. *Cur. Eye Res.* **13**, 547-550.
- Ohno, S., Kawasaki, H., Imajoh, S., Suzuki, K., Inagaki, M., Yokokura, H., Sakoh, T., and Hidaka H. (1987) Tissue specific expression of three distinct types of rabbit protein kinase C. *Nature* **325**, 161-166
- Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989) A common language for physical mapping of the human genome. *Science* **245**, 1434-1435.

- Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G. M., Helms, C., Frank, M., MacCollin, M., Scheinman, R. and Frank, T. (1986) A random-clone strategy for restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* **83**, 7826-7830.
- Osborne, N. N., Barnett, N. D., Morris, N. J., and Huang, F.L. (1992) The occurrence of three isoenzymes of protein kinase C (α , β and γ) in retinas of different species. *Brain Res.* **570**, 161-166.
- Osborne, N. N., Wood, J. and Groome, N. (1994) The occurrence of three calcium-independent protein kinase C sub species (δ , ϵ and ζ) in retina of different species. *Brain Res.* **637**, 156-162.
- Pak, W.L. (1979) Study of photoreceptor function using *Drosophila* mutants in Neurogenetics: Genetic approaches to the nervous system. *X.O. Breakfield, eds (New York: Elsevier north- Holland)*, 67-99.
- Pak, W.L. (1991) Molecular genetic studies of photoreceptor function using *Drosophila* mutants. *Prog. Clin. Biol. Res.* **362**, 1-32.
- Palczewski, K., Subbaraya, I., Gorczyca, W.A., Helekar, B.S., Ruiz, C.C., Ohguro, H., Huang, J., Zhao, X., Crabb, J.W., Johnson, R.S., Walsh, K., Gray-Keller, M.P., Detwiler, P.B. and Baehr, W. (1994) Molecular cloning and characterisation of retinal photoreceptor guanylate cyclase activating protein. *Neuron* **13**, 395-404.
- Parimoo, S., Patanjali, S.R., Shukle, H., Chaplin, D.D. and Weisman, S.M. (1991) cDNA selection: efficient PCR approach for the selection of cDNA encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA* **88**, 9623-9627.
- Parker, P. J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., Stabel, S., Waterfield, M.D, and Ullrich, A. (1986) The complete primary structure of protein kinase C- the major phorbol ester receptor. *Science* **233**, 853-859
- Parra, I. and Windle, B. (1993) High resolution visual mapping of stretched DNA by fluorescent hybridisation. *Nat. Genet. ics* **5**, 17-21.
- Paton, D., Hyman, B.N. and Justice, J. (1976) Introduction to Ophthalmology. *The Upjohn Company. Michigan*.
- Payne, A.M., Downes, S.M., Bessant, D.A.R., Taylor, R., Holder, G.E., Warren, M.J., Bird, A.C. and Bhattacharya, S.S. (1998) A mutation in guanylate cyclase activator 1A (GUCA1A) in an autosomal dominant cone dystrophy pedigree mapping to a new locus on chromosome 6p21.1. *Hum Mol. Genet.* **7**, 273-277.
- Payne, R. (1986) Phototransduction by microvillar photoreceptors of invertebrates: mediation of visual cascade by inositol triphosphate. *Photobiochem. Photobiophys.* **13**, 373-397.
- Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A* **89**, 10882-10886.

- Pepose, J. S. and Leib, D.A. (1994) Herpes simplex viral vectors to for therapeutic gene delivery to ocular tissues. *Invest. Ophthalmol. Vis. Sci.* **35**, 2662-2666
- Perrault, I., Rozet, J.M., Calvas, P., Gerber, S., Camuzat, A., Dollfus, H., Chatelin, S., Souied, E., Ghazi, I., Leowski, C., Bonnemaïson, M., Le Paslier, D., Frezal, J., Dufier, J-L., Pittler, S., Munnich, A. and Kaplan, J. (1996) Retinal specific guanylate cyclase gene mutations in Leber's congenital amaurosis. *Nat. Genet. ics* **14**, 461-464.
- Petersen, M.B., Economou, E.P., Slangenaupt, S.A., Chakravarti, A. and Antonarkis, S.E. (1990) Linkage analysis of the human HMG14 gene on chromosome 21 using a GT dinucleotide repeat as polymorphic marker. *Genomics* **7**, 136-138.
- Pierce, J.C., Sauer, B. and Sternberg, N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficiency. *Proc. Natl. Acad. Sci. USA* **89**, 2056-2060.
- Pittler, S.J. and Baehr, W. (1991) Identification of a nonsense mutation in the rod photoreceptor cGMP phosphodiesterase β subunit gene of the *rd* mouse. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8322-8326.
- Pinkel, D., Straume, T. and Grey, J.M. (1986) Cytogenetic analysis using quantitative high sensitivity fluorescent hybridisation. *Proc. Natl. Acad. Sci. USA* **83**, 2934-2938.
- Polans, A.S., Burton, M.D., Haley, T.L., Crabb, J.W. and Palczewski, K. (1993) Recoverin, but not visinin, is an autoantigen in the human retina identified with a cancer-associated retinopathy. *Invest. Ophthalmol. Vis. Sci.* **34**, 81-90.
- Polans, A.S., Baehr, W. and Palczewski, K. (1996) Turned on by Ca^{2+} . The physiology and pathology of Ca^{2+} binding proteins in the retina. *Trends in Neurosci.* **19**, 547-554.
- Polymeropoulos, M.H., Xiao, H., Glodek, A., Gorski, M., Adams, M.D., Moreno, R.F., Fitzgerald, M.G., Venter, J.C. and Merrill, C.R. (1992) Chromosomal assignment of 46 brain cDNAs. *Genomics* **12**, 492-496
- Polymeropoulos, M.H., Xiao, H., Sikela, J.M., Adams, M., Venter, J.C. and Merrill, C.R. (1993) Chromosomal distribution of 320 genes from a brain cDNA library. *Nat. Genet.* **4**, 381-386.
- Portera-Cailliau, C., Sung, C.H., Nathans, J. and Adler, R. (1994) Apoptotic photoreceptor cell death in mouse models of retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA* **91**, 974-978.
- Ragoussis, J., Monaco, A., Mockridge, I., Kendall, E., Campbell, R.D. and Trowsdale, J. (1991) Cloning of the HLA class II region in yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **88**, 3753-3757.

- Randon, J., Boulanger, L., Marechal, J., Garbarz, M., Vallier, A., Ribeiro, L., Tamagnini, G, et al (1994) A variant of spectrin low-expression allele α LELY carrying a hereditary elliptocytosis mutation in codon 28. *Br. J. Haematol.* **88**, 534-540.
- Ranganathan, R., Harris, G.L., Stevens, C.F. and Zuker, C.S. (1991) A *Drosophila* mutant defective in extracellular calcium-dependent photoreceptor deactivation and rapid desensitisation. *Nature* **367**, 639-642.
- Rao, V.R., Cohen, G.B. and Oprian, D.D. (1994) Rhodopsin mutation G90D and a molecular mechanism for congenital night blindness. *Nature* **367**, 639-642.
- Rehmtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J. and Swaroop, A. (1996) The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 191-195.
- Reuter, J.H. and Sanyal, S. (1984) Development and degeneration of the retina in rds mutant mice: the electroretinogram. *Neurosci. Letts.* **46**, 231.
- Richards, R.I., Holman, K., Friend, K., Kremer, E., Hillen D., Staples, A. et al (1992) Evidence of founder chromosome in fragile X syndrome. *Nat. Genet.* **1**, 257-260
- Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-2890
- Robinson, P.R., Cohen, G.B., Zhukovsky, E.A. and Oprian, D.D. (1992) Constitutively active mutants of rhodopsin. *Neuron* **9**, 719-725.
- Roof, D.J., Adamian, M. and Hayes, A. (1994) Rhodopsin accumulation at abnormal sites in retinas of mice with a human P23H rhodopsin transgene. *Invest. Ophthalmol. Vis. Sci.* **35**, 4049-4062.
- Romeo, G. and McKusick, V.A. (1994) Phenotypic diversity, allelic series and modifier genes. *Nat. Genet.* **7**, 451-453.
- Rosenfield, P.J., Cowley, G.S., McGee, T.L., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1992) A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. *Nat. Genet.* **1**, 209-213.
- Rosenfield, P.J. and Dryja, T.P. (1995) Molecular genetics of retinitis pigmentosa and related retinal degenerations. In: *Molecular genetics of ocular diseases*. p99-126. Ed: J.L. Wiggs. Wiley-Liss publication.
- Rosenthal, A., Rhee, L., Yadegari, R., Paro, R., Ullrich, A., and Goeddel, D. V. (1987) Structure and nucleotide sequence of a *Drosophila melanogaster* protein kinase C gene. *EMBO. J.* **6**, 433-441

- Rozmahel, R., Wilschanski, M., Matin, A., Plyte, S., Oliver, M., Auerbach, W., Moore, A., Forstner, J., Durie, P., Nadeau, J., Bear, C. and Tsui, L.C. (1996) Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nat. Genet.* **12**, 280-287.
- Saito, K., Kikka, U., Nishizuka, Y. and Tanaka, C. (1988) Distribution of protein kinase C like immunoreactivity neurons in rat brain. *J. Neurosci.* **8**, 369-382.
- Saari, J.C., Bredberg, D.L. and Noy, N. (1994) Control of substrate flow at a branch in the visual cycle. *Biochemistry* **33**, 3106-3112.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A* **74**, 5463-5467.
- Sanyal, S. (1987) Cellular site of expression and genetic interaction of the *rd* and *rd5* loci in the retina of the mouse. In Degenerative retinal disorders: clinical and laboratory investigations (J.G. Hollyfield, R.E. Anderson, and M.M. LaVail, eds). Alan R. Liss, New York p175-194.
- Saxon, P.J., Srivatsan, E.J., Leipzig, V., Sameshima, J.H. and Stanbridge, E.J. (1985) Selective transfer of individual human chromosomes to recipient cells. *Mol. Cell. Biol.* **5**, 140-146
- Schaeffer, E., Smith, D., Mardon, G., Quinn, W. and Zuker, C. (1989) Isolation and characterisation of two new Drosophila protein kinase C genes, including one specifically expressed in photoreceptor cells. *Cell.* **57**, 403-412
- Schuler, G.D. (1997) Pieces of the puzzle: expressed sequence tags and the catalogue of human genes. *J. Mol. Med.* **75**, 694-698.
- Schuler, G.D. and 103 others (1996) A gene map of the human genome. *Science* **274**, 540-546.
- Schwartz, D.C. and Cantor, C.R. (1984) Preparation of yeast chromosome-sized DNAs by pulse field gel electrophoresis. *Cell* **37**, 67-75.
- Screaton, G.R., Bell, M.V., Bell J.I. and Jackson, D.G. (1993) The identification of a new alternative exon with highly restricted tissue expression in transcripts encoding the mouse Pgp-1 (CD44) homing receptor. Comparison of all 10 variable exons between mice, human, and rat. *J Biol Chem* **268**, 12235-12238.
- Sedlacek, Z., Konecki, D.S., Siebenhaar, R., Kioschis, P. and Poustka, A. (1993) Direct selection of DNA conserved between species. *Nucleic Acids Res.* **21**, 3419-3425.
- Sheffield, V.C., Carmi, R., Kwitek-Black, A., Rokhlina, T., Nishimura, D., Duyk, G.M., Elbedour, K., Sunden, S.L. and Stone, E.M. (1994) Identification of a Bardet-Biedel syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. *Hum Mol. Genet.* **3**, 1331-1335.

- Shizuya, H., Birren, B., Kim, U-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300- kilobase-pair fragments of human DNA in E coli using an F-factor based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794-8797.
- Shichi, H. and Somers, R.L. (1978). Light dependent phosphorylation of rhodopsin. Purification and properties of rhodopsin kinase. *J. Biol. Chem.* **253**, 7040-7046.
- Shugarts, Y.Y., Banerjee, P., Knowles, J.A., Lewis, C.A., Jacobson, S.G., Matisse, T.C., Penchaszadeh, G.K., Gilliam, T.C. and Ott, J. (1995) Fine genetic mapping of a gene for autosomal recessive retinitis pigmentosa on chromosome 6p21. *Am. J. Hum. Genet.* **57**, 499-502.
- Shyjan, A.W., de Sauvage, F.J., Gillet, N.A., Goeddel, D.V. and Lowe, D.G. (1992) Molecular cloning of a retina-specific membrane guanylyl cyclase. *Neuron* **9**, 727-737.
- Siegel, S. and Castellan, N.J. (1988) Non parametric statistics: for behavioural sciences. Second edition. McGraw-Hill international editions.
- Silverman, G.A., Ye, R.D., Pollock, K.M., Sadler, J.E. and Korsmeyer, S.J. (1989) Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci. USA* **86**, 7485-7489.
- Small, K.W., Syrquin, M., Mullen, L. and Gehrs, K. (1996) Mapping of autosomal dominant cone degeneration to chromosome 17p. *Am. J. Ophthalmol.* **121**, 13-18.
- Smith, D.P., Stamnes, M.A. and Zuker, C.S. (1991) Signal transduction in the visual system of Drosophila. *Annu. Rev. Cell Biol* **7**, 161-190.
- Smith, M.W., Holmsen, A.L., Wei, Y.H., Peterson, M. and Evans, G.A. (1994) Genomic sequence sampling: a strategy for high-resolution sequence-based physical mapping of complex genomes. *Nat. Genet. ics* **7**, 40-47.
- Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.
- Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilo base pairs. *Proc. Natl. Acad. Sci. USA* **87**, 103-107.
- Sternberg, N. (1994) The P1 cloning system-past and future. *Mammalian Genome* **5**, 397- 404.
- Strachan, T., Abitol, M., Davidson, D and Beckmann, J.S. (1997) A new dimension for the human genome project: towards comprehensive expression maps. *Nat. Genet.* **16**, 126-132.
- Stryer, L. (1986) Cyclic GMP cascade of vision. *Annu. Rev. Neurosci.* **9**, 87-119.

- Stryer, L. (1988). *Biochemistry*; third edition. W.H. Freeman and Company, New York.
- Stubbs, L. and 11 others (1996) Detailed comparative map of human chromosome 19q and related region of the mouse genome. *Genomics* **35**, 499-508.
- Subbaraya, I., Ruiz, C.C., Helekar, B.S., Zhao, X., Gorczyca, W.A., Pettenati, M.J., Rao, P.N., Palczewski, K. and Baehr, W. (1994) Molecular characterisation of human and mouse photoreceptor guanylate cyclase-activating protein (GCAP) and chromosomal localisation of the human gene. *J. Biol. Chem.* **269**, 31080-31089.
- Suber, M.L., Pittler, S.J., Qin, N., Wright, G.C., Holcombe, V., Lee, R.H., Craft, C.M., Lolley, R.N., Baehr, W. and Hurwitz, R.L. (1993) Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase β -subunit gene. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3968-3979.
- Sun, H. and Nathans, J. (1997) Stargardt's ABCR is localised to the disc membrane of retinal rod outer segments. *Nat. Genet.* **17**, 15-16.
- Sung, C-H., Davenport, C.M., Hennessey, J.C., Maumenee, I.H., Jacobson, S.G., Heckenlively, J.R., Nowakowski, R., Fishman, G., Gouras, P. and Nathans, J. (1991) Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6481-6485.
- Sung, C-H., Davenport, C.M. and Nathans, J. (1993) Rhodopsin mutations responsible for autosomal dominant retinitis pigmentosa. Clustering of functional classes along the polypeptide chain. *J. Biol. Chem.* **268**, 26645-26649.
- Sung, C-H., Mackino, C., Baylor, D. and Nathans, J. (1994) A rhodopsin gene mutation responsible for autosomal dominant retinitis pigmentosa results in a protein that is defective in localisation to the photoreceptor outer segment. *J. Neurosci.* **14**, 5818-5833.
- Suzuki, S. and Kaneko, A. (1990) Identification of bipolar cell subtypes by protein kinase C-like immunoreactivity in the gold fish retina. *Vis. Neurosci.* **5**, 223-230.
- Swaroop, A., Xu, J.Z., Pawar, H., Jackson, A., Skolnick, C. and Agarwal, N. (1992) A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 266-270.
- Szylyk, J.P., Fishman, G.A., Alexander, K.R., Peachey, N.S., and Derlacki, D.J. (1993) Clinical subtypes of cone-rod dystrophy. *Arch. Ophthalmol.* **111**, 781-788.
- Terasawa, M., Hagiwara, M., Hachiya, T., Kobayashi, R., Obata, S., Wakabayashi, T. and Hidaka, H. (1991) Identification and characterisation of protein kinase C-related enzymes in the frog retina. *Arch. Biochem. Biophys.* **287**, 213-217.
- Tergwilliger, J.D. and Ott, J. (1994) *Handbook of human genetic linkage* (Baltimore: Johns Hopkins University Press).

- The Utah Marker Development Group (1995) A collection of ordered tetranucleotide-repeat markers from the human genome. *Am. J. Hum. Genet.* **57**, 619-628.
- Trask, B., Christensen, M., Fertitta, A., Bergmann, A., Ashworth, L., Branscomb, E., Carrano, A. and Van den Engh, G. (1992) Fluorescence in situ hybridisation mapping of human chromosome 19: mapping and verification of cosmid contigs formed by random restriction fingerprinting. *Genomics* **14**, 162-167.
- Trask, B.J., Pinkel, D. and Van den Engh, G. (1989) The proximity of DNA sequences in interphase nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710-717.
- Trask, B.J., Massa, H., Kenwick, S. and Gitschier, J. (1991) Mapping of human chromosome Xq28 by two-colour fluorescence in situ hybridisation of DNA sequences to interphase cell nuclei. *Am. J. Hum. Genet.* **48**, 1-15.
- Travis, G.H., Brennan, M.B., Danielson, P.E., Kozak, C.A. and Sutcliffe, J.G. (1989) Identification of a photoreceptor-specific mRNA encoded by the gene responsible for retinal degeneration slow (*rds*). *Nature* **338**, 70-73.
- Travis, G.H., Sutcliffe, J.G. and Bok, D. (1991) The retinal degeneration slow (*rds*) gene product is a photoreceptor disc membrane associated glycoprotein. *Neuron* **6**, 61-70.
- Tripathy, R.C. and Tripathy, B.J. (1984) *The eye* (ed. Davson, H.), Academic press.
- Tsilfidis, C., MacKenzie, A.E., Mettler, G., Barcelo, J. and Korneluk, R.G. (1992) Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nat. Genet.* **1**, 192-195.
- Tunnacliffe, A., Parker, M., Povey, S., Bengtsson, B.O., Stanley, K., Solomon, E. and Goodfellow, P. (1983) Integration of Eco-gt and SV40 early region sequences into human chromosome 17: a dominant selection system in whole cell and microcell human-mouse hybrids. *EMBO J.* **2**, 1577-1584
- Udovichenko, I.P., Cunnick, J., Gonzales, K. and Takemoto, D.J. (1993) Phosphorylation of bovine photoreceptor cyclic GMP phosphodiesterase. *Biochem. J.* **295**, 49-55.
- Udovichenko, I.P., Cunnick, J., Gonzales, K. and Takemoto, D.J. (1994) Functional effect of phosphorylation of the photoreceptor phosphodiesterase inhibitory subunit by protein kinase C. *J Biol. Chem.* **269**, 9850-9856.
- Udovichenko, I.P., Newton, A.C. and Williams, D. (1997) Contribution of protein kinase C to the phosphorylation of rhodopsin in intact retinas. *J Biol. Chem.* **272**, 7952-7959.
- Van Nie, R., Ivanyi, D. and Demant, P. (1978) A new H-2 linked mutation, *rds*, causing retinal degeneration slow in mouse. *Tissue Antigens.* **12**, 106-108.

- Van Soest, S., et al (1994) Assignment of a gene for autosomal recessive retinitis pigmentosa (RP12) to chromosome 1q31-q32.1 in an inbred and genetically heterogeneous disease population. *Genomics* **22**, 499-504.
- Vithana, E., Al-Magthteh, M., Bhattacharya, S.S. and Inglehearn, C.F. (1998) RP11 is the second most common locus for dominant retinitis pigmentosa. *J. Med. Genet.* **35**, 174-175.
- Wald, G. (1968) The molecular basis of visual excitation. *Nature* **219**, 800-807.
- Walter, M.A., Spillet, D.J., Thomas, P., Weissenbach, J. and Goodfellow, P.N. (1994) A method for constructing radiation hybrid maps of whole genomes. *Nat. Genet.* **7**, 22-28.
- Warburg, M., Sjo, O. and Fledelius, H.C. (1991) Deletion mapping of a retinal cone-rod dystrophy: assignment to 18q21.1. *Am. J. Med. Genet.* **39**, 288-293.
- Warburton, D., Gersend, s., Yu, M.T., Jackson, C., Handelin, B. and Housman, D. (1990) Monochromosomal rodent-human hybrids from microcell fusion of human lymphoblastoid cells containing an inserted dominant selectable marker. *Genomics* **6**, 358-366.
- Wässle, H., Yamashita, M., Greferath, U., Grünert, U. and Müller, F. (1991) The rod bipolar cell of the mammalian retina. *Vis. Neurosci.* **7**, 99-112.
- Weber, J.L. (1990) Informativeness of human (dC-dA)_n. (dG-dT)_n polymorphisms. *Genomics* **7**, 524-530.
- Weber, J.L. and May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388-396.
- Weber, J.L., Wang, Z., Hansen, K., Stephenson, M., Kappel, C., Salzman, S., Wilkie, P.J., Keats, B., Dracopoli, N.C., Brandriff, B.F. et al. (1993) Evidence for human meiotic recombination interference obtained through construction of a short tandem repeat-polymorphism linkage map of chromosome 19. *Am. J. Hum. Genet.* **44**, 1079-1095.
- 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.
- Weleber, R.G. and Eisner, A. (1988) Cone degeneration ('Bull's-eye dystrophies') and colour vision defects. In Newsome, D.A (ed.), *Retinal Dystrophies and Degenerations*. Raven Press, New York, pp233-256.
- Weleber, R.G., Carr, R.E., Murphy, W.H., Sheffield, V.C. and Stone, E.M. (1993) Phenotypic variation including retinitis pigmentosa, pattern dystrophy, and fundus flavimaculatus in a single family with a deletion of codon 153 or 154 of the peripherin/RDS gene. *Arch. Ophthalmol.* **111**, 1531-1542.

- Wetsel, W. C., Kahn, W. A., Merchenthaler, I., Rivera, H., Halpern, A.E., Phung, H.M., Negro-Vilar, A. and Hannun, Y. (1992) Tissue and cellular distribution of the extended family of protein kinase C isoenzymes. *J. Cell Biol.* **117**, 121-133.
- Whittaker, P.A., Campbell, A.J.B., Southern, E. and Murray, N. (1988) Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. *Nucleic Acids Res.* **16**, 6735-6736.
- Wilcox, A.S., Khan, A.S., Hopkins, J.A and Sikela, J.M. (1991) Use of 3' untranslated sequences of human cDNA for rapid chromosome assignment and conversion to STSs: implication for an expression map of the genome. *Nucleic. Acids. Res* **19**, 1837-1843.
- Wilden, U., Hall, S.W. 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. U.S.A.* **83**, 1174-1178.
- Willmotte, R., Marechal, J., Morle, L., Baklouti, F., Philippe, N., Kastally, R., Kotula, L., et al (1993) Low expression allele α LELY of red cell spectrin is associated with mutations in exon 40 (α V/41 polymorphism) and intron 45 and with partial skipping of exon 46. *J. Clin. Invest.* **91**, 2091-2096.
- Wolbring, G. and Cook, N.J. (1991) Rapid purification and characterisation of protein kinase C from bovine retinal outer segments. *Eur. J. Biochem.* **201**, 601-606.
- Woo, S.S., Jiang, J., Gill, B.S., Patterson, A.H. and Wing, R.A. (1994) Construction and characterisation of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Res.* **22**, 4922-4931.
- Wood, J. P. M., McCord, R. J. and Osborne, N. N. (1997) Retinal protein kinase C. *Neurochem. Int.* **30**, 119-136
- Wright, A.F., Bhattacharya, S.S., Clayton, J.F., et al. (1987) Linkage relationships between X-linked retinitis pigmentosa and nine short arm markers; exclusion of the disease locus from Xp21 and localisation to between DXS7 and DXS14. *Am. J. Hum. Genet.* **41**, 635-644.
- Wright, A.F. (1997a) A searchlight through the fog. *Nat. Genet.* **17**, 132-134.
- Wright, A.F. (1997b) Gene therapy for the eye. *Br. J. Ophthalmol.* **81**, 620-623.
- Wu, C., Zhu, S., Simpson, S. and de Jong, P.J. (1996) DOP- Vector PCR: a method for rapid isolation and sequencing of insert termini from PAC clones. *Nucleic Acids Research* **24**, 2614-2615
- Xu, S., Nakazawa, M., Tamai, M. and Gal, A. (1995) Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J. Med. Genet.* **32**, 915-916

- Xu, S.Y., Schwartz, M., Rosenburg, T. and Gal, A. (1996), A ninth locus (RP18) for autosomal dominant retinitis pigmentosa maps in the pericentromeric region of chromosome 1. *Hum. Mol. Genet.* **5**, 1193-1197.
- Yagasaki, K., and Jacobson, S.G. (1989) Cone-rod dystrophy: phenotypic diversity by retinal function testing. *Arch. Ophthalmol.* **107**, 701-708.
- Yahraus, T., Braverman, N., Dodt, G., Kalish, J.E., Morrell, J.C., Moser, H.W., Valle, D. and Gould, S.J. (1996) The peroxisome biogenesis disorder group 4 gene, PXAAA1, encodes a cytoplasmic ATPase required for stability of the PTS1 receptor. *EMBO. J.* **15**, 2914-2923.
- Yamashita, M. and Wässle, H. (1991) Responses of rod bipolar cells isolated from the rat retina to the glutamate agonist 2-amin-4 phosphonobutyric acid (APB). *J. Neurosci.* **11**, 2372-2382.
- Yamamoto, S., Sippel, K.C, Berson, E.L. and Dryja, T.P (1997) Defects in the rhodopsin kinase gene in the Oguchi form of stationary night blindness. *Nat. Genet.* **15**, 175-178.
- Yang-Feng, T.L. and Swaroop, A. (1992) Neural retina-specific leucine zipper gene NRL (D14S46E) maps to human chromosome 14q11.1-q11.2. *Genomics* **14**, 491-492.
- Yarfitz, S. and Hurley, J.B. (1994) Transduction mechanisms of vertebrate and invertebrate photoreceptors. *J Biol. Chem.* **269**, 14329-14332
- Yau, K. W. and Baylor, D.A. (1989) Cyclic GMP-activated conductance of retinal photoreceptor cells. *Annu. Rev. Neurosci.* **12**, 289-327.
- Young, R.W. and Bok, D. (1969) Participation of the retinal pigment epithelium in the rod outer segment renewal process. *J. Cell Biol.* **42**, 392-402.
- Young, R.W. (1984) Cell death during differentiation of the retina in the mouse. *J. Comp. Neurol.* **229**, 362-373.
- Ziegle, J.S., Su, Y., Corcoran K.P., Nie, L., Mayrand, P.E., Hoff, L.B., McBride, L.J., Kronick, M.N. and Diehl, S.R. (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* **14**, 1026-1031.
- Zuker, C.S. (1996) The biology of vision of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 571-576.

Am. J. Hum. Genet. 62:1248–1252, 1998

Segregation of a *PRKCG* Mutation in Two RP11 Families

To the Editor:

Retinitis pigmentosa (RP) is a group of inherited neurodegenerative disorders of the retina. RP patients experience night blindness and tunnel vision (constricted visual fields) at an early stage and may become completely blind in the advanced stages of the disease (Bird 1995). RP is both clinically and genetically heterogeneous. The autosomal dominant subgroup of RP (adRP) can be caused by mutations in at least two genes, *rhodopsin* (3q21) and *peripherin* (6p12), and in seven other loci mapped by linkage analysis (Xu et al. 1996). We mapped the chromosome 19q locus (also known as "RP11" [MIM 600138; <http://www.ncbi.nlm.nih.gov/htbin-post/Omim>]; McKusick 1994) in four British families (Al-Magthteh et al. 1996). A Japanese family (Xu et al. 1995) and three American families (McGee et al. 1997) also had linkage to this locus, suggesting that RP11 is a major locus for adRP (Al-Magthteh et al. 1996). Interestingly, the phenotype in these families is characterized by "bimodal expressivity"; symptomatic RP patients have a relatively early onset of the disease, whereas some obligate disease-gene carriers are indistinguishable from normal individuals. Recombination events in the families with linkage have refined the localization of the RP11 gene to a 5-cM interval between markers D19S572 and D19S418, in the telomeric region of chromosome 19q13.4 (McGee et al. 1997). Among the positional candidate genes and expressed sequence tags mapping in this region is the gene *PRKCG*, a member of the protein kinase C (PKC) gene family (Hug and Sarre 1993).

PKC is a multifunctional family of closely related serine/threonine protein kinases. PKCs function in a wide variety of cellular processes, such as membrane-receptor signal transduction and control of gene expression. Various PKC isoenzymes are expressed in a tissue-specific manner (Hug and Sarre 1993). The PKC gene family has been shown to be expressed widely in the retina, although it is not clear which isoenzymes are involved in any particular retinal cell type (Newton and Williams

1993; Ohki et al. 1994). The observation by Newton and Williams (1991; 1993) that PKC phosphorylates rhodopsin in a light-dependent manner suggests the involvement of these kinases in desensitization of the photopigment. In *Drosophila* phototransduction, a PKC isoenzyme known as "eye-PKC" was found to be exclusively expressed in photoreceptor cells (Schaeffer et al. 1989). A mutation in eye-PKC has been shown to be responsible for the recessive *inaC* (inactivation no after potential C) *Drosophila* mutant, which exhibits photoreceptor deactivation and retinal degeneration (Smith et al. 1991). However, the specific-tissue expression and function of the PKC γ isoenzyme in the human retina is not yet clearly defined. Several reports have shown PKC γ to be expressed in rabbit, rat, frog, and goldfish retinas (Osborne et al. 1992). This expression of PKC γ is confined to amacrine and ganglion cells. No expression has been detected in the photoreceptors. This does not exclude the possibility that PKC γ expression in photoreceptors is at levels below the sensitivity of the detection method. Therefore, in addition to its being a positional candidate, there is circumstantial evidence to implicate *PRKCG* as a candidate for the RP11 gene. We have recently characterized the genomic structure of the *PRKCG* gene. The gene has 18 exons, and the intron-exon boundaries, including splice-site consensus sequences, have been defined. Amplification primers for each exon are listed in table 1.

In this study, we screened the *PRKCG* gene for mutations, in RP11 families. All 18 exons and the 1.5-kb 5' UTR were screened by both heteroduplex analysis (Keen et al. 1991) and direct genomic sequencing. We identified a point mutation that segregates with RP in two families, RP1907 (Al-Magthteh et al. 1996) and ADRP24 (denoted as "family 2" by Moore et al. 1993; also see the legend to fig. 3). Another two isolated RP patients with a family history indicating dominant disease were also found to have the same mutation. This mutation was a C→A transversion, which substitutes a serine for an arginine residue at codon 659 of the *PRKCG* gene within the C4 catalytic domain (fig. 1). This residue is conserved in all known *PRKCG* genes, including human, bovine, and rat genes. The presence and segregation of this mutation (the sequence of which is shown in fig. 2) was also confirmed by digestion of the PCR product of exon 18 by the restriction enzyme

Table 1

Primer Sequences (5'-3') Used in Amplification of the 18 *PRKCG* Exons, and PCR Conditions for each Amplification

Forward Primer	Reverse Primer	[Mg ⁺⁺], Annealing Temperature
1F, agaaaggcaggatcctggtc	1R, cggcgtgataggagtctgca	1.5, 65°C
2F, ttggacacctggccctgc	2R, ctgagggtcccaggagcc	1.5, 65°C
3F, gctggactaatccatgctc	3R, aggagaaattgggacggagc	1.3, 60°C
4F, gctgacctagagagcaaggc	4R, gctttggaaggccctggca	1.5, 65°C
5F, tgagggtctaccgcagctt	5R, acaagtgccttgggtcagcc	1.5, 64°C
6F, ctctaaccctgcaactctt	6R, tctgacagctgtcattgctt	1.5, 60°C
7F, gccatgagctcggctctgca	7R, gtaatattgcctccatcccc	1.5, 65°C
8F, tgcctctccatgggtgc	8R, aaggccagctctgaacctt	1.5, 60°C
9F, ctatctatgccatggct	9R, aactgctccattcaacg	1.5, 58°C
10F, gagcatttccctatcggtg	10R, aaccagaaatctgaccttccc	1.3, 55°C
11F, aggtcctgtaccactgggtt	11R, atcccaacgcagatgtccag	1.5, 65°C
12F, gtatgttgatcccgcctcta	12R, acgtcagaaggctcagtggtt	1.0, 58°C
13F, agccaactgacctctgactt	13R, gtgttgagttcagcagttctag	1.5, 60°C
14F, ctgactgctgaactcaaac	14R, taagggatctcaaagcgtg	1.5, 56°C
15F, gcacttaacgtgggtagcg	15R, tagccaagccagcttctcc	1.5, 56°C
16F, gcattgcccctgactcttat	16R, agtgacttcaggaatggggag	1.5, 60°C
17F, atgtacctgtccggcact	17R, accaggtttttgttgctgg	1.5, 55°C
18F, ctggagctgcttaactttcc	18R, acgttggggacacctagtgg	1.5, 64°C

Ac1 (fig. 3). This mutation was absent from 500 normal control chromosomes. Haplotype analysis in these two families and in the two single adRP cases with markers flanking the *PRKCG* gene reveals a founder effect, with disease-gene carriers of each pedigree sharing the same haplotype over an interval >5 cM, including the entire RP11 region (this haplotype would be expected to have a frequency of 10^{-5} , on the basis of allele frequencies in the U.K. population; see fig. 3). This could imply that the mutation on the ancestral chromosome took place

relatively recently, although no genealogical link exists between these families during the past 160 years.

Close genetic association between the RP11 phenotype in these families and the Arg659Ser mutation is clearly evident. In conjunction with the involvement of an eye-PKC gene in retinal degeneration in *Drosophila* (Smith et al. 1991), this suggests that this mutation could be a cause of RP in these families. As stated above, one form of PKC is known to phosphorylate rhodopsin, perhaps as part of an adaptation mechanism. If the enzyme

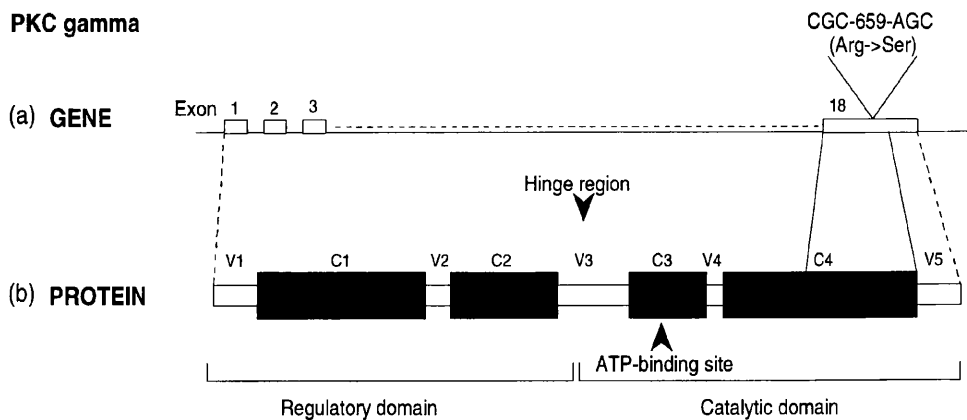


Figure 1 Diagrammatic representation of the position of the Arg659Ser mutation within exon 18 of (a) the DNA sequence and (b) the domain structure of PKC γ , as described by Hug and Sarre (1993). C1-C4 are the conserved domains, which are flanked and separated by the variable domains, V1-V5. V1-C2 represents the regulatory domain, which is separated from the catalytic domain (C3-V5) by V3 or the hinge region. C3 contains the ATP-binding site.

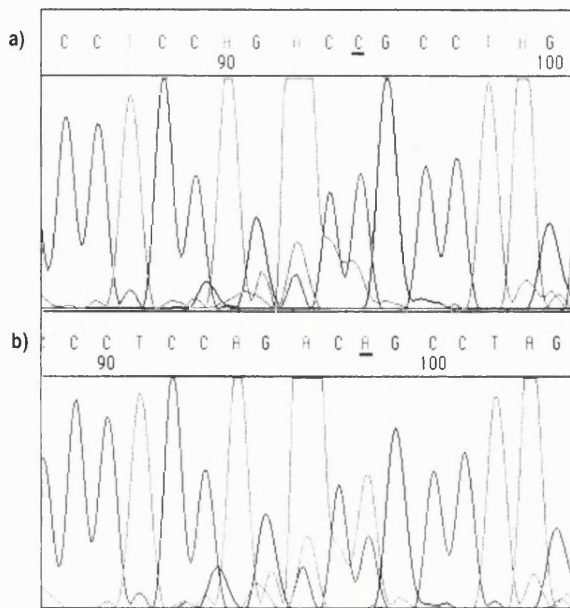


Figure 2 Comparison of sequencing data from (a) a normal and (b) a heterozygous mutant *PRKCG* exon 18 sequence. A C→A transition (underlined bases) changes codon 659 from arginine (CGC) to serine (AGC).

involved were *PRKCG*, a mutation in this gene might be expected to have an effect on photoreceptors' adaptation or recovery in response to light flashes or to high or low light levels. Mutations in rhodopsin kinase and arrestin, both similarly involved in the restoration of resting potential in photoreceptors after light stimulus, have been shown to cause a rare recessive form of congenital stationary night blindness (CSNB), known as "Oguchi disease" (Fuchs et al. 1995; Yamamoto et al. 1997). As in the case of Oguchi disease, one might expect a mutation in *PKC γ* , an enzyme, to cause a recessive rather than a dominant phenotype. However, a mutation that affects the ability of the photoreceptor to restore rhodopsin sufficiently to an inactive state might lead to a background level of constitutive activation, which has been hypothesized, for several rhodopsin mutations, as being the cause of a dominant form of CSNB (Rao et al. 1994). Furthermore, a mutation in the β -subunit of phosphodiesterase (*PDE β*), an enzyme, was also found to cause a dominant form of CSNB (Gal et al. 1994). Mice deficient in *PKC γ* have been created by use gene-knockout technology. These animals exhibit mild deficits in spatial and contextual learning, but no mention is made, in the published description, of defective vision (Abeliovich et al. 1993; Chen et al. 1995). If *PRKCG* is the RP11 gene in these families, then it could be argued that these mice are not a true model for a human *PRKCG* missense mutation—in which the presence of

the abnormal protein is likely to cause the disease, rather than the absence of normal enzyme activity. Alternatively, given the incomplete penetrance of the RP11 phenotype in some patients, this may be the result of a genetic background that protects against retinal degeneration in the mouse inbred strain.

Nevertheless, the entire *PRKCG* gene has been sequenced in three other families with RP11 linkage, and no disease-causing mutation has been found. Southern blots digested with various restriction enzymes and hy-

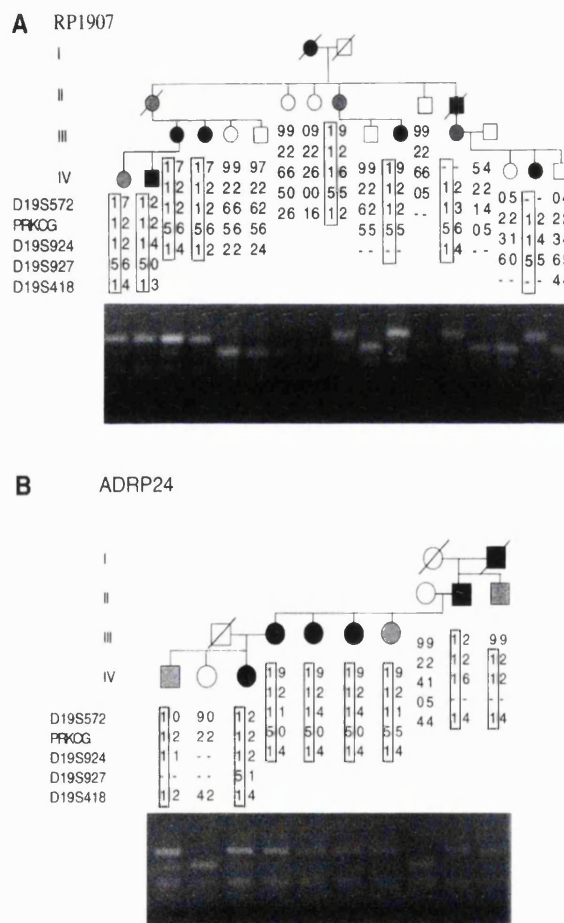


Figure 3 Pedigrees of (a) RP1907 and (b) ADRP24. (On the basis of the characteristic bimodal expressivity phenotype, this family was thought to be an RP11 family, but it gave a nonsignificant LOD score [0.6 at recombination fraction 0] with 19q markers.) Blackened symbols denote affected individuals; unblackened symbols denote normal individuals; and gray-shaded symbols denote asymptomatic disease-gene carriers, on the basis of haplotype analysis. The linked haplotype for 19q markers, including the *PRKCG* mutation, are shown below the pedigrees. Allele 1 for *PRKCG* represents the C→A mutation at codon 659, whereas allele 2 represents the normal sequence. Restriction digests of exon 18 PCR products, which demonstrate absence of the *Acl* site in mutated alleles, are also shown, with fragment sizes of 110 bp (for the mutated allele), 81 bp (for the normal allele), and 49 bp.

bridized with a *PRKCG* cDNA probe also failed to disclose any rearrangements in these families. It is possible that, in these families, another, probably common, disease-causing mutation could have been missed or could lie in either upstream or downstream promoter/regulatory regions or introns that have not yet been fully characterized. A search for other RP patients carrying the Arg659Ser or other mutations in the *PRKCG* gene, in other laboratories involved in RP research, would further substantiate the involvement of this gene in this common form of dominant RP. Alternatively, the absence of causative mutations in a proportion of RP11 families might imply microheterogeneity, a hypothesis proposed to explain the apparent lack of mutations in *RPGR* in >50% of families with RP3 linkage (Fujita et al. 1997). Nevertheless, it also remains possible that the Arg659Ser is a rare allelic variant of the *PRKCG* normal sequence and is in linkage disequilibrium with the disease allele in these families. A search for other RP patients carrying this change may reveal a wider founder effect, which could further refine the locus for this common form of dominant RP.

Finally, a number of apparently non-disease-causing *PRKCG* polymorphisms were also detected by this study. One particularly interesting change in exon 18 (TTT647TTC) was found to segregate perfectly with the disease phenotype in another RP11-linked family (data not shown). Despite the fact that this change was absent from the 500 normal control chromosomes, it does not alter the amino acid sequence. Other neutral nucleotide-sequence changes (AAT189AAC and GGC411GGT), non-disease-causing amino acid changes (R141C, H415Q, and A523D), and a T-nucleotide insertion in intron 16, 39 bp downstream of exon 16, were also found. None of these changes were found to segregate with the disease phenotype in the families in which they were found, and each was also present in normal individuals.

Acknowledgments

We thank Drs. Jude Fitzgibbon and Martin Warren for helpful discussions. This work was supported by Wellcome Trust grants 043006/Z/94, 042375/Z/94, and 035535/Z/92; the Foundation Fighting Blindness USA; and the British Retinitis Pigmentosa Society.

MAI AL-MAGHTHEH,¹* ERANGA N. VITHANA,¹*
CHRIS F. INGLEHEARN,² TONY MOORE,¹
ALAN C. BIRD,¹ AND SHOMI S. BHATTACHARYA¹

¹Department of Molecular Genetics, Institute of Ophthalmology, University College London, London; and

²Molecular Medicine Unit, St. James's University Hospital, Leeds

References

- Abeliovich A, Paylor R, Chen C, Kim JJ, Wehner JM, Tonegawa S (1993) PKC γ mutant mice exhibit mild deficits in spatial and contextual learning. *Cell* 75:1263–1271
- Al-Maghteh M, Vithana E, Tarttelin E, Jay M, Evans K, Moore T, Bhattacharya SS, et al (1996) Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11), and association with a unique bimodal expressivity phenotype. *Am J Hum Genet* 59:864–871
- Bird AC (1995) Retinal photoreceptor dystrophies. *Am J Ophthalmol* 119:543–562
- Chen C, Kano M, Abeliovich A, Chen L, Bao S, Kim JJ, Hashimoto K, et al (1995) Impaired motor coordination correlates with persistent multiple climbing fiber innervation in PKC γ mutant mice. *Cell* 83:1233–1242
- Fuchs S, Nakazawa M, Maw M, Tamai M, Oguchi Y, Gal A (1995) A homozygous 1-base pair deletion in the arrestin gene is a frequent cause of Oguchi disease in Japanese. *Nat Genet* 10:360–362
- Fujita R, Buracznska M, Gieser L, Wu WP, Forsythe P, Abrahamson M, Jacobson SG, et al (1997) Analysis of the *RPGR* gene in 11 pedigrees with the retinitis pigmentosa type 3 genotype: paucity of mutations in the coding region but splice defects in two families. *Am J Hum Genet* 61:571–580
- Gal A, Orth U, Baehr W, Schwinger E, Rosenberg T (1994) Heterozygous missense mutation in the rod cGMP phosphodiesterase beta-subunit gene in autosomal dominant stationary night blindness. *Nat Genet* 7:64–68
- Hug H, Sarre TF (1993) Protein kinase C isoenzymes: divergence in signal transduction. *Biochem J* 291:329–343
- Keen J, Lester DH, Inglehearn CF, Curtis A, Bhattacharya SS (1991) Rapid detection of single base mismatches as heteroduplexes on hydrolink gels. *Trends Genet* 7:5
- McGee TL, Devoto M, Ott J, Berson EL, Dryja TP (1997) Evidence that the penetrance of mutations at the RP11 locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous RP11 allele. *Am J Hum Genet* 61:1059–1066
- McKusick VA (1994) Mendelian inheritance in man: a catalog of human genes and genetic disorders. Johns Hopkins University Press, Baltimore
- Moore AT, Fitzke F, Jay M, Arden GB, Inglehearn CF, Keen TJ, Bhattacharya SS, et al (1993) Autosomal dominant retinitis pigmentosa with apparent incomplete penetrance: a clinical, electrophysiological, psychophysical and molecular genetic study. *Br J Ophthalmol* 77:473–479
- Newton AC, Williams DS (1991) Involvement of protein kinase C in the phosphorylation of rhodopsin. *J Biol Chem* 266:17725–17728
- (1993) Rhodopsin is the major *in situ* substrate of protein kinase C in rod outer segments of photoreceptors. *J Biol Chem* 268:18181–18186
- Ohki K, Yoshida K, Imaki J, Harada T, Matsuda H (1994) The existence of protein kinase C in cone photoreceptors in the rat retina. *Curr Eye Res* 13:547–550
- Osborne NN, Barnett ND, Morris NJ, Huang FL (1992) The occurrence of three isoenzymes of protein kinase C (α , β , and γ) in retinas of different species. *Brain Res* 570:161–166
- Rao VR, Cohen GB, Oprian DD (1994) Rhodopsin mutation

- G90D and a molecular mechanism for congenital night blindness. *Nature* 367:639–641
- Schaeffer E, Smith D, Mardon G, Quinn, W, Zuker C (1989) Isolation and characterization of two new drosophila protein kinase C genes, including one specifically expressed in photoreceptor cells. *Cell* 57:403–412
- Smith DP, Ranganathan R, Hardy RW, Marx J, Tsuchida T, Zuker CS (1991) Photoreceptor deactivation and retinal degeneration mediated by a photoreceptor-specific protein kinase C. *Science* 254:1478–1484
- Xu S, Nakazawa M, Tamai M, Gal A (1995) Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J Med Genet* 32:915–916
- Xu SY, Schwartz M, Rosenberg T, Gal A (1996) A ninth locus (RP18) for autosomal dominant retinitis pigmentosa maps in the pericentromeric region of chromosome 1. *Hum Mol Genet* 5:1193–1197
- Yamamoto S, Sippel KC, Berson EL, Dryja TP (1997) Defects in the rhodopsin kinase gene in the Oguchi form of stationary night blindness. *Nat Genet* 15:175–178

Address for correspondence and reprints: Dr. Shomi S. Bhattacharya, Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, United Kingdom. E-mail: smbcssb@ucl.ac.uk

* These authors contributed equally to this research.

© 1998 by The American Society of Human Genetics. All rights reserved.
0002-9297/98/6205-0030\$02.00

LETTERS TO THE EDITOR

Homozygosity for Asn86Ser mutation in the CuZn-superoxide dismutase gene produces a severe clinical phenotype in a juvenile onset case of familial amyotrophic lateral sclerosis

A 13 year 9 month old girl, the daughter of consanguineous (first cousin) Pakistani parents, presented with pain in her right calf of two months duration and increasing weakness in her right leg with an inability to bear weight. Neurological deficit was initially confined to the right lower limb. Examination showed an asymmetrical lower motor neurone pattern with distal hypotonia and weakness, wasting in the right quadriceps and peroneal muscles, and absent reflexes and flexor plantar responses. After one week of inpatient observation muscle weakness progressed and lower amplitude compound muscle potentials were also seen in the upper limb, confirming an ascending picture of disease progression. The muscle weakness extended to involve facial muscles with definite weakness distally in the right upper limb and profound weakness in the lower limbs. Further progression led to reduced expiratory air flow and pneumonia with respiratory failure and death 14 weeks after presentation.

The predominant clinical features in this patient were progressive asymmetrical lower motor neurone weakness and wasting with more marked involvement of lower limbs and distal musculature, together with bulbar involvement. A paternal uncle had also recently died, aged 34, following an 11 month illness with rapidly progressive motor neurone disease with initial spastic features and later a flaccid symmetrical pattern. These observations suggested a diagnosis of familial motor neurone disease of the progressive muscular atrophy type.

Blood samples were collected from this patient, her parents, her paternal grandmother, and a paternal aunt and uncle. Genomic DNA was isolated and screened for mutations in the superoxide dismutase (SOD1) gene by PCR amplification, single strand conformation polymorphism analysis, and DNA sequencing. An aberrant band shift was identified in amplified DNA in exon 4 of the SOD1 gene. The amplified DNA was sequenced and a single base change was identified (nucleotide 257A→G) causing an amino acid change of asparagine (Asn) (AAT) to serine (Ser) (AGT) at codon 86. The proband was homozygous for this mutation, both of her parents (III.1, III.2) (fig 1), and her paternal uncle (III.7) were heterozygous, while her aunt (III.4) and grandmother (II.4) were homozygous for the normal allele.

In order to check that this sequence change was not a harmless polymorphism, 67

Scottish amyotrophic lateral sclerosis (ALS) patients, 60 anonymous Scottish controls, and 84 anonymous healthy Pakistani controls were also analysed using allele specific oligonucleotides and dot blot hybridisation. All were homozygous for the normal allele at codon 86. This codon is extremely well conserved throughout the animal and plant kingdom, specifying Asn in 54 different species in the EMBL-SWISS-PROT alignment database.¹ This fact alone implies that this amino acid is functionally important.

The family described here is extensive with many first cousin marriages and potentially many heterozygotes or homozygotes for this mutation. To date two people have died of ALS. The proband was homozygous for the mutation Asn86Ser, while her paternal uncle was probably heterozygous (fig 1). The other heterozygotes in this family appear to be clinically healthy. Therefore, at present, it is not possible to predict the ALS risk factor associated with heterozygosity of Asn86Ser. Recently, a Japanese family with two affected members heterozygous for the Asn86Ser mutation has been reported.² The clinical phenotypes of these two subjects were very different.³ The father died of respiratory failure at the age of 56 years, four years after onset of symptoms but with only upper body involvement. His daughter, diagnosed at the age of 36 years with lower limb weakness and slow progression of the disease to a general muscular atrophy nine years after onset of the first symptoms, survived for more than 11 years.

There has only been one previous report of homozygosity for a SOD1 mutation in ALS. An autosomal recessive form of FALS in a Swedish/Finnish population has been described in nine families.⁴ Homozygosity for the Asp90Ala mutation caused ALS in 30/37 subjects identified. There have been no reports of heterozygosity for the mutation associated with ALS in this population. However, the heterozygous form of the same mutation identified in Belgian subjects has led to the development of ALS.⁴

The clinical phenotype associated with a given mutation in the SOD1 gene is clearly not only dependent on the mutation itself, but may also be influenced by the genetic background of the patient, and possibly by environmental factors. The Asn86Ser mutation has only been identified in a small number of people as yet. In our Pakistani family there were two patients with ALS. The proband was homozygous for Asn86Ser and had a more severe form of disease than her uncle, who was probably heterozygous. Along

with the possible complication of variable penetrance, the true relationship between this mutation and ALS may not be discovered for some time.

CAROLINE HAYWARD
DAVID J H BROCK

Human Genetics Unit, Molecular Medicine Centre,
University of Edinburgh, Western General Hospital,
Edinburgh EH4 2XU, UK

ROBERT A MINNS

Department of Paediatric Neurology, Royal Hospital
for Sick Children, Edinburgh EH9 1LF, UK

ROBERT J SWINGLER

Department of Neurology, Dundee Royal Infirmary,
Dundee DD1 9ND, UK

- Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its new supplement TrEMBL. *Nucleic Acids Res* 1996;24:21-5.
- Radunovic A, Leigh PN. Cu/Zn superoxide dismutase gene mutations in amyotrophic lateral sclerosis: correlation between genotype and clinical features. *J Neurol Neurosurg Psychiatry* 1996;61:565-72.
- Kurahashi K, Okushima T, Kimura K, Narita S, Matsunaga M. Different phenotype and clinical course in a family with motor neurone disease. *Med J Aomori* 1993;38:142-6.
- Anderson PM, Nilsson P, Ala-Hurala V, et al. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in Cu/Zn superoxide dismutase. *Nat Genet* 1995; 10:61-6.
- Anderson PM, Forsgren L, Binzer M, et al. Autosomal recessive adult-onset amyotrophic lateral sclerosis associated with homozygosity for Asp90Ala CuZn-superoxide dismutase mutation. A clinical and genealogical study of 36 patients. *Brain* 1996;119:1153-72.
- Robberecht W, Aguirre T, Van den Bosch L, Tilkin P, Cassiman JJ, Matthijs G. D90A heterozygosity in the SOD1 gene is associated with familial and apparently sporadic amyotrophic lateral sclerosis. *Neurology* 1996;47: 1336-9.

RP11 is the second most common locus for dominant retinitis pigmentosa

Autosomal dominant retinitis pigmentosa (ADRP), an inherited retinal degeneration, is caused by mutations in two known genes, rhodopsin and peripherin/RDS, and seven loci identified only by linkage analysis, on chromosomes 1cen, 7p, 7q, 8q, 17p, 17q, and 19q. This high level of locus heterogeneity

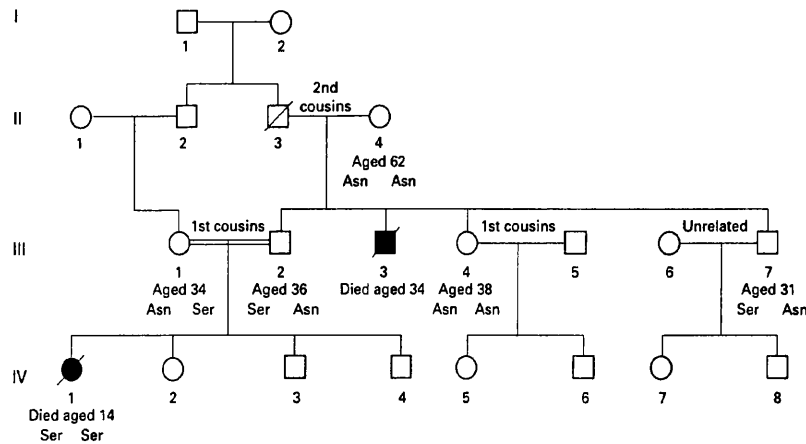


Figure 1 Pedigree of family segregating the Asn86Ser mutation. Members with ALS are represented by solid symbols. Age and Asn86Ser status are shown beneath the symbols.

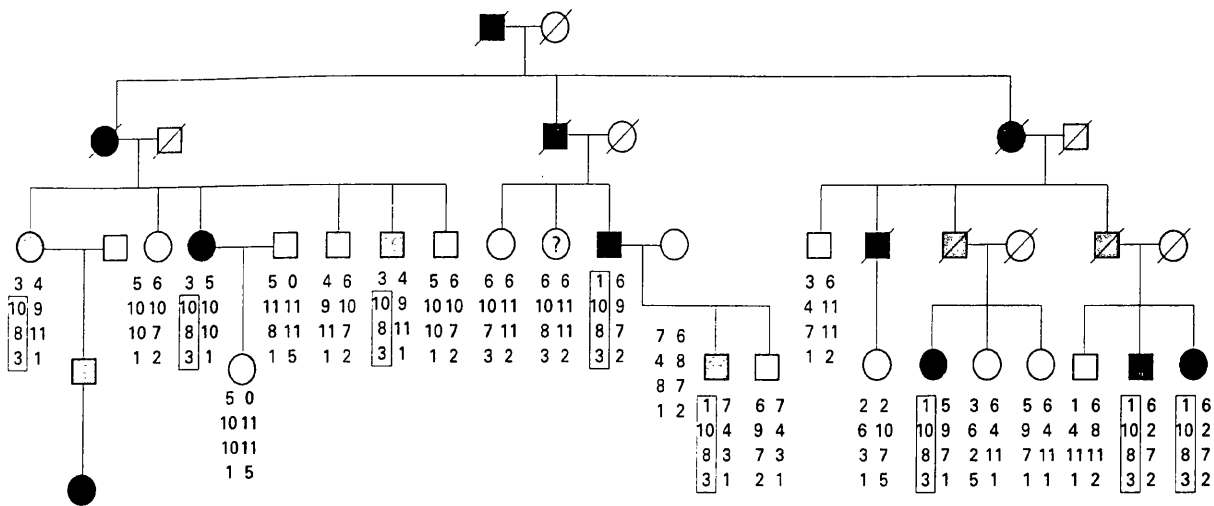


Figure 1 New RP11 linked UK family with haplotypes for 19q13.4 markers shown beneath. Alleles are for, from top to bottom, D19S921, D19S572, D19S927, and D19S924. Filled symbols represent affected subjects and shaded symbols represent gene carriers. Lod scores were calculated with these subjects marked as unaffected, but with penetrance set at 0.7. The symbol containing a question mark represents a subject in whom there has been a crossover between D19S921 and the more distal markers. This crossover makes it impossible to determine genetically whether she is a gene carrier or is normal. Analysis of this patient could potentially have refined the locus further, but it is not possible to distinguish carriers from normal subjects in the clinic beyond doubt.

poses considerable difficulties to any laboratory proposing to undertake genetic counselling in dominant RP families. We recently described four families with dominant RP which mapped to the RP11 locus on 19q.¹ Locus refinement based on linkage analysis in these families placed the RP11 gene between D19S180 and D19S926, and though inheritance was clearly dominant, each pedigree was noted to include non-expressing carriers. Of a total of 20 large ADRP pedigrees analysed in our laboratory at that time, 20% were therefore shown to map to the RP11 locus, a frequency second only to the rhodopsin locus, which in this data set accounted for approximately 50%.² This observed high frequency for the RP11 locus was further substantiated by a report of a Japanese RP11 family, also with variable penetrance.³

We have now tested for linkage to 19q markers in three new UK pedigrees and a family of Spanish origin. This led to the identification of one further RP11 linked UK family, shown in fig 1. Multipoint analysis with markers D19S572, D19S927, and D19S924, all mapping within the interval for the RP11 gene, gave a maximum multipoint

lod score of 4.38 at D19S924/927. A new proximal crossover in this family, together with a change in the published Genethon marker order, further refines the RP11 locus to within the interval D19S921-D19S418, with markers D19S572, D19S927, and D19S924 showing no crossovers with the RP phenotype in affected subjects (fig 2). Once again unaffected family members could not be included in the haplotype analysis since the family includes several carriers who show no symptoms. The linked haplotype in this family is unique, indicating that it is unrelated to other UK 19q linked families. Our frequency estimate for RP11 as a proportion of all ADRP is therefore still approximately 20%, but with increased significance resulting from a slightly larger sample size. We also note with interest that McGee *et al.*⁴ reported an additional three RP11 linked families from the USA, once again with reduced penetrance. This brings to a total of nine the number of published RP11 pedigrees, well above the number identified for the six other linked loci. A much larger number of peripherin/RDS mutations have been identified in ADRP patients, but where the gene is known it is possible to identify the causative mutation in the absence of any family structure, so such mutations will inevitably be over-represented. Current estimates of the frequency of peripherin/RDS mutations as a cause of ADRP are less than 5%.^{5,6}

These data therefore further refine the RP11 locus. They also serve to underline the conclusions of our previous study, namely that 19q13.4 is the site for a major locus for dominant RP, probably the second most common after rhodopsin, and that partial penetrance is a consistent observation in families with this form of RP. An investigation of the mechanism of variation in severity of the disease phenotype in these families could have important implications in the search for a cure for RP. In the mean time, analysis of a new ADRP family should begin by determining the level of penetrance. A fully penetrant phenotype implicates rhodopsin as the most likely site for mutations, which could then be screened for by strategies we described in a previous letter to the Journal.⁷ Evidence of non-penetrance would suggest the RP11 locus. Though the 7p and 8q ADRP loci do also exhibit variation in severity of disease,^{8,9}

it is increasingly evident that the locus on 19q13.4 is more common, both in the UK and in other populations.

We gratefully acknowledge the Wellcome Trust (grant numbers 043006/Z/94, 042375/Z/94, and 035535/Z/92) for funding.

ERANGA VITHANA
MAI AL-MAGHTHEH
SHOMI S BHATTACHARYA
CHRIS F INGLEHEARN*
Department of Molecular Genetics, Institute of Ophthalmology, University College London, London, UK

*Present address: Molecular Medicine Unit, Clinical Sciences Building, St James's University Hospital, Leeds LS29 7TF, UK

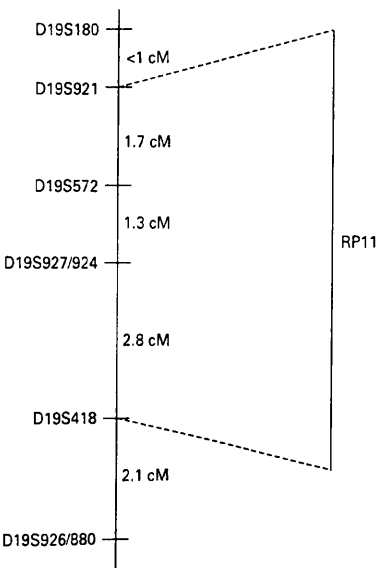


Figure 2 The RP11 locus and surrounding markers.

- Al-Magtheth M, Vithana E, Tarttelin EE, *et al.* Evidence for a major retinitis pigmentosa locus on 19q13.4 (RP11) and association with a unique bimodal expressivity phenotype. *Am J Hum Genet* 1996;59:864-71.
- Inglehearn CF, Tarttelin EE, Plant C, *et al.* A linkage survey of 20 dominant retinitis pigmentosa families: frequencies of the nine known loci and evidence for further heterogeneity. *J Med Genet* 1998;35:1-5.
- Xu S, Nakazawa M, Tamai M, Gal A. Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J Med Genet* 1995;32:915-16.
- McGee TL, Devoto M, Ott J, Berson TP, Dryja TP. Dominant retinitis pigmentosa with reduced penetrance: refinement of the RP11 locus to a 4.9 cM interval on 19q. *Invest Ophthalmol Vis Sci* 1997;38(suppl):1139.
- Wells J, Wroblewski J, Keen TJ, *et al.* Mutations in the human retinal degeneration slow (RDS) gene can cause either retinitis pigmentosa or macular dystrophy. *Nat Genet* 1993;3:213-18.
- Souied EH, Rozet JM, Gerber S, *et al.* Screening for mutations within the rhodopsin, peripherin-RDS and ROM1 genes in autosomal dominant retinitis pigmentosa (ADRP) in pedigrees from France. *Invest Ophthalmol Vis Sci* 1995;36(suppl):890.
- Tarttelin EE, Al-Magtheth M, Keen TJ, Bhattacharya SS, Inglehearn CF. Simple tests for rhodopsin involvement in retinitis pigmentosa. *J Med Genet* 1996;33:262-3.
- Inglehearn CF, Carter SA, Keen TJ, *et al.* A new locus for autosomal dominant retinitis pigmentosa (adRP) on chromosome 7p. *Nat Genet* 1993;4:51-3.
- Blanton SH, Heckenlively JR, Cottingham AW, *et al.* Linkage mapping of autosomal dominant retinitis pigmentosa (RP1) to the pericentric region of human chromosome 8. *Genomics* 1991;11:857-69.

Evidence for a Major Retinitis Pigmentosa Locus on 19q13.4 (RP11), and Association with a Unique Bimodal Expressivity Phenotype

Mai Al-Magthteh,¹ Eranga Vithana,¹ Emma Tarttelin,¹ Marcelle Jay,² Kevin Evans,¹ Tony Moore,² Shomi Bhattacharya,¹ and Chris F. Inglehearn¹

¹Department of Molecular Genetics, Institute of Ophthalmology, and ²Department of Clinical Ophthalmology, Moorfields Eye Hospital, London

Summary

Retinitis pigmentosa (RP) is the name given to a heterogeneous group of retinal degenerations mapping to at least 16 loci. The autosomal dominant form (adRP), accounting for ~25% of cases, can be caused by mutations in two genes, rhodopsin and peripherin/RDS, and by at least six other loci identified by linkage analysis. The RP11 locus for adRP has previously been mapped to chromosome 19q13.4 in a large English family. This linkage has been independently confirmed in a Japanese family, and we now report three additional unrelated linked U.K. families, suggesting that this is a major locus for RP. Linkage analysis in the U.K. families refines the RP11 interval to 5 cM between markers D19S180 and AFMc001yb1. All linked families exhibit incomplete penetrance; some obligate gene carriers remain asymptomatic throughout their lives, whereas symptomatic individuals experience night blindness and visual field loss in their teens and are generally registered as blind by their 30s. This “bimodal expressivity” contrasts with the variable-expressivity RP mapping to chromosome 7p (RP9) in another family, which has implications for diagnosis and counseling of RP11 families. These results may also imply that a proportion of sporadic RP, previously assumed to be recessive, might result from mutations at this locus.

Introduction

Retinitis pigmentosa (RP) is a heterogeneous form of inherited retinal degeneration characterized by night blindness (nyctalopia) and constricted visual fields in the early stages, often progressing to registrable blindness. RP affects 1–2 in every 5,000 births in the Western

world (Boughman et al. 1980). Clinical manifestations include pigment deposition in the retina and attenuation of retinal blood vessels, with later depigmentation or atrophy of the retinal pigment epithelium. Electroretinogram (ERG) abnormalities are recordable in the early stages, with reduced amplitude of rods and, in some cases, cones. In advanced RP both rod and cone ERG responses are extinguished.

RP can be inherited in an autosomal dominant, autosomal recessive, or X-linked fashion, and the effects of the mutant gene may be confined to the eye or may be part of a multisystem disorder such as Usher or Bardet Biedl syndrome. So far, at least 16 different loci have been implicated in nonsyndromic RP (reviewed by Inglehearn and Hardcastle 1996). The autosomal dominant form (adRP), which accounts for ~25% of patients (Jay 1982), is itself both genetically and clinically heterogeneous. It can be caused by mutations in the rhodopsin (Dryja et al. 1990) and peripherin/RDS (Farrar et al. 1991; Kajiwarra et al. 1991) genes on chromosomes 3q and 6p, respectively. Linkage analysis has implicated another six loci—on 7p (Inglehearn et al. 1993), 7q (Jordan et al. 1993), 8q (Blanton et al. 1991), 17p (Greenberg et al. 1994), 17q (Bardien et al. 1995), and 19q (Al-Magthteh et al. 1994)—in its causation. Initially, the chromosome 19q adRP locus was localized, in a single large family, to a 22-cM interval between markers D19S180 and D19S214 (Al-Magthteh et al. 1994). This linkage has recently been confirmed with the identification of a linked family of Japanese origin (Xu et al 1995). The locus has been given the number RP11 (MIM 600138; McKusick 1992). The original 19q family, known as “ADRP5,” was diagnosed as having type II/R adRP as defined by Lyness et al. (1985), with apparent incomplete penetrance (Moore et al. 1993). A more detailed analysis in the light of known carrier status led Evans et al. (1995) to assign the term “bimodal expressivity” to the phenotype in this family.

We now report three new adRP families in which the phenotype maps to the RP11 locus. Like the original RP11 family, each has type II/R adRP with the “bimodal expressivity” phenotype. Typing of new 19q markers in all four linked families refines the locus to a 5-cM interval as described below.

Received April 10, 1996; accepted for publication July 24, 1996.

Address for correspondence and reprints: Dr. Chris F. Inglehearn, Department of Molecular Genetics, Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, United Kingdom. E-mail: cinglehe@hgmrc.ac.uk

© 1996 by The American Society of Human Genetics. All rights reserved.
0002-9297/96/5904-0017\$02.00

Subjects and Methods

Ascertainment of Phenotype in adRP Families

The adRP families presented in this study were ascertained through Moorfields Eye Hospital, London. ADRP5 and ADRP29 are from the north of England and the south of Wales, respectively, and have a type II/R form of RP with bimodal expressivity. A detailed clinical description of these families has been reported elsewhere (Moore et al. 1993; Evans et al. 1995). RP1907 and ADRP2 are newly identified British RP families that display a similar phenotype. Affected members of all these pedigrees experience night blindness in their teens and are generally registered as blind in their 30s. Some patients have posterior subcapsular lens opacities, and later in the disease process there is also macular involvement. This macular degeneration is more pronounced in members of the ADRP2 family than in the other three families, which may be the result of allelic variation such as that seen at the rhodopsin or RDS/peripherin loci.

Bimodal expressivity is defined by the presence of asymptomatic individuals who have both affected parents and affected children. Such individuals are shown as shaded symbols in figure 1. In this figure there also are shown a number of individuals who carry the affected chromosome 19 haplotype yet who do not have affected children. Since these individuals cannot be proved to be disease carriers they are shown as normal; but, given the linkage results obtained, it is likely that these individuals are carrying the mutated gene. Nevertheless, in linkage analysis these cases were treated as normal while the known carriers (shaded in fig. 1) were considered affected.

On examining families ADRP5 and ADRP29, Evans et al. (1995) concluded that 65% of gene carriers were affected whereas 35% remained asymptomatic. A similar analysis of families ADRP2 and RP1907 reveals 13 affected individuals among 19 haplotype carriers (including both categories of haplotype carrier described above), or ~68%, which correlates well with the previous estimate. This count does not include deceased family members, for whom clinical status can be inferred only anecdotally from living relatives.

DNA Isolation and Microsatellite Analysis

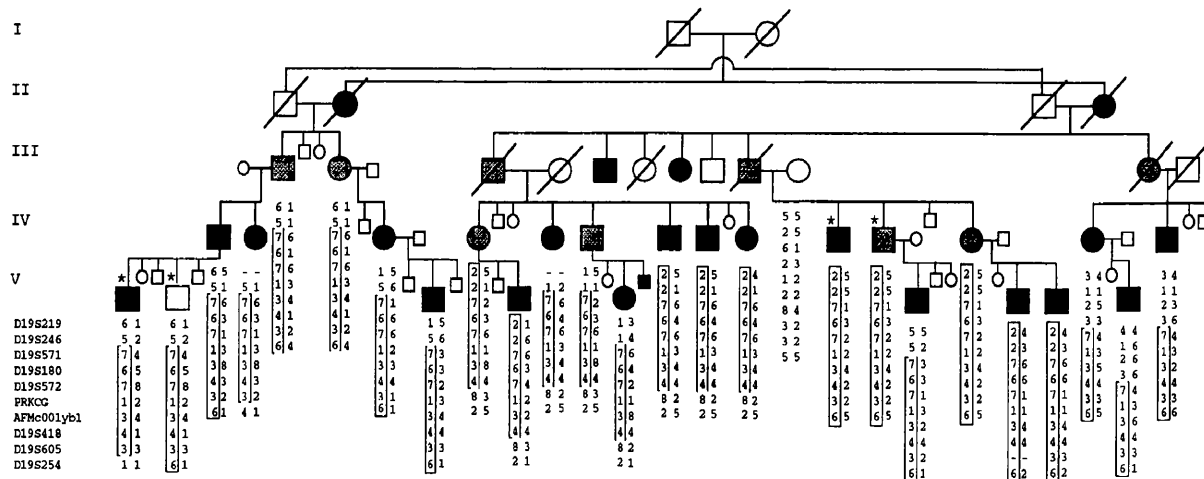
DNA was isolated from peripheral blood lymphocytes by means of standard protocols. PCR reactions were carried out in 96-well microtiter plates from Omnigene. PCR reactions were performed in a total volume of 12.5 μ l, composed of 2 μ l diluted genomic DNA (~100 ng), 1 \times *Taq* buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, and 0.1% nonionic detergent), 0.2 mM each of dNTPs, 2 pmol each of both forward (fluorescently end-labeled) and reverse primers, and 0.5 units *Taq* polymerase. Twenty-five cycles of PCR amplification were per-

formed, at 94°C for 1 min, 55°C for 1 min (unless specified otherwise), and 72°C for 1 min, with an Omnigene thermal cycler. One microliter of each reaction was then diluted in 3 μ l of loading buffer (deionized formamide with dextran blue) and 0.5- μ l size standard (Rox 2500 standard; Applied Biosystems), denatured at 95°C for 2 min, and kept on ice until being loaded on a 6% denaturing polyacrylamide gel. Electrophoresis was performed in 1 \times Tris-borate-EDTA buffer by means of an Applied Biosystems 373 DNA sequencer. Genotype data were collected and analyzed by means of Gene Scan 672 software supplied by Applied Biosystems. Markers D19S180, D19S572, and AFMc001yb1 were genotyped in all families, since they are the closest markers within and flanking the RP11 interval. Other markers were genotyped, where necessary, to confirm linkage or to establish disease haplotype. An *Msp*I polymorphism in the protein kinase C (PRKCG) gene was analyzed by standard Southern blotting techniques. Data for this marker are only shown in ADRP5; in ADRP29 it proved uninformative; in RP1907 the technique was not done, since no recombinant meioses were evident with flanking markers; and for ADRP2 there was sufficient DNA for only partial analysis.

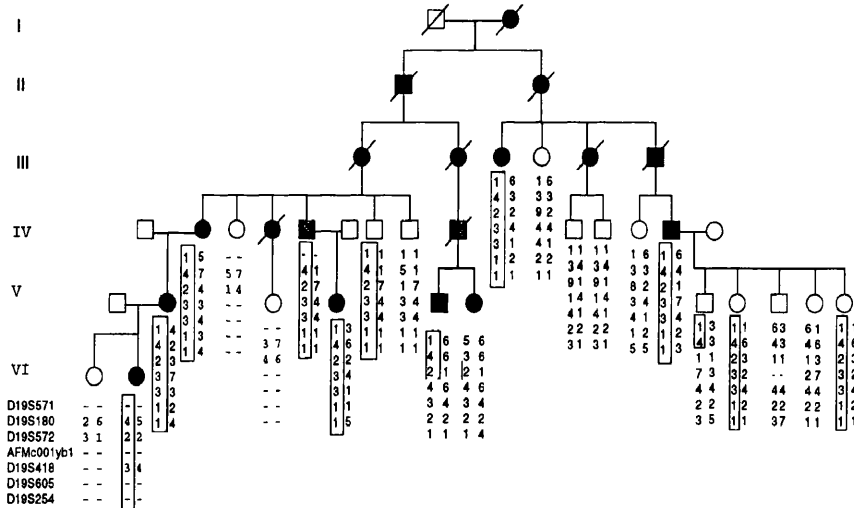
Linkage Analysis

Pedigree information, allele frequencies, and genotype data were processed by means of the Linksys data-management package LS4 (Attwood and Bryant 1988). The adRP disease-allele frequency was assumed to be .0001. Allele frequencies for all markers were estimated on the basis of data for the normal spouses of affected family members. Only affected individuals and normal individuals >30 years old were incorporated in the analysis. However, a penetrance value of .7 (the estimated proportion of symptomatic individuals relative to all disease-gene carriers) was used to account for the incomplete penetrance seen at this locus. A value of .0001 was used to account for possible phenocopies or misdiagnoses. Both this and the penetrance were defined as values of liability classes in the data file generated by LS4. Sex-averaged recombination frequencies were also used. LS4 output files were used as input for the LINKAGE programs (Lathrop and Lalouel 1984). Pairwise and multipoint analyses were performed by means of the MLINK and LINKMAP programs, respectively, from LINKAGE package version 5.1 (Lathrop et al. 1984). Maximum-likelihood recombination fractions (θ 's) between pairs of markers were assumed to be as published (Gyapay et al. 1994). For markers derived from different genome maps, these θ 's were calculated, in ADRP5 and other families studied in our laboratory, for mapping purposes, by means of the ILINK program from the LINKAGE package. Multipoint analyses were per-

ADRP5



ADRP29



formed on the HGMP Resource Centre computing facility (Rysavy et al. 1992).

Results

ADRP5 is the family that initially showed linkage to markers D19S214 and D19S180 from the 19q13.4 region (Al-Magthteh et al. 1994). Markers D19S219, D19S246, D19S571, D19S572, PRKCG, AFMc001yb1, D19S418, D19S605, and D19S254 from the 19q13.4 region (fig. 2) were also genotyped in this family. Pairwise linkage analyses between each marker and disease phenotype were performed. LOD scores obtained from these analyses are shown in table 1. Markers D19S572, PRKCG, and AFMc001yb1 gave the highest LOD scores with no recombination. Haplotype analysis with these data in ADRP5 (see fig. 1) localizes the RP11 disease gene to an 8-cM interval between markers D19S180 and D19S605. Critical recombination events with these

markers are shown in figure 2. Given the number of alleles for each marker, multipoint analysis with all the markers was not possible. However, a series of multiple three-point analyses for successive combinations of two markers and disease in the region were performed (fig. 3). A maximum LOD score of 9.3 was obtained with markers D19S572 and AFMc001yb1 (fig. 3). This is consistent with the localization obtained by haplotype analysis.

In order to estimate the frequency of adRP caused by mutations at this and other loci and to further refine this localization, other adRP families from the Moorfields Eye Hospital genetic register and other sources were genotyped for various markers between D19S571 and D19S254. In total, 20 large adRP families have been screened for linkage to each of the eight known adRP loci (C. Inglehearn, unpublished data). Three unrelated adRP families, denoted "ADRP29," "ADRP2," and "RP1907," were found to be linked to this locus. Each

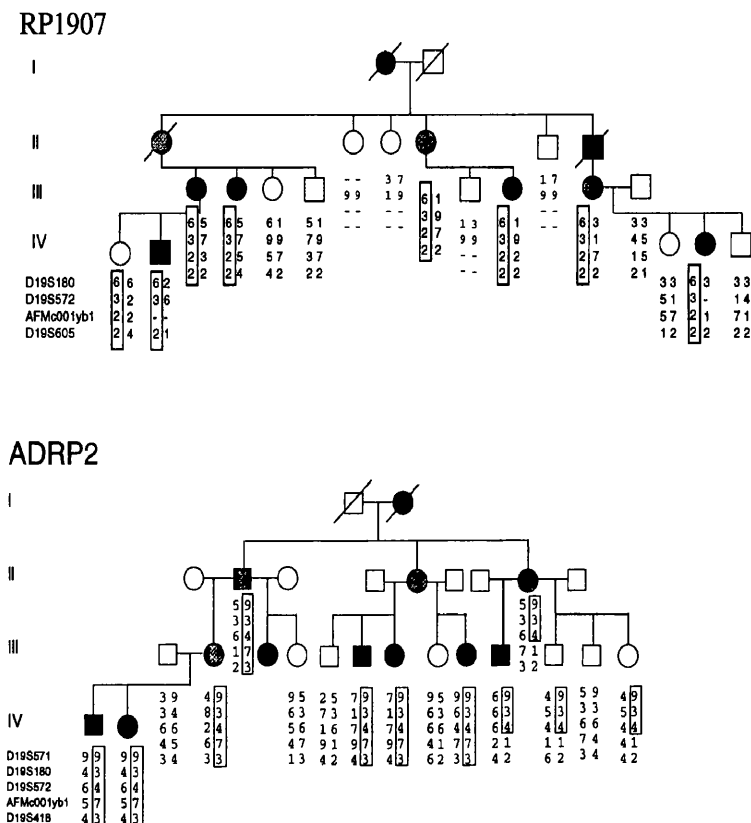


Figure 1 Haplotype analysis of the 19q13.4 markers in ADRP5, ADRP29, RP1907, and ADRP2. The disease haplotypes are boxed. Unblackened-unshaded symbols represent normal individuals; blackened symbols represent individuals with RP; and gray-shaded symbols represent asymptomatic obligate disease-gene carriers who have affected children. In each of the bottom three families there are individuals marked as normal who nevertheless have inherited the disease haplotype. These are almost certainly also asymptomatic gene carriers, but, since they do not have affected children, they were considered as normal individuals in the linkage analysis. The penetrance factor of .7 used in the linkage analysis allows for the ambiguous status of such individuals. Individuals V-7, V-8, V-9, V-10, and V-11 in ADRP29 were not included in linkage analysis, since they were too young to be diagnosed reliably. A dash (-) denotes an unknown genotype, whereas asterisks (*) denote siblings from ADRP5 who inherit the same haplotypes from both parents, although one sibling is affected and the other is an asymptomatic carrier (see the Phenotype of Chromosome 19–Linked [RP11] Families subsection in the Discussion section, above). In the analysis of ADRP5, only affected meioses were used, so only these haplotypes are given. Symbols for other typed individuals are reduced in size but are nevertheless shown, and numbering of the pedigree includes these samples.

of these shows crossovers with markers from all other previously known adRP loci (data not shown), and each has a phenotype similar to that of ADRP5 (type II/R with bimodal expressivity). The RP11 locus is linked, in each family, to a different haplotype of markers in the 19q region, excluding a founder effect.

In family ADRP29 a LOD score of 3.19 was obtained with marker D19S572, with no recombination ($\theta = 0$) (table 1). Multipoint analysis with markers D19S180 and D19S572 by means of the program LINKMAP gave a maximum LOD score of 3.39 at D19S572. Haplotype analysis in individuals V-5 and V-6 (figs. 1 and 2) of this family suggested a recombination event that could have taken place in either II-1 or III-2 and then could have been passed to V-5 and V-6 through IV-9. This recombination occurred between AFMc001yb1 and D19S572, refining the locus to a 5-cM interval between D19S180 and AFMc001yb1 (fig. 2).

In family RP1907 (see fig. 1) maximum LOD scores

of 3.04, 2.1, and 2.8 were obtained with markers D19S572, AFMc001yb1, and D19S180, respectively, at $\theta = 0$ (table 1). Multipoint analysis with D19S572 and D19S180 in this family gave a LOD score of 3.12. No critical recombination events with the flanking markers for RP11 were detected.

Pairwise LOD-score analysis in ADRP2 demonstrated linkage to markers D19S571 and D19S572, with LOD scores of 2.2 and 2.3, respectively, at $\theta = 0$ (table 1). Multipoint analysis did not significantly alter these LOD scores. However, a LOD score of 2 is generally accepted as sufficient evidence for linkage of a disease to a previously known locus with similar phenotype, an understanding that often is referred to as “posterior probability.” Haplotype analysis in this family, shown in figure 1, demonstrated another recombination event between D19S572 and AFMc001yb1, which confirms the 5-cM refinement obtained in ADRP29.

Previously, another retinal dystrophy locus has been

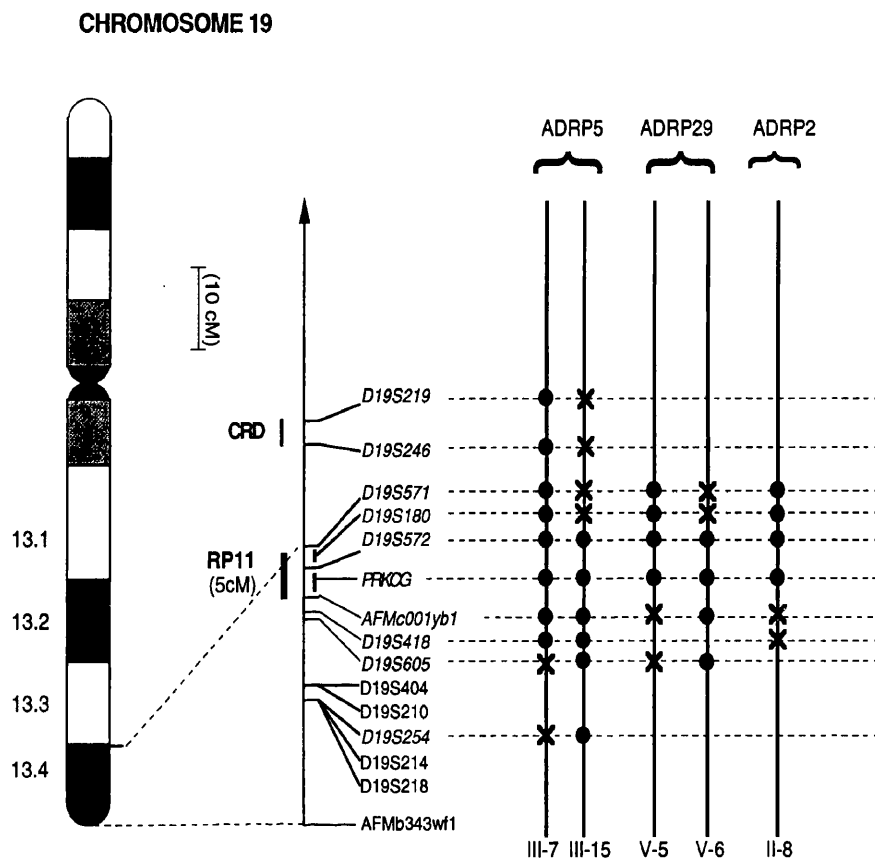


Figure 2 Linkage map of chromosome 19q13.3-q13.4, showing localization of the RP11 gene according to haplotype analysis in ADRP5, ADRP29, and ADRP2. Markers used in this analysis are italicized, and distances are those estimated by Gyapay et al. (1994). The markers D19S180 and PRKCG are placed in an interval by means of haplotype analysis, since they have not been incorporated into the existing Génethon map. Haplotypes of individuals with critical recombination events are denoted by the thicker vertical lines on the right-hand side, with nonrecombinant alleles shown as blackened circles and with recombinant alleles shown as X's. The figure also shows the localization of the CRD locus relative to the RP11 region.

mapped to the interval D19S219-D19S246 on 19q13.3-13.4 (Evans et al. 1994). Multipoint and haplotype analysis in ADRP5 by use of these markers clearly demonstrate the exclusion of the ADRP phenotype from the cone/rod dystrophy interval. Furthermore, haplotype and linkage analysis in ADRP5 and other families made it possible to integrate markers from other genome maps into the current Génethon poly-CA map. These data are shown in figures 2 and 3.

Discussion

RP11: A Major Locus for adRP

So far, four British families from this laboratory—ADRP5, ADRP29, RP1907, and ADRP2—and a Japanese family (Xu et al. 1995) have been linked to the RP11 locus. Haplotype analysis in the four families described here showed no evidence of a founder effect (fig. 1). These families were identified as part of a genetic survey of 20 adRP families with >11 meioses. All of these families now have been linked to or excluded from the known adRP loci (C. Inglehearn, unpublished data).

From these data it is possible to estimate that ~20% of adRP results from mutations at the RP11 locus, making it the second most common locus for dominant RP, after rhodopsin.

This observation, together with the bimodal phenotype consistently associated with this locus, raises the interesting possibility that a significant proportion of sporadic or apparently recessive RP also may derive from mutations at this locus. The disease could then appear without any apparent family history, confusing the diagnosis. Sporadic RP accounts for ~50% of all cases and is often assumed to be recessive in origin (Jay 1992).

Linkage Analysis and Further Refinement of the RP11 Locus

Haplotype analysis using newly identified markers in the 19q13.4 region has refined the initial localization in ADRP5 from 22 to 8 cM between markers D19S180 and D19S605. Haplotype analysis in the newly linked ADRP29 and ADRP2 families confirms and further refines this to a 5-cM interval between D19S180 and AF-

Table 1

Two-Point LOD Scores between 19q13.4 Markers and Disease Phenotype in Linked Families

	LOD SCORE AT $\theta =^a$							MAXIMUM LOD SCORE ($\hat{\theta}$)
	.00	.01	.05	.10	.20	.30	.40	
ADRP5:								
D19S219	-13.0	-4.4	-.8	.52	1.32	1.27	.79	1.32 (.20)
D19S246	-7.9	-3.5	-1.4	-.4	-.23	.33	.20	.33 (.30)
D19S571	3.06	4.98	5.16	4.81	3.82	2.72	1.43	5.20 (.03)
D19S180	1.99	4.06	4.34	4.08	3.16	1.98	.64	4.35 (.04)
D19S572	7.31	7.19	6.71	6.07	4.67	3.09	1.37	7.31 (.00)
PRKCG	5.71	5.62	5.23	4.72	3.62	2.41	1.12	5.71 (.00)
AFMc001yb1	7.67	7.54	7.00	6.30	4.80	3.15	1.40	7.67 (.00)
D19S418	5.91	5.81	5.40	4.86	3.72	2.46	1.08	5.91 (.00)
D19S605	.98	3.86	4.14	3.90	3.05	2.00	.87	4.15 (.04)
D19S254	-.8	2.88	4.34	4.55	3.98	2.88	1.49	4.55 (.10)
ADRP29:								
D19S571	-6.8	-3.4	-1.7	-1.0	-.4	-.2	-.1	.0 (.50)
D19S180	.6	1.75	2.17	2.13	1.70	1.10	.48	2.17 (.05)
D19S572	3.19	3.13	2.90	2.59	1.94	1.26	.60	3.19 (.00)
AFMc001yb1	-2.7	-.6	.00	.15	.17	.09	.02	.17 (.20)
D19S418	2.26	2.21	2.00	1.73	1.18	.65	.23	2.26 (.00)
D19S605	-2.7	-.5	.13	.29	.29	.18	.07	.29 (.10)
RP1907:								
D19S180	2.80	2.75	2.54	2.27	1.66	.98	.32	2.80 (.00)
D19S572	3.04	2.98	2.76	2.47	1.82	1.10	.37	3.04 (.00)
AFMc001yb1	2.12	2.08	1.92	1.70	1.21	.70	.23	2.12 (.00)
D19S605	.26	.25	.21	.17	.10	.04	.01	.26 (.00)
ADRP2:								
D19S571	2.18	2.14	2.01	1.82	1.40	.91	.38	2.18 (.00)
D19S180	.87	.85	.78	.70	.52	.35	.17	.87 (.00)
D19S572	2.30	2.26	2.12	1.92	1.47	.95	.39	2.30 (.00)
AFMc001yb1	-1.5	.26	.83	.96	.86	.59	.26	.96 (.10)
D19S418	-1.5	.26	.83	.96	.86	.59	.26	.96 (.10)

^a LOD scores were calculated at a disease penetrance of .7, under the assumption of sex-averaged θ 's.

Mc001yb1 (fig. 2). Refinement of the locus should facilitate identification of the disease gene through either a positional cloning or a positional candidate approach. Another retinal degeneration locus for cone/rod dystrophy (CORD2) has been mapped to 19q13.3-13.4 (Evans et al. 1994). This refinement excludes the possibility of these two retinal degeneration phenotypes being allelic variants of the same genetic locus and instead confirms the presence of two distinct retinal degeneration loci, separated by ~15 cM (see fig. 3).

Phenotype of Chromosome 19-Linked (RP11) Families

In 1993 Moore et al. published a study of four adRP families with incomplete penetrance. They suggested, on the basis of analysis of these families, that two different forms of incomplete-penetrance RP could be seen. In one form, typified by family 2, gene carriers exhibited a spectrum of phenotypes, including asymptomatic, mild, and severe disease. In contrast, in families 1, 3, and 4 patients were either severely affected or asymptomatic.

Family 2 has since been linked to chromosome 7p (RP9), whereas families 3 and 4, referred to here as "ADRP5" and "ADRP29," respectively, are now linked to 19q (RP11). More detailed clinical analysis of these families, in the light of known genotypes, provided further support for the two different forms of incomplete penetrance (Evans et al. 1995; Kim et al. 1995). We now report two new adRP families, ADRP2 and RP1907, both linked to 19q and both showing the "all or nothing" form of incomplete penetrance referred to here as "bimodal expressivity." Furthermore, family 1, described by Moore et al. (1993) as falling into this category, shows a segregation pattern suggestive of linkage to 19q, although LOD scores are not significant. Two other new small families that also appear to have the bimodal inheritance pattern also showed weak linkage to 19q markers. No family with this phenotype has yet been excluded from linkage to the RP11 region. It can therefore be concluded that, so far, at least five and perhaps as many as eight RP families have mutations at

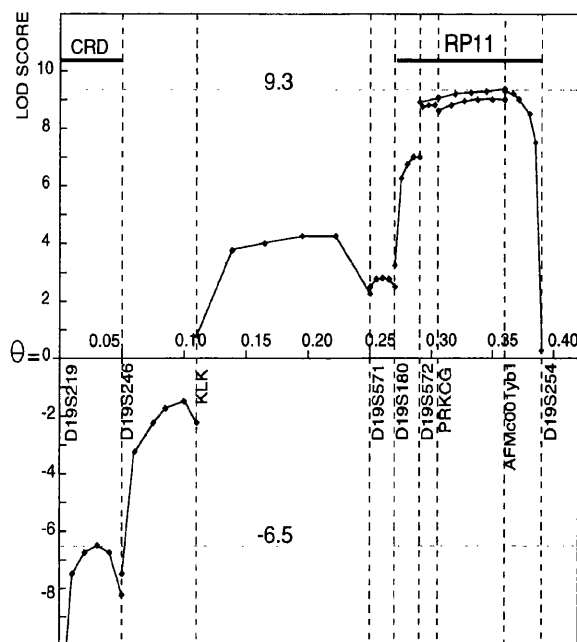


Figure 3 Multiple three-point analyses between marker pairs and disease phenotype in ADRP5, by use of 19q13.4 markers (Gyapay et al. 1994; NIH/CEPH 1992). Distances are based on our own data rather than on the Génethon map, since two of the markers used were from the NIH/CEPH map. These distances were calculated by means of the ILINK program from LINKAGE package version 5.1. Two-point LOD scores for marker KLK in family ADRP5 have been published previously (Al-Magthteh et al. 1994).

this locus. On this basis it appears likely that this relatively common locus for dominant RP is uniquely associated with a bimodal RP phenotype.

Environmental factors seem unlikely to play a major role in this characteristic phenotype, since their influence would be expected to be dose dependent and should therefore produce a graded, rather than an "all or nothing," phenotype. The term "digenic inheritance" has been used to describe the situation in which simultaneous mutations of two genetic loci are involved in disease pathogenesis (Kajiwara et al 1994). It is possible that digenic (or multigenic) inheritance could explain bimodal expressivity at the 19q RP locus. One hypothesis put forward by Evans et al. (1995) is that the second component in disease causation could be the apparently normal allele of the RP11 gene in each patient. However, in ADRP5 there are two examples of pairs of siblings who appear to have inherited, from both parents, the same haplotypes at the RP11 locus; yet one is affected whereas the other is an asymptomatic carrier (fig. 1, individuals marked with an asterisk [*]). Since double recombinants in such a small, well-mapped region are unlikely, this appears to exclude an allelic effect as the only factor, although it still could be involved as a component in a "multigenic" effect. The other component(s) would result from polymorphisms or mutations at a different locus or loci and could be inherited from either

parent, by recessive, dominant, or X-linked modes. Such an effect therefore could be mapped only by model-independent linkage techniques such as sib-pair analysis and would be practical only with a very large sample of patients.

An alternative hypothesis that could be invoked to explain the bimodal expressivity phenotype is expansion or contraction of a triplet-repeat codon, such as has been associated with other neurodegenerative disorders (Ross 1995). A cursory examination of the pedigrees in figure 1 does appear to support an increase in severity in subsequent generations, a phenomenon known as "anticipation," which has been associated with triplet-repeat-expansion disorders. However, this is unlikely to be significant, since normal younger members of the families are less likely to consent to give blood, reducing the apparent frequency of asymptomatic carriers in the lower generations. Nevertheless, we have tested for and excluded expansion of CAG repeats as a cause of RP in a number of adRP families, including ADRP5, using the repeat-expansion-detection (RED) technique (T. J. Keen and C. Inglehearn, unpublished data).

Prospects

Informed genetic counseling using a haplotype-based diagnosis is now available for most family members. Symptomatic individuals can be given a clearer picture of the expected progression of the disease, while their risk of having affected children can be more accurately estimated, as ~35% (on the basis of penetrance value) rather than 50%. Similarly, asymptomatic gene carriers >30 years of age will probably remain unaffected for the rest of their lives, although, once again, their children will be at a 35% risk of developing symptomatic RP. Identification of the disease gene and mutation spectrum leading to retinal degeneration in RP11 families would further improve counseling, increase our understanding of RP causation and normal eye function, and could lead to an understanding of the biological basis of bimodal expressivity in these families.

Acknowledgments

This research was supported by Wellcome Trust grants 043006/Z/94/Z, 042375/Z/94/Z, and 035535/Z/92/Z, the British Retinitis Pigmentosa Society, and the Foundation Fighting Blindness USA.

References

- Al-Magthteh M, Inglehearn CF, Keen TJ, Evans K, Moore AT, Jay M, Bird AC, et al (1994) Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. *Hum Mol Genet* 3:351-354
- Attwood J, Bryant S (1988) A computer programme to make

- analysis with LIPED and LINKAGE easier to perform and less prone to input errors. *Ann Hum Genet* 52:259
- Bardien S, Ebenezer N, Greenberg J, Inglehearn CF, Bartmann L, Goliath R, Beighton P, et al (1995) An eighth locus for autosomal dominant retinitis pigmentosa is linked to chromosome 17q. *Hum Mol Genet* 4:1459–1462
- Blanton SH, Heckenlively JR, Cottingham AW, Freidman J, Sadler LA, Wagner M, Freidman LH, et al (1991) Linkage mapping of autosomal dominant retinitis pigmentosa (RP1) to the pericentric region of human chromosome 8. *Genomics* 11:857–869
- Boughman JA, Confally PM, Nance WE (1980) Population genetic studies of retinitis pigmentosa. *Am J Hum Genet* 32:223–235
- Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, Sandberg MA, et al (1990) A point mutation of rhodopsin gene in one form of retinitis pigmentosa. *Nature* 343:364–366
- Evans K, Al-Magthteh M, Fitzke FW, Moore AT, Jay M, Inglehearn CF, Arden GB, et al (1995) Bimodal expressivity in dominant retinitis pigmentosa genetically linked to chromosome 19q. *Br J Ophthalmol* 79:841–846
- Evans K, Fryer A, Inglehearn CF, Duvall-Young J, Whittaker J, Gregory CY, Ebenezer N, et al (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. *Nat Genet* 6:210–213
- Farrar GJ, Kenna P, Jordan SA, Kumar-Singh R, Humphries MM, Sharp EM, Shiels DM, et al (1991) A three base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature* 354:478–480
- Greenberg J, Goliath R, Beighton P, Ramesar R (1994) A new locus for autosomal dominant retinitis pigmentosa on the short arm of chromosome 17. *Hum Mol Genet* 3:915–918
- Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Génethon human genetic linkage map. *Nat Genet* 7:246–339
- Inglehearn CF, Carter SA, Keen TJ, Lindsay J, Stephenson AM, Bashir R, Al-Magthteh M, et al (1993) A new locus for autosomal dominant retinitis pigmentosa on chromosome 7p. *Nat Genet* 4:51–53
- Inglehearn CF, Hardcastle AJ (1996) Nomenclature for inherited diseases of the retina. *Am J Hum Genet* 58:433–435
- Jay M (1982) Figures and fantasies: the frequencies of the different genetic forms of retinitis pigmentosa. *Birth Defects* 18:167–173
- Jordan SA, Farrar GJ, Kenna P, Humphries MM, Sheils D, Kumar-Singh R, Sharp EM, et al (1993) Localisation of an autosomal dominant retinitis pigmentosa gene to 7q. *Nat Genet* 4:54–58
- Kajiwarra K, Berson EL, Dryja TP (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* 264:1604–1608
- Kajiwarra K, Hahn LB, Mukai S, Travis GH, Berson EL, Dryja TP (1991) Mutations in the human retinal degeneration slow gene in autosomal dominant retinitis pigmentosa. *Nature* 354:480–483
- Kim RY, Fitzke FW, Moore AT, Jay M, Inglehearn CF, Arden GB, Bhattacharya SS, et al (1995) Autosomal dominant retinitis pigmentosa mapping to chromosome 7p exhibits variable expression. *Br J Ophthalmol* 79:23–27
- Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 36:460–465
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446
- Lyness AL, Ernst W, Quinlan MP, Clover GM, Arden GB, Carter RM, Bird AC, et al (1985) A clinical, psychophysical and electroretinographic survey of patients with autosomal dominant retinitis pigmentosa. *Br J Ophthalmol* 69:326–339
- McKusick VA (ed) (1992) *Mendelian inheritance in man*. Johns Hopkins University Press, Baltimore
- Moore AT, Fitzke F, Jay M, Arden GB, Inglehearn CF, Keen TJ, Bhattacharya SS, et al (1993) Autosomal dominant retinitis pigmentosa with apparent incomplete penetrance: a clinical, electrophysiological, psychophysical and molecular genetic study. *Br J Ophthalmol* 77:473–479
- NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. *Science* 258:67–86
- Ross AR (1995) When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. *Neuron* 15:493–496
- Rysavy FR, Bishop MJ, Gibbs GP, Williams GW (1992) The UK Genome Mapping Project online computing service. *Comput Appl Biosci* 8:149–154
- Xu S, Nakazawa M, Tamai M, Gal A (1995) Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J Med Genet* 32:915–916