

# New Strategies in Molecular Genetic Studies of X-Linked Retinitis Pigmentosa

MARCELLE JAY<sup>1</sup>, CHRIS INGLEHEARN<sup>2</sup>

London and Newcastle

## Summary

**New techniques developed in the past few years are being used to isolate the genes responsible for X-linked retinitis pigmentosa. The use of hypervariable probes in linkage, the construction of genetic and physical maps, and the search for putative gene sequences of interest are described.**

The search for the gene for X-linked retinitis pigmentosa (RP) is the task of locating a DNA sequence, perhaps only a few thousand base pairs long, among the 200 million base pairs which make up the X chromosome. It is of interest to describe the progress made in this search, and the means used to achieve it. For a more technical description of the methods discussed below, the reader should refer to Davies.<sup>1</sup>

DNA may be manipulated in the laboratory in many ways. The most widely used of these involves cleavage in a reproducible manner by restriction enzymes (reviewed by Roberts).<sup>2</sup> These were first identified in bacterial cells as a defence mechanism to destroy invading plasmid or viral DNAs. They are endonucleases which recognise short sequences of DNA of between four and ten base pairs in length and cleave the DNA chain at or close to that sequence. Different bacterial strains were found to produce enzymes with different recognition sequences, so that enzymes have now been purified which recognise many different sequence motifs. The recognition sequence is termed a cleavage, or restriction, site. For example, the cleavage site of the res-

triction enzyme EcoRI (short for *Escherichia coli* restriction enzyme I) is between guanine and adenine in the sequence GAATTC (guanine, adenine, adenine, thymine, thymine, cytosine). The fragments resulting from cleavage are measured in units of base pairs (bp), kilobases (kb) which are thousands of base pairs, and megabases (Mb) which are thousands of kb or millions of bp.

DNA extracted from whole blood can be digested (cleaved) with a restriction enzyme and separated according to size by gel electrophoresis.<sup>3</sup> Fragments of a low molecular weight travel faster through the gel, while fragments of high molecular weight travel more slowly and are therefore retained near the top of the gel. The size of fragments is estimated by comparing mobilities with marker DNA fragments of known length. The logarithm of the distance travelled through the gel is directly proportional to the length of the fragment.

The results of the digest can be visualised in the gel by staining with the dye ethidium bromide, which binds between the separate strands of the double helical DNA molecule. This can be made to fluoresce under ultra-

From: <sup>1</sup>Department of Clinical Ophthalmology, Institute of Ophthalmology, University of London. <sup>2</sup>Molecular Genetics Unit, University of Newcastle-upon-Tyne.

Correspondence to: Dr Marcelle Jay, Department of Clinical Ophthalmology, Moorfields Eye Hospital, City Road, London EC1V 2PD.

violet light. The products of electrophoresis, once checked in this way, can be denatured by treatment with alkali which renders the fragments single stranded. They are then blotted to a nylon membrane still retaining their disposition according to size and this technique is known as Southern blotting.<sup>4</sup> The membrane is then challenged with a short sequence of radioactively labelled single stranded DNA known as a probe. This fragment will have been produced in large quantities by cloning into a plasmid vector which is then replicated many times over in suitable host bacterial cells. Used in this way, the probe has the property of binding to DNA fragments with sequences that are complementary to its own. Since all the cells from the individual whose DNA is being analysed contain the same chromosomal complement, and since many cells are represented in the DNA on the membrane, there will be many copies of the complementary DNA fragment. These will be of the same size, and will have migrated together as a band on the gel, which thus shows as a radioactive band on an autoradiograph.

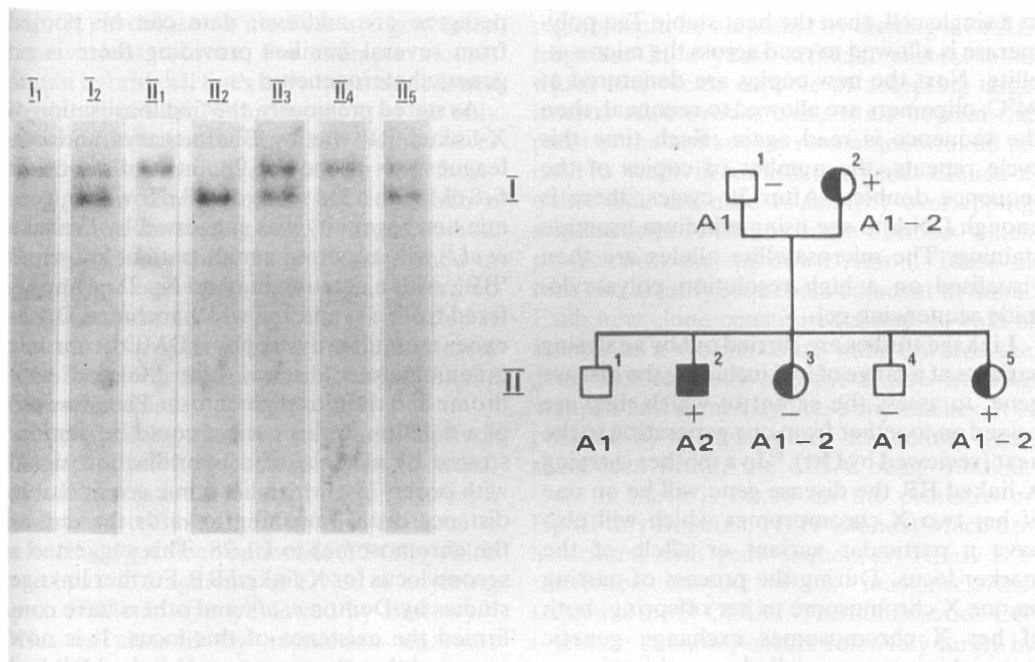
The choice of probes depends on the disorder studied, and probes can be divided into three groups, though these are not mutually exclusive. The first group comprises gene specific probes, which are used if the gene product is known, as for example in beta-globin.<sup>5</sup> The presence, absence, or presence in an altered form of the gene product can be detected directly by a gene specific probe. A second group of probes comprises those which recognise variations in DNA sequences such as those arising from mutations causing loss or gain of a cleavage site. The third group consists of oligonucleotides which recognise point mutations. Oligonucleotide probes are short sequences generally of between 15 and 30 bases in length which correspond to unique sequences of DNA already known to occur in the gene. These probes will detect single-base mismatches.

In the study of X-linked RP, since the gene product is not known and the genetic lesion has not yet been found, the probes chosen are those which recognise variation. Normal DNA shows considerable variation between individuals, and, in particular, cleavage sites can be gained or lost through mutation of a

single base pair in the recognition sequence. A site may be therefore present or absent, resulting in fragments of different lengths in different individuals. This is known as a Restriction Fragment Length Polymorphism or RFLP. RFLPs are transmitted in a mendelian fashion and they can be detected by suitable probes which demonstrate that individuals are either homozygous for one or other fragment length, or they are heterozygous and produce one fragment length from one chromosome and a different fragment length from the homologous chromosome. This is illustrated in Figure 1 where the left hand side shows the Southern blots obtained from a family with X-linked RP shown on the right. The unaffected father (I.1) is hemizygous for the 12 kb allele, while the carrier mother is heterozygous and has both 12 kb and 9 kb alleles denoted as A1–A2. The son who has inherited the 9 kb allele (II.2) is affected, while the sons who have inherited the 12 kb allele (II.1 and II.4) are unaffected. Clearly in this family, the 9 kb allele is segregating with the disease locus.

RFLPs are used extensively in linkage studies, and they constitute valuable markers when they are situated sufficiently close to the location of the gene being studied. The first report of linkage in X-linked RP was to a probe L1.28 which maps to the region Xp11.3 of the X chromosome.<sup>6</sup> Probe names derive variously from the names of researchers who isolated them, the techniques used to clone them, or a variety of trivial sources. These names tell us nothing about their location or usefulness in linkage mapping. However the particular chromosomal location detected by each probe, known as a locus, is assigned a so-called 'D' number. This consists of the letter D (for DNA) followed by the number of the chromosome of origin (or an X or Y if the probe derives from a sex chromosome), then an S or a Z to denote single copy or repeat sequence respectively, and finally a number assigned in order of discovery to each probe found on a particular chromosome.

Thus for example, the locus detected by probe L1.28 is DXS7, since it was the seventh single copy probe to be assigned to the X chromosome. Similarly locus D1Z2 contains a highly repeated sequence on chromosome 1.



**Fig. 1.** On the left hand side are the Southern blots obtained from the family with X-linked RP shown on the right. In this family, the 9 kb allele, designated as A2, segregates with the disease locus.

If a sequence is subsequently proved to be part of a gene, it is assigned a G number (for gene) composed in a similar manner.

In order to detect linkage, carriers must be informative, that is, they must be heterozygous for the RFLP so that one can distinguish between their X chromosomes and identify which carries the abnormal gene. In practice, the frequency for each particular version or allele of an RFLP varies in the population. The most useful RFLPs are those for which no one allele is very much more common in the population than the other or others. Thus more of the carriers tested are likely to be heterozygous and therefore informative for that probe. In addition to the more common 'dimorphic' RFLPs (those for which there are only two alleles), another type of probe has been used successfully in studies of X-linked RP. Scattered throughout the genome, there are blocks of short sequences of DNA of between 2 and 80 base pairs which are repeated in tandem, head to tail, a variable number of times. These sequences are known as VNTRs (Variable Number Tandem Repeats) or mini satellites and, as a result of

the variation in repeat block size, the loci which contain mini satellites are hypervariable.<sup>7</sup> Probes to these sequences detect bands of different lengths on Southern blots depending on the number of tandem repeats, and as such they constitute a multi-allele system for which a high proportion of individuals are heterozygous. One example of such a hypervariable probe is M27 $\beta$  which maps to the region between the centromere and Xp11.<sup>8</sup> Another, which also maps to this region, is a microsatellite, the name given to a region of the DNA sequence which contains runs of a variable number of the dinucleotide CA. Since the difference between the many different versions or alleles found at this locus may be as little as two basepairs (one repeat unit), alleles cannot be resolved by conventional methods. Instead, the sequence containing the repeat is amplified millions of times using the newly discovered polymerase chain reaction (PCR).<sup>9</sup> Short sequences (oligomers) of around 20 bp from either side of the microsatellite are used to prime polymerase synthesis across the region of interest. These oligomers are mixed with DNA from potentially as little

as a single cell, then the heat stable Taq polymerase is allowed to read across the microsatellite. Next the new copies are denatured at 94°C, oligomers are allowed to reanneal, then the sequence is read again. Each time this cycle repeats, the number of copies of the sequence doubles. After 30 cycles, there is enough DNA to see using ethidium bromide staining. The microsatellite alleles are then visualised on a high resolution polyacrylamide sequencing gel.

Linkage studies are carried out by analysing variants at a range of loci including the disease gene, to assess the extent to which they are passed on together from one generation to the next (reviewed by Ott).<sup>10</sup> In a mother carrying X-linked RP, the disease gene will be on one of her two X chromosomes which will also have a particular variant or allele of the marker locus. During the process of passing on one X chromosome to her offspring, both of her X chromosomes exchange genetic material in a process called recombination or crossing-over. As a result, the closer the marker locus is to the disease gene, the more likely they are to be passed on together. Thus loci are described in terms of a genetic distance which is the recombination distance between the disease locus and the marker (or RFLP) system. This genetic distance is given either in centimorgans (cM) or as a recombination fraction with a distance of 1 cM corresponding to 1% of recombination. The closer the linkage, the smaller the recombination fraction. A genetic distance of 1 cM has been found to correspond approximately to a physical distance of 1 Mb. Linkage is measured by determining recombinants within a pedigree; these are the individuals in whom crossing over has taken place between the marker and the disease locus. The probability of linkage, which depends on the number of recombinants, is calculated for a given recombination fraction. Two loci are either linked at a certain recombination fraction or they are unlinked. The ratio of these two probabilities is the odds that the two loci are linked, and the logarithm to the base 10 of this ratio is the lod score. A lod score of three corresponds to odds of 1000:1 in favour of linkage, while negative values of the lod score are against linkage. Since lod scores for each

pedigree are additive, data can be pooled from several families providing there is no genetic heterogeneity.

As stated previously, the first localisation of X-linked RP was by Bhattacharya and colleagues<sup>6</sup> to probe L1.28, at a distance of 6–8 cM with a lod score of 6.8. However, genetic heterogeneity was suggested by Francke *et al.*,<sup>11</sup> who reported a male patient known as 'BB', with a deletion in band Xp21, who suffered from a syndrome which included Duchenne muscular dystrophy (DMD), chronic granulomatous disease, the McLeod syndrome and retinitis pigmentosa. The presence of a deletion in this patient could be demonstrated by absence of a hybridisation signal with probe 754, which lies some considerable distance distal (meaning towards the end of the chromosome) to L1.28. This suggested a second locus for X-linked RP. Further linkage studies by Denton *et al.*<sup>12</sup> and others have confirmed the existence of this locus. It is now accepted that there are two X-linked RP loci known as RP2 and RP3. RP2 is the locus linked to L1.28 at a distance of 6–8 cM and a lod score of 9.35,<sup>13</sup> and is referred to as the proximal locus (meaning close to the centromere). RP3 is the distal locus and is linked to probe 754 at distance of 4 cM and a lod score of 5.44.<sup>12</sup> The two loci are further distinguished clinically by the presence of a tapetal reflex in heterozygotes with RP3, but no tapetal reflex has been found so far in heterozygotes with RP2.<sup>14</sup> Recent unpublished studies suggest that 55% of families with X-linked RP have the RP2 form, and 45% have the RP3 form. (Wright AF, personal communication).

The distance of RP2 to its nearest markers is 6–8 cM, but to clone the gene, researchers need to be within distances of the order of 1 cM (which is approximately one million base pairs). The first step is to find markers closely flanking the gene and to construct a genetic map of the region round the gene. Various markers are used in linkage as individuals in families may be informative for one marker but not for another. The results of linkage to a set of markers are analysed by computer. This process, known as multipoint linkage, results in an ordering of markers relative to the gene locus and termed a genetic

map. X chromosome markers are generated from various sources. For example mouse-human hybrid cell lines have been established which contain the X chromosome as the only human chromosome. Another method is to use a fluorescence activated cell sorter to flow-sort human chromosomes and to obtain a preparation containing only X chromosomes. The material is then cleaved and cloned in appropriate vectors to make what is termed an X chromosome 'library'. Clones selected at random from such a library should contain DNA fragments from various locations on the X chromosome.

Once flanking markers are found, a physical map of markers corresponding to actual map distances is then constructed by using a number of techniques. The first of these is pulsed field gel electrophoresis (PFGE) which can separate fragments of DNA of anything up to 10 Mb in size (reviewed by Barlow).<sup>15</sup> DNA is cleaved by restriction enzymes chosen to cut at rare cleavage sites, thus generating large fragments. These are then subjected to electrophoresis on a gel, but the voltage is applied in pulses which alternate in two perpendicular directions. This variation in direction combined with a variation in the duration of the pulse results in size separation of very large fragments. The size of fragments is calibrated by size markers. These gels can be subjected to Southern blotting in the usual way. The object is to find two probes which hybridise to different ends of the same fragment, as this means that the probes are separated by a distance which is no more than the length of this fragment. A map can thus be built up which places the markers in order and determines the physical distances between them (that is, the distance in base pairs rather than the distance in units of recombination). If two markers are found which flank the gene and which are separated by a distance of less than 1 Mb (1000 kb) as demonstrated by PFGE, then that fragment can be cut out of the gel, eluted, and used to create a 'library' of clones from a specific region of the X chromosome.

The relatively 'fine structure' map generated by this approach over perhaps a megabase around the disease locus will contain 15 to 20 genes as well as the gene of interest. This

region could be amplified by cloning a PFGE fragment in a yeast artificial chromosome which is a vector capable of accepting large inserts. Also probes close to or beyond the point of the last known recombination can be used to 'walk' in towards the gene. Clones at either side are used to identify overlapping sequences in X chromosome libraries which project into the unknown region. These in turn can identify sequences adjacent to them. Each new clone constitutes a step in what is known as a chromosome walk towards the gene of interest. While the region is being mapped and cloned by these approaches, it is possible by various means to identify gene sequences within the DNA being studied. One of these is based on the presence of HpaII tiny fragment (HTF islands) which are associated with gene sequences. HpaII is a restriction enzyme which recognises the sequence motif CCGG. The dinucleotide CG (cytosine guanine) occurs relatively rarely in human DNA, and is usually methylated. Since HpaII will not cut where the CG in the middle of its cleavage site is methylated, it rarely cuts human DNA and gives mainly very large fragments. However, in HTF islands, which often occur at the beginning of genes, there are clusters of unmethylated CG dinucleotides, which give rise to the so-called HpaII tiny fragments. These islands are therefore distinguishable in genomic DNA which is depleted in CG and methylated. They can be located by means of HpaII, and also by other restriction enzymes, such as SacII and SStII, which also have CG in their cleavage sites.<sup>16</sup> Some genes do not have HTF islands, and therefore clones are selected which hybridise to DNA from different species, the aim being to find conserved sequences as these are likely to be gene sequences. These putative gene sequences and the HTF island sequences are tested against RNA from various tissues such as brain, liver, and retina with the aim of finding which are tissue specific. Any gene which is expressed in the retina would clearly be a strong candidate. The success of this method however, will depend to some extent on the expression and concentration of the gene product involved, both in affected and unaffected individuals.

Progress so far, at the time of writing, indi-

cates that there are at least two loci for X-linked RP, and there may be other X-linked retinal dystrophies. Although the genes responsible for X-linked RP have not yet been characterised, the existing technology has one practical application. Prenatal diagnosis using chorion villi biopsy at about eight weeks' gestation is available for some heterozygotes in some families. The proximal or distal locus has to be assigned and this is possible if the tapetal reflex has been observed. It may also be feasible by comparing linkage data to one or other set of flanking markers, and in practice, this service can be offered to a third of families studied so far.

### Conclusions

Linkage studies have proved that there are at least two locations on the X chromosome for genes implicated in X-linked RP. With this information it is possible to provide informed counselling and prenatal testing in some cases. Ultimately the approaches described should lead to the isolation of the genes which, when defective, cause the RP phenotype. The effort which is going into this task both in this country and abroad is such that the genes may be isolated very soon, and should certainly be found within the next three to five years. Once they are found it will be possible to determine exactly what defects are involved, and how the disease process manifests itself. From a greater understanding of what the RP defect is, it may in turn become possible to intervene in some way to alleviate the symptoms of affected individuals.

We wish to thank Dr Shomi Bhattacharya for the use of a previously published figure and for helpful suggestions. The authors gratefully acknowledge grants from the British Retinitis Pigmentosa Society, and the American RP Foundation Fighting Blindness, and thank them for their support. CFI is the recipient of a Wellcome Research Fellowship.

### References

- <sup>1</sup> Davies KE, ed: *Genome analysis; a practical approach*. Oxford: IRL Press, 1988.
- <sup>2</sup> Roberts RJ: *Restriction enzymes*. In: *Nucleic acid hybridisation a practical approach*. Hames BD and Higgins SJ, eds. Oxford: IRL Press, 1985; 203–10.
- <sup>3</sup> Sharp PA, Sugden B, Sambrook J: Detection of two restriction endonuclease activities in *Haemophilus parainfluenzae* using analytical agarose. *Biochemistry* 1973, **12**: 3055–63.
- <sup>4</sup> Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975, **98**: 503–17.
- <sup>5</sup> Weatherall DJ and Cleff JB: Recent developments in the molecular genetics of human haemoglobin. *Cell* 1979, **16**: 467–79.
- <sup>6</sup> Bhattacharya SS, Wright AF, Clayton JF, Price WH, Phillips CI, McKeown CME, Jay M, Bird AC, Pearson PL, Southern EM, Evans HJ: Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature* 1984, **309**: 253–5.
- <sup>7</sup> Jeffreys AJ, Wilson V, Thein SL: Hypervariable minisatellite regions in human DNA. *Nature* 1985, **314**: 67–73.
- <sup>8</sup> Meitinger T, Fraser NA, Lorenz B, Zrenner E, Murken J, Craig IW: Linkage of X-linked retinitis pigmentosa to the hypervariable DNA marker M27B (DXS255). *Hum Genet* 1989, **81**: 283–6.
- <sup>9</sup> Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988, **239**: 487–91.
- <sup>10</sup> Ott J: A short guide to linkage analysis. In: *Human genetic diseases, a practical approach*. Davies KE ed. Oxford: IRL Press, 1986; 19–32.
- <sup>11</sup> Francke U, Ochs HD, de Martinville B, Giacalone J, Lindgren V, Distèche C, Pogo RA, Hofker MH, van Ommen GJB, Pearson PL, Wedgwood RJ: Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. *Am J Hum Genet* 1985, **37**: 250–67.
- <sup>12</sup> Denton MJ, Chen JD, Serraville S, Colley P, Halliday FB, Donald J: Analysis of linkage relationships of X-linked retinitis pigmentosa with the following Xp loci: L1.28, OTC, 754, XJ-1.1, pERT87, and C7. *Hum Genet* 1988, **78**: 60–4.
- <sup>13</sup> Wright AF, Bhattacharya SS, Clayton JF, Dempster M, Tippett P, McKeown CME, Jay M, Jay B, Bird AC: 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* 1987, **41**: 635–44.
- <sup>14</sup> Nussbaum RL, Lewis RA, Lesko JG, Farrell R: Mapping X-linked ocular diseases: II. Linkage of X-linked retinitis pigmentosa to X chromosomal short arm markers. *Hum Genet* 1985, **70**: 45–50.
- <sup>15</sup> Barlow D: Pulsed field gel electrophoresis. *Genome* 1989, **31**: 465–7.
- <sup>16</sup> Lindsay S and Bird AP: Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* 1987, **327**: 336–8.