Nephrol Dial Transplant (1996) 11 [Suppl 6]: 2-4

Nephrology
Dialysis
Transplantation

Structural motifs of the PKD1 protein

C. Löhning[1], M.Pohlschmidt[1], M. A. Glücksmann-Kuis[2], G. Duyk[2], P. Bork[3], M. O. Schneider[4], S. T. Reeders[4] and A.-M. Frischauf[1]

[1]Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK, [2]Millennium Pharmaceuticals, Incorporated, Cambridge, MA 02139, USA, [3]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg and Max Delbrück Centre for Molecular Medicine, D-13125 Berlin-Buch, Germany, and [4]Renal Division, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

Abstract. The complete sequence of the polycystic kidney disease gene (PKD1) and its transcript have been described. The predicted protein is not a member of a previously described gene family, but contains several structural motifs that are present in proteins of known function. Most of these domains are present in the extracellular parts of proteins involved in interactions with other proteins and carbohydrates. The PKD1 gene product also contains potential transmembrane sequences. The molecule is likely to be involved in cell–cell or cell–matrix interactions, which is consistent with the different manifestations of polycystic kidney disease.

Key words: cell–cell interaction; cell–matrix interaction; polycystic kidney disease

Background

The gene for polycystic kidney disease (PKD1) was one of the first disease genes to be localized by linkage analysis with DNA markers [1], but was only identified in 1994 [2]. There were several reasons for the long delay between the two events. First, the candidate gene approach of guessing the right gene on the basis of the cellular abnormalities of cystic epithelia such as increased proliferation, changed polarity and basement membrane abnormalities [3] was very difficult. In retrospect, the candidate gene approach could not have worked anyway, since the PKD1 gene represents a novel gene with no related family members previously described. Second, there was no help from a genetically equivalent animal model [4], since all described animal models of polycystic kidney disease involve different loci or have not been mapped [5]. The 'positional cloning' approach which eventually led to the identification of the PKD1 gene was slow and laborious because of the high density of genes in the candidate region, all of which were candidates for PKD1 and had to be tested for mutations in DNA from PKD1 patients. A breakthrough was achieved when the European PKD1 Consortium identified the gene with the help of a translocation that segregated with the disease in one family. Several mutations leading to a predicted truncated protein clearly showed that the search for the gene was finished [2]. However, an unforeseen problem arose from the fact that the major part of the PKD1 gene was present not only in a single copy, but was duplicated several times in a more proximal location on chromosome 16p. Since those copies appeared to be transcribed and very similar in sequence to PKD1, a large part of the gene remained uncharacterized and with it the mutations located there. In view of the difficulties in distinguishing cDNAs corresponding to the different loci, the genomic DNA of the PKD1 locus was the obvious source of unambiguous sequence information since overlapping cosmids could be clearly assigned to the PKD1 locus. Three groups based their characterization of the gene on the sequence from the genomic cosmid clones from the PKD1 locus, but used different approaches to identify the transcribed sequence. The American PKD1 Consortium [6] relied exclusively on sequence interpretation programs to predict coding parts in the genomic sequence. This led to a largely correct overall picture of the gene, though a long exon was split up and the 5′ end was not found. The International Polycystic Kidney Disease Consortium [7] used the sequence of the transcripts not only of PKD1 but also of the related loci to identify exons on the genomic DNA sequence. This produced a predicted PKD1 transcript that was very similar to the one described by Hughes *et al.* [8], who used a radiation hybrid that contained only the PKD1 locus but not the related loci as a source of PKD1 transcript, again comparing the results to the genomic sequence.

Correspondence and offprints requests to: A.-M. Frischauf, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK.

Motifs in the PKD1 gene product

The predicted PKD1 protein is not a member of a previously described gene family. It does, however, contain a number of peptide motifs found in characterized proteins of known function and cellular location. Most of these motifs are associated with extracellular proteins. A signal sequence at the 5′ end precedes two leucine-rich repeats (LRRs) that are flanked by cysteine-rich regions.LRRs are found in a tandem arrangement in many different proteins, all of which seem to be involved in protein–protein interactions [9]. Not all LRRs are associated with cysteine-rich regions. An example for which the crystal structure in the presence of ligand has been solved is porcine RNase inhibitor, which consists of 15 LRRs [10]. The subgroup containing conserved cysteine-rich regions include molecules involved in cell adhesion, such as human platelet glycoproteins I and V [11], and signal transducing proteins, such as the protooncogene Trk, the high affinity nerve growth factor receptor [12] and toll, a transmembrane receptor mediating dorsal–ventral polarity in the *Drosophila* embryo [13]. For toll it has been shown that the leucine-rich repeats regulate the intrinsic signalling activity of its cytoplasmic domain [14].

The next functionally defined motif of the PKD1 protein is a C-type lectin domain. C-type lectins are Ca^{2+}-dependent carbohydrate-recognition proteins with a wide range of specificities and functions [15,16]. They are involved in interactions at the cell surface, such as endocytosis (asialoglycoproteins), cell adhesion (selectins) and response to bacterial and fungal pathogens [macrophage mannose receptor and mannose-binding proteins ('collectins')]. All lectin domains (carbohydrate recognition domains) are extracellular.

Proceeding in a C-terminal direction along the PKD1 polypeptide is a single copy of a motif related to the repeated motifs involved in ligand binding in LDL-receptor and LDL receptor-related protein [17]. Related modules also occur in the complement factors C8 and C9 [18].

The next recognizable motif is present in extracellular parts of previously characterized proteins and was named PKD repeat [7]. Closely related domains were found in a transmembrane protein of melanocytes (Pmel17) and the melanosomal matrix protein MMP115 [19]. Surprisingly, such domains are also found in some bacterial collagenases, a protease [20] and a surface layer protein from *Methanothermus* [21]. Harpaz and Chothia emphasized that this motif is related to the previously described I-set of the immunoglobulin superfamily, which includes cell adhesion molecules and receptors [22]. Proteins with immunoglobulin domains have a diversity of functions with a role in recognition at the cell surface as a common feature. In the PKD1 gene product a single PKD1 (immunoglobulin-like) domain is found between the LRRs and the C-type lectin domain. The remainder of the repeats are located in a cluster in the middle part of the protein. The sequence of these domains is widely diverged, which explains the identification of 13 repeats by The International Polycystic Kidney Disease Consortium [7] and 15 repeats by Hughes *et al.* [8]. The latter group also describes the presence of four type III fibronectin related repeats.

The next striking feature of the PKD1 protein is a large region containing several hydrophobic stretches with the potential to form transmembrane domains. There is, however, no known protein with such an arrangement of transmembrane sequences. Hughes *et al.* [8] suggested the presence of 11 membrane-spanning segments on a basis of a hydrophobicity plot. This region is followed by a region with no obvious structural domains until the C-terminal end of the protein.

Outlook

Knowing what we know now about the PKD1 gene product, what does this tell us about its function and its potential role in polycystic kidney disease? Most of the domains of the protein are found in the extracellular part of characterized proteins and point to a role in cell–cell and/or cell–matrix interaction. By analogy, PKD1 is likely to be glycosylated, and many potential glycosylation sites have been found. PKD1 may be a transmembrane protein and could also be involved in signal transduction. Future studies will aim to identify proteins and other ligands of PKD1 both outside and possibly inside the cell. In a general sense, this picture of PKD1 fits well with the different manifestations of polycystic kidney disease. Thickening of the basement membrane is an early event in PKD1 and different glycoprotein composition of the extracellular matrix in cystic epithelia has been observed [3]. Another observation is the mislocalization of Na^+/K^+-ATPase to the apical membrane in cystic epithelia [23]. Since the interaction between epithelial cells and extracellular matrix is known to play a role in the differentiation of the cells, it is plausible that these defects reflect changes in the differentiation state of the cells brought on by changes in the extracellular matrix. The same argument could be used to explain the abnormal proliferation in the lining of the cysts. Even though a connection between the overall structure and the putative role of the PKD1 protein to the disease phenotype can be postulated, it is not clear how exactly the mutations in the PKD1 gene cause the defect. All mutations described so far are located in the unique C-terminal part of the protein and lead to premature termination [24–26]. Since mutations in the main part of the PKD1 molecule have not yet been identified due to the presence of the closely related transcripts, it is not possible to say whether all disease-causing mutations are due to truncation. In this context it is intriguing that both The International Polycystic Kidney Disease Consortium [7] and Hughes *et al.* [8] report differentially spliced mRNAs in unaffected individuals that would lead to premature termination of the protein. Again, the analysis is complicated by the

PKD1-related loci. Some of these questions will be more easily addressed in the mouse since there is no evidence for duplication of this region in the mouse genome [Olsson, Löhning, Horsley, Kearney, Harris and Frischauf, unpublished data], and preliminary data show a high degree of sequence conservation between human and mouse PKD1 transcripts [Löhning, unpublished data]. In summary, the identification and characterization of the PKD1 gene has been an essential step in the molecular analysis of the disease. Even though mutation analysis is likely to be problematic in the near future, analysis of the functional properties of the gene product has begun and should lead to insight into the molecular and cellular mechanism of polycystic kidney disease.

Acknowledgements. We thank all other members of the International Polycystic Kidney Disease Consortium on whose work this review is based. C. L. was a fellow of the Deutsche Forschungsgemeinschaft.

References

1. Reeders ST, Breuning MH, Davies KE *et al.* A highly polymorphic marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 1985; 317: 542–544
2. The European Polycystic Kidney Disease Consortium. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 1994; 77: 881–894
3. Carone FA, Bacallao R, Kanwar YS. The pathogenesis of polycystic kidney disease. *Histol Histopathol* 1995; 10: 213–221
4. Himmelbauer H, Pohlschmidt M, Snarey A *et al.* Human mouse homologies in the region of the polycystic kidney disease gene (pkd1). *Genomics* 1992; 13: 35–38
5. Aziz N. Animal-models of polycystic kidney-disease. *BioEssays* 1995; 17: 703–712
6. The American PKD1 Consortium. Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat. *Hum Mol Genet* 1995; 4: 575–582
7. The International Polycystic Kidney Disease Consortium. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell* 1995; 81: 289–298
8. Hughes J, Ward CJ, Peral B *et al.* The polycystic kidney disease 1 (pkd1) gene encodes a novel protein with multiple cell recognition domains. *Nature Genet* 1995; 10: 151–160
9. Kobe B, Deisenhofer J. The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* 1994; 19: 415–421
10. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* 1995; 374: 183–186
11. Roth GJ. Developing relationships: arterial platelet adhesion, glycoprotein 1b, and leucine-rich glycoproteins. *Blood* 1991; 77: 5–19
12. Schneider R, Schweiger M. A novel modular mosaic of cell-adhesion motifs in the extracellular domains of the neurogenic trk and trkb tyrosine kinase receptors. *Oncogene* 1991; 6: 1807–1811
13. Hashimoto C, Hudson KL, Anderson KV. The Toll gene of *Drosophila*, required for dorsal–ventral embryonic polarity, appears to encode a transmembrane protein. *Cell* 1988; 52: 269–279
14. Winans KA, Hashimoto C. Ventralization of the *Drosophila* embryo by deletion of extracellular leucine-rich repeats in the toll protein. *Mol Biol Cell* 1995; 6: 587–596
15. Drickamer K. Two distinct classes of carbohydrate recognition domains in animal lectins. *J Biol Chem* 1988; 263: 9557–9560
16. Weis WI, Quesenberry MS, Taylor ME *et al.* Molecular mechanisms of complex carbohydrate recognition at the cell surface. *Cold Spring Harbor Symp Quant Biol* 1992; 57: 281–289
17. Krieger M, Herz J. Structures and functions of multiligand lipoprotein receptors—macrophage scavenger receptors and ldl receptor-related protein (lrp). *Annu Rev Biochem* 1994; 63: 601–637
18. Sudhof TC, Goldstein JL, Brown MS *et al.* The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 1985; 228: 815–822
19. Kwon BS. Pigmentation genes: the tyrosinase gene family and the pmel-17 gene family. *J Invest Dermatol* 1993; 100: S134–S140
20. Yoshihara K, Matsushita O, Minami J *et al.* Cloning and nucleotide sequence analysis of the colH gene from *Clostridium histolyticum* encoding a collagenase and a gelatinase. *J Bacteriol* 1994; 176: 6489–6496
21. Brockl G, Behr M, Fabry S *et al.* Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogens *Methanothermus fervidus* and *Methanothermus sociabilis*. *Eur J Biochem* 1991; 199: 147–152
22. Harpaz Y, Chothia C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol* 1994; 238: 528–539
23. Wilson PD, Sherwood AC, Palla K *et al.* Reversed polarity of Na(+) K(+) ATPase: mislocation to apical plasma membranes in polycystic kidney disease epithelia. *Am J Physiol* 1991; 260: 420–430
24. Schneider MC, Zhang F, Geng L *et al.* Identification of the autosomal dominant polycystic kidney disease gene. *Am J Hum Genet* 1994; 55: 37
25. Turco AE, Rossetti S, Bresin E *et al.* A novel nonsense mutation in the PKD1 gene (C3817T) is associated with autosomal dominant polycystic kidney disease (ADPKD) in a large 3 generation Italian family. *Hum Mol Genet* 1995; 4: 1331–1335
26. Peral B, Gamble V, Millan JLS *et al.* Splicing mutations of the polycystic kidney disease 1 (PKD1) gene induced by intronic deletion. *Hum Mol Genet* 1995; 4: 569–574