Polycystic Kidney Disease: The Complete Structure of the *PKD1* Gene and Its Protein

The International Polycystic Kidney Disease Consortium*

Summary

Mutations in the PKD1 gene are the most common cause of autosomal dominant polycystic kidney disease (ADPKD). Other PKD1-like loci on chromosome 16 are approximately 97% identical to PKD1. To determine the authentic PKD1 sequence, we obtained the genomic sequence of the PKD1 locus and assembled a PKD1 transcript from the sequence of 46 exons. The 14.5 kb PKD1 transcript encodes a 4304 amino acid protein that has a novel domain architecture. The amino-terminal half of the protein consists of a mosaic of previously described domains, including leucinerich repeats flanked by characteristic cysteine-rich structures, LDL-A and C-type lectin domains, and 14 units of a novel 80 amino acid domain. The presence of these domains suggests that the PKD1 protein is involved in adhesive protein-protein and protein-carbohydrate interactions in the extracellular compartment. We propose a hypothesis that links the predicted properties of the protein with the diverse phenotypic features of ADPKD.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common monogenic disorders in humans, affecting 1 in 1000 individuals. Its major manifestation is progressive cystic dilatation of the renal tubules, leading to renal failure in half of affected individuals by age 50. ADPKD is also associated with hepatic, pancreatic, and splenic cysts, cardiac valve abnormalities, and an increased incidence of cranial aneurysms and subarachnoid hemorrhage (Gabow, 1990).

Despite intensive investigation, the underlying biochemical defect in ADPKD remains unknown. A series of apparently unrelated abnormalities has been detected at the cellular and tissue levels both in ADPKD and in other forms of renal cystic disease. The most carefully documented of these findings are abnormalities in the composition of the tubular basement membrane, proliferation of tubular epithelial cells, and a reversal of the normal polarized distribution of cell membrane proteins such as the Na⁺/K⁺ ATPase (Wilson et al., 1986).

Phenotypically indistinguishable forms of ADPKD are caused by mutations in three separate loci (Reeders et al., 1985; Kimberling et al., 1993; Peters et al., 1993). Two of these loci, PKD1 and PKD2, have been mapped to the short arm of chromosome 16 and chromosome 4, respectively. The third locus has not been mapped (Fossdal et al., 1993; Daoust et al., 1995). Mutations in PKD1 account for approximately 90% of ADPKD cases. This locus previously had been mapped to a gene-rich 500 kb interval in band 16p13.3 (Germino et al., 1992) that includes the TSC2 locus for tuberous sclerosis (TS) (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). Some TS patients are known to develop renal cystic lesions that resemble those of ADPKD, which led investigators to examine families with TS for positional segregation of ADPKD. One unusual family had members with polycystic

*The International Polycystic Kidney Disease Consortium comprises the following groups: *Group 1:*

M. Alexandra Glücksmann-Kuis,† Olga Tayber,‡ Elizabeth A. Woolf,‡ Lydie Bougueleret, Nanhua Deng, Geoffrey D. Alperin, François Iris, Fiona Hawkins, Cheryl Munro, Nathan Lakey, and Geoffrey Duyk Millennium Pharmaceuticals, Incorporated Cambridge, Massachusetts 02139 Group 2:

Michael C. Schneider,[†] Lin Geng, Fan Zhang, Zhihui Zhao, Steven Torosian, Jing Zhou, and Stephen T. Reeders Renal Division Department of Medicine

[†]Correspondence should be addressed to M. Alexandra Glücksmann-Kuis, Michael C. Schneider, or Anna-Maria Frischauf. [‡]These authors contributed equally within this group. Brigham and Women's Hospital and Harvard Medical School Boston, Massachusetts 02115 Group 3: Peer Bork European Molecular Biology Laboratory Meyerhofstrasse 1 D-69117 Heidelberg Federal Republic of Germany and Max Delbrück Center for Molecular Medicine D-13125 Berlin-Buch Federal Republic of Germany Group 4: Marita Pohlschmidt,[‡] Corinna Löhning,[‡] Barbara Kraus, Urszula Nowicka, Alan L. S. Leung, and Anna-Maria Frischauf[†] Imperial Cancer Research Fund Lincoln's Inn Fields London WC2A 3PX England



Figure 1. Schematic Representations of the *PKD1* Genomic Region and the Full-Length *PKD1* cDNA

(A) Schematic representation of the *PKD1* genomic region. The large bold numbers indicate the numbers of selected exons. The numbers outside of the semicircle denote the size of the introns. The small numbers inside the circle represent the size of the exons. Solid lines inside the semicircle represent the cosmids from this region. N54T denotes a NotI site in cGGG10. The distance between exons 1 and 2 is larger than 15,377 bp by at least 1000 bp. The distance between exons 22 and 23 is larger than 796 bp by at least 2000 bp. These represent the genomic gaps not sequenced.

(B) Schematic representation of the full-length PKD1 cDNA. The numbers above the rectangles indicate selected exons. Below is a physical representation of the partial cDNAs described in the Experimental Procedures.

kidney disease who were found to have inherited a balanced translocation near the *TSC2* locus, with a breakpoint in a novel gene named the polycystic breakpoint protein gene (*PBP*). *PBP* codes for a 14 kb mRNA (European Polycystic Kidney Disease Consortium, 1994). Deletions and point mutations confined to the *PBP* gene confirmed its identity as the *PKD1* gene (European Polycystic Kidney Disease Consortium, 1994; Schneider et al., 1994, Am. J. Hum. Genet., abstract). cDNA clones comprising the terminal 5.6 kb of the *PKD1* transcript were found to contain an open reading frame (ORF) of 4.8 kb. Analysis of the deduced peptide encoded by the last third of the gene did not reveal any homologies to known proteins and, therefore, did not suggest a biochemical function for the product of the *PKD1* gene.

A major problem in the isolation and sequencing of the remaining part of the PKD1 gene has been the presence of several transcriptionally active copies of closely related PKD1-like sequences that map centromeric to PKD1 on chromosome 16p13.1 (European Polycystic Kidney Disease Consortium, 1994). This has posed great difficulty in distinguishing the PKD1 locus transcript from those of the PKD1-like loci.

Here we describe a strategy leading to the identification of the complete *PKD1* gene sequence. We also provide the genomic structure of the gene and show that the mRNA transcripts are alternatively spliced. One form of the *PKD1* transcript encodes a 4304 amino acid polypeptide with five distinct extracellular peptide domains that are likely to be involved in protein–protein and protein–carbohydrate interactions. Although the PKD1 protein shares domains with a number of extracellular proteins, the combination of domains found in PKD1 has not been found in any known protein.

Results

A series of overlapping cosmid clones spanning the predicted *PKD1* genetic interval has been described (Germino et al., 1992). The integrity of the cosmid contig was confirmed by long-range restriction mapping and genetic linkage analysis of polymorphic sequences derived from the cosmids. Three cosmids (cGGG1, cGGG10, and cDEB11 from centromere to telomere) form a contig that includes the 3' end of *TSC2* (cDEB11) and extends over 80 kb centromeric to it. At the proximal end of cGGG10, there is a CpG island represented by a NotI site, N54T (Himmelbauer et al., 1991) (Figure 1A).

To identify transcripts from the region, the cosmid clones were hybridized to a set of five cDNA libraries. KG8, a cDNA clone containing the last 3.2 kb of the *PBP* sequence and located on cDEB11, was mapped by use of a panel of somatic cell hybrids and found to hybridize to a single locus on chromosome 16p13 (data not shown). Sequence analysis showed that KG8 contains the polyadenylated 3' end of a gene and has an ORF of 2100 bp and a 1019 bp 3' untranslated region. KG8 was also found to contain a polymorphic microsatellite repeat (Snarey et al., 1994). Analysis of this repeat in a large number of *PKD1* kindreds revealed no recombination in the disease locus (S. Somlo, unpublished data).

To obtain clones extending 5' of KG8, the cosmids cGGG10 and cDEB11 were hybridized to a series of different cDNA libraries (see Experimental Procedures). In contrast with KG8, when some of the resulting cDNA clones were analyzed using somatic cell hybrid panels (data not shown), they were found to hybridize strongly to several loci on chromosome 16 as well as to the *PKD1* region. The restriction maps of the hybridizing loci were so similar

ng agai

that we concluded that a series of recent duplications of part of the *PKD1* gene had occurred (excluding the KG8 segment) and had given rise to several *PKD1*-like genomic segments.

Because of the high degree of similarity between *PKD1* and *PKD1*-like loci and because they all are transcriptionally active, it was not possible to determine the correct full-length *PKD1* cDNA sequence by assembling overlapping partial cDNA clones. To determine the sequence of the authentic *PKD1* transcript, we sequenced cGGG10 entirely and parts of cDEB11 containing *PKD1* exons.

Sequence of the Genomic Region of the *PKD1* Locus

The duplicated portion of the *PKD1* gene is largely contained within the cGGG10 cosmid. Prior to sequencing cGGG10, we established the integrity of the clone in several ways. First, the restriction map of cGGG10 was compared with a map of the genomic DNA from the *PKD1* region and was found to be identical. Second, restriction maps of the overlapping portions of cGGG1 and cDEB11 were compared with cGGG10 and were also found to be identical. Third, sequences derived from cGGG10 and overlapping portions of cDEB11 showed 100% identity. Finally, a P1 phage, PKD 1521, was obtained by screening a genomic P1 library with primers from the *TSC2* gene. No sequence differences were observed between PKD 1521 and cGGG10 in the regions sampled.

Several approaches were taken to obtain the sequence of cGGG10 (see Experimental Procedures). A final 10-fold sequence redundancy was achieved for this cosmid in order to compare the genomic sequence accurately with that of the *PKD1*-specific and *PKD1*-like cDNAs (homologous to this cosmid). The cGGG10 sequences were assembled into three contigs of 8 kb, 23 kb, and 4.4 kb, separated by 1 kb and 2.2 kb gaps (Figure 1). The cosmid cDEB11 was also sequenced and assembled to a 2-fold redundancy and compared with *PKD1*-specific cDNAs in order to obtain intron/exon boundaries of the unique 3' end of the gene.

cDNAs from the PKD1 and the PKD1-like Loci

To identify putative coding regions and intron/exon boundaries, genomic and cDNA sequences were compared (see Experimental Procedures). When the sequences of overlapping cDNAs were assembled, a transcript length of 14.5 kb was obtained. The predominant transcript detected by Northern blot analysis using the unique sequence KG8 probe is ~14 kb (data not shown), suggesting that the cDNA clones represent the full length of the *PKD1* transcript. Restriction and sequence analyses indicate that a CpG island overlaps the 5' end of the sequence. CpG islands have been found to mark the 5' ends of many genes (Antequera and Bird, 1993). The most 5' cDNA clones (UN53, UN54, and UN59) all have identical 5' ends, providing further evidence that no additional upstream exons were missed (see Experimental Procedures).

The cDNAs used to assemble the PKD1 transcript, along

with genomic exon/intron structure, are shown in Figures 1A and 1B. By comparing the sequences of overlapping cDNAs with the *PKD1* genomic sequence, *PKD1*-specific cDNAs were distinguished from those encoded by the homologous loci (see Experimental Procedures). We identified 46 exons and their exon/intron boundaries. The full-length transcript constructed from the genomic sequence of the exons produces a large continuous ORF of 12,912 bp.

Alternative splicing of the primary PKD1 transcript is apparent from the sequences of the cDNAs. For this reason, we sought to isolate a minimum of two cDNAs containing each exon, thereby increasing the probability that all exons that contribute to the PKD1 transcript were detected. Despite this degree of coverage, it is possible there are PKD1 transcripts containing exons that are not present in any of the cDNAs we sampled.

Exon 17 was found in two cDNAs (UN34 and BK156) and in cGGG10, but was not included in the final transcript for a number of reasons. First, the cDNAs in which this exon is found differed in sequence from the cosmid and are likely to represent PKD1-like genes (see Experimental Procedures). Second, this exon is not found in FK7 (a cDNA that was cloned by using a PKD1-specific probe; see Experimental Procedures), whose sequence is identical to the genomic sequence. Finally, when included in the full-length cDNA, this exon introduces a stop codon (743 nt downstream of exon 17) that would produce a truncated protein of 2651 amino acids. We have recently identified an ADPKD patient with a heterozygous mutation that introduces a stop codon at position 10,601 of the ORF (Schneider et al., 1994, Am. J. Hum. Genet., abstract). Other mutations that truncate the PKD1 protein downstream of this exon have also been reported by the European Polycystic Kidney Disease Consortium. Therefore, it is unlikely that transcripts that include exons 17 are predominant forms in the kidney. Further studies are needed to determine whether this exon is included in other spliced forms of the PKD1 transcript.

Sequence Analysis of the Predicted PKD1 Protein

The assembly of 46 exons yields a predicted transcript of 14.5 kb in length with 228 nt of 5' untranslated and 1019 nt of 3' untranslated sequence. This transcript differs from the *PBP* sequence (European Polycystic Kidney Disease Consortium, 1994) because of the presence of two extra cytosines at positions 12,873 and 12,874 of the ORF described in this paper (position 4563 of *PBP*). The polypeptide encoded by the assembled transcript is 4304 amino acids in length, with a predicted molecular weight of 462 kDa.

The nucleotide sequence encompassing the putative Met-1 codon, CTAACG<u>ATG</u>C, is an uncommon translation start site (Kozak, 1984). Nevertheless, this methionine is chosen as the probable start site because it is preceded by an in-frame stop codon 63 bases upstream. The PKD1coding region begins with a 23 amino acid sequence with many of the properties of a signal peptide and is followed

1	NPPAAPARLA	LALGLGLWLG	ALAGGPORGC	GPCEPPCLCG	PARGAACRVN	CSGRGLRTLG	PALRI PADAT	ELDVSHNLLR	80
81	ALDVGLLANL	SALAELDISN	NKISTLEEGI	FANLENLSEI	NLSGNPFECD	CGLAWLPOWA	EECOVRVVOP	EAATCAGEGS	160
161	LRR1 LAGOPLLGIP	LLDSGCGEEY	LRR2 VACLPDNSSG	TVAAVSFSAA	HEGLLOPEAC	SAFCFSTGQG	LAALSEQGWC	LCGAAQPSSA	240
241	SFACLSLCSG	PPAPPAPTCR	GPTLLQHVFP	ASPGATLVGP	HGPLASGOLA	AFHIAAPLPV	TDTRWDFGDG	SAEVDAAGPA	320
321	ASHRYVLPGR	YHVTAVLALG	AGSALLGTDV.	OVEAAPAALE	LVCPSSVQSD	ESLOLSIONR	GGSGLEAAYS	IVALGEEPAR	400
401	AVEPLOPSOT	EIFPGNGHCY	RLVVEKAAWL.	OAOEOCOANA	GAALANVDSP	AVORFLVSRV	TRSLOWWIGE	STVOGVEVGP	480
481	APOGEAFSLE	SCONWL POEP	HPATAEHCVR	LGPTGWCNTD	LCSAPHSYVC	ELOPGGPVQD	AENLLVGAPS	GDLQGPLTPL	560
561	AQQDGLSAPH	EPVEVMVFPG	LRLSREAFLT	TAEFGTOELR	RPAQLRLOVY	RLLSTAGTPE	NGSEPESRSP	DNRTQLAPAC	640
641	MPGGRWCPGA	LDL-A	HPOALANGET	BOPGLPGAPT	ADWREPLESY	PAGPPAQISV	1000000000	PODDVODQRD	720
721	AGPGALLHCS	PAPGHPGPRA	PYLSANASSW	LPHLPAQLEG	TWGCPACALR	LLAQREQUIV CONNATATAR	MDCCSLSARE	FNUCPAL/VAT	880
991	EVENGYSERN	DTIFCULLD	WISEGERWIT	WVENSASBA	NUSLEWPARE	PICGLBATPS	PEARVLOGVL	VRYSPVVEAG	960
961	SDMVFRWTIN	DKOSLTFONV	VENVIYOSAA	VFKLSLTASN	HVSNVTVNYN	VTVERMNRMQ	GLOVSTVPAV	LSPNATLALT	1040
1041	AGVLVDSAVE	VAFLWTEGDG	EQALHOFOPP	YNESFPVPDP	SVAOVLVEHN	VTHTYAAPGE	YLLTYLASNA	FENLTOOVPV	1120
1121	SVRASLPSVA	PRD R2 VGVSDGVLVA	GRPVTFYPHP	LPSPGGVLYT	WDFGDGSPVL	TOSOPAANHT	YASRGTYHVR	LEVNNTVSGA	1200
		PKD	83						
1201	AAOADVRVEE	ELRGLSVDMS	LAVEOGAPVV.	VSAAVOTGDN	ITWIEDMODG	TVLSGPEATV.	EHVYLRAONC	TVTVGAGSPA	1280
1281	GHLARSLHVL	VEVLEVLEVE	PAACIPTOPD	ARLTAYVIGN	PAHYLEDWIF	GDGSSNTTVR	GCPTVTHNET	REGTEPLALY	1360
1361	LSSRVNRAHY	FISICVEPEV	GNVTLOPERO	FVOLGDEAWL	VACAWPPFPY	RYTWDFGTEE	AAPTEARGPE	VTFIYRDPGS	1440
1441	YLVTVTASNN	ISAANDSALV	EVOEPVLVTS	IKVNGSLGLE	LCOPYLESAV	GRGRPASYLW	DIGDGGWLEG	PEVTHAYNST	1520
1521	GDFTVRVAGW	NEVSRSEAWL	<u>NVTVK</u> RRVRG	LVVNASRTVY	PLNGSVSFST	SLEAGSDVRY	SWVLCDRCTP	IPGGPTISYT	1600
1601	FRSVGTFNII	VIAENEVGSA	ODSIEVYVLQ	LIEGLQVVGG	GRYEPTNETV	OLOAVVRDGT	NVSYSWTAWR	DRGPALAGSG	1680
1681	KGFSLTVLEA	GTYHVOLRAT	NMLGSAWADC	THDEVEPVOW	LMVAASPN <u>PA</u>	AVNTSVTLSA	ELAGGSGVVY	TWSLEEGLSW	1760
1761	ETSEPFTTHS	FPTPGLHLVT	MTAGNPLGSA	NATVEVOVOV	PVSGLSIRAS	PRD R1 EPGG <u>SFVAAG</u>	SSVPEWGOLA	TGTNVSWCWA	1840
1761 1841	ETSEPETTHS VPOGSSKRGP	FPTPGLHLVT HVTMVFPDAG	MTAGNPLGSA TESIBLNASN	<u>NATVEVDVO</u> V AVSWVSATYN	PV5GLSIRAS LTAE®PIVGL	PKD K1 EPGG <u>SFVAAG</u> VLWASSK <u>VVA</u>	SSVPENGOLA PRD #11 PGOLVHFOIL	TGTNVSWCWA LAAGSAVTFR	1840 1920
1761 1841 1921	ETSEPFTTHS. VPGGSSKRGP. LOVGGANPEV	FPTPGLHLVT HVTMVFPDAG LPGPRFSHSF	MTAGNPLGSA TFSIRLNASN PRVGORVVSV	<u>natvevdvo</u> v avswysatyn rgknhyswao	PV5GLSIRAS LTAE®PIVGL AOVRIVVLEA	PKD K1 EPGG <u>SFVAAG</u> VLWASSK <u>VVA</u> VSGLQVPNCC	0 SSVPEWGOLA PKD R11 PGOLVHFOIL PKD R EPGIATGTER	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS	1840 1920 2000
1761 1841 1921 2001	ETSEPETTHS VPOCSSKRGE LOVGGANEEV RVAYAWYFSL	FPTPGLHLVT. HVTMVFPDAG LPGPRFSHSF OKVOGDSLVI	MTAGNPLGSA TESIRLNASN PRVGDHVVSV LSGRDVTYTP	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV	PVSGLSIRAS LTAEEPIVGL AQVRIVVLEA RAFNALGSEN	PKD R1 EPGG <u>SEVAAG</u> VLWASSK <u>VVA</u> VSGLQVPNCC <u>RTLVLEVQ</u> DA	Ø SSVPFWGQLA PKD #11 PGOLVHFOIL PKD # EPGIATGTER VQYVALQSGP	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS PRD 823 CFTNRSAOFE	1840 1920 2000 2080
1761 1841 1921 2001 2081	ETSEPFTTHS. VPOUSSKRGP. LOVGGANPEV RVAYAWYFSL. AATSPSPRRV.	FPTPGLHLVT. HVTWVFPDAG LPGPRFSHSF OKVOGDSLVI AYHWDFJGS	MTAGNPLGSA TFSIRLNASN PRVGORVVSV LSGRDVTYTP PGODTDEPRA	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV, EMSYLRPGDY	PV5GLSIRAS LTAE®PIVGL AOVRIVVLEA RAFNALGSEN RVOVNASNLV	PRD R1 EPGG <u>SFVAAG</u> VLWASSK <u>VVA</u> VSGLQVPNCC <u>RTLVLEVQ</u> DA DA	0 SSVPFWGQLA PRD #11 PGOLVHFOIL PKD # EPGIATGTER VQYVALQSGP VQVVALQSGP	TGTNVSWCWA LAAGSAVTFR 13 NFTARVORGS PRD R13 CFTNRSAQFE VDVVLPLQVL	1840 1920 2000 2080 2160
1761 1841 1921 2001 2081 2161	ETSEPETTHS. VPOUSSKRGE LOVGGANPEV RVAYAWYFSL AATSPSPRRV MRRSORNYLE	FPTPGLHLVT. HVTWVFPDAG LPGPRFSHSF OKVOGDSLVI AYHVDFGDGS PX0 X14 AHVDLRDCVT	MTAGNPLGSA TFSIRLNASN PRVGDRVVSV LSGRDVTYTP PGODTDEPRA YOTEYRWEVY	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV, EMSYLRPGDY RTASCORFGR	PV5GLSIRAS LTAE®PIVGL AOVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALPGVD	PRD X1 EPGG <u>SFVAAG</u> VLWASSK <u>VVA</u> VSGLQVPNCC <u>RTLVLEVO</u> DA <u>SFEVAOATVT</u> VSRPRLVLPR	0 SSVPFWGQLA PRD #11 PGOLVHFOIL PKD # PKD # P	TGTNVSWCWA LAAGSAVTFR 13 NFTARVORGS PRD R13 CFTNRSAQFE VDVVLPLQVL VFVVSFGDTP	1840 1920 2000 2080 2160 2240
1761 1841 1921 2001 2081 2161 2241	ETSEPETTHS. VPOGSSKRGP. LOVGGANPEV RVAYAWYFSL AATSPSPRRV MRRSORNYLE LTOSIQANVT	EPTPGLHLVT. HVTWVEPDAG LPGPRESHSE OKVOSDSLVI AYHUDFGDGS PSp x14 AHVDLRDCVT VAPERLVPII	MTAGNPLGSA TFSIRLNASN PRVGQHVVSV LSGRDVTYTP PGODTDEPBA YQTEYRWEVY EGGSYRVWSD	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV EHSYLREGDY RTASCORPGR TRDLVLDGSE	PV5GLSIRAS LTAEEPIVGL AOVRIVVLEA BAFNALGSEN RVOVNASNLV PARVALFGVD SYDPNLEDGD	PRD R1 EPGG <u>SEVAAG</u> VLWASSK <u>IVA</u> VSGLQVPNCC <u>RTLVLEVO</u> DA SEPVAQATVT VSRPRLVLPR QTPLSFHWAC	0 SSVPFWGQLA PRD #11 PGOLVHFOIL PKD # EPGIATGTER VQYVALQSGP VQVLACREPE LALPVGHYCF VASTOREAGG	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS PRD 813 CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALNEGPRGS	1840 1920 2000 2080 2160 2240 2320
1761 1841 1921 2001 2081 2161 2241 2321	ETSEPFTTHS. VPOGSSKRGP. LOVGGANPEV. RVAYAWYFSL. AATSPSPRRV MRRSORNYLE LTQSIQANVT STVTIPRERL	EPTPGLHLVT. HVTMVEPDAG. LPGPRESHSE DKVOSDSLVT AVMDFGDGS PRD X14 ANUDEDVT VAPERLVPTI AAGVEYTESL	MTAGNPLGSA TFSIRLNASN PRVGDRVVSV LSGRDVTYTP PGODTDEPBA YQTEYRWEVY EGGSYRVWSD TVWKAGRKEE	NATVEVDVOV AVSWVSATYN PGKNHVSWAO VAAGLLEIOV, EMSYLRPGDY RTASCORPGR ATNOTVLIGSE ATNOTVLIRS	PVSGLSIRAS LTAE®PIVGL ACVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALFSVD GRVPIVSLEC	PRD RJ EPGG <u>SFVAAG</u> VLMASSK <u>VVA</u> VSGLQVPNCC RTLVLEV0DA SFPVAQATVT VSRPRLVLPR QTPLSFMAQ	0 SSVPFWGOLA PRD #11 PRD #11 PRD # EPGIATGTER VQVVALQSOP _VOVLACREPE LALPVGHYCF VASTOREAGO VSRSSVVVLE	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS PRD 323 CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALNEGPRGS GRCLNCSSGS	1840 1920 2000 2080 2160 2240 2320 2400
1761 1841 1921 2001 2081 2161 2241 2321 2401	ETSEPFTTHS. VPOGSSKRGP. LOVGGANPEV. RVAYAWYFSL AATSPSPRRY. MRRSORNYLE LTOSIQANVT STVTIPREVARTF	EPTPGLHLVT. HVTWVFPDAG. LPGPRESHSE OKVORDSLVT AYHNDFGDAS HVDLRDCVT VAPERLVFII AAGVEYTFSL SNKTLVLDET	MTAGNPLGSA TFSIRLNASN PRVGORVVSV LSGRDVTYTP PCODTDEPRA YQTEYRWEVY EGGSYRVWSD TVWRAGRKEE TSTUSAGMR	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV, EMSYLREGDY RTASCORFGR TRDUVLDGSE ATNOTVLIRS LVURRGURD	PVSGLSIRAS LTABEPIVGL AOVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALFGVD GRVPIVSLEC GEGVTPTLTV	PRD RJ EPGGSEVAAO VLWASSK <u>WA</u> VSGLQVPNCC RTLVLEVQDA SFPVAQATVT VSRPRLVLPR QTPLSFIMAC VSCKAQAVYE VSCKAQAVE	o <u>SSVPFWGOLA</u> <u>SSVPFWGOLA</u> <u>PRD #11</u> <u>PRD #1</u> <u>PRD #</u> <u>PRD </u>	TGTNVSHCWA LAAGSAVTFR 12 NFTARVORGS PRD 823 CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALNEGPRGS GRCLNCSSGS GRCLNCSSGS	1840 1920 2000 2080 2160 2240 2320 2400 2480
1761 1841 1921 2001 2081 2161 2241 2321 2401 2481	ETSEPFTTHS. VPGGSSKRGP. LOVGGANFEV RVAYAWYFSL. AATSPSPRRY. MRRSQRNYLE LTQSIQANVT STVTIFRERL LGAVHALTTK	EPTPGLHLWT. HVTMVFPDAG. LPGPRESHSE OKVOXDSLWI AYMNDFGDXS PXD X14 AHVDLRDCVT VAPRRLVPII AAQVEYTSSL SNKTLVLDES SNKTLVLDES	MTAGNPLGSA TPSIRLNASN PRVGQHVVSV LSGROVTYTP PGODTDEPRA YQTEYRWEVY EGGSYRVWSD TVWKAGRKEE TTSTCSAGMR AEDAGAELVY	NATVEVDVOV AVSWVSATYN FGKNHVSWAO VAAGLLEIOV, EMSYLRPGDY RTASCORFGR TRDLVLOSGE ATNOTVLIKS LVLRRGVLRD ALLLRRCRQG	PVSGLSIRAS LTAESPIVGL AQVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALFSVD SYDPNLEDOD GRVPIVSLEC GEGTTPTLTV HCEPCVVKG	PRD RJ EPGGSEVAAG VLWASSK <u>WA</u> VSGLQVPNCC <u>RTLVLEVQ</u> DA SFFVAQATVT VSRPRLVLPR QTPLSFMAC VSCKQQVYE LGRSGEEBCC SLSSYGAVLP	SSYPEWGOLA PRO #11 PROLYHEOIL PRD # EPGIATGTER VQYVALQSGP _VQVLACREPE LALPVGHYCF VASTQREAGG VSRSSYVILE ASIRLSPARP PGFPRHEVG	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS PXD A1J CFTNRSAOFE VDVVLPLQVL VDVVLPLQVL VFVVSFGDTP CALMFGPRGS GRCLNCSSGS PLOGSCRLFP PLOGSCRLFP	1840 1920 2000 2160 2240 2320 2400 2480 2560
1761 1841 1921 2001 2081 2161 2241 2321 2401 2481 2561	ETSEPFTTHS. YPOGSSKRGP. LOVGSANEEV. RVAYAWYFSL AATSPSPRRY. MRRSORNYLE LTOSIGANVT. KRGRWAARTP KAVVALNRSL	EPTPGLHLWT. HVTMVFPDAG. LFGPRFSHSE OKVO3DSLWI AYHWDFGCSS PXD X14 AHVDDRCVT VAPERLVPTI AAGVEYTFSL SNKTLVDET AJTEPEPKOS	MTAGNPLGSA TPSIRLNASN PRVGORVVSY. LSGRDVTYTP PCODTDEPRA YQTEYRWEVY EGGSYRVWSD UVWRAGRKEE TTSTUSAGRR AEDAGAPLYY ATGLIVWLHG	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV, EHSYLRPGDY, RTASCQRFGR ATNQTVLIRS LVLRRGVLRD LLLURRCNQG LLASVLPGLL	PVSGLSIRAS LTAERPIVGL ACVRJVVLFA RAFNALGSEN RVOVNASNLV PARVALFSVD GRVPIVSLEC GEGTYFTLTV HCEBFCVYKG ROADOOHVIE	PRD KJ EPGGSEVAAG VLWASSK <u>WA</u> VSGLQVPNCC <u>RTLVLEVO</u> DA SFEVAQATVT VSRPRLVLPR USRQAUPE LGRGGEEGC VSCKAQAVPE SLSSYGAVLP	SSYPFWGOLA SSYPFWGOLA PGOLVHFOIL PGD X11 PGD X10 PGD	TGTNVSWCWA LAAGSAVTFR 12 NFTARVORGS PRD 823 CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP GRCLNCSSGS GRCLNCSSGS GRCLNCSSGS ELGGSCRJFP ELGGSCRJFP ELGGSCRJFP	1840 1920 2000 2080 2160 2240 2320 2400 2480 2560 2560
1761 1841 1921 2001 2081 2161 2241 2321 2401 2481 2561 2641	ETSEPFTTHS. VPGGSSKRGP. LOVGGANEEV. RVAYAWYFSL AATSPSPRRV. MRRSORNYLE LTOSIOANVT STVTIPRERL KRGRWAARTP LGAVHALITK AAVVALISE QIRKNITECL OF UK AFCL	EPTPGLHLVT. HVTMVFPDAG LPGPRFSHSE OKVO3DSLVI AYKMDFGDSS PXD X14 ANVDLRDCVT VAPERLVFII AAGVEYTFSL SIKTLVLDET VHECTGMAD AITLPEPMOS VELEWHTVDD	MTAGNPLGSA TFSIRLNASN PRVGDHVVSV LSGRDVTYTP PGODTDEPBA YQTEVRWEVY EGGSYRVWSD TVWKAGRKED TVWKAGRKED TSTUSAGMR AEDAGAPLVV ATGITVWLMG	NATVEVDVOV AVSKVSATYN RGKNHVSWAO VAAGLLEIOV, EMSVLREGDY RTASCOREGR TRDLVLDGSE ATNQTVLIRS LVLRRGVLRD ALLLRRCRQG LTASVLPOLL UMSRGVLRD ALLLRRCRQG OMSTETAL	PVSGLSIRAS LITAESPIVGL AQVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALFGVD GRVPIVSLEC GEGVTPTLTV HCEBFCVTKG ROADPQHVIE RSCLKQTLKK	PRD KI EPGGEVAAG VLWASSKUVA VSGLQVPNCC RTLVLEVQDA SFEVAQATVT VSCRAQAVYE VSCKAQAVYE USCKAQAVYE SLALVTVLN LEAMMLILQA	SSYPEWGOLA PRD #11 PRO #11 PRD # EPGIATGTER VQYVALQSGP VOVLACREPE LALPVGHYCF VASTOREAGO VSRSSVYLE ASIRLSFWRP PGFRPHFEVG EYERALDVAA ETTAGTVTPT	TGTNVSHCWA LAAGSAVTFR 12 NFTARVORGS PMD 813 VDVVLPLQVL VFVVSFGDTP CALNEGPRGS GRCLNCSSGS GRCLNCSSGS GRCLNCSSGS LAVVVQDQLG EPKHERORRA AIGOSILAIT	1840 1920 2000 2080 2160 2240 2320 2400 2480 2560 2560 2720
1761 1841 1921 2001 2081 2161 2241 2401 2401 2481 2561 2564 2641 2721	ETSEPFTTHS. VPGGSSKRGP. LOVGGANFEV. RVAYAWYFSL. AATSPSPRRY. MRRSORNLE LTOSIGANVT STVTIPRERL KRGRWAARTF LGAVHALTFK AAVVALNESL GIRKNITETL GDLIHLASSD GIRKNITETL	EPTPGLHLVT. HVTWVEPDAG LPGPRESHSE OKVOGDSLVI AYHNFGCXSS PKD XI4 AHVDLRDCVT VAPERLVPT AAGVEYTFSL SNKTLVLDET VHEECTGMAD AITLPEPNOS SULEVHTVDO VRAPOPSELG APCOL ANI S	MTAGNPLGSA TFSIRLNASN PRVGORVVSY. LSGRDVTYTP PGODTDEPRA VOTERWEVY EGGSYRVSD TVWKAGRKEE TTSTUGAGMR AEDAGAPLOY ATGLTVWLAG QQIAALAQ DESPERWAS DIMOLIELND	NATVEVDVOV AVSWVSATYN RGKNHVSWAO VAAGLLEIOV EMSVLRCGDY RTASVORFGR ATNOTVLIRS LVLRRGVLRD LLLRRCRQG LTASVLCOL CMGPSRELVC QAYNLTSALM	PVSGLSIRAS LTAESPIVGL AOVRIVVLEA RAFNALGSEN RVGVNASNLV PARVALEOGD GRVPIVSLEC GEGVFVTJLSLEC GEGVFTJLTV HCEBFCVYKG ROADPOHVIE RSCLKOTLEK RILKRSRVLA	PRD KI EPGGSEVAAG VLWASSKWA VSGLQVPNCC RTLWLEVQDA SFFVAQATVT VSCRQAVYE LGRGCEEGC VSCRQAVYE SLSSYGAVLP YSLAUTVILN EEAMLIQA EEAMLIQA	SSVEPENGOLA PKD #11 PGOLVHFOIL PKD # EPGIATGTER VQYVALQSGP VQVLACREPE VASTOREAGO VSRSSVVLE ASIRLSPARP GEFRPHFEVG EYERALDVAA ETTAGTVTPT IVAGCKRSDP	TGINVSWCWA LAAGSAVTFR 13 NFTARVORGE PATP RJJ CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALMFORGSCRJFP CALMFORGSCRJFP LGGSCRJFP LGGSCRJFP LGGSCRJFP ALGOS ILMIT RSLLCYGGAP	1840 1920 2000 2160 2240 2320 2400 2480 2560 2560 2560 2560 2560 2560
1761 1841 1921 2001 2081 2161 2321 2401 2481 2561 2561 2561 2561 2721 2801 2881	ETSEPFTTHS. VPOGSSKRGF. LOVGGANEEV. RVAYAWYFSL. AATSPSPRRY. MRRSORNYLE LTQSIQANVT STVTIPERRL KRGRWAARTF LGAVHALTTK AAVVALNSL QILHLASSD GPGCHFSIPE EMAAGEHESS	EPTPGLHLVT. HVTMVEPDAG. LPGPRESHSE OKVOSDSLVI AVHUDRGXIS PRD X14 AHVDIRDCVT VASPELVETI VHEBCTGHHD ATTLPEPNGS VSLEWHTVDD VRAPOPSELG APSGALANLS ANSANSVVVO	MTAGNPLGSA TPSIRLNASN PRVGORVVSV PGODTDEPRA YQTEYRWEVY EGGSYRWSD TVWRAGRKEE EGGSYRWSD TVWRAGRKEE AEDAGAPLVY ATGLTVWLHG IQQIAALAQ AESPSRWVAS DVVQLIFLVD DVASGAVT	NATVEVDVOV AVSWVSATYN RGKNHVSHAO VAAGLLEIOV. EHSYLREGDY RTASCORFGR TRDLVLDSGE LVLRRGVLRD ALLLRRCRGG LTASVLFGLL CMGESRELVC QAYNLTSALM SNFFPFQYIS LDSSNPAAGL	PVSGLSIRAS LITAESPIVGL AOVRIVVLEA RAFNALGSEN RVOVNASNLV PARVALFGVD SYDFINLEGOD GRVFIVSLEC GEGTFFLIV HCEEFCVYKG RQADFQHVIE RSCLKQTLKK RILMRSRVLN NYTVSTKVAS	PRD XI EPGGSPVAAG VLWASSK <u>WA</u> VSGLQVPNCC <u>RTLVLEVQ</u> DA SFPVAQATVT VSCRQAVYE LGRGCEEGC SLSSYGAVLP YSLALVTVLN LEAMLILQA EEPLTLAGEE HAFQTQAGAQ GHVLSEFEFE	SSYPFWGQLA PKD R11 PKD R11 PKD R EPGIATGTER VQYVALQSGP VQVALQSGP VASTOREAGG VASTOREAGG VASTOREAGG VASTOREAGG VASTOREAGG PGFRPHPEVG P	TGINVSWCWA LAAGSAVTFR ij NFTRVORGS PRD RIJ CFINRSAOFE VDVVLPLQVL VFVVSFGDTP CALNFGPRGS RUGSCRLFP LAVVVQDQLG EPKHERONRA AIGOS LUIT RSLLCYGGAP AITVKVPNNS RSLLCYGGAP	1840 1920 2000 2160 2240 2480 2480 2560 2560 2560 2560 2560 2800 2800 2800
1761 1941 1921 2001 2081 2161 2241 2321 2401 2561 25641 2721 2801 2801 2881 2861	ETSEPFTTHS. VPOGSSKRGF. LOVGGANFEV. RVAYAWYFSL. AATSPSPRRV. MRRSORNLE LTOSIOANVT STVTIFRERL LGAVHALTKK AAVVALNRSL GIRKNITETL GDLIHLASSD UIRKNITETL GDLIHLASSD WAARGHRSS RUBPFSLOG	EPTPGLHLVT. HVTWVFPDAG. LPGPRESHSE OKVOODSLVI. AYHWFFGOSS SHKTLVLDES SHKTLVLDES SHKTLVLDES VHEGCTOWHD AITLPEPMOS PSGRLANLS ANSANSVVVQ DMBSVTFFI ADMBSVTFFI	MTAGNPLGSA TPSIRLNASN PRVGDKVVSV. LSGRDVTYTP PCODTDEPEA. YQTEVRWEVY CGGSYRWSD TVWRAGRKEB AEDAGADUY ATGLTVWILMG AESPSADAUS AESPSADAUS PVVQLIFLVD PQASVGAVVT	NATVEVDVOV AVSWVSATYN RGKNHVSHAO VAAGLLEIOV EMSVLRPGDY RTASCORFGR ATNOTVLTRS LVLRRGVLRD ALLLRRCRQG LTASVLFOLL LVLRRGVLRD ALLLRRCRQG WULRSREP SNFFFFGVIS LDSSNFAAGL	PVSGLSIRAS LITASEPIVGL AOVRIVVLEA RAFNALGSEN RVDTVIASNLV PARVALFSVD GRVPIVSLGC GEGTTFLTV HCEBPCVVKG ROADPOHVIE RSCLRQTLHK RILMRSRVLM NTVSTKVAS HLQLNYTLLD	PRD K1 EPGGSEVAAG VLWASSK <u>WA</u> VSGLQVPNCC <u>RTLVLEVQDA</u> <u>SFFVAQATVT</u> VSRPRLVLPR UGRSGEBEGC VSCKAQAVPE LGRSGEBEGC SLSSYGAVLP YSLAUTVILN EEPUTLAGEE HAFQTQAGAQ GHYLSEEPEP VTSLC0VPSE	SSVEPENGOLA SSVEPENGOLA PRD #11 PGOLVHEOIL SCALAGE PGIATGTER VQYVALQSGP VQVALQSGP VQVALQSGP VASTQREAGO VSRSSVVLA SIELSENER PGFRPHFEVG EYERALDVAA ETTAGTVTPT IVAGGKRSDP LIVENLASSER VLAVVLHSEP PDMUMPTEGL	TGTNVSWCWA LAAGSAVTFR JJ NFTARVORGS PRD BJJ CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALNFGPRGS GRCLNCSRGS PLOGSCRLSP ELOSSCRLSP AIGOS LIMIT RSLLCYGANS AIGUS LIMIT RSLLCYGANS AIGUS LIMIT	1840 1920 2080 2160 2240 2320 2400 2480 2560 2640 2720 2800 2880 2960
1761 1941 1921 2001 2081 2161 2241 2321 2401 2561 2561 2561 2721 2801 2821 2901 3041	ETSEPFTTHS. VPOGSSKRGP. LOVGGANEEV. AATSPSPRY. HRRSORNILE LTOSIGANVILE STVTIPRTS. KRGRWAARTF LGAVHALTTK AAVVALNRSL QILHLASSD GPGCHFSIPS GDLHLASSD GPGCHFSIPS AVCLTRHITA	EPTPGIHLVT. HVTWVEPDAG. LPGPRESSE OKVOGDSLVI AYHNDFGCSS PX0 AI4 AHVDIRDCVT VAPERLVPII AAGVITYBSL SIKTLVLDST WHECTGMHD AITLPEPKOS VSLEWHTVDD VRAPOPSELG APSGALANLS ANSANSVVVQ ADHRRYTFFI	MTAGNPLGSA TFSIRLNASN PRVGDRVVSV LSGRDVTYTP PCODTDEPFA YOTEYRWEVY EGGSYRVWSD TVWRAGRKED TVWRAGRKED TSTUSAGMR AEDAGAPLVY AGLTYWLHG IQQIAALLAQ AESPSRWVAS DVVQLIFLVD PQASVGAVVT SPGSRDPAGS	NATVENDUOV AVSWVSATYN FGKNHVSHAO VAAGILEIOV, EMSYLREGDY, FTASOORFOR TRDIVLDSSE ATNQTVLIRS ATNQTVLIRS ALLURCHQG LIXSVFAGLL UNIRGVLRD ALLURCHQG QAYNLTSALM SNFFPFGYIS SNFFPFGYIS LDSSNFAAGL UDSSNFAAGL MUNSHFR ADVNYIVMLT	PVSGLSIRAS LITAEEPIVGL AQVRIVVLEA RAFNALGSEN RVOVNASNLV FARVALFGVD SYDPNLEDGD GRVPIVSLEG GGVTFTLTV HCEEFCVYKG ROADPOHVIE RSCLKQTLMK NYTVSTKVAS WSALQVSVGL	PRD K1 EPGGSPVAAG EPGGSPVAAG VSGLQVPNCC RTLVLEVQDA SEPVAQATVT VSRPRLVLER QTPLSFHWAC VSCRNQAVYE USCRNQAVYE SLALVTVLN LEAWLILQA EEPLTLAGEE HAFQTQAGAQ GHYLSEEPEP YTSLCQYFSE MAAILHKLDO	SUPPRICAL SEVERACIA PRD #11 PRD #15 PRD #1 PRD #1 P	TSTNVSWCWA LAAGSAVTFR 1 MFTARVORGS VDVVLPLQVL VDVVLPLQVL VFVVSFGDTP CALNFGPRGS GRCLNCSGGS GRCLNCSGGS PLGGSCRLPF LAVVVQQDLG EPKHERQNRA AIGOS LIMIT RSLLCYGGAP AITVKVPNNS RSLLCYGGAP AITVKVPNNS PCGGORGREY	1840 1920 2000 2160 2240 2320 2400 2480 2560 2640 2720 2800 2800 2960 3040 3120
1761 1941 1921 2001 2081 2401 2401 2401 2481 2721 2721 2801 2721 2801 2721 3041 3041	ETSEPETTHS. VPOGSSKRGE- LOVGGANEEV. RVAYAWYFSL AATSPSPRRV. MRRSORNYLE LTGSIGANVT LTGSIGANVT LTGSIGANVT LGAVHALTKK AAVVALNRSL GLIHLASSD URANITETL GDLIHLASSD WAARGHRSS RTIBPESLOG AVCLTRHLTA	EPTPGLHLVT. HVTMVFPDAG. LPGPRFSHSF. OKVOGDSLVI AYHNFCGOS. PKD X14 ANVDLRDCVT VAPBRLVPII VASEVITSL SNKTLVLDET VHFECTGMHD AITLPEPNGS VREAVFTVD VREAVPSLLANLS ANSANSVVVQ GSGTTAHVGI GSGTTAHVGI	MTAGNPLGSA TFSIRLNASN PRVGQRVVSY LSGRDWTYTF PCODTDERA VQTEYRWEVY EGGSYRWSD KEDAGARKEE TTSTGSAGME AEDAGAPLVY ATGLTWILMG QQIAALAQ AESPSRVASS VVVQLIFLVD PQASVGAVT LQQTAALAQ AESPSRVASS HVRFVFPEPT SPGSRDPAGS	NATVEVUVOV AVSWVSATYN RGKNHVSHAO VAAGILEIOV. EMSVLRGDY. RTASCQRFGR TRDLVLDGSE LVLRRGVLRD LVLRRGVLRD LULRRGVLGL CMGFSRELVC QAYNLTSALM SNFFFFGVIS LDSSNFAAGL LDSSNFAAGL HRHLDGDRAF	PVSGLSIRAS LTAEEPIVGL ACVRJVVLEA RAFNALGSEN RVGVNASSLV FARVALFGVD SYDPNLECGD GRVFIVSLEC GEGTFFITV HCEBFCVYKG ROADPOHVIE RSCLKQTLKK RILMKSRVLM NYTVSTKVAS HLQLNYTLLD CAVCLVTTW WSALQVSVGL CAVCLVTHW	PRO RI EVGSEYLAGO VLWASSKVIVA VSGLQVPNCC RTLVLEVODA RTLVLEVODA RTLVLEVODA VSGLQVPEL LGRSGEEEGC SLSSVGAVL SLSSVGAVL REAMELICA LEAMELICA LEAMELICA RESPERTANCE REAMELICA RESPERTANCE REAMELICA REAMELI	SSVPFWGOLA SSVPFWGOLA FRD #11 FRD #1 FRD #1 FRD #1 FRD #1 FRD #1 FRD #1 FRD #1 FRD #1 VQYVALQSOP VQYVALQSOP VQVVACREPE LALPVGHYCF VASTOPLAG SSVFVVLE SSVFVV SSVFVV SSVFVV SSVFVVVLE SSVFVVVLE SSVFVVVLE SSVFVVV	TGINVSWCWA LAAGSAVTFR 12 NFTRVORGS PRD 813 CFTNRSAOFE VDVVLPLQVL VFVVSFGDTP CALMFGPRGS RCS CACKSSGS PLGGSCRLFP CALMFGPRGS RCS LAVVVQDQLG EPKHBRQHRA AIGOS ILMIT RSLLCYGASP AIGOS ILMIT RSLLCYGASP CGGQGRFKY LSPAMFLGHV	1840 1920 2000 2160 2240 2320 2400 2480 2560 2560 2560 2560 2560 2880 2960 3040 3120
1761 1941 1921 2001 2081 2161 2321 2401 2401 2561 2641 2721 2801 2801 2801 29041 3041 3041 3041	ETSEPETTHS VPOGSSKRGE LOVGGANEEV RVAYAWYFSL AATSESPERKY MRRSGRNLLE LGAVAMYTSL UTOSIOANVT STVTIPRERL LGAVHALTKK KRGRWAARTF LGAVHALTKK GDLIHLASSD GPGCHFSIPE GDLIHLASSD GPGCHFSIPE UWAARGHRSS RRIHPESLOG AVCLTRHLTA EILVKTGWGR	EPTRGLALUT. HYTWYFPDAG. LPGPRFSHSF. OKYOMDSLWI AYMNPTCDS: SWRTUVLDFY VAPERLUPII SWRTUVLDFY VHESCTONHO SHATUPENOS SWRTUVLDFY VHESCTONHO ANGANYTFI ASGANEVYPS GGGTTAHVGI ADGREYTFI GGGTTAHVGI	MTAGNFLGSA TFSIRLNASM ERVGGNVVSV. LSGRDVTYFP EGGSTDEFRA YQTEYRHEYY EGGSYRWSD TGSTGSAGRR AEDMAPLUY AFGLTYWLING IQTIAALAQ AESPSNW35 DVVGLIFLWD FQASGAUV SCOSDDAGS DVVGLIFLWD FQASGAUV SCOSDDAGS	NATVENDADY AVSWYSATYN REKNHYSKAO YAAGLLEIOV, EMSYLREGY RTASCOREG ANDYLDGE CMC93REI/C GWRUF5AL SUPEROVIS SUPEROVIS UNENGVER ALLURKEROF SUPEROVIS UNENFRAD SUPEROVIS UNENFRAD SUPEROVIS	PYSGLSIRAS LITABEPIVGL AGVIRIVYLEA BAFNALGSEN. RAFNALGSEN. FNAVALGSEN. FNAVALGSEN. GEOTPIVLE/G GEOTPI/L/Y HCEEPCV/KG GEOTPI/L/Y HCEEPCV/KG RSCLAC/TLKK	PRO RI EPGGEYJAAG VLMASSKVIA VSGLQVPNCC ETLVLEVQDA SEEVAOATVT VSGLQVPNCV VSGRPLUER VSG	D SSVPFNGOLA PROJENTOL PROJENTOL PROJENTOL PROJECT VOJVACLOSOF VASTOPRENGO VASTOPRENCO VAS	TUTNUSWCHA LAAGSAVTER J INFLANVORGE NJ CFTNRSAOFE VDWULPLQVL VDWULPLQVL VDWULPLQVL VDWULPLQVL VDWULPLQVL VDWULPLQVL VDWULPLQVL VDWULPLQVL RAUNGPRGS RGLOCSGGR PLOSCELLAT RAUNGPRGS RGLOCSGGR LAVVQDDLG EVENSROHA ALGUSTLAT RENENCSSA	1840 1920 2000 2160 2320 2400 2480 2560 2640 2720 2800 2800 3040 3120 3280
1761 1941 1921 2001 2041 2041 2041 2041 2051 2051 2051 2051 2051 2051 2051 205	ETSEPETTHS. VPOOSSKRGE. LOVGENPEY. RVAYAWYESL AATSESERREY. MRRSORNYLE LTOSIGANT STYTIPERL KORWAARTY AAVALINSL GOLHFLASSD GOLHFSIED DWAARGURGS RITHESLO DWAARGURGS AUCHLASSD GOLHFSIES LUXKTWER LIXKTWER IVEDLOTARS	ЕРТРСИЦИТ: НУТИЧЕРОВО LPGPRFENSF ОКУОПОВЦИТ АУИМЛЕТСКА РЯБ 214 АУИОЛЯСИСТ ЗИКТИЧЛЯ ЗИКТИЧЛЯ АТТРЕРИЗ ЗИКТИЧЛЯ ИНЕСТОНИВ ИНЕСТОНИВ АТТРЕРИЗ ЗИКТИЧЛЯ АЛТРЕРИЗ ЗИКТИЧЛЯ АЛТРЕРИЗ ЗИКТИЧЛЯ НОВЕТСТОНИ ИНЕСТОНИВ АЛТРЕРИЗ ЗИКТИЧЛЯ СВОТОТИТИТИТИТИТИТИТИТИТИТИТИТИТИТИТИТИТИ	MTAGNPLGSA TFSIRLNASH. FRYGDHVYSY. LSGRUYTYPP PGODTIBEPBA YQTEYRWEVY EGGSYRWBD TISTCSAGRR AEDAGABLUY. ATGLTWHIAG IQQIAALAG AESPSRWAS DVVQLIFLWJ QASVGAVIT SQGSRDPAG VUTPENGLY WLGYDFSRG	NATVENUVOV AVSWSATYN RGRNHYSHAO VAAGLEIGY ENSYLREGDY RTASCOREGR TRDJULDGE LULRROGU LULRROGU ANGYDLINS LULRROGU QAYLINSALM SURPFPOYIS LUSSNPAAGL UNSURPAAGL SURPFPOYIS HINLDGDRAF	PVSGLSTRAS LTABEPTVGL AQURJVVLEA BAFNALGSEN RVOTNASNUL PARVALPGVD SYDFNLEDGD GGOTPTVSLCG GGOTPTLPL/V HCEBCCVKG ROADGAUTS ROADG	PRO X1 EFGGEFYAGD VUMASSKVIA VSGLQVPNCC RETVILEVOR SFPLADATYT VSRPLULPR GTFLSFPMAC VSCKQAVYE VSCKQAVYE VSCAUVEL VSCKQAVYE VSLAUVEL MAGTQAGQ GYULSEEPEP MAAILARED ATHELGSW AEUGREFEK VVYEVULA	D SSVPFMGULA PRD #1 PRD #1 PRD # PRD	TUTNVSWCMA LAAGSAVTER J MTDARVORGS PRD NJJ CETNRSAOFF RD NJJ CETNRSAOFF RGORCHPROS RCLNCSGO R	1840 1920 2000 2160 2160 2320 2400 2480 2560 2640 2720 2880 2880 2960 3120 3120 3200 3230
1761 1841 1921 2001 2081 2241 2321 2481 2561 2561 2561 2561 2961 3291 3291 3291 3291 3281 3281 3281	ETSEPETTHS. VPGGSSKRGE- LOVGGANEEV. RVAYAMYESL AATSESEREV. RVAYAMYESL AATSESEREV. VTGSTQANT UTGSTQANT UTGSTQANT UTGSTQANT UTGSTQANT CLIENLASS GCGCHESIES GCGCHESIES GCGCHESIES CULHICASS GCGCHESIES AVCLIRILA VCLIRILA SVCLIRI SVCLI	EPERGLAUXT HYTRWFEDAG LPGPRESHEF OKVORDSLVI AVBRELVEI AVDRELVEI AV	MTAGNELGSA. TFSIRLNASH. ERUGDHVUSY. LSGRUVTYTF FGODTDEFRA VOTENRETY. GGSYRWSD. TVWRAGNERE GGSYRWSD. XORALARADA AEDMARADAY. AEDMARADAY. AEDMARADAY. NGTUTVILIG. SCOSEDDAGS. HVRVTPEPT. KLYGVDSRG	NATVENUVGV AVSWYSATYN FRIKNYVSHAO VAAGLIEIOV, EHSYLREGDY RTASCORFGR TRDIVLORS ANNOTVLIRS JULRRGVLGA ALLURKCROG JULRRGVLGA ALLURKCROG JULRSUSHEN SNPFPEYIS JUSSIPRAGL YNLNISHER JOYNY IMIL JUSSIPRAGL HINGLISFLEY HRILDORAF ERVLANSD	PV9GLSIRAS LTABEPIVGL AOVRIVVLEA BAFNALGSEN RVOVDASNLV PASVALGVD STOPRLEDG GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFPIVSLEC GRVFVIVSLEC GRVFVIVSLEC GRVFVIVSLEC AUXOUNTS NULDERSLUDIENI AUXOUNTS	PRO X1 EFGSETVAGA VSGLQVPNCC RTLVLEVQDA SFFVAQATVT VSGLQVPNCC RTLVLEVQDA SFFVAQATVT VSGRQUEVQ VSGKQQVFS USGRQUEVQ VSGLQVVSL RAGVEVG VSGLQVSFS MAJLIKLOSV ATPHELOSV ATPHELOSV	0 SVYERMOLA PRD #1 PRD #1 PRD # PQUARTER PQUARTER VQYVALOSGP VQVALOSGP VQVALOSGP VQSVSYVIE ASTRESMOR VSRSYVIE ASTRESMOR PGPRIPREVQ VSRSYVIE STRATVTPF IPIERASKR VLAVVLISP EDMVMTEGL LOASIGNAT LIVACKNSP EDMVMTEGL LOASIGNAT PLISSPSIU	TSTNVSKYMA LAAGSAVTFS J NFTARVORGS PRF MJ CTTNRSOFF VDVVLPLQVL VFVVSFGDTF CALMFQFRG GRCLNCSGS GRCLNCSGS GRCLNCSGS FUCOSCRLPF LAVVVQDUG EVKHSRCHMA LAVVVQDUG EVKHSRCHMA SILCVGQUG NESHLSTS LSLCVGQUG LSPAFFDAV FSGFTFIG SGNRGUAR SGNRGUAR SGNRGUAR	1840 1920 2000 2080 2160 2320 2480 2480 2480 2550 2550 2580 2950 3040 3120 3280 3280 3280 3340 3280 3360 3440
1761 1941 1921 2001 2081 2241 2321 2401 2561 2561 2641 2721 2901 2801 3041 3121 3201 3121 3281 3261 3121 3281 3261 3261 3261 3261 3261 3261 3261 326	ETSEPETTIS. VPOGSSKRGE. LOVGCINPEV. RVAYAWYESI. ATTERSORNYLE JOSIOANYL SVVIIPREN. KRSORNYLE JOSIOANYL GUINNITEL COLINE COCHFIELS DWAAGGIRSS RITHPELO WAAGGIRSS RITHPELO AVGLUTASS OCCULIC CONTAINER VECTOR	EPERGLAUXT HYTMYEPDAG LPGPREPHER ORVOGDSLUT MYMDFGDGS PZD X4 AVVDLRCVUDET VARBELVPLI SHKTUVLDET VARBELVPLI SHKTUVLDET VARDFGDGSLUT ANGENTAVG VAROPESLG ANGENTAVG VAROPESLG ANGENTAVG VAROPESLG ANGENTAVG CONTA	MTAGNFLGSA TFSIRLNASH ERVGORVYEY EGOTTERFRA FGOTTERFRA GOTTERFRA GOTTERFRA GOTTAL GOTTAL ADDI TWO ADDI TWO ADDI TWO ADDI ADDI ADDI ADDI ADDI ADDI ADDI ADD	NETVENTUNITY AVSWYSATYN RGKNIYYSHAO VAAGLEIOV LENSYLREGDY RTASCORFGR TRDIVLOGBE JVLRRGUN LUGSE ATWOTYLLRG LUGSENAACL VANDARAACL DOSINFAACL SMPFPGYIS LUGSENAACL DOSINFAACL JVLNISENFE EKEVLAASDA AFWOMKSDL LUGVIAEG	PVSGLSTRAS LTABEPTVGL AQVRJYVLEA RAPNALGSEN RVDVNASNLY PARVALPGVD SYDFNLEDD GGCTFFL/TV HCEBCVTVS GGCTFFL/TV HCEBCVTVS ROADOHVIE RSCLW7TLKD RSCLW7TLKD RSCLW7TLKD RSLW7	PRO RJ PEGSEYJAGO VULNASSKVIA VSGLQVPNCO SPEVAOATVT VSRPLVLPR OTPLSFMMC VSGLQVPNC LGRSGEEEGC SLSSYGAVU VSLAUTVLN ESPUTLAGEE MAAILHELDO AFTHELGSW AELGRGFDK WFSGGRFDK WFSGGRFDK	SSVPFNGULA FXF #11 FXGLMHOLT FXGLMHOLT VAD X PAD X PAD X PAD X PAD X PAD X PAD X PAD X VASTOREAGE VASTOREAG	TUTNVSWCMA LAAGSAVTER J MTTARVORGS PRD NJJ CLTNESAOET VDVVLPLQVL VFVVSPGDTP CALINGFRGS RGCLNCSSGC RGCSCLLP RGCLNCSGG RGCLNCSGG RC RCLNCSG RCLNCSGG RC RCLNCSGG RCLNCS	1840 1920 2000 2080 2160 2320 2490 2560 2560 2580 2800 3280 3040 3120 3280 3360 33200 33200 3360 3360
1761 1941 1921 2001 2001 22401 2321 2401 2561 2721 2901 3041 3041 3041 3281 3291 3281 3241 3241 3241 3241 3261 3241 3261 3261 3261 3261 3261 3261 3261 326	ETSEPETTHS. VEGGSSKRGE- LOVGGSNEEV. RVAYAMYESL AATSESERRY. NRRSGRNYLE LGAVHALTYK AVVALNSE GULHILASS GULHILASS GULHILASS CULHILASS CULHICASS AVCLTRILT HVELQTARS ACCULICL OMBIGLIEFE ACCULICL OMBIGLIEFE COMBIGLIEF	EPERGLAUXT HYTAVFEDAG LPGPRESSE OKVOGDSLVI AVMOTEDAS HYD 214 AVVDLRCVT VABERLAVI AGVEYTFSL SKNTLVLDES KNTLVLDES VSLRVHTVDO VSLRVHTVDO VSLRVHTVDO VSLRVHTVDO VSLRVHTVDO VSLRVHTVDO SGRTARVGT ASSANSWVO ADHREYTFI FGASLFVPS GRGTARVGT GRGTARVGT GRGTARVGT GRGTARVGT GRGTARVGT GRGTARVGT GRGTARVGT	MTAGNPLGSA. TFSIRLNASH. ERUGDRVVSV. LSGRUVTYTP PGODTDEFRA VOTEVRMETY. GGSYRMSD TVMRAGNERE TVMRAGNERE TVMRAGNERE TVMRAGNERE AEDMADASLAVY. ATGLTVPLIME SCASBODAGS DVVGLIFLVD PASUGAVY. SCASBODAGS DVVGLIFLVD PASUGAVY. MLTVGDFRAG UNTEVPLETE NLSVDSRAG VETENAGU,	NATVENUVOV AVSMVSATVN RGKNHVSHAC VAAGILEIOV, EHSVLREGDY, RTASCORFGR TRDIVLOSS ANNOTVLIRS ALLIARCROG ALLIARCROG ALLIARSHER ADVINTIMUL USSIPPAGI BIRLOGRAF EREVLAASSA DJONYTIMUL BIRLOGRAF HISBLISPLEV AFVOGHKSU DLIOOVLAG	PV9GLSIRAS LTABEPIVGL AOURIVULEA BAFNALGSEN RVOVRASNLV PARVALGSUN FARVALGSUN FARVALGSUN FARVALGSUN FARVALGSUN FARVALGSUN HCBEFCVWG GRUPTVSLDC G	PRO RI PEGSETAGN VLMASSKVIA VSGLQVPNCC RTLVLEVQDA SFPVAOATVT VSRPLVLPR OTPLSFWAG VSCKOAVP VSCKOAV	0 SSVPENGOLA PRD #1 PRD #1 PRD #1 PRD # VQVVALOSGP VQVALOSGP VQVALOSGP VQVALOSGP VQSTQREXQ VQSTQREXQ VQSTQREXQ VQSTQREXQ EVENTVVFF IPISTANE EVENTVFF IPISTANE EDMARTECL LOSSIGAT VLAVILISE EDMARTECL LOSSIGAT POLISDRIN SSTQCEXTEF POSVANLI	TSINVERCHA LAAGSAVTYE I INTERVORGS PRF HJ CETNISAOGE VDWVLPLQVL VFWVSFGDTF CALNFGFRG GRCLNCSGG GRCLNCSGG GRCLNCSGG GRCLNCSGG GRCLNCSGG HAUNTONIC LAWVODQLG EFWHERGINA AIGNS LINIT RSLCYGQG AITWKVPINS RSLCYGQG HAUNTONIC SGSSFILAG LAGSSFTA AITWKVPINS	1840 1920 2000 2080 2160 2400 2560 2560 2560 2880 2960 2960 2960 3120 3120 3200 3240 3120 3240 3240 3240 3240 3240 3240 3240 32
1761 1941 1921 2001 2001 2001 2001 2001 2001 200	ETSEPETTIS. VPOGSSKRGE. LOVGCINPEV. RVAYAWYESL ATTERSERRY. MRRSORNYLE JOSIOANYL SVVIIPREN. KOGVIANYE AVVALNESL GUINELASSD GOCCHFIELS MAAGGIRSS RITHPELQ MAAGGIRSS RITHPELQ AVGCULIC COLVIDE MAAGGIRSS RITHPELQ AGCULICTASSD GOCCHFIELS AGCULICTASSD GOCGRESIES MAAGGIRSS RITHPELQ AGCULICTASSD GOCGRESIES MAAGGIRSS NCCLING NCCASSD ACCULICTASSD COCHFIELS AGCULICTASSD COCHFIELS COCHFI	EPERGLAUNT HYTMYEPDAG LEGERPENSE OKYMMDELVI NYMPTCDES MAVDLRDCVT VAPERLVPI MAGVEYTFEL SKRTLVI,DE KRTLVI,DE	MTAGNFLGSA. TFSIRLNASH. ERVGORVYSY. LSGRDYTYTE FGOTDEFFA. GGSTRWEY GGSTRWEY GGSTRWEY GGSTRWEY ADDATA	NATVENTUNIY AVSMYSATYM RGKNHYSHAO YAAGILEIGY, EHSYLREGY, EHSYLREGY, ULTRAYLAG ATNOTYLINS JULRRAYLAG MONDBERLY, ULTRAYLAG SNPFPFOYJ HANLSGIRFA SNPFPFOYJ HINLSGIRFA LUSSIPPFAGL JULNISSIF EKEVLASSDA HYSLLSFELASSDA	PVSGLSIRAS LITABEPIVGL AQURIVVLEA RAFNALGSEN RUTVRAGNU, PARVALGSEN RUTVRLEY GRVTFIVSLEY GRVTFIVSLEY GRVTFIVSLEY GRVTFIVSLEY GRVTFIVSLEY RUTVSLEY RU	PRO RJ PEGOSETAGO VUMASSKUMA VSGLQVPNCC RTLVLEVODA SFEVAOATVT VSGRPLULRR OTFLSEPING VSGKAQAVE VSGKAQAVE VSGKAQAVE VSGKAQAVE VSGKAQAVE VSGKAQAVE VSGKAQAVE VSGKAGAVE VS	SSVPFNGULA FXF #11 FXGLMHOLT FXGLMHOLT VGYVALQSGP VQSQVALQSGP VQSQVALQSQVALQSGP VQSQVALQSQVA VQSQVALQSQVALQSGP VQSQVALQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQSQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQVA VQVALQSQV	TUTNVERCHA LAAGSAVTYF I INTERVORGS PRO HIJ VEVVLPLQVL VEVVERGDTP CALMEGREG RCLINESSES GRCLNESSES GRCLNESSES GRCLNESSES HAVEVORGE LAVVVQDQG CALMEGREG RCLNESSES HAVEVORGE HAVEVORGE RSHLFVGDQ FGGRGRFKY HESEFTHIG SSNAFTH	1840 1920 2000 2080 2160 2240 2400 2480 2560 2720 2800 2800 2900 3040 3120 3280 3120 3280 3120 3280 3340 3520 3560 3520
1761 1841 1921 2001 2241 2401 2481 2561 2561 2643 2721 2803 2721 2803 2903 2903 2903 2904 3041 3041 3041 3051 3051 3051 3051 3051 3051 3051 305	ETSEPETTIS VPOGSSKRGE LOVGGNREEV RVATWYFEL AATSEYEREV HRRSONVIC SYVTIPERL VGGRAARTE LOVGICHEISE VGCHEEJEE DER PESLGO VCCFELTE NVCLFELTASSD GCCHEGIES NVCLFELTASSD GCCHEGIES NVCLFELTASSD GCCHEGIES NVCLFELTASSD GCCHEGIES DER PESLGO NVCLFELTASSD GCCHEGIES DER PESLGO DER PESLGO D	EPERGLAUNT ENTIMPEDAG LPGPRESHSE OKVORDSLVI AVMORDSLVI AVMORDSLVI AVMORDSLVI AVADEVITS SHKTUNDEF VHECTOWNE SHKTUNDEF VHECTOWNE SHKTUNDEF SHKT	MTAGNFLGSA. TESIRLNAGN. ERVGONVUSV. LSGRDVTYTE PROTDEPRA. VGCEVNEVY EGGSYRWED VGCEVNEVY EGGSYRWED TUSTGSAGNR ABENGABLUY. ATGLTVRLIG IQITAALAG. AESJERRWAS DVVGLIFLW ATGLTVRLIG IQITAALAG. HIJ YGCLAGAN MICHAELAG. HIJ YGCLAGAN VGCEVNEVY AUGUSTAN HIJ YGCLAGAN VGCEVNEVY AUGUSTAN HIJ YGCLAGAN HIJ	NATVENTUNIY ANSMYSATYN FGKNHYSHAO YAAGILEIGY, ENSYLREGDY, ENSYLREGDY, ENSYLREGDY, THDIYLLOGE ATMOYULRS ALLIARCROUG ANTOYIL IS ALLIARCROUG ANTOYIL IS ALLIARCROUG ANTOYIL IS ANTOYIL IS ANTOYIC IS ANTOYIC IS ANTOYIL IS ANTOYIC IS ANTO	PVSGLSIRAS LITAESPIVGL AQURIUVLEA BAFNALGSEN ROTMAASNUU FARVALEDO STOPHLEDO GEOTPFLOW HCESECVING ROADOAUTE ROADOAUTE ROADOAUTE ROADOAUTE BALGRENUL UTVSTKUS SCLARTLEU DIVAOUVSS FLDDEKSLU VSSPAPPODT SLALERRHUN ANUFWNECH BELIAWAAN	PRO RJ PEGOSETAGO VLMASSRUMA VSGLQVPHCC RTLVLEYODA SEPUNDATU VSDBULLAR SEPUNDA VSDRULAR SERVENT VSLQVFB VSLAVVIN SLSSYGAUP SLSSYGAUP SERVENT MATCOAGG MATCAGE REPORTAGE VVVYPVLAI WFSGRULAR MATCOAGG VVVYPVLAI VSSGRULAR MATCAGE VVVYPVLAI VSSGRULAR MATCAGE VVVYPVLAI	SSVFFNGLA SSVFFNGLA FCOLMHEDIL FCOLMHEDIL FCOLMHEDIL FCOLMHEDIL FCOLMHEDIL FCOLMHEDIL VOYALOSGP VQVALOSGP	TUTNUSKUMA LAAGSAVUTES JATAVURGS JATAVURGS JATAVURGS JATAVISS CETNRSAGER VIDVULLQUI VITVUSEDLET VIDVULLQUI VITVUSEDLET RELEVISE REGESCELIPP LAINTORORGS REGESCELIPP LAINTORORGS RENESSION LESPANFLOHV LESPANFLOHV LESPANFLOHV LESPANFLOHV LESPANFLOHV LESPANFLOHV LESPANFLOHV LESPANFLOHV	1840 1920 2080 2160 2480 2480 2480 2480 2550 2550 2880 2960 3120 3120 3120 3120 3120 3120 3120 312
1761 1941 1921 2001 2081 2161 2321 2401 2561 2561 2641 3041 3041 3041 3281 3201 3281 3361 3521 3661 3761	ETSEPETTIS VPOGSSKRGE LOVGGNNEEV. ATSESSERV ATSESSERV ATSESSERV TOTI PRESL VTOTI PRESL VTO	EPTRIGUIUM ENTRAFEDAG LEGEREGHSE ONYOROBLUT NUMERICUNA NUMERICUNA SINGUA NUMERICUNA SINGUAL NUMERICUNA SINGUAL NUMERICUNA SINGUAL SINGUA NUMERICUNA SINGUAL SINGUA NUMERICUNA SINGUA	MTAONFLGSA TESIRLINAGN ERVGQHVVSV LSGRUVYSV EGGYTDEFRA SCOUTDEFRA SCOUTDEFRA SCOUTDEFRA SCOUTDEFRA SCOUTDEFRA SCOUTDEFRA SCOUTDEFRA NOTSKILL SCOUTDEFRA SC	NATVENTAVO AVSMVSATYN FGRNHVSHAD YAAGLEFIOY EHSYLREGDY HENSYLREGDY HENSYLREGDY HENSYLREGDY HENSYLREGDY HENSYLREGDY HENSSHI SWEPPERYS SWEPPEYS HENSEN	PV9GLSIRAS LITABEPIVGL AQURIVVLEA RAFNALGIEN RVUTNAGNIV PARVALDGUE SUPRILEDG GRVTFILV HCEBECVING GRVTFILV HCEBECVING GRVTFIVSLEC GRVTFIVSLEC GRVTFIVSLEC GRVTFIVSLEC GRVTFIVSLEC GRVTFIVSLEC SUPRILEV VSLOV	NO XI PEOSETIAGI VUMASSKVINA VSGLQVPNCC RTLVLEVODA SFEVAOATVT VSGRQVENCO VSGKQAVE VSGKQA	SSVEPRIGULA SEVERINGULA FEGLIAREDILL FEGLIAREDILL EPGLIAREDILL SEVERING SEVERING SEVERIT SEVER SEVERIT SEVERI	TUTNVERCHA LAAGSAVTFFF JJ NFTARVORGE JPD NJ VEVVLPLQVL VEVVEFODTP CALNEGFREG VEVVLPLQVL VFVVEFODTP CALNEGFREG RECLEVERCHA RESLEVERCHA SEAST VEVVEFNEG RESLEVERCHA RESLEVENCAS SEASTLASF SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE SEASTLASF CORREPLATE CORRE	1840 1920 2080 2080 22400 2480 2560 2560 2880 2880 2880 3120 3200 3140 3120 3280 3280 3280 3280 3280 3280 3280 32
1761 1841 1921 2001 2261 2241 2321 2401 2321 2641 2361 2841 2261 2841 3041 3121 3201 3201 3201 3201 3201 3361 3521 3661 3761 3761 3761	ETERPTTIA VPOGSSKRGE LAVGGINERV RVAVAMYESL ATTERSTRUK VVAVAMYESL VVS CONT VVS CONT V	EPERGIAILAT ENTIMPEDAG LPGENESHSE OKYOODSLAT AVMONDSLAT AVMONDSLAT AVMONDSLAT AVMONDSLAT SHKTLVLDEF VIESCTOWNO VIESCTOWNO VIESCTOWNO VIESCTOWNO VIESCTOWNO VIESCTOWNO ANDEN PERS SIGNALASPYS POARLESPYS POAR	MTAGNFLGSA. TESIRLNASH. ERVGORVVSV. LSGRDYTYTE PROTDEPRA. VGCETNRCV EGGSYRWED TUSKIGARES ABMCAPLUY. ATGLTWLIG IQITAALAG. AESJFRIVAS DVVGLIFLWLIG IQITAALAG. AESJFRIVAS DVVGLIFLWLIG IQITAALAG. AESJFRIVAS DVVGLIFLWLIG IQITAALAG. HLIPSOLAUG. PRAESFASDE PRAESFASDE PRAESFASDE VETENDGUV. VGCETRUGUV. UTESULAUG.	NATVENTUNIY ANSMYSATYN FGKNHYSHAO YAAGILEIGY, ENSYLREGDY, ENSYLREGDY, ENSYLREGDY, ENSYLREGDY, TRIJULGGE ATMOTYLINS ALLINKCROG QATNITSALM SINFFERYIS LOSINFAAGL DIJOUYLSGL DLIOUYLSGL DLIOUYLSGL SAFUATYN SAFUATYN VERVLANSD LEFPAAGRLI VERVLANSD	PVSGLSIRAS LITAESPIVGL AQURIVVLEA BAPNALGSEN ROTVAANANU, FARVALGSEN ROTVAANANU, FARVALGSEN STOPHLEDO GEOTPFLOW HEESECVING ROADOHVIE ROADOHVIE ROADOHVIE ROADOHVIE HINSLOLFHI HINSLOLFHI HINSLOLFHI HINSLOLFHI JALSPREILU, SCLORVALW, SC	NO XI PEOSETIAGO VUMASSRIVIA VSGLOVPINCC ETLIVILEYODA SEPUNDATU VSDEDLINE SERVINA VSDEDLINE SERVINA SERVINA SERVINA MATOCAMO MATOCAMO MATOCAMO MATOCAMO MATOCAMO MATOCAMO SERVINA MATOCAMO MATOCAMO SERVINA MATOCAMO MATOCAMO MATOCAMO SERVINA SERVINA MATOCAMO SERVINA SERVINA MATOCAMO SERVINA MATOCAMO SERVINA MATOCAMO SERVINA MATOCAMO SERVINA MATOCAMO SERVINA SERVINA SERVINA MATOCAMO SERVINA	SSVFFMOLA SSVFFMOLA FOLMATCH FOLMATCH FOLMATCH FOLMATCH VQYVALOSGP VQVLACSGP VQVLACSGP VQVLACSGP VQVALOSGP	TUTNUSKUMA LAAGSAVUTES J TUTNUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKUSKU	1840 1920 2000 2160 2400 2400 2400 2400 2400 2400 2400 2800 3040 3120 3280 33200 3280 3340 3360 3360 3360 3360 3360 3760 3920
1761 1941 1921 2001 2031 2241 2241 2241 2241 2241 2241 2241 22	ETSEPETTIS VPOGSSKRGE LOVGGNNEEV. RVAYWAYESL AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. AATSESEREV. COLINIASSE COCUMENTIS C	EPTROLAUNT. HYTRYFPDAG. LPSPRFGHSF OKYGORGLVT. VYMRPTCDAS. FKD 714 AVVLDRCT. VAPRHLVPII AVVDRCT. SHKTUVLDET VAPRHLVPII ANGVEYTSJ. SHKTUVLDET VAPRHLVPII ANGVEYTSJ. SHKTUVLDET VAPRHLVPII ANGVEYTSJ. DISCHART DI	MTAONFLGSA TESIRLINAGN ERVGOHVVSV LSGRUVYSV EGGSTDEFRA VGTEVREVY EGGSTMWSD VGTEVREVY EGGSTMWSD TSTVSAGNR ESDAGAPLV SCSSDDAG DVVOLFLVD FOASGAVF SCSSDDAG DVVOLFLVD FOASGAVF LVFSCREV FOASGAVF LVFSCREV ELVESCREV LVFSCREV LV	NATVENTAVA AVSAVSATYN FGRNHVSHAO YAAGLEIOV, EMSYLREGDY FRASCIREGG TRDJVLOGB ATMGYVLRS ALLINRCROM SNPFPEYIS SNPFPE S	PVSGLSIRAS LITABEPIVGL AQVRJVVLEA BAFNALGSEN RVOVNASNLJ RVOVNASNLJ RVOVNASNLJ RVOVNASNLJ RVOVNASNLJ RVOVNASNL RVOVNA	NED X1 EFGGETYAGU VLMASSKVIVA VSGLQVPNCC EFTUAUEVGDA SEFYAQATVT VSGNQVVFNC GTPLESFIMAC VSGNQVVFNC GTPLESFIMAC VSGNQVVFNC VSGNQVFNC VSGNQVVFNC VSGNQVFNC VSQ	SSVFFMGLA FROM JELL FROM J	TUTNVENCHA LANGENVTER INFLANVENCE PRO BIJ CTINESADET VDWLPLQVL VDWLPLQVL VDWLPLQVL VDWLPLQVL VDWLPLQVL VDWLPLQVL VDWLPLQVL VDWLPLQVL LEVANVORDE REGESCRLPF LEVANVORDE REGESCRLPF LEVANVORDE REGESCHAPPEN LEVANVORDE SESASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF LEVANVORDE SSSASTLAFF	1840 1920 2000 2160 2400 2400 2560 2720 2720 2800 3040 3120 3200 3440 3360 3440 3520 3680 3760 3840 403 3920 4000
1761 1941 1921 2001 2031 2241 2321 2321 2351 23641 23641 23641 23641 23641 23641 23641 23641 3121 3041 3121 3641 3561 3561 3561 3561 3561 3561 3761 3761 4031	ETERPTIME VPOGSSKRGE LAVGGINERV. RVAVAMYESL ANTERSERV MERSORNYLE LAVGINERV KRISKART UNDITECT GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS GUILITASS MULDIRESPONDIC DEPS CLAWGE DEPS CLAWGE MULDIRESPONDIC DEPS CLAWGE MULDIRESPONDIC DEPS CLAWGE MULDIRESPONDIC DEPS CLAWGE CANADA DEPS CLAWGE	ЕРТРОНИЦИТ ВИТИЧЕРДАС LIGORIECHST ОКУСОВЕЦИТ МИМОТСОВЕ РЯБ В14 МИОДЕХССКИР ИНДЕХСКИРСТ ИНДЕХСКИРСТ ВИТИЧЕР ПОЛУТИРИСТОВИ ПОЛУТИРИ ПОЛИТИРИ ПОЛУТИ ПОЛУТИРИ ПОЛУТИРИ ПОЛУТИРИ ПОЛУТИ ПОЛУТИРИ ПОЛИ П	MTAGNFIGSA TFSIRLNASM FRYGGHYVSY. LSGRUVYTF FGODDEFRA YOTSYNWSD FGODDEFRA YOTSYNWSD TOWNACHNES SASTANSD YOLLFLY SASTANSD VYGLIFLYD FSSIRLAS VYGLIFLYD FGSSIDASC VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENNGLY VYTENGLAST SASTANSD VYTENGLAST SASTANSD VYTENGLY V	NATVENDADY ANSWYSATYN RCRENTYSAG YSAGUT EIGY. EMSYLREGY RTASCOREG RHYLREGY ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ALLENKCHOG ANNUTSAL ALLENKCHOG ANNUTSAL AUGUNTS SAFEAL AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND	PVSGLSIRAS LITABEPIVGL ACURIVYLFA ACURIVYLFA BAFNALGSEN FVXVRASHL PARVALGSEN FVXVRASHL PARVALGSEN FVXVRASHL PARVALGSEN FVXVRASHL RCLAVTLAN FVXVRASHL FVXVXVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	NED XI PEOSEYADA VLMASSKVIMA VSGLQVPNCC RTLVIEWOPA SFEYADATVT VSGRPLUEWOPA SFEYADATVT VSGRPLUEWOPA LANGTUNG VSGRPLUEWOPA NATVISION VSGRVDATV NATVISION	S SEVERAGUA NET ALL SEVERAGUA COMPOSE SEVERAGUA SEVERAL SEVERA	TOTIVENCIA LANGSAVTES JJ INTLAVURGES PATO ALL CETINESADE VDWULFLQVL VDVULFLQVL VDVULFLQV	1840 1920 2000 2160 2240 2240 2560 2560 2800 2800 2800 3010 32200 32200 33200 33200 3340 3420 36600 3450 36600 34600 36600 36600 3920 4000 4000 4000
1761 1941 1921 2001 2031 2161 2321 2321 2321 23641 2721 32641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 34641 3461	ETERPETUS VPOGGSKRGE LOVGGNNERV RVAVAMYESL ATTERERNEV MRRSONVILE JTOSIONVIL JTOSIONVIL JTOSIONVIL GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO GOLINITASIO COLUMICASIO CASICICARI VILLASIO GOLINITASIO CASICICARI VILLASIO GOLINITASIO CASICICARI VILLASIO VILL	ЕРТРСИНИТ НУТИЧЕРОАС LICOPHENET ОКУОНДЕНТ ОКУОНДЕНТ РБО 214 МУОДЕВССИ РБО 214 МУОДЕВСИ РБО 214 МУОД	ITTAINFIGSA TESIFLINASI ERUGUNUSI ERUGUNUSI EGOTDEFEN EGOTDEFEN CONTRETA EGOTDEFEN CONTRETA EGOTDEFEN CONTRETA EGOTTA TVIKKORKE EGOSTRINES TVIKKORKE EGOSTRINES TVIKKORKE EGOSTRINES TVIKKORKE EGOSTRINES EGOSTRINES TVIKKORKE EGOSTRINES EGOSTRINES EGOSTRINES FOR CONTRETA EGOSTRINES EGOSTRINES EGOSTRINES EGOSTRINES TVIKKORKE EGOSTRINES EGOSTRINES EGOSTRINES TVIKKORKE EGOSTRINES EGOSTRIN EGOSTRINES EGOSTRIN EGOSTRINES EGOSTRINES E	NATVENDAGY AVSAVSATYN PROKINYSHOL VAAGILEIOV EMSYLREGY REMYLREGY ATMONYSHOL VAAGILEIOV EMSYLREGY ATMONYSHOL VILNES	PVSGLSIRAS LITASEPIVGL ACVFILVVLFA BAFNALGSEN TOVTNABALLY PARVALGSUN FORMLESCO GRVFILVSLCC GRVFILVS GR	VED X1 VUMASSKVIA VUMASSKVIA VISALQVINCC ETUTIENDA SEPNIATIV VSRPLIVLER OFSCRADING SLESVANT VSRPLIVLER OFSCRADING SLESVER MATILELSE	SSUPERATION FOR A SUPERATION FOR A SUPERATION FOR A SUPERATION FOR A SUPERATION FOR A SUPERATION SU	TOTNVENCHA LANGSAVTER JJ JJ CTETURAGOTE VOVULELQU/ VTVU SPECULT VTVU SPECULT VTVU SPECULT VTVU SPECULT VTVU SPECULT VTVU SPECULT VTVU SPECULT SPECULT VTVU SPECULT VTVU SPECULT VTVVV SPECULT VTVV SPECU	1840 1920 2000 2080 2160 2320 2400 2560 2720 2800 2720 2800 3120 3200 3120 3200 3200 3550 3560 3560 3560 3560 3560 3760 3920 4000 4000

Figure 2. PKD1 Open Reading Frame The different domains are underlined. See text.

by a predicted cleavage site (von Hejne, 1986) (Figure 2). In addition to the signal sequence, the identification of five domains that have been identified in other extracellular proteins strongly suggests the extracellular location of at least the amino-terminal half of the protein.

Immediately downstream of the signal sequence, there are two leucine-rich repeats (LRRs) (Figure 2). These LRRs (in exons 2 and 3) are flanked on both sides by cysteine-rich regions that have homology to the flanking regions of a subset of other LRRs. LRRs occur in numerous proteins (Figure 3) and have been shown to be involved in diverse forms of protein-protein interaction. The number of LRRs within the respective proteins varies between 2 and 29 (Kobe and Deisenhofer, 1994). Adhesive platelet glycoproteins form the largest group in the LRR superfamily (Kobe and Deisenhofer, 1994). The structure of the array of 15 LRRs in porcine ribonuclease inhibitor (RI) has recently been determined (Kobe and Deisenhofer, 1995); the LRRs form a horseshoe-like structure that surrounds and binds to RNase A (Kobe and Deisenhofer, 1995). It has been suggested that proteins containing only a few LRRs, such as the PKD1 protein, interact with other proteins via the LRRs in order to form the horseshoe-like superstructure for protein binding (Kobe and Deisenhofer, 1994).

Although LRRs occur in various locations in different proteins, the additional flanking cysteine-rich domains de-

PKD1 exon 2	LOVSHNLLRALDVGLLANLSALAE
PKD1 exon 3	LDISNNKISTLEEGIFANLFNLSE
ALS/Human	LHLERNQLRSLSAGTFAHTPALAS
CBP8/Human	LSLQWNMLRVLPAGLFAHTPCLVG
TrkA/Human	LVLSGNOLHCSCALRWLQRWEEEG
A2GL/Human	LOLSGNRLRKLPPGLLANFTLLRT
GP1A/Human	LOVSFNRLTSLPLGALRGLGELQE
HSGPV/Human	LFLDHNALRGIDQNMFQKLVNLVN
Garp/Human	DLSGNQLRSILASPLGFYTSLRF
OMPG/Human	VDLSNNSLTQILPGTLINLTNLTN
PGS2/Human	LGLSFNSISAVDNGSLANTPHLRE
5T4G/Human	LELASNHFLYLPRDVLAQLPSLRH
Slit/Drosl	LYLESNEIEQIHYERIRHLRSLTF
Toll/Dros	LOLSNNRLTHLPDSLFAHTTNLTI
	TOSNIC GANLLE

Figure 3. LRRs

LRRs are coded by exons 2 and 3 on the PKD1 transcript. Examples of proteins that also contain LRRs are human insulin-like growth factorbinding protein complex acid-labile chain precursor (ALS); human carboxypeptidase 83 kb chain (CBP8); human high affinity nerve growth factor receptor (trkA); leucine-rich a-2-glycoprotein (A2GL); platelet membrane glycoprotein 1B a chain precursor (GP1A); platelet glycoprotein V precursor (HSGPV); human garp gene product (garp); human oligodendrocyte-myelin glycoprotein precursor (OMPG); human bone proteoglycan, decorin (PGS2); human 5T4 oncofetal antigen (5T4G); and Drosophila slit and Toll proteins. Conserved amino acids are represented in the bottom line of the figure.

fine a subgroup of extracellular proteins (Kobe and Deisenhofer, 1994). Only a few proteins contain both the distinct amino-terminal and carboxy-terminal flanking cysteinerich domains (Figures 4 and 5). Among this group are Toll, slit, Trk, TrkB, and TrkC. This set of proteins all have intracellular domains that could relay signals to the cytoplasm. For example, the Drosophila Toll protein is required for mediating dorsoventral patterning (Hashimoto et al., 1988). The Drosophila slit protein is believed to mediate interactions between growing axons and the surrounding extracellular matrix (Rothberg et al., 1990). In vertebrates, these domains are found in the Trk family of tyrosine kinase receptors (Schneider and Schweider, 1991); the platelet glycoproteins I and V, which mediate the adhesion of platelets to sites of vascular injury (Roth, 1991); and the 5T4 oncofetal trophoblast glycoprotein, which appears to be highly expressed in metastatic tumors (Myers, 1994).

The PKD1 protein also contains a single domain with homologies to C-type (for calcium-dependent) lectin proteins (Drickamer, 1988) (see Figure 2). These domains are believed to be involved in the extracellular binding of carbohydrate residues for diverse purposes, including internalization of glycosylated enzymes (asialoglycoprotein receptors) and cell adhesion (selectins) (Weis, 1992). The classification of C-type lectins has been based on exon organization and the nature and arrangement of domains within the protein (Bezouska et al., 1991). For example, class I (extracellular proteoglycans) and class II (type II transmembrane receptors) all have three exons encoding the carbohydrate recognition domain (CRD), whereas in classes III (collectins) and IV (lectin cell adhesion molecules [LEC-CAMs]), the domains are encoded by a single exon. The C-type lectin CRD in the PKD1 protein does not fit into the above classification, because it has a novel combination of protein domains and because it is encoded by two exons (exons 6 and 7; Figure 6). Previous analysis has failed to establish a correlation between the type of carbohydrate bound to each C-type lectin and the primary structure of its CRD (Weis, 1992).

PKD1 exon 1	PCEPPCLCGP	APGAACRVNCSG	RGLRTLGPALRI
ALS/Human	ACPAACVCSYDD	DADELSVFCSSRNLT-	-RLPDGVPGGTQA
TrkA/Human	QCPALCECS	-EAARTVKCVNRNLT-	-EVPTDLPAYVRN
GP1B/Human	GCPAPCSCA	GTLVDCGRRGLTW	IASLPTAFPVDTTE
OMGP/Human	ICPLQCICTE	RHRHVDCSGRNLS-	-TLPSGLQENIIH
HSGPV/Human	PCPPACKCV	FRDAAQCSGGDVA-	-RISA-LGLPTNK
PGS2/Human	VCPFRCQCH	LRVVQCSDLGLD-	-KVPKDLPPDTTL
Slit/Dos	SCPHPCRCA	DGIVDCREKSLT-	-SVPVTLPDDTTD
Toll/Dros	KCPRGCNCHVRT	DKALVINCHSGNLT-	-HVPR-LPNLHKN
	с с	r r	Ι.

Figure 4. Amino-Terminal Cysteine-Rich Domain

This repeat is encoded by exon 1 in the *PKD1* transcript. Examples of proteins that also contain these repeats are human insulin-like growth factor-binding protein complex acid-labile chain precursor (ALS); human high affinity nerve growth factor receptor (trkA); platelet membrane glycoprotein 1B a chain precursor (GP1B); human oligodendrocyte-myelin glycoprotein precursor (OMPG); platelet glycoprotein V precursor (HSGPV); human bone proteoglycan, decorin (PGS2); and Drosophila slit and Toll proteins. Conserved amino acids are represented in the bottom line of the figure (conserved amino acids are not aligned).

Exon 10 encodes a low density lipoprotein A (LDL-A) module (see Figure 2), a cysteine-rich domain of about 40 amino acids in length. This module was originally identified in the LDL receptor (Sudhof et al., 1985), but it is also present in the extracellular portion of many other proteins, often in tandem arrays (Bork and Bairoch, 1995) (Figure 7). Because of their hydrophobic nature, these domains have been implicated as ligand-binding regions in LDL receptor-related proteins (Krieger and Herz, 1994).

In addition to the five extracellular protein modules that have been recognized previously, the PKD1 protein contains 14 copies of a novel domain of approximately 80 amino acids (see Figure 2). We named this domain the PKD domain. The first such domain is encoded by exon 5 between the LRRs and the C-type lectin module. The other 13 PKD domains are arrayed in tandem, starting at amino acid 1031 and ending at amino acid 2142 and contained in exons 13, 14, and 15. Profile and motif searches (see Experimental Procedures) identified several other extracellular proteins that also contain one or more copies of this novel domain. The PKD domains are unusual in that they are found in the extracellular parts of proteins from higher organisms, eubacteria, and archaebacteria. In general, extracellular modules of proteins from higher organisms are not found in bacteria. The few exceptions appear to be the result of horizontal gene transfer (Doolittle and Bork, 1993) (Figure 8). The animal proteins containing an individual PKD domain are heavily glycosylated, melanoma-associated cell surface proteins, such as melanocyte-specific human Pmel17 (Kwon, 1993), the

KD1 exon 4	SGNPFECDCGLAWLPQWAEE-QQVRVVQPEAATCAGPGSLAGQPLLGIP~LLDSGCG
LS/Human	EGNPWDCGCPLKALRDFALQNPSAVPRFVQAICEGGDDCQPFAYTYNNITCA
rkA/Human	SGNPLHCSCALRWLQRWEEEGLGGVPEQKLQCHGQGPLAHMPNASCG
2GL/Human	SGNPWICDQNLSDLYRWLQA-QKDKMFSQNDTRCAGPEAVKGQTLLAVVAKSQ
SGPV/Human	GHNSWRCDCGLGPFLGWLRQHLGLVGGEEPPRCAGPGAHAGLPLWALPG-GDAECP
T4G/Human	DNNPWVCDCHMADMVTWLKETEVVQGKDRLTCAYPEKMRNRVLLELNS-ADLDCD
lit/Dros1	SDNPFACDCHLSWLSRFLRSATRLAPYTRCQSPSQLKGQNVADLHD-QEFKCS
oll/Dros	NDNPLVCDCTLLWFVQLVRGVHKPQYSRQFKLRTDRLVCSQPNVLEGTPVRQIEP-QTLICP
	ttNPh CDCtL h hht C tPt htt Ct

Figure 5. Carboxy-Terminal Cysteine-Rich Domain

This repeat is encoded by exon 4 in the transcript. Examples of other proteins that contain these domains are also listed, such as human insulin-like growth factor-binding protein complex acid-labile chain precursor (ALS); human high affinity nerve growth factor receptor (trkA); leucine-rich α -2-glycoprotein (A2GL); human heparin sulfate glycoprotein V (HSGPV); human 5T4 oncofetal antigen (5T4G); and Drosophila slit and Toll proteins. The hydrophobic and turn-like or polar amino acids are denoted by h and t, respectively. Conserved amino acids are represented in the bottom line of the figure.

melanosomal matrix protein (MMP) 115 protein (Mochii et al., 1991), and the Nmb protein (Weterman et al., 1995). The physiological functions of these glycoproteins remain to be elucidated. Four eubacterial extracellular enzymes, including three distinct collagenases (Yoshihara et al., 1994) and lysine-specific Achromobacter protease I (API), also contain a single copy of the domain adjacent to their catalytic domains. Four copies of the PKD domain are also present in the heavily glycosylated surface layer protein (SlpB) from Methanothermus (Brockl et al., 1991; Yao et al., 1994).

The PKD domain is predicted to be a globular domain that contains an antiparallel ß sheet. Although the PKD domains do not contain conserved cysteines, we believe they are extracellular domains, first because all homologous domains are extracellular; second, because the first such domain in PKD1 (amino acids 281-353) is located between other known extracellular modules; and third, because there are no predicted transmembrane regions between the other identified (extracellular) modules in PKD1 and the 13 remaining PKD domains. Whereas the PKD domains in SlpB are very similar (Brockl et al., 1991), pointing to a rather recent duplication, the 14 domains in PKD1 are quite divergent. Even the most conserved (WDFGDG) motif is considerably modified in some of the PKD domains (Figure 8). Therefore, it is unlikely that unequal recombination between genomic sequences encoding these domains is a common source of mutations in this disease.

In the carboxy-terminal half of the protein, we found regions of similarity to a putative Caenorhabditis elegans protein (GenBank accession number Z48544) encoded by

PKD1 exons 6/7 RegII/Human Botb/Ruman Manr/Human Lec/Rat Protc/Chicken Lec3/Megro

DCPP	-DWSSYEGHCYRFF	KEWMHWD.	DAE	-EFCTE	QQTGAI	HLVSI	QSK	EEADEVI	RSLTS	EMLKGI	0V-V
HCPS	QWWPYAGHCYKI-	HRDEKKI	QRDAI	LTTCR-	-KEGGI	DLTS	HTI	EELDFI	ISQL-	-GYEPN	IDEL
CCPI	-NWVEYEGSCYWFS	SSVKPWT	EAD	-KYCQ-	-LENA	HLVV	/TSWI	EQRFV	QQHM-	GI	LNT
NCEE	-GWIKFQGHCYR	HFEERET	MDAI	ESRCR-	-EHQAI	HLSS:	ITPI	SEQEFVI	ISH	^AÇ	DYQ
TCPGNI	LDWQEYDGHCYWAS	TYQVRWN	DAQ	-LACQT	VHPGA	YLAT:	[QSQ]	LENAFIS	SETV-	SP	INRL
CP	GHCY	tW	А	Ct	ttA	Lì	10	tFh	t	t	

PKD1 exons 6/7 RegII/Human Botb/Human Manr/Human Lec/Rat Protc/Chicken Lec3/Megro

COI VQGVC	VGE.	HLOGE	ur.o	050	COMPT	92110	Indicendor	100001011	DCDIN	L HOI VCHUQI	
LHDPKRNF	RWH	WSSGS.	LFL	YK-	SWAT	GSPNS	SS-NRG-YCVSLTSN	TGYKKWKDD	NCDA(QYSFVCKFKG	
LSDVWNKC	RFE	TDGM	EFD	YD-	DYYL	IAEY	ECVASKP-	TNNKWWII	PCTRE	FKNFVCEFQA	
LNDIKIQM	IYFE	WSDGT	PVT	FT-	KWLR	GEPSI	HENNRQEDCVVMKG-	KDGYWADR	GCEWI	PLGYICKMKS	
LTDONGPW	IKWV	DGTDY	ETG	FK-	NWR P	GQPDI	DWYGHG-LGGGEDCA	HFTTDGHWNDD	VCRRI	PYRWVCETEL	
LSDRAVEN	DFR	SDGH	SLQ	FE-	NWRPI	NQPD	VFFSAGEDCVVMIWH	EQGEWNDV	PCNYI	IL PFTCKKGT	
LNDIDLEC	HYV	SNGE.	ATD	FT-	YWSSI	NNPNI	WENQDCGVVNYD	TVTGQWDDD	DCNK	NKNFLCKMPI	
Ltt.	h	ttG	h	ht	W	tΡ	CV	GW	Ct	tahCtht	

Figure 6. C-Type Lectin Domain

This repeat is encoded by exons 6 and 7. Examples of other proteins that contain these repeats are also listed, such as human regenerating islet cell factor (RegII); human botrocetin (Botb); human mannose receptor (Manr); rat C-lectin (Lec); chicken proteoglycan core protein (Protc); and barnacle lectin BRA3 (lec3). The hydrophobic and turn-like or polar amino acids are denoted by h and t, respectively. Conserved amino acids are represented in the bottom line of the figure. A lowercase letter or represents aromatic amino acids.

PKD1 exon 10	ACMPGGRWCPGANICLPLDASCHPQ-ACANGCTS
Entero1/Pig	ECLPGSRPCADALKCIAVDLFCDGELNCPDGSDEDSKICAT
AM2/mouse	RCPPNEHSCLGTELCVPMSRLCNGIQDCMDGSDEGA-HCRE
LDLrel1/Ceano.	RCPPGKWNCPGTGHCIDQLKLCDGSKDCADGADEQQ-~CSQ
HSPG/Human	PCGPQEAACRNGH-CIPRDYLCDGQEDCEDGSDELD-~CGP
C8com/Human	RCEGFVCAQTGRCVNRRLLCNGDNDCGDQSDEAN-~CRR
	tCt Ctt Ch LCtG tC DGsDEt Ct

Figure 7. LDL-A Domain

The LDL-A domain is encoded by exon 10. Examples of other proteins that contain these domains are also listed, such as pig enteropeptidase, mouse AM2 receptor, C. elegans LDL-related receptor, human basement membrane proteoglycan (HSPG), and human C8 complement β chain. The turn-like or polar amino acids are denoted by t. Conserved amino acids are represented in the bottom line of the figure.

chromosome III (Wilson et al., 1994). The differences in the amino termini of these proteins imply that they are paralogs (i.e., not the equivalent genes in different species) and thus hint at the presence of a multigene family.

Between positions 3986 and 4040 there are several hydrophobic regions that might represent possible transmembrane domains, but without any clear resemblance to other such domains. Since the overall architecture of the PKD1 protein does not resemble other known proteins in which membrane domains are present, future independent data will be required to determine whether this segment spans the cell membrane.

Discussion

We report the DNA sequence and deduced protein sequence of *PKD1*, the gene that is commonly mutated in autosomal dominant polycystic kidney disease (ADPKD). The sequence presented in this paper extends the previously published partial sequence (by 2689 amino acids) and shows that the *PKD1* gene encodes a 4304 amino acid protein whose amino-terminal portion is made up of a series of extracellular protein domains. Since these domains are present in a combination that has not been seen in other multidomain proteins, the product of the *PKD1* gene cannot be assigned to an existing protein family. The recognizable modules include two leucine-rich repeats flanked by cysteine-rich domains, a C-type lectin carbohydrate recognition domain, an LDL-A domain, and a novel 80 amino acid domain present both as a single unit and separately as a tandem array (Figure 9).

In view of its enormous length and the presence of multiple adhesive domains, the PKD1 protein appears to be a multifunctional protein that is involved in various proteinprotein and protein-carbohydrate interactions in the extracellular compartment. The presence of several distinct binding domains suggests that the PKD1 protein binds to more than one molecule or to several parts of a large extracellular molecule. It is unclear whether the PKD1 protein contains a cytoplasmic segment or whether the protein is wholly extracellular. Nevertheless, the structure of the PKD1 protein suggests that it binds to components of the extracellular matrix or to cell membrane-associated proteins. Therefore, the PKD1 protein may mediate cellcell or cell-matrix interactions, or may itself be an intrinsic component of the extracellular matrix.

Although a number of defects have been observed at both the cellular and the tissue levels in ADPKD and in rodent models of renal cystic disease, it has been unclear whether they represent primary or secondary events (Calvet, 1993; Carone et al., 1994). One such defect is the abnormal distribution of cell membrane components between the apical and the basolateral surfaces of the polarized tubular epithelium. For example, Na⁺/K⁺ ATPase is found on the basolateral surface of normal tubules and nondilated tubules in ADPKD samples, but is present on the apical surface of cystic epithelia (Wilson et al., 1991). It has been suggested that mislocalization of membrane proteins results from a defect in protein sorting. However, the structure reported here makes it unlikely that the PKD1 protein is involved in the primary sorting of proteins (Carone et al., 1994).

Another hypothesis is that the *PKD1* gene encodes a growth factor or growth factor receptor and that mutations in the *PKD1* gene result in epithelial proliferation (Gran-tham, 1990; Wilson et al., 1986). An increase in the number of cells lining the tubules inevitably leads to dilatation in all

PKD1 exon 5 R1	GPLASGQLAAFHIAAPLPVTDTRWDFGDGSAEVDAAGPAASHRYVLPGRYHVTAVLALGAG-SALLGTDVQVE
PKD1 exon 13 R2	SPNATLALTAGVLVDSAVEVAFLWTFGDGEQALHQFQPPYNESFPVPDPSVAQVLVEHNVTHTYAAPGEYLLTVLASNAFE-NLTQQVPVSVR
PKD1 exon 14 R3	VAGRPVTFYPHPL-PSPGGVLYTWDFGDGSPVLTQSQPAANHTYASRGTYHVRLEVNNTVS-GAAAQADVRVF
PKD1 exon 15 R4	AVEQGAPVVVSAAVQTGDNITWTFDMGDGTVLSGPEATVEHVYLRAQNCTVTVGAGSPAG-HLARSLHVLVF
PKD1 exon 15 R5	IPTQPDARLTAYVTGNPAHYLFDWTFGDGSSNTTVRGCPTVTHNFTRSGTFPLALVLSSRVN-RAHYFTSICVE
PKD1 exon 15 R6	FVQLGDEAWLVACAWPPFPYRYTWDFGTEEAAPTRARGPEVTFIYRDPGSYLVTVTASNNIS-AANDSALVEVQ
PKD1 exon 15 R7	LGLELQQPYLFSAVGRGRPASYLWDLGDGGWLEG-PEVTHAYNSTGDFTVRVAGWNEVS-RSEAWLNVTVK
PKD1 exon 15 R8	VVPLNGSVSFSTSLEAGSDVRYSWVLCDRCTPIPGGPTISYTFRSVGTFNIIVTAENEVG-SAQDSIFVYVL
PKD1 exon 15 R9	YFPTNHTVQLQAVVRDGTNVSYSWTAWRDRGPALAGSGKGPSLTVLEAGTYHVQLRATNMLG-SAWADCTMDFV
PKD1 exon 15 R10	PAAVNTSVTLSAELAGGSGVVYTWSLEEGLSWETSEPFTTHSFPTPGLHLVTMTAGNPLG-SANATVEVDVQ
PKD1 exon 15 R11	FVAAGSSVPFWGQLATGTNVSWCWAVPGGSSKRGPHVTMVFPDAGTFSIRLNASNAVS-WVSATYNLTAE
PKD1 exon 15 R12	VVAPGQLVHFQILLAAGSAVTFRLQVGGANPEVLPGPRFSHSFPRVGDHVVSVRGKNHVS-WAQAQVRIVVL
PKD1 exon 15 R13	GIATGTERNFTARVQRGSRVAYAWYFSLQKVQGDSLVILSGRDVTYTPVAAGLLEIQVRAFNALG-SENRTLVLEVQ
PKD1 exon 15 R14	FTNRSAQFEAATS~PSPRRVAYHWDFGDGSPGQDTDEPRAEHSYLRPGDYRVQVNASNLVS-FFVAQATVTVQ
Pmel/Human	PLTFALQLHDPSGYLAEADLSYTWDFGDSSGTLISRAPVVTHTYLEPGPVTAQVVLQAAIP-LTSCGSSPVPG
Pmel/Bovine	PLTFALQLHDPSGYLAGADLSYTWDFGDSTGTLISRALTVTHTYLESGPVTAQVVLQAAIP-LTSCGSSPVPG
Nmb/Human	PIMFDVLIHDPSHFLNYSTINYKWSFGDNTGLFVSTNHTVNHTYVLNGTFSLNLTVKAAAPGPCPPPPPP
Coll/Vibal	VGESITFSSENSTDFNGKIVSVLWDFGDGSTSTQTKPTHQYGSEGEYSVSLSVTDSEG-LTATATHTVVI
Slpb/Meth1	TSGTAPLNVLFTDTSTGSPTTWKWNFGDGTSSTQKSPTHAYSTAGTYTVTLTVTINSAGSNTATKTNYVTV
Slpb/Meth2	RSGIAPLTVTFKDNSSGSPTAWNWSFGDGAYSNEKYPKHTYTAPGSYTISLTASNAAGSNTLTKSNYIVV
Slpb/Meth4	RSGTAPLTVTFKDNSSGSPTAWNWSFGDGAYSNEKYPKHTYMAPGSYTISLTASNAAGSNTLIKNNYIVV
Coly/Clope	VEEEINFDGTESKDEDGEIKAYEWDFGDGEKSNEAKATHKYNKTGEYEVKLTVTDNNG-GINTESKKIKV
Api/Achly	SGLTATFT-DSSTDSDGSIASRSWNFGDGSTSTATNPSKTYAAAGTYTVTLTVTDNGG-ATNTKTGSVTV
	hh t h h tttth a WDhGDGt tt t htH aht G h h h htN ht h h h

Figure 8. PKD Domains

These domains are present in exons 5, 13, 14, and 15. Other proteins that contain PKD domains are also listed, such as human and bovine melanoma antigen Pmel17, human Nmb protein, Clostridium perfringens collagenase, four domains of the Methanothermus fervidus Slpb protein, and Achromobacter lyticus protease (API). The hydrophobic and turn-like or polar amino acids are denoted by h and t, respectively. Conserved amino acids are represented in the bottom line of the figure.



Figure 9. A Schematic Representation of the PKD1 Protein

The figure depicts the location of the LRRs, flanked by cysteine-rich domains, the C-lectin domain (C-LEC), the LDL-A domain, and the fourteen PKD domains represented by numbers. The amino acid sites for predicted cysteine disulfide bonds in the C-lectin domain are identified with arrows. The cysteines represented by the numbers 298, 387, 401, and 409 should be represented instead by the numbers 419, 507, 522, and 530. The carboxyl half of the protein has no identifiable domains.

forms of cystic disease, including ADPKD. The structure of *PKD1* protein makes it unlikely that the primary defect in ADPKD involves a growth factor or classical receptor-mediated signal transduction pathway.

A number of reports have suggested that defects in extracellular matrix components or cell-matrix interactions are involved in the pathogenesis of ADPKD (Haverty and Neilson, 1988). First, it is known that the ADPKD phenotype encompasses nonrenal abnormalities such as vascular aneurysms and cardiac valve defects (Gabow, 1990). These abnormalities are also prominent features of diseases such as Ehlers-Danlos syndrome and Marfan syndrome that result from mutations in extracellular matrix components such as collagen and fibrillins (Kontusaari et al., 1990; Tsipouras et al., 1992). Analysis of ADPKD tissue reveals variably thickened basement membrane with decreased amounts of proteoglycans and increased amounts of fibronectin relative to those of normal individuals (Carone et al., 1989). A well-studied rodent model of acquired cystic disease provides further evidence that extracellular matrix abnormalities are involved in cyst formation (Butowski et al., 1985).

One hypothesis linking extracellular matrix abnormalities and ADPKD postulated that the mechanical compliance of the tubular basal membrane is altered, allowing tubules to distend (Welling and Grantham, 1972). However, in this study, altered elasticity in cystic tubules was not detected. A second hypothesis relies on a growing body of evidence that the extracellular matrix influences the development and morphology of the cells in contact with it (Rodriguez-Boulan and Nelson, 1989). The extracellular matrix interacts with the cell cytoplasm through cell adhesion molecules known as integrins (Hynes, 1992). During normal tubular development, coordinated interactions between the extracellular matrix and epithelial cells, mediated by cell-extracellular matrix adhesion, are critical for tubular morphogenesis. These interactions lead to a specific pattern of gene expression that in turn results in normal cell differentiation. Cell differentiation in turn modulates the synthesis of matrix components.

Our data suggest that the PKD1 protein presents several adhesive domains to the extracellular space. We hypothesize that these domains bind matrix or cell membraneassociated ligands, and that these interactions mediate normal epithelial differentiation. One property of the differentiated state is the maintenance of tubular architecture and an epithelial cell morphology that is appropriate for each tubular segment. Mutation of the PKD1 protein leads to failure of these interactions, so that epithelial differentiation does not occur or is not maintained.

The abnormal state of differentiation accounts for the range of features of the ADPKD phenotype; although cystlining cells are arranged in a monolayer and have welldefined tight junctions and other features of a differentiated epithelium, gene expression is abnormal for a mature epithelium (Carone et al., 1993). Epithelia fail to acquire, or maintain, critical properties of the differentiated state, such as the synthesis of extracellular matrix components that define the mature tubular basement membrane (Klingel et al., 1993; Dvergsten et al., 1994). The increased cell proliferation and occasional micropolyp formation seen in ADPKD (Evan et al., 1979) also reflect a less developed state. Another characteristic of cystic epithelium that more closely resembles the undifferentiated state is the distribution of cell membrane markers: the apical location of Na⁺/K⁺ ATPase, for example, is similar in cystic and fetal kidneys.

A number of experimental approaches can be used to test the hypotheses that arise from our predicted protein. First, antibodies raised against the PKD1 protein can be used to determine the cellular localization of the protein and will also help to dissect the molecules interacting with the extracellular domains of the protein. Second, these antibodies may block interactions in cell culture, where properties such as cell adhesion, cell polarity, synthesis of matrix components, and morphology could be assessed. A cell culture approach has been used successfully to determine the functions of integrins (Ruoslahti and Pierschbacher, 1987). Identification of the PKD1 protein ligands will allow the definition of the pathways that lead to normal epithelial morphogenesis.

Further study is also needed to determine the spectrum of mutations and the basis for the dominant inheritance of ADPKD. The distribution of mutations reported to date is biased by the relative ease of analyzing the unique 3' end of the *PKD1* gene. The great majority of *PKD1* mutations remain uncharacterized and may be distributed throughout the gene. The genomic sequence did not provide clues that might account for the high new mutation rate observed in *PKD1* (Snarey et al., 1994). The intron/exon structure reported here will expedite the analysis of mutations in the duplicated part of the *PKD1* gene. ADPKD is only one of many genetic forms of renal cystic disease. Mutations in at least fifteen nonallelic loci in human and mouse have been shown to cause the disease (Reeders, 1992; Calvet, 1994). It is clear that a large number of genes, involved in one or more biochemical pathways, are responsible for maintaining normal tubular morphology. Determination of the structure of the *PKD1* gene and its protein provides an important entry point for the dissection of these pathways.

Experimental Procedures

Genomic Clones

The P1 phage named PKD 1521 was obtained from a human genomic library (Genome Systems, St. Louis, MO) and was isolated by use of primers from both the 5' end of the TSC2 gene and the 5' end of KG8. The cosmid cGGG10 has been described (Germino et al., 1992). A random library of the cosmid was constructed by cloning sheared DNA fragments into the Smal site of pUC 19. Initial sequence assembly for the cosmid cGGG10 was performed with forward and reverse sequences from approximately 1000 random cloned fragments. A preliminary map was constructed by using the restriction map of the cosmid. Directed subclones of cGGG10 were made in pBluescript in order to create sequencing islands anchored to specific restriction enzymes. These large subclones from cGGG10 were then restricted with more frequent cutter enzymes and cloned into M13mp19 and mp18. Directed sequencing employing primer walks to form large anchored contigs was also performed by using the appropriate subclones of cGGG10. A contig of 34.3 kb was constructed, with two gaps that contain highly repetitive regions with no identifiable coding sequence. cDEB11 has been described (Germino et al., 1992). A random library was constructed with sheared cDEB11 DNA and cloned into the Small site of pUC 19. This cosmid was sequenced to obtain at least 2-fold coverage.

The products of cycle sequencing were separated on automated sequencers (Applied Biosystems, Incorporated) according to the instructions of the manufacturer, with modifications described below. Because of the difficulty of sequencing certain regions, the standard chemistry needed to be modified. We used both dye terminator and dye primer methods when appropriate for sequencing different regions. We also used a range of polymerases, different melting temperatures, and polymerization conditions to optimize the quality of the sequence. When sequencing across the CpG island at the 5' end of the *PKD1* gene, we obtained the best sequencing results by adding 5% dimethyl sulfoxide to the polymerization step and sequencing single-stranded templates.

cDNA Library Screening

cDNA clones were identified in two ways. First, fragments of cosmids cGGG10 and cDEB11 were hybridized to five cDNA libraries (lymphoblast, fetal kidney, adult kidney, brain, and teratocarcinoma). Second, each cDNA clone was hybridized to fetal kidney and lymphocyte cDNA libraries to obtain overlapping clones to extend the sequence. The first cDNA used to screen libraries was KG8, which maps to the unique region of the *PKD1* locus and was recovered from an adult lymphocyte library. To obtain the rest of the *PKD1* transcript, 14 new cDNAs were sequenced to completion, 4 cDNAs were partially sequenced, and an additional 20 cDNAs were mapped with respect to cGGG10. Additional data were obtained from polymerase chain reaction (PCR) products of the renal cell carcinoma library as a template (American Type Culture Collection).

Overlapping partial cDNAs described below were isolated from lymphocyte and fetal kidney libraries. In this way, we assembled a 14.5 kb transcript starting from the 3' end until we reached the CpG island. We assumed we had reached the 5' end of the transcript, not only because of the presence of the CpG island, but because three cDNAs isolated (UN53, UN54, and UN59, described below) all had the same 5' end. No further upstream clones were recovered upon further screening (with UN53) the libraries that provided most of the cDNAs used to build the full-length cDNA.

FK7 and FK11 were recovered from a 14-16 week fetal kidney cDNA library by using KG8. This library was oligo(dT) primed and constructed with the Superscript Lambda System (GIBCO BRL). FK7 and FK11 were recovered as Sall inserts. BK156, BK194, and UN49 and UN52 were recovered from a Jurkat library by using FK7 as a probe. UN34 was recovered from the same library by hybridizing a Scal-Sall probe from the 5' end of FK7. UN53, UN54, and UN59 were recovered from the same library by double screening for clones that were both negative when probed with an FK7 and positive when screened with BK156 and UN52. This Jurkat library was a gift of the laboratory of M. Owen at the Imperial Cancer Research Fund. NKG11 was recovered from a lymphocyte library screened with cGGG10 and has been described previously (Germino et al., 1992). Fhkb21 was isolated from a fetal kidney library by using BK156 as a probe (Clontech). MSK3 was obtained by probing an adult kidney library (Clontech) with the 5' end of KG8. MSK4 was obtained by nested reverse transcription-PCR with primers spanning from exons 7 to 8 and exons 13 to 14, followed by a second round of PCR with internal primers in exons 8 and 13.

cDNA Sequencing

The cDNAs were sequenced to 5-fold coverage by primer walking and by subcloning small fragments into M13 or pBluescript. All cDNA sequences were compared with cGGG10 sequence to assess whether they were from the correct locus and to determine intron/exon boundaries. Regions of discrepancy were sequenced again to determine whether the differences were genuine. Some of the cDNAs described above were unequivocally different from the genomic sequence (more than 3 bp difference for every 100 bp), suggesting that these cDNAs were encoded by another locus.

MSK3, FK7, and FK11, obtained by using a PKD1-specific probe (KG8), were found to be 100% identical to the genomic sequence. UN49 showed 99% identity and is possibly PKD1 specific. BK241, BK194, UN52, UN53, UN54 and UN59, BK156, Fhkb21, and NKG11 were 97%-98% homologous to the cGGG10-defined exon sequence and therefore were assumed to have originated from the duplicated loci. In general, differences between genomic and cDNA sequences were nucleotide substitutions scattered throughout the cDNA sequence. One exception is BK194, which has an extra CAG at position 1863 of the published sequence and arose from alternative splicing of exon 33. Another exception is BK241, which has a tandem repeat of TTATCAATACTCTGGCTGACCATCGTCA inserted at position 1840 of the previously published sequence (European Polycystic Kidney Disease Consortium, 1994) and was not included in the full-length cDNA because it arose from a duplicated locus. Except for BK241, cDNAs in the UN and BK series that overlap each other are more similar to themselves than to the cGGG10 sequence.

All sequence assembly was performed by using the Staden package, XBAP (Dear and Staden, 1991).

Protein Homology Searches

The PKD1 sequence was subjected to a number of sequence analysis approaches (Koonin et al., 1994; Bork et al., 1994). To identify homologs, initial (SwissProt, PIR, GenPept, TREMBL, EMBL, GenBank, and NRDB) database searches were performed by use of the BLAST series of programs (Altschul and Lipman, 1990) by applying a filter for compositionally biased regions (Altschul et al., 1994). By default, the BLOSUM62 amino acid exchange matrix was used (Henikoff and Henikoff, 1993). To detect additional candidate proteins that might be homologous to PKD1, the BLOSUM45 and PAM240 matrices were also applied. Putative homologs with a BLAST p value below 0.1 were studied in detail. Multiple alignments of the candidate domains were carried out using CLUSTALW (Thompson et al., 1994), and patterns (Rohde and Bork, 1993), motifs (Tatusov et al., 1994), and profiles (Gribskov et al., 1987) were derived. With all these constructs, iterative database searches were performed. Results of these database searches were used for improving the multiple alignments that were then used for the next round of database searches. The final multiple alignment, containing all retrieved members of a module family, was then used as input for the secondary structure predictions (Rost and Sander, 1994).

Acknowledgments

M. Owen and J. Dunne provided the Jurkat library. The human fetal kidney was obtained from the Medical Research Council fetal tissue bank with permission of the ethics committee. M. C. S. and S. T. R. are funded in part by the National Institute of Diabetes and Digestive and Kidney Diseases. We would like to thank G. G. Germino, K. Harris. N. J. Barton, S. Somlo, D. Weinstat-Saslow, and G. Waterbury for the isolation of KG8 and NKG11. The laboratory of A.-M. F. was funded by the Imperial Cancer Research Fund. M. P. was funded by a Boehringer Incelheim Foundation fellowship, C. L. has a fellowship from the Deutsche Forschungsgemeinschaft. The Imperial Cancer Research Fund group would like to thank K. Zerres, B. Wirth, A. Poustka, B. Korn, G. Neild, and S. Wood. We also thank the members of the Millennium Pharmaceuticals sequencing core (T. Brody, S. Ellis, P. Paglierani, T. Revett, P. Shu, and S. Yang), L. Berkemeier, J. Pulido, Y.-H. Fu, A. Shyjan, J. Shao, and L. Holmgren for technical assistance. B. M. Brenner, F. Lee, R. Tepper, and L. Reid are thanked for critical reading of this manuscript. M. Levin and T. Clark are thanked for helpful discussions. D. Allison is thanked for her graphic designs.

Received March 30, 1995; revised April 3, 1995.

References

Altschul, S. F., and Lipman, D. J. (1990). Protein database searches for multiple alignments. Proc. Natl. Acad. Sci. USA *87*, 5509–5513. Altschul, S. F., Boguski, M. S., Gish, W., and Wooton, J. C. (1994). Issues in searching molecular sequence databases. Nature Genet. *6*, 119–129.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. USA 90, 11995–11999. Bezouska, K., Crichlow, G. V., Rose, J. M., Taylor, M. E., and Drickhamer, K. (1991). Evolutionary conservation of intron position in a subfamily of genes encoding carbohydrate-recognition domains. J. Biol. Chem. 226, 11604–11609.

Bork, P., and Bairoch, A. (1995). Extracellular modules. Trends Biochem. Sci. 20(3), poster.

Bork, P., Ouzounis, C., and Sander, C. (1994). From genome sequences to protein function. Curr. Opin. Struct. Biol. 4, 393–403.

Brockl, G., Behr, M., Fabry, S., Hensel, R., Kaudewitz, H., Biendl, E., and Konig, H. (1991). Analysis and nucleotide sequences of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogen *Methanothermus fervidus* and *Methanothermus sociabilis*. Eur. J. Biochem. *199*, 147–152.

Butowski, R. J., Carone, F. A., Grantham, J. J., and Hudson, B. G. (1985). Tubular basement membrane changes in 2-amino-4,5-diphenylthiazol induced polycystic kidney disease. Kidney Int. 28, 744–751.

Calvet, J. P. (1993). Polycystic kidney: primary extracellular matrix abnormality or effective cellular differentiation? Kidney Int. 43, 101–108.

Calvet, J. P. (1994). Injury and development in polycystic kidney disease. Curr. Opin. Nephrol. Hyper. 3, 330–348.

Carone, F. A., Hollenberg, P. F., Nakamura, S., Punyarit, P., Glogowski, W., and Flouret, G. (1989). Tubular basement membrane change occurs pari passu with the development of cyst formation. Kidney Int. *35*, 1034–1040.

Carone, F. A., Nakamura, S., Bacallao, R., Nelson, W. J., and Kanwar, Y. S. (1993). Cell polarity in human renal cystic disease. Lab. Invest. 70, 648–655.

Carone, F. A., Bacallao, R., and Kanwar, Y. S. (1994). Biology of polycystic kidney disease. Lab. Invest. 70, 437–448.

Daoust, M. C., Reynolds, D. M., Bichet, D. G., Somlo, S. (1995). Evidence for a third genetic locus for autosomal dominant polycystic kidney disease. Genomics 25, 733–736.

Dear, S., and Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. Nucl. Acids Res. *19*, 3907–3911.

Doolittle, R. F., and Bork, P. (1993). Evolutionary mobile modules in proteins. Sci. Am. 269, 50–56.

Drickamer, K. (1988). Two distinct classes of carbohydrate-recognition domains in animal lectins. J. Biol. Chem. 263, 9557–9561.

Dvergsten, J., Manivel, J. C., Correa-Rotter, R., and Rosenberg, M. E. (1994). Expression of clustering in human renal diseases. Kidney Int. *45*, 828–835.

European Chromosome 16 Tuberous Sclerosis Consortium (1993). Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75, 1305–1315.

European Polycystic Kidney Disease Consortium (1994). The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Cell 77, 881–894.

Evan, A. P., Gardner, K. D., and Grantham, J. J. (1979). Polypoid and papillary hyperplasia: a potential cause of ductal obstruction in adult polycystic kidney disease. Kidney Int. *16*, 743–750.

Fossdal, R., Bothvarsson, M., Amundsson, P., Ragnarsson, J., Peters, D., Breuning, M. H., and Jensson, O. (1993). Icelandic families with autosomal dominant polycystic kidney disease: families unlinked to chromosome 16p13.3 revealed by linkage analysis. Hum. Genet. 91, 609–613.

Gabow, P. A. (1990). Autosomal dominant polycystic kidney disease: more than a renal disease. Am. J. Kidney Dis. *16*, 403–413.

Germino, G. G., Weinstat-Saslow, D., Himmelbauer, H., Gillespie, G. A. J., Somlo, S., Wirth, B., Barton, N., Harris, K. L., Frischauf, A.-M., and Reeders, S. T. (1992). The gene for autosomal dominant polycystic kidney disease lies in a 750 kb CpG-rich region. Genomics *13*, 144– 151.

Grantham, J. J. (1990). Polycystic kidney disease: neoplasia in disguise. Am. J. Kid. Dis. 15, 110–116.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA 84, 4355–4358.

Hashimoto, C., Hudson, K. L., and Anderson, K. V. (1988). The *Toll* gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell *52*, 269–279.

Haverty, T. P., and Neilson, E. G. (1988). Basement membrane gene expression in polycystic kidney disease. Lab. Invest. 58, 245–248.

Henikoff, S., and Henikoff, J. G. (1993). Performance evaluation of amino acid substitution matrices. Proteins 17, 97-61.

Himmelbauer, H., Germino, G. G., Ceccherini, I., Romeo, G., Reeders, S. T., and Frischauf, A. M. (1991). Saturating the region of the polycystic kidney disease gene with Notl linking clones. Am. J. Hum. Genet. 48, 325–334. Erratum: Am. J. Hum. Genet. 48, 1017.

Hynes, R. O. (1992). Integrins: versatility, modulation, and signals in cell adhesion. Cell 69, 11-25.

Kimberling, W. J., Kumar, S., Gabow, P. A., Keynon, J. B., Connolly, C. J., and Somlo, S. (1993). Autosomal polycystic kidney disease: localization of the second gene to chromosome 4q13-q23. Genomics *18*, 467–472.

Klingel, R., Ramadori, G., Schuppan, D., Knittel, T. K. H., Meyer zum Buschenfelde, K. H., and Kohler, H. (1993). Co-expression of extracellular matrix glycoproteins undulin and tenascin in human autosomal dominant polycystic kidney disease. Nephron 65, 111–118.

Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. Trends. Biochem. Sci. 19, 415–421.

Kobe, B., and Deisenhofer, J. (1995). A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374, 183–186.

Kontusaari, S., Tromp, G., Kuivaniemi, H., Romanic, A. M., and Prockop, D. J. (1990). A mutation in the gene for type III procollagen (COL3A1) in a family with aortic aneurysms. J. Clin. Invest. *86*, 1465– 1473.

Koonin, E. V., Bork, P., and Sanders, C. (1994). Yeast chromosome III: new gene functions. EMBO J. *13*, 493–503.

Kozak, M. (1984). Compilation and analysis of sequences upstream

from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. 12, 857–872.

Krieger, M., and Herz, J. (1994). Structures and functions of multiligand lipoprotein receptors. Annu. Rev. Biochem. 63, 601–637.

Kwon, B. S. (1993). Pigmentation genes: the tyrosinase gene family and the pmel 17 gene family. J. Invest. Derm. (Suppl.) *100*, 134–140. Mochii, M., Agata, K., and Eguchi, G. (1991). Complete sequence and expression of a cDNA encoding a chicken 115 kD melanosomal matrix protein. Pigment Cell Res. *4*, 41–47.

Myers, K. A., Rahi-Saund, V., Davison, M. D., Young, J. A., Cheaters, A. J., and Stern, P. L. (1994). Isolation of a cDNA encoding 5T4 oncofetal trophoblast glycoprotein. J. Biol. Chem. 269, 9319–9324.

Peters, D. J. M., Spruit, L., Saris, J. J., Ravine, D., Sandkuijl, L. A., Fossdal, R., Boersma, J., Vaneijk, R., Norby, S., and Constantinoudeltas, C. D. (1993). Chromosome 4 localization of a second gene for autosomal-dominant polycystic kidney disease. Nature Genet. 5, 359– 362.

Reeders, S. T. (1992). Multilocus polycystic disease. Nature Genetics 1, 235–237.

Reeders, S. T., Breuning, M. H., Davies, K. E., Nicholls, D. R., Jarman, A. J., Higgs, D. R., Pearson, P. L., and Weatherall, D. J. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature *317*, 542–544.

Rodriguez-Boulan, E., and Nelson, W. J. (1989). Morphogenesis of the polarized epithelial cell phenotype. Science 247, 718-725.

Rohde, K., and Bork, P. (1993). A fast, sensitive pattern-matching approach for protein sequences. Comput. Appl. Biosci. 9, 183–189. Rost, B., and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19, 55–72.

Roth, G. J. (1991). Developing relationships: arterial platelet adhesion, glycoprotein lb, and leucine-rich glycoproteins. Blood 77, 5–19.

Rothberg, J. M., Jacobs, J. R., Goodman, C. S., and Artavanis-Tsakonas, S. (1990). Slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. Genes Dev. 4, 2169–2187.

Ruoslahti, E., and Pierschbacher, M. D. (1987). New perspectives in cell adhesion: RGD and integrins. Science 238, 491-539.

Schneider, R., and Schweider, M. (1991). A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogeneic trk and trk B tyrosine kinase receptors. Oncogene 6, 1807–1811.

Snarey, A., Thomas, S., Schneider, M. C., Pound, S. E., Barton, N., Wright, A. S., Somlo, S., Germino, G. G., Harris, P. C., Reeders, S. T., and Frischauf, A.-M. (1994). Linkage disequilibrium in the PKD1 region. Am. J. Hum. Genet. 55, 365–371.

Somlo, S., Wirth, B., Germino, G. G., Weinstat-Saslow, D., Gillespie, G. A., Himmelbauer, H., Steevens, L., Coucke, P., Willems, P., Bachner, L., Coto, E., Lopez-Larrea, C., Peral, B., San Millan, J. L., Lavinha, J., Saris, J. J., Breuning, M. H., Frischauf, A.-M., and Reeders, S. T. (1992). Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (*PKD1*) with respect to physically mapped markers. Genomics *13*, 152–158.

Sudhof, T. C., Goldstein, J. L., Brown, M. S., and Russell, D. W. (1985). The LDL receptor gene: a mosaic of exons shared with different proteins. Science 228, 815–822.

Tatusov, R. L., Altschul, S. F., and Koonin, E. V. (1994). Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks. Proc. Natl. Acad. Sci. USA *91*, 12091–12095.

Thompson, J. D., Higgins, D. G., and Gibson, T. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680.

Tsipouras, P., Del Mastro, R., Sararazi, M., Lee, B., Vitale, E., Child, A. H., Godfrey, M., Devereux, R. B., Hewett, D., Steinmann, B., Viljoen, D. L., Sykes, B., Kilpatrick, M., Ramirez, F., and the International Marfan Syndrome Collaborative Study (1992). Linkage analysis demonstrates that Marfan syndrome, dominant ectopia lentis and congenital contractural arachnodactyly are linked to the fibrillin genes on chromosomes 15 and 5. N. Engl. J. Med. 326, 905–909.

von Hejne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucl. Acids Res. 14, 4683-4690.

Weis, W. I., Quesenbery, M. S., Taylor, M. E., Bezouska, K., Hendrickson, W. A., and Drickhamer, K. (1992). Molecular mechanisms of complex carbohydrate recognition at the cell surface. Cold Spring Harbor Symp. Quant. Biol. *57*, 281–289.

Welling, L. W., and Grantham, J. J. (1972). Physical properties of isolated perfused renal tubules and tubular basement membranes. J. Clin. Invest. 51, 1063–1075.

Weterman, M. A. J., Ajubi, N., van Dinter, I., Degen, W., van Muijen, G., Ruiter, D. J., and Bloemers, H. P. J. (1995). *nmb*, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. Int. J. Cancer 60, 73–81.

Wilson, P. D., Schrier, R. W., Breckon, R. D., and Gabow, P. A. (1986). A new method for studying human polycystic kidney disease epithelia in culture. Kidney Int. *30*, 371–378.

Wilson, P. D., Sherwood, A. C., Palla, K., Du, J., Watson, R., and Norman, J. T. (1991). Reversed polarity of Na⁺, K⁺-ATPase: mislocation to apical plasma membranes in polycystic kidney disease epithelia. Am. J. Physiol. *260*, F420–F430.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature 368, 32–38.

Yao, R., Macario, A. J. L., and Conway, E. (1994). An archaeal S-layer gene homolog with repetitive subunits. Biochim. Biophys. Acta *1219*, 697–700.

Yoshihara, K., Matsushita, O., Minami, J., and Okabe, A. (1994). Cloning and nucleotide sequence analysis of the *colH* gene from *Clostridium histolyticum* encoding a collagenase and a gelatinase. J. Bacteriol. *176*, 6489–6496.

GenBank Accession Numbers

The accession numbers for the sequences reported in this paper are as follows: for the open reading frame, U24499; for the cDNA, U24497; and for the genomic region, U24498.