Identification of a gene for autosomal dominant polycystic kidney disease: implications for understanding the pathogenesis and treatment of the disease

P. C. Harris

MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, UK

Introduction

The renal changes that occur in autosomal dominant polycystic kidney disease (ADPKD) are well documented. The end-stage kidney is often massively enlarged, consisting of multiple fluid-filled cysts that destroy the precise structure of the entire kidney. It is now clear, however, that ADPKD is not just a renal disease with cysts and connective-tissue abnormalities detected in many other organs [1]. Nevertheless, since the kidney is the major diseased organ, attempts to understand the pathogenesis of ADPKD have focused on the cell biology and biochemistry of renal cystic tissue.

The earliest stages of cyst formation are characterized by a splitting, thickening and disorganization of the tubular basement membrane [2,3], which appears to occur simultaneously with changes of the associated tubular epithelia to a cystic phenotype [4]. Cystic epithelia have altered growth characteristics and responses [5]; they express a range of proteins normally seen only in the fetus [6,7]; and the polarity of some membrane proteins is altered [8,9]. These changes illustrate the unique differentiation state of cystic epithelia which is probably central to its role in cyst development. However, it has not been possible from these analyses to determine the primary event which triggers these changes in the ADPKD kidney.

One advantage of studying a genetic disease is that the primary defect can be identified, by its position in the genome, without understanding its function. This positional cloning method involves the identification of a genetic marker which always segregates with the disease when traced through disease families. This approach appeared attractive in ADPKD because the mapping of the PKD1 locus to chromosome 16 by genetic linkage with α globin, in 1985 [19]. More recently, the PKD2 locus has been mapped to the chromosome region 4q12–22 [20,21].

Genetic heterogeneity of ADPKD

Most cases of ADPKD are due to the polycystic kidney disease 1 (PKD1) locus (~85% in the European population; [11]) with the majority of the remainder due to a second locus, PKD2. However, a number of families with disease not linked to either of these loci have recently been described [12–14]. Comparisons of PKD1 and PKD2 families have shown that PKD1 is a more severe disease with an average age of end-stage renal failure of 54.2–59 years compared with 66.2–71.5 years for PKD2 [15–18]. The first step toward identifying an ADPKD gene was the mapping of the PKD1 locus to chromosome 16 by genetic linkage with α globin, in 1985 [19]. More recently, the PKD2 locus has been mapped to the chromosome region 4q12–22 [20,21].

Identifying the PKD1 gene

Once linkage has been established a stage of physical and genetic mapping is required to determine the precise position of the disease locus. In the case of PKD1, 7 years after initial linkage, the gene had been localized to a 600 kb interval in 16p13.3 [22–24]. However, because this was a gene rich area (containing ~20 different genes; [22]) identification of the PKD1 gene still proved difficult and an additional clue to pinpoint its location was required. In this case the clue turned out to be a rare chromosome rearrangement. Interestingly more than half of all positional cloning successes have required such a clue to aid identification of the gene [25].

A translocation involving chromosomes 16 and 22 was identified in a Portuguese family (family 77) with both ADPKD and tuberous sclerosis (TSC) [26] (a
tuberous sclerosis locus, TSC2, was mapped to the PKD1 region of chromosome 16 in 1992; [27]). Individuals with the balanced exchange had ADPKD, while the son with TSC had unbalanced products, which resulted in deletion of a small region of chromosome 16 (and part of chromosome 22). One possible explanation for these findings was that the translocation breakpoint on chromosome 16 interrupted the PKD1 gene while the deleted region in the son contained the TSC2 gene. The translocation breakpoint on chromosome 16 was cloned and characterized and was found to interrupt a gene that encoded a large transcript (~14 kb) (Figure 1) [26]. Ultimately the identification of further mutations of this gene in PKD1 patients proved that this was the PKD1 gene. Meanwhile, by characterizing a number of large and overlapping deletions detected in other TSC2 patients, and subsequently the identification of intragenic mutations, the TSC2 gene was identified [28]. Remarkably the TSC2 and PKD1 genes lie immediately next to each other, in a tail to tail orientation, with only 60 bp between their polyadenylation sites [29] (Figure 1).

The PKD1 gene lies in a duplicated genome region

Full characterization of the PKD1 gene has proved difficult because all but 3.8 kb at the 3' end of the transcript is encoded by a genomic region reiterated several times on the same chromosome (in 16p13.1) [26] (Figure 1). This arrangement is further complicated because the duplicate area encodes three genes (the HG loci) that share substantial homology with PKD1 (>95%) over most of their length. The degree of homology between the different genomic areas, and between the transcripts means that when analysed by hybridization methods or PCR (including RT-PCR) both areas are generally visualized simultaneously. This complication has impeded mutation detection and delayed characterisation of the entire transcript. In the initial description of the PKD1 gene only 5.7 kb at the 3' end of the transcript was cloned and sequenced; the predicted protein showed no clear homologies with other proteins [26], so its function remained unclear.

The PKD1 protein, polycystin

Analysis of the cDNA sequence showed a predicted PKD1 protein, polycystin, of 4302 aa. In contrast to the initial analysis of the C-terminal region of the protein [26], study of the whole molecule revealed several clear homologies with known motifs (Figure 2) [31,32]. At the extreme N-terminal end of the protein is a signal peptide which is immediately followed by two leucine-rich repeats (LRRs), surrounded by cysteine containing amino and carboxy flanking regions [33]. Following this is a C-type lectin domain with many of the residues that are normally involved in binding carbohydrate and calcium conserved [34]. A cysteine-rich LDL-A domain has also been predicted next to the lectin [31]. Approximately one-third of the entire protein consists of 14–16 copies of a repeated motif of approximately 85 aa, 13–15 of which are arranged in a tandem array (Figure 2) [31,32]. Secondary structure predictions of this area shows that

---

**Fig. 1.** Map of the PKD1 region of chromosome 16 showing the positions of the PKD1 and TSC2 genes. The intron/exon structure of PKD1 and the genomic extent (bracket) and transcript size (solid box) of TSC2 are shown. The area of the PKD1 gene affected by the various PKD1 mutations so far described [26,40–42]; and the smallest (solid line) and largest (dashed line and arrows) regions lost in TSC2/PKD1 deletions [43], are shown. The breakpoint of the chromosome translocation detected in family 77, which pinpointed the location of the gene, is also illustrated.
it forms a β-strand and turn structure that is likely to form a sandwich of two β-sheets. This structure is also found in immunoglobulin (Ig) domains [35] and similarity between the PKD1 repeat and Ig motifs has been described [32]. Immediately after the Ig-like domains are four repeats with sequence similarity to fibronectin type III (FN-III) domains [36]. The presence of the signal peptide and these various protein domains indicate that the N-terminal two-thirds of the protein is extracellular. The motifs identified are generally involved in protein–protein or protein–carbohydrate binding and indicate that polycystin may be involved in cell–cell or cell–matrix interactions [33,37–39].

The C-terminal one-third of the protein has no clear homologies with other proteins and so its function is more difficult to determine. However, it has been noted that this area contains several hydrophobic regions which might represent transmembrane (TM) domains [30,32]. A model that considered hydrophobicity and charged residues in the C-terminal region predicted a topology of 11 TM domains with an extracellular N-terminus and cytoplasmic C terminus (Fig. 2) [32]. It is not clear if this membrane-associated area functions as a receptor or forms a pore or channel. It is possible that the cytoplasmic tail enables polycystin to signal between extracellular targets and the cells interior.

**PKD1 mutations**

Study of mutations of the *PKD1* gene should allow the mechanism of disease and important functional regions of the protein to be identified. Up to this time approximately a dozen mutations have been characterized in typical PKD1 patients [26,40–42]. Most of these are stop or frameshifting mutations, or deletions, in the single copy, 3' part of the gene (Figure 1), which is most readily analysed. In each of these cases a transcript has been detected from the mutated gene, suggesting that an abnormal protein, which would have a defective C-terminal region, is produced, causing disease through a gain of function. Some of these mutations identify small regions of the protein which must be functionally important [41]. A second type of *PKD1* mutation has also been described, which are large deletions that disrupt *PKD1* and the adjacent *TSC2* gene [43]. In these contiguous gene syndrome cases a specific phenotype of TSC and severe, childhood onset polycystic kidney disease is seen. These are clear inactivating mutations of *PKD1* (*PKD1* is often completely deleted; Figure 1) and indicate that loss of a *PKD1* allele, at least in combination with a *TSC2* mutation, causes severe polycystic disease.

As there is evidence of both null alleles and possible gain of function mutations causing cystic development, the mechanism of disease causation is still uncertain. It is possible, however, that these two different types of mutation may be reconciled if *PKD1* generates more than one protein product. There is some evidence that alternative splicing of *PKD1* may occur, resulting in the production of a much shorter, secreted product [31,32]. It is therefore possible that the two different types of *PKD1* mutation disrupt different splice forms, so that those in typical adult onset cases disrupt just the full length polycystin, while the null mutations inactivate both forms. The search for further mutations especially, in the 5' end of *PKD1*, and characterization of the splicing variants of *PKD1*, is required to fully establish the mutational mechanism in this disease.

**Polycystin and the pathogenesis of polycystic kidney disease**

How does cloning and characterizing a gene that is mutated in a complex condition, such as ADPKD, aid in the understanding of that disorder? Cloning and sequencing of *PKD1* has revealed the likely structure (Figure 2) and possible function of the defective protein in this disease. Previous studies have highlighted the simultaneous occurrence of basement membrane abnormalities [2,3] and changes in the differentiation state of tubular epithelia [44] at the point of cystic
dilation. It has been unclear, however, whether the primary defect was an extracellular matrix protein [45], a cellular growth factor [46] or if the basic defect lay in communication between cell and matrix [44]. It is well documented that cell–matrix interactions are important both for generating an organized basement membrane [47] and to control the status of cellular differentiation [48–50]. We can see that polycystin may play a role in this interaction, directly linking the basement membrane and cytoplasm of associated epithelial cells. Disruption of this interaction by reduced or aberrant polycystin expression may be the primary defect that triggers the dedifferentiation of cystic epithelia and basement membrane disorganization, resulting in cyst development.

Various parts of this hypothesis can be tested. Polycystin would be predicted to be located on the basolateral surface of epithelial cells in kidney tubules. Studies of polycystin expression that are presently under way, by mRNA in situ hybridization and immunohistochemistry, will help determine its cellular location. The significance of a recent study showing the PKD1 protein in the mesenchyme and vasculature [51] is now unclear, since the epitope to these antibodies is now not thought to form part of polycystin because of an error in the original sequence [52]. It will be possible to refine the model of the topology of the proposed membrane-bound region of polycystin using specific antibodies and by comparing to the sequence of PKD1 proteins from other species. Determining the targets that polycystin binds to both extra- and intracellularly could identify other proteins involved in the development of a cystic phenotype; a cystic pathway could potentially be defined. These other molecules may be the primary cause of the other renal cystic diseases of man, including those associated with the uncloned ADPKD genes.

Significance to clinical practice

Identification of the PKD1 gene improves the prospects for presymptomatic diagnosis in families where the mutation has been detected. However, as yet only a few such mutations have been described and each of these has been identified in just a single family [41], indicating that many different mutations cause PKD1. The rate of mutation detection will have to improve (especially in the duplicated area of the gene) before such testing can become widely available. Microsatellite markers around PKD1 remain useful for genetic diagnosis at this time [53,54].

Further understanding of cyst formation obtained from studying polycystin and the elucidation of steps in a cystic pathway may suggest points for potential therapeutic intervention. However, if the aim is to prevent cyst formation, intervention may have to occur early, because microcysts have been noted in PKD1 kidney as early as 12 weeks [55–57]. The prospects for gene therapy are not clear. The PKD1 transcript is large and would need to be precisely targeted in the kidney at an early stage. Furthermore, it is not known if introduction of more normal product will prevent disease; blocking the action of the abnormal protein may be necessary.

It is possible that the identification of factors that modify disease severity may be important as prognostic indicators and will further our understanding of the disease process. Modifying factors are known to affect disease severity radically in mouse models of polycystic kidney disease [58] and that considerable unexplained variability in disease severity is found in PKD1, even within families [59]. Finally we must remember that many individuals live clinically normal lives with ADPKD; not developing ESRD [15]. If the rate of cyst expansion could be limited in all patients, even if the development of cysts could not be prevented, this may be an effective treatment. Consequently, further understanding of the process of cyst expansion and other factors affecting the rate of renal destruction [60] may be therapeutically important.

Acknowledgements. The author would like to thank the many individuals whose work is summarized here, and the support of the Medical Research Council and the Wellcome Trust.

References

15. Gabow PA, Johnson AM, Kaehny WD et al. Factors affecting


41. Milunovic J, Rust PF, Fialkow PJ et al. Intrafamalial pheno-