Gazetteer

L'Association pour la recherche sur le cancer

Otherwise known as … l'ARC (pronounced 'lark').

What is it famous for? Being one of France's largest medical research charities and the best known, as far as the French public is concerned, because of its strong media presence.

Does it have much money to give away? L'ARC's income in 1999 was FFr317 million (£32 million), three quarters of which went into funding research. Although this might not seem a lot, l'ARC — together with the other major French cancer charity, La Ligue Nationale Contre le Cancer (LNCC) — has an important role in plugging the gaps left by the stagnant budgets of CNRS and INSERM, France's public research agencies. Indeed, there are few biomedical scientists in France who have not benefited from l'ARC funding at one time or another.

So it's not all 'cancer research' then? L'ARC is considered very progressive and is known for its willingness to fund basic research that might seem far-removed from the cancer patient. To give an idea of l'ARC's breadth of coverage, this journal published 15 research articles that originated from French labs last year; 11 of these, with subject material ranging from neuron specification in frogs to blastomere fate in the zebrafish, acknowledged l'ARC for part-funding.

What else does l'ARC spend its money on? Last year, 410 stipends for graduate students and postdocs were disbursed by l'ARC to the tune of FFr50.6 million (£5.2 million). These stipends, or *bourses*, are hard to come by in France and l'ARC's support in this area is seen as a lifeline, particularly as French graduate students almost invariably take four years to complete their doctoral studies, and government support lasts only three years.

Where does the money come from? L'ARC raises almost all its money from public donations and legacies, relying heavily on its media presence. The charity is firmly associated in the public mind with the founder of l'ARC, Jacques Crozemarie (or Crozy as he is familiarly known), who until recently made regular televised appeals (in a white lab coat, although he has never been a physician or scientist). Advertisements in the press have been accompanied by signed photographs of Crozy.

How did l'ARC begin? L'ARC's forerunner, created in 1962 by Crozemarie, was the fund-raising body for the Institut Gustave-Roussy and several other cancer research labs based in the Paris suburb of Villejuif. It became a national charity in 1979 and since then has been the subject of some well-publicised controversies.

Why is it in the news? Crozemarie is alleged to have embezzled FFr300 million (£30.6 million) of the charity's funds. He appeared in court in June — along with 25 others — accused of diverting the funds into Swiss bank accounts through an elaborate scheme of false and inflated invoices. The scandal, which was made public in December 1994, has been described by the French media as "the biggest racket involving a charity this century." In 1993, before its name fell into disrepute, l'ARC's income was FFr581.2 million (£59.3 million). But according to France's national audit commission, which has been investigating the charity's accounts, less than one-third of l'ARC's income was spent funding research in 1993. Public generosity towards medical charities in general dropped sharply when the scandal broke.

Has l'ARC recovered? The installation of Michael Lucas — the former head of the watchdog body that exposed the financial irregularities of the charity — as the new president of l'ARC in 1996 and a purge of l'ARC's administrative council, seem to have helped rebuild public confidence, if the steady increase in donations is anything to go by.

What will happen to Crozemarie? He could face imprisonment for up to five years but his fate hangs in the balance as the French judiciary, like the rest of the country, grinds to a halt for the summer holidays.

Primer

Human memory

Alan J. Parkin

At one level we can say that we know exactly what human memory is — the learning of new information involves changes in the synaptic connections between neurons specialised for the storage of information. This is known from *in vivo* studies of simple invertebrates and there is no reason to suppose that these fundamental mechanisms do not also apply to humans. We also know which regions of the brain are specialized for memory and that various neurotransmitters are involved in memory, notably acetylcholine, the neurotransmitter known to be massively depleted in Alzheimer's disease.

But how does this anatomical and physiological knowledge enable us to understand the different types of memory, or how any given piece of knowledge is derived from a particular combination of synaptic connections? Quite simply we have no idea at present and for this reason we have to think of memory at two levels: biological and psychological. Eventually, the two sources of

knowledge might merge into a single account of memory but at present the gap between the two areas of research is enormous.

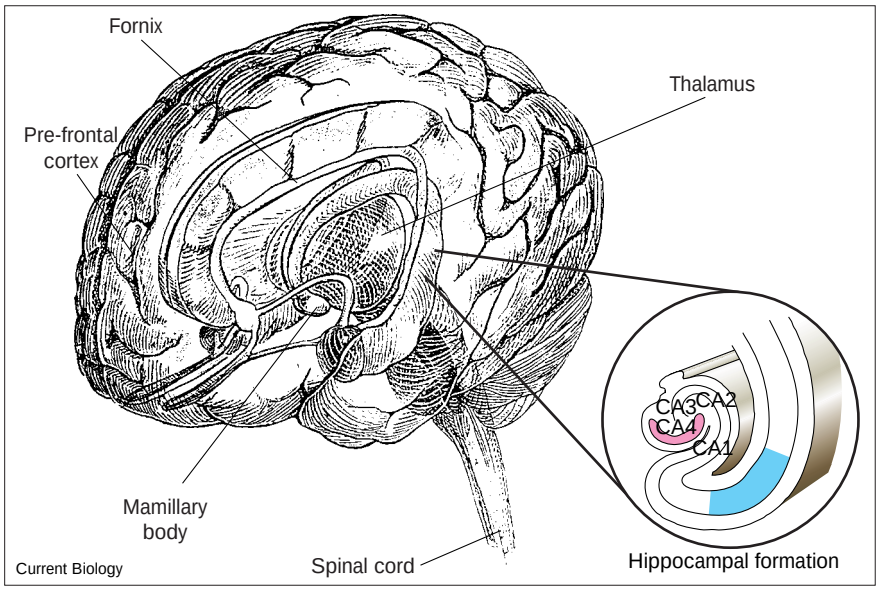
Neuroanatomy of human memory

Damage to two main regions of the brain can cause severe loss of memory, or amnesia. The first of these is the hippocampal formation (see Figure 1). Operations in which its structures have been removed to cure epilepsy cause amnesia. More convincingly, patients who have suffered small strokes affecting only the hippocampal formation experience severe amnesia.

A second brain region concerned with memory is the diencephalon. Haemorrhaging in the diencephalon (in Korsakoff's Syndrome) results in dramatic loss of memory; strokes in this region can also have the same effect. The hippocampus has a large efferent pathway, known as the fornix, which terminates in the mamillary bodies. Lesions of the fornix have also caused significant loss of memory.

The various structures mentioned so far are all part of a circuit known as the limbic system (Figure 1). The fact that damage to any of the structures can cause amnesia has led many to propose that memory is mediated by a 'circuit' in which, rather like an electrical circuit, damage anywhere will have the same effect.

Figure 1



The limbic system of the brain. The hippocampal formation comprises the hippocampus itself (which has four fields, CA1 to CA4) and associated structures including the dentate gyrus (pink) and the entorhinal cortex (blue). These structures all lie within the medial temporal lobe of the cerebral cortex. The other areas concerned with memory lie in an area called the diencephalon, these are the mamillary bodies, the dorso-medial nucleus of the thalamus and the mamillothalamic tract. The fornix links the hippocampus and the mamillary bodies. The circuitry of the limbic system is completed by projections from the dorso-medial nucleus to the pre-frontal cortex, from which efferent pathways return to the hippocampal formation. (Brain diagram adapted with permission from Bloom FE, Lazerson A: *Brain, Mind and Behavior*, 2nd edn. New York: WH Freeman; 1988.)

A psychological model of memory

Although there are some disagreements as to the psychological basis of memory, there is broad agreement on a 'modal' model for memory. The model is essentially based on information processing, where memory is conceived of as a series of 'stores' between which information flows. Three types of basic store are identified: sensory store, short-term store (STS), and long-term store (LTS) (see Figure 2). (These terms should not be confused with the commonly used phrases, 'short-term memory' and 'long-term memory', which memory theorists avoid because they are used very imprecisely to indicate periods of time over which information is held. In contrast, STS and LTS refer to hypothetical brain structures with assumed properties. 'Working memory' is another term that often crops up — for the purpose of this article working memory is best considered to be a more detailed aspect of STS.)

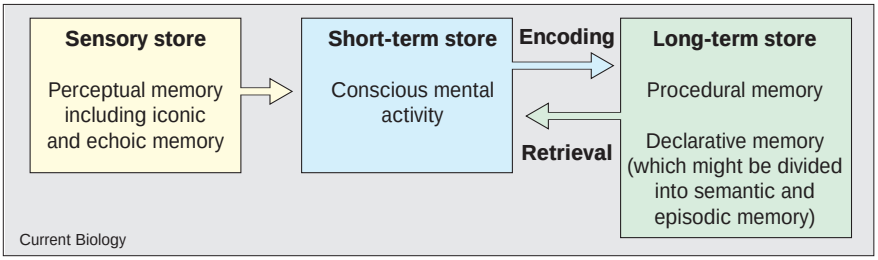
The sensory store provides short-lived storage of the perceptual input — visual information, known as 'iconic memory', is thought to last about half a second, whereas auditory information, 'echoic memory', lasts several seconds. Information then passes into STS; this represents the locus of conscious mental activity in which incoming information interacts with previously stored knowledge to direct a response.

Previously stored knowledge resides in LTS. Information enters via a process known as 'encoding' and, once in LTS, is accessed by a process called 'retrieval'. STS has a very limited capacity — this is indicated by the digit-span test in which subjects are required to repeat back a series of random digits immediately. Adults typically manage about seven digits before they start making mistakes. In addition the information in STS is actively stored, rather like the memory of a calculator, and can be lost if there is some form of distraction. Information in LTS is thought to be passively stored because it has undergone the process of 'consolidation' in which new memories come to be represented in terms of new patterns of synaptic interconnections.

Biological evidence

Patients with amnesia provide good evidence for the distinction between

Figure 2



The current modal model for memory, in which information flows between a series of stores.

STS and LTS . Their performance in the digit-span test, which requires only STS, is normal but performance in the story recall test, with its demands on LTS is very poor. Similar effects can be demonstrated in normal people given the drug scopolamine, which impairs action of acetylcholine; patients given scopolamine perform normally on digit-span but have markedly impaired story recall.

Biological evidence also bears on the distinction between encoding and retrieval. Amnesia takes two basic forms: 'anterograde', in which there is an inability to remember new information, and 'retrograde', in which the patient fails to remember information acquired before their brain injury. If encoding and retrieval were linked in some essential way we would expect these two forms of amnesia to be correlated with one another but they are not, suggesting independent mechanisms. Data from normal humans also support this. The drug imipramine when given to normal volunteers impairs the learning of new information but does not interfere with retrieving information learned before the drug was administered.

The nature of LTS

LTS has been described as if it were some single, large structure containing all our stored knowledge. In fact, there is abundant evidence that LTS has different independent components. There is a proposed division between so-called 'procedural' and 'declarative' memory. Procedural memory is defined as information we possess but which we cannot describe verbally. Procedural memory is typified by skills such as typing, and playing a musical instrument. Thus when a typist is asked how to type the word 'caterpillar' he or she will make the finger movements to provide the answer.

It is important to note that procedural memory is not a single entity; rather, a number of memory skills have a procedural quality. We know this because comparisons of people's ability on different forms of procedural memory task do not correlate — which they should if procedural memory were a single structure.

Declarative memory represents all knowledge that can be consciously accessed and expressed symbolically through speech and writing. There is a lively debate as to whether declarative memory is a single system or whether it has two separate components: 'semantic memory' — our knowledge of language, concepts or facts — and 'episodic memory' — our personal autobiographical knowledge of the past. Two lines of evidence bear on this issue: evidence from the relatively new field of functional neuroimaging (see green box) and the breakdown of LTS in human amnesia.

Amnesia and LTS

Studies of amnesic patients have shown that they perform well on various tests of procedural memory. Thus, as well as showing the preservation of existing skills, amnesic patients have also been shown to acquire new skills (for example, typing). The perceptual learning of amnesic patients — the ability to analyse and identify new visual forms — also seems to be normal. As amnesia is typically caused by damage to the limbic and frontal lobe structures, this evidence strongly indicates that memory systems with procedural characteristics are not mediated by structures in the limbic system.

By contrast, declarative memory is impaired in amnesic patients; they have immense problems retaining any form of factual or event-based knowledge. The limbic and frontal structures must therefore be directly concerned with declarative memory.

But what about the finer distinction between episodic and semantic declarative memory? It is well known that amnesic patients fail to acquire new vocabulary (the much studied amnesic patient HM has learned only six new words since his temporal operation in 1953) as well as failing to remember ongoing personal events. This seems to support the idea of a single declarative memory system. But it might be that episodic memory helps us to acquire new semantic memory. For example, your recall of the word 'hippocampus' might depend on your remembering reading this article.

Crucial evidence in this debate has centred on the nature of retrograde amnesia. Various researchers have claimed that retrograde amnesia is principally a failure of episodic memory. Although it is true that memory for episodes is severely impaired in amnesia there is now good evidence that the semantic type of memory is also affected. Thus, retrograde amnesia results in a failure to recognise people from the past and failure to define words and concepts introduced before the brain injury. Interestingly both these kinds of impairment become more severe for information acquired closer to the brain injury — a phenomenon

known as the 'temporal gradient'. Explanation of this phenomenon remains a mystery although some have suggested that it reflects the time-course of consolidation. The fact that temporal gradients can extend back 20 years makes this a difficult explanation to accept.

The roles of different structures

Limbic and frontal structures are best thought of as mediating declarative memory but what role do these different structures play in the process of memory? The hippocampus is divided into four 'fields' and one of these, CA1, receives inputs from all parts of the prefrontal cortex. Moreover, neurons in CA1 have a very high degree of interconnectivity. Memory for an event is thought to represent the integration of different aspects of information and, although there is no direct proof, it is thought that CA1 provides the anatomical basis and that other parts of the hippocampal formation 'fix' a particular event and then mediate storage of that event in the temporal cortex.

Far less is known about the function of diencephalic structures in memory but one idea is that these are involved in encoding information which enables the time-based characteristics of a memory to be established. The frontal lobes are generally thought to be involved in more 'strategic' aspects of memory. These strategic processes involve the operations used to determine what aspects of an event are encoded and those that control how memories are retrieved. Thus, studies of patients with frontal lesions have shown that their free recall of information is often poor but improves dramatically with prompting. Also their learning processes often reveal a failure to attend to the most relevant aspects of an event. There is also evidence that the pre-frontal cortex is crucial for determining the truth of memories because, quite often in frontal lobe damage, patients suffer from 'illusory memory' or 'confabulation' in which they tend to remember things that have not actually happened.

It should be obvious that the study of human memory is complex, not least because the problems can be approached on two fundamentally different levels — the biological and the psychological. There is progress in both these domains but it will be a long time before they come together to provide a single explanation of human memory.

Functional neuroimaging

In functional neuroimaging, a range of different techniques are used to indicate which parts of the brain are receiving most blood flow and are, hence, most active. Normal people are given memory tasks while undergoing functional neuroimaging. Such studies have confirmed much of what was already known about the neuroanatomical basis of memory but the evidence has also thrown up some new issues and addressed some controversies. First, neuroimaging studies have indicated different patterns of activation when people are answering questions about episodic and semantic memory – contradicting the evidence from amnesia. Studies have also highlighted the role of the frontal lobes of the brain in encoding and retrieval, although the original view that the left frontal region mediated encoding and the right mediated retrieval is now in some doubt. A more robust finding is the demonstration of frontal lobe involvement in the setting up of new links between objects (known as 'association') and the hippocampal involvement in detecting that an object is novel ('stimulus novelty'). Most interesting, perhaps, has been the discovery that a little-known parietal lobe structure, the right precuneus, is active during retrieval – a finding that is stimulating various ideas about how the various elements of a memory might be brought together during the process of retrieval.

Key references

Baddeley AD: *The Psychology of Human Memory*. Hove: Psychology Press; 1997.
Baddeley AD: *Essentials of Human Memory*. Hove: Psychology Press; 1999.
Parkin AJ: The hippocampus is the key. *Curr Biol* 1996, 6:1583-1585.
Parkin AJ: *Memory and Amnesia: an Introduction*. Hove: Psychology Press; 1997.
Parkin AJ: Novelty, association and the brain *Curr Biol* 1997, 7:R768-R769.
Squire LR: *Memory and Brain*. New York: Oxford University Press; 1987.

Address: Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Correspondence

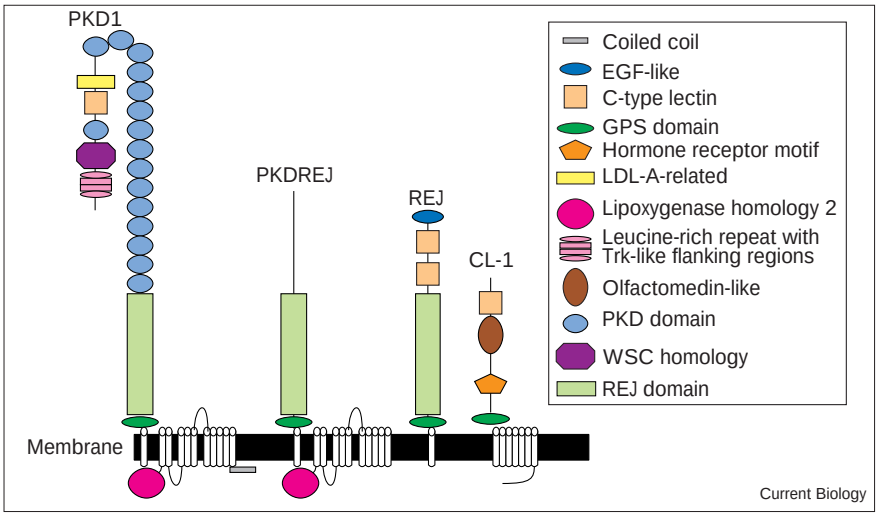
A latrophilin/CL-1-like GPS domain in polycystin-1

C.P. Ponting*, K. Hofmann† and P. Bork‡,§

Autosomal dominant polycystic kidney disease (ADPKD) is a common inherited disorder (incidence of 1 in 1,000), the cardinal manifestations of which are renal and liver cysts and intracranial aneurysm. The gene defective in the most common and severest form of ADPKD, *PKD1*, is broadly expressed and encodes a 4,302 amino acid plasma-membrane protein, polycystin-1 (PKD1) [1–3]. Despite detailed knowledge of the domain architecture for the majority of the PKD1 sequence (Figure 1), little is known of PKD1 function. Identification of *PKD1* mutations has been hampered by the presence of numerous *PKD1* homologues elsewhere on chromosome 16 [2] and *in vitro* expression of full-length PKD1 is yet to be reported.

Here, we report the identification of three previously unrecognised domains in PKD1 that are likely to possess distinct carbohydrate-, lipid- and protein-binding functions. These domains were identified using PSI-BLAST database searches [4,5] with an expect-value (E-value) inclusion threshold of $E < 0.01$. Independent evidence was provided using a generalised profile analysis method [6], in which the significances of findings were better than $p < 0.01$ in

Figure 1



A schematic representation of the domain architectures of PKD1, human PKDREJ, sea urchin REJ and human CL-1 (not to scale). Domain positions are taken, in part, from reference [18]. The large amino-terminal extracellular region of PKD1 has been recognised to contain several domain types. An extended region (the REJ domain) before the first putative transmembrane sequence is similar to part of REJ [19] and PKDREJ [22]; a report that this region contains four fibronectin type 3 domains is not substantiated by detailed sequence analysis [19]. PKD1 is proposed to contain 11 transmembrane (TM) regions; the carboxy-terminal six of these are contained in a region homologous to the product of a second ADPKD gene, *PKD2*, and pore-forming α-subunits of Ca^{2+} and Na^{+} channels [18,23]. By contrast to REJ, which contains only a single TM sequence [19], human PKDREJ also contains 11 predicted TM regions [21]). A predicted coiled-coil structure at the PKD1 carboxyl terminus is involved in its interaction with PKD2 [24]. The GPS, WSC and lipoxygenase homology 2 domains are described in the text and Supplementary material.

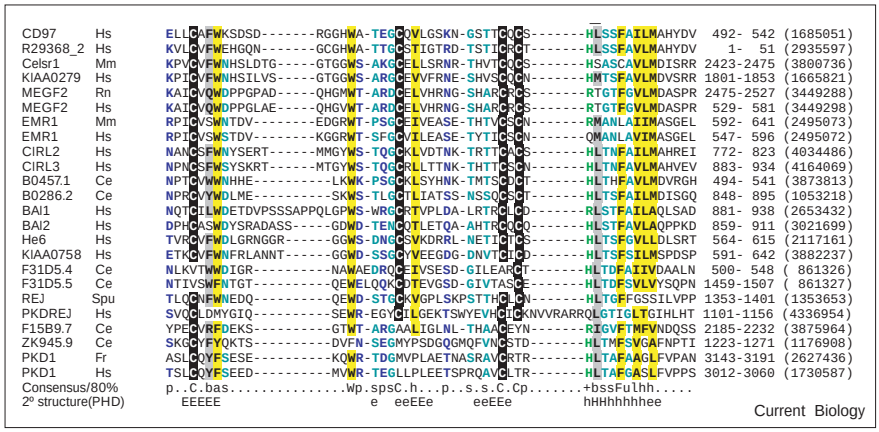
all cases. An extracellular region lying between the leucine-rich repeats (LRRs) and the amino-terminal PKD domain was found to contain a WSC domain that is also present in tandem in a fungal β-1,3-exoglucanase [7] and in *Saccharomyces cerevisiae* cell-wall integrity and stress-response component proteins [8]. An intracellular region between the first two transmembrane regions of PKD1 was found to contain a β-barrel LH2 domain homologous to a non-catalytic domain of lipoxygenases [9]. This domain has been proposed to facilitate binding of lipase and lipoxygenase substrates to the enzymes' active sites [9]. This indicates that the PKD1 LH2 domain is likely to possess a lipid-binding function. Multiple alignments and information relating to these domains may be found in the Supplementary material, and are available from the SMART database [10,11].

The third novel domain of PKD1 was found to be a member of a family that was recently identified in G-protein-coupled receptors (GPCRs) including the Ca^{2+}-independent α-latrotoxin receptor (latrophilin or CL-1) and in other membrane-associated proteins such as the sea urchin receptor for egg jelly protein, REJ [12,13]. A PSI-BLAST search with the sequence of this domain from REJ (residues 1,353–1,401) demonstrates significant similarity with PKD1 (residues 3,002–3,050) within one iteration using an *E*-value threshold of 0.01 (Figure 2). Latrophilin/CL-1 is known to be cleaved within this domain [14], so it has been termed the GPCR proteolytic site (GPS) domain [15].

GPS domains in human, *Fugu* and *Caenorhabditis elegans* PKD1 proteins contain only one of the two putative disulphide bridges found in CL-1; the second disulphide bridge appears to be missing in GPS domains of *C. elegans* proteins F31D5.4 and F31D5.5 (Figure 2). Others have pointed out that CL-1 and three other GPS-domain-containing proteins, CL-2, CL-3 and CD97, are all proteolytically processed endogenously [14,16] and that the amino acids surrounding the putative cleavage site are highly conserved among GPS domain sequences [12,15]. The cleavage site in the CL-1 GPS domain is highly conserved in PKD1 orthologues: rat CL-1 is cleaved at the Leu–Thr peptide bond of the sequence His^{836}-Leu-Thr-Asn^{839} [14], which corresponds to His^{3047}-Leu-Thr-Ala^{3050} in human PKD1. Consequently, we predict that PKD1 is endogenously cleaved to produce an amino-terminal chain of 3,048 amino acids and a carboxy-terminal chain of 1,255 amino acids. By analogy, the PKD1 amino-terminal chain is unlikely to be released after cleavage, as the two subunits of cleaved rat CL-1 have unexpectedly been shown to be strongly associated [14].

The exogenous ligand for CL-1, α-latrotoxin, requires both the GPS domain and the hormone receptor domain of CL-1 (the latter is also in secretin receptors) for binding [12]. Unfortunately, the endogenous ligand of CL-1 remains unknown. Given that homologous molecules often bind homologous receptors, however, the as-yet-unknown ligands for PKD1 and CL-1 may yet prove to be homologues. It may be significant that a single base-pair change resulting in substitution of Val^{3008} with Leu, just amino-terminal to the GPS domain, has been observed in a patient with ADPKD, although it is unproven whether this mutation causes ADPKD [17].

Figure 2



Multiple alignment of GPS domains containing the suggested cleavage site (marked by an arrow). Amino acid residues are coloured according to a 80% consensus (calculated using the program Consensus [25]: + indicates positively charged residues (H, K and R, green); – indicates negatively charged residues (D and E, green); a indicates aromatic residues (F, H, W and Y, highlighted in yellow); b indicates big residues (E, F, I, K, L, M, Q, R, W and Y, grey or yellow); c indicates charged residues (D, E, H, K and R, green); h indicates hydrophobic residues (A, C, F, H, I, L, M, V, W and Y, highlighted in yellow); l indicates aliphatic residues (I, L and V, highlighted in yellow); o indicates alcohol residues (S and T, pink), p indicates polar residues (D, E, H, K, N, Q, R, S and T, dark blue); s indicates small residues (A, C, S, T, D, N, V, G and P, light blue); u indicates tiny residues (A, G and S, light blue). Residues that are predicted to form disulphide bridges are shown as white-on-black. Predicted [26] secondary structures are indicated below the alignment (e/E, extended or β-strand structure; h/H, helix); lowercase letters represent predictions that have expected accuracies of > 72% and uppercase letters represent predictions that have expected accuracies of > 82%. Residue numbers and GenBank identifiers are shown following the alignment. Abbreviations: Ce, *Caenorhabditis elegans*; Fr, *Fugu rubripes*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Spu, *Strongylocentrotus purpuratus*; MEGF2, multiple EGF domain-containing protein 2; EMR1, EGF domain-containing mucin-like hormone receptor 1; CIRL, calcium-independent receptor of alpha-latrotoxin (latrophilin); BAI1, brain-specific angiogenesis inhibitor 1.

PKD1, CL-1 and REJ possess functional similarities that are likely to be due to their common GPS domains and transmembrane regions. Each of these molecules is suggested to mediate transmembrane influx of Ca^{2+} [13,18,19]. However, this function of PKD1 has been ascribed to its carboxy-terminal PKD2-like and voltage-gated Ca^{2+} channel-like region [17], which is lacking in CL-1 and REJ. It is notable that REJ, which contains a single transmembrane region, and a CL-1 truncation variant containing only its GPS-domain-containing extracellular region and the single amino-terminal transmembrane region, both support Ca^{2+} influx [13,20]. Consequently, we predict that the PKD1 transmembrane region 1 functions similarly in supporting Ca^{2+} influx.

In conclusion, the three previously unrecognised WSC, GPS and LH2 domains in human PKD1 are tentatively suggested to bind three different ligand types, namely carbohydrate, protein and lipid, respectively. Most of the amino-terminal region (amino acids 1–3,466) of human PKD1 has now been assigned domain homologues. However, mice heterozygous for a *PKD1* mutation that results in a protein that lacks its carboxy-terminal region (amino acids 3,497–4,293) and yet retains all the newly identified domains have been shown to possess kidney and pancreas defects [21]. This demonstrates that further experiments are required to determine the functions of different regions and domains of this large and complex molecule.

Supplementary material

Supplementary material including multiple alignments and descriptions of the three domains is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

C.P.P. is a National Research Council Senior Associate.

References

1. The International Kidney Disease Consortium: Polycystic kidney disease: the complete structure of the ***PKD1*** gene and its protein. *Cell* 1995, 81:289-298.
2. The European Polycystic Kidney Disease Consortium: The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 1995, 77:881-894.
3. Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan J: The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* 1995, 10:151-159.
4. Altschul SF, Madden TL, Shaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389-3402.
5. Altschul SF, Koonin EV: Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci* 1998, 23:444-447.
6. Bucher P, Karplus K, Moeri N, Hofmann, K: A flexible search technique based on generalized profiles. *Comput Chem* 1996, 20:3-24.
7. Cohen-Kupiec R, Broglie KE, Friesem D, Broglie RM, Chet I: Molecular characterization of a novel β-1,3-exoglucanase related to mycoparasitism of ***Trichoderma harzianum***. *Gene* 1999, 226:147-154.
8. Verna J, Lodder A, Lee K, Vagts A, Ballester R: A family of genes required for maintenance of cell wall integrity and for the stress response in ***Saccharomyces cerevisiae***. *Proc Natl Acad Sci USA* 1997, 94:13804-13809.
9. Gillmor SA, Villasenor A, Fletterick R, Sigal E, Browner MF: The structure of mammalian 15-lipoxygenase reveals similarity to the lipases and the determinants of substrate specificity. *Nat Struct Biol* 1997, 4:1003-1009.
10. Schultz J, Milpetz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signalling domains. *Proc Natl Acad Sci USA* 1998, 95:5857-5864.

11. SMART database: http://coot.embl-heidelberg.de/SMART/
12. Sugita S, Ichchenko K, Khvotchev M, Südhof TC: α-Latrotoxin receptor CIRL/latrophilin 1 (CL1) defines an unusual family of ubiquitous G-protein-linked receptors. *J Biol Chem* 1998, 273:32715-32724.
13. Krasnoperov V, Bittner MA, Holz RW, Chepurny O, Petrenko AG: Structural requirements for α-latrotoxin binding and α-latrotoxin-stimulated secretion. *J Biol Chem* 1999, 274:3590-3596.
14. Krasnoperov V, Bittner MA, Beavis R, Kuang Y, Salnikow KV, Chepurny OG, Little AR, Plotnikov AN, Wu D, Holz RW, *et al.*: α-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron* 1997, 18:925-937.
15. Ichtchenko K, Bittner MA, Krasnoperov V, Little AR, Chepurny O, Holz RW, Petrenko AG: A novel ubiquitously expressed α-latrotoxin receptor is a member of the CIRL family of G-protein-coupled receptors. *J Biol Chem* 1999, 274:5491-5498.
16. Gray JX, Haino M, Roth MJ, Maguire JE, Jensen PN, Yarme A, Stetler-Stevenson MA, Siebenlist U, Kelly K: CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. *J Immunol* 1996, 157:5438-5447.
17. Watnick TJ, Piontek KB, Cordal TM, Weber H, Gandolph MA, Qian F, Lens XM, Neumann HPH, Germino GG: An unusual pattern of mutation in the duplicated portion of ***PKD1*** is revealed by use of a novel strategy for mutation detection. *Hum Mol Genet* 1997, 6:1473-1481.
18. Sandford R, Sgotto B, Aparicio S, Brenner S, Vaudin M, Wilson RK, Chissoe S, Pepin K, Bateman A, Chothia C, *et al.*: Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. *Hum Mol Genet* 1997, 6:1483-1489.
19. Moy GW, Mendoza LM, Schultz JR, Swanson WJ, Glabe CG, Vacquier VD: The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *J Cell Biol* 1996, 133:809-817.
20. Trimmer JS, Schackmann RW, Vacquier VD: Monoclonal antibodies increase intracellular Ca^{2+} in sea urchin spermatozoa. *Proc Natl Acad Sci USA* 1986, 83:9055-9059.
21 Lu W, Peissel B, Babakhanlou H, Pavlova A, Geng, L, Fan X, Larson C, Brent G, Zhou J: Perinatal lethality with kidney and pancreas defects in mice with a targetted Pkd1 mutation. *Nat Genet* 1997, 17:179-181.
22. Hughes J, Ward CJ, Aspinwall R, Butler R, Harris PC: Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Hum Mol Genet* 1999, 8:543-549.
23. Mochizuki T, Wu G, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, Reynolds DM, Cai Y, Gavow PA, Pierides A, *et al.*: PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 1996, 272:1339-1342.
24. Tsiokas L, Kim E, Arnould T, Sukhatme VP, Walz G: Homo- and heterodimeric interactions between the gene products of ***PKD1*** and ***PKD2***. *Proc Natl Acad Sci USA* 1997, 94:6965-6970.
25. Brown N, Lai J: Consensus. http://www.bork.embl-heidelberg.de/Alignment/consensus.html
26. Rost B, Sander C: Prediction of protein structure at better than 70% accuracy. *J Mol Biol* 1993, 232:584-599.

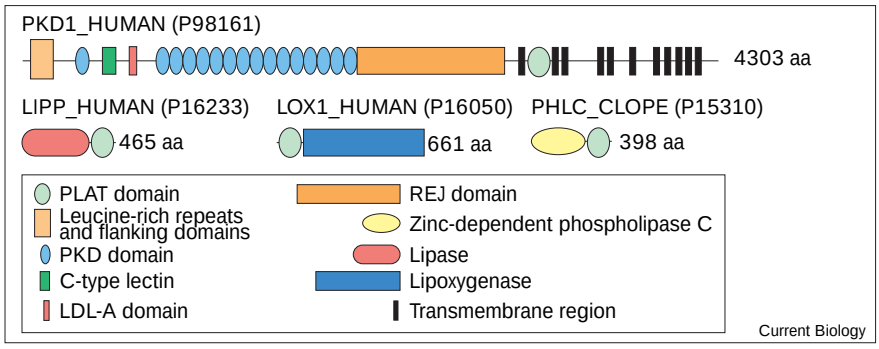
Addresses: *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda Maryland 20894, USA. †MEMOREC Stoffel GmbH, Stöckheimer Weg 1, D-50829 Köln, Germany. ‡European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany. §Max-Delbrück-Center for Molecular Medicine, D-13125 Berlin-Buch, Germany.
E-mail: ponting@ncbi.nlm.nih.gov

The PLAT domain: a new piece in the PKD1 puzzle

Alex Bateman* and Richard Sandford†

Autosomal dominant polycystic kidney disease (ADPKD) has a prevalence of 1 in 800 of the world's population and accounts for 10% of individuals who require renal replacement therapy, either dialysis or transplantation. Renal cyst formation occurs as part of a 'two-hit' process in which inactivation of both alleles of ADPKD genes leads to abnormalities of cell proliferation, apoptosis and differentiation [1]. Of ADPKD cases, 85% are due to mutations in the *PKD1* gene, which encodes a 4,302 amino acid protein, polycystin-1 (PKD1), of unknown function. Comparison of the PKD1 sequence with homologous sequences from mouse and *Fugu* predicts polycystin-1 to have a large extracellular region of 3,000 amino acid residues, a region containing 11 putative transmembrane segments and a short intracellular tail [2]. A well-defined extracellular domain structure is apparent; the presence of amino-terminal leucine-rich repeats, a C-type-lectin domain and multiple PKD repeats suggests a role in cell–cell or cell–matrix interactions (Figure 1) [3]. So far, no extracellular ligands of polycystin-1 have been identified. Of the intracellular regions of PKD1, functional properties have been defined only for the short 198 amino acid carboxy-terminal region, which contains a predicted coiled-coil domain. These include a direct interaction with the carboxyl terminus of the protein encoded by PKD2, polycystin-2 [4], activation of

Figure 1

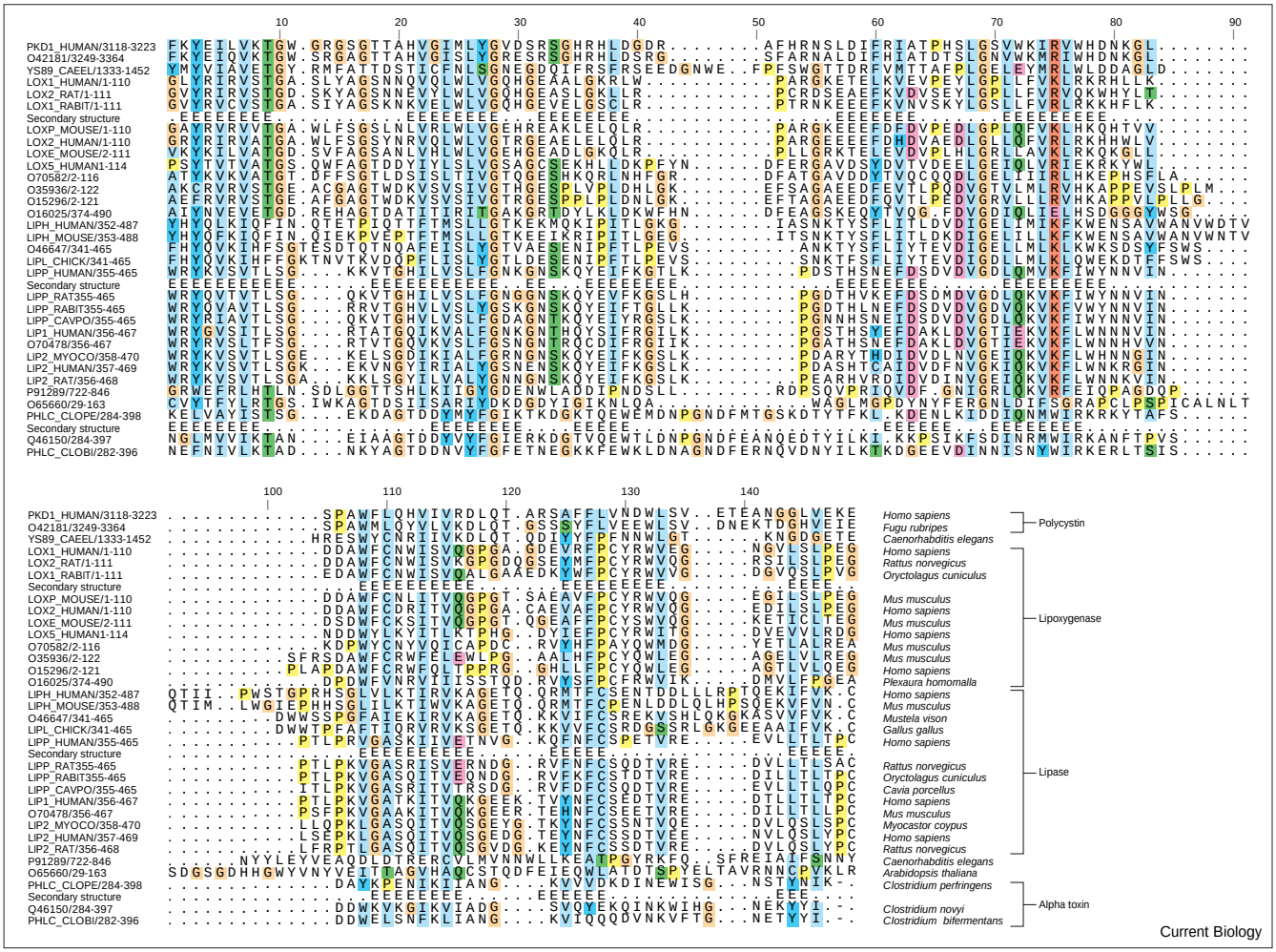


Schematic diagram of the domain organisation of PLAT domain proteins. Information and alignments for each of the domains can be found in the Pfam database [14,15] using the following identifiers or accession numbers: leucine-rich repeats, PF00560; C-type lectin, PF00059; LDL-A, PF00057; PKD domains, PF00801; PLAT domain, PF01477; lipoxygenase, PF00305; Zn^{2+}-dependent phospholipase C, PF00882; lipase, PF00151. Abbreviation: aa, amino acids.

transcription factor AP-1 [5] and activation of heterotrimeric G proteins [6]. The last of these occurs via a motif present in one of polycystin-1's most highly conserved regions. Polycystin-1 may therefore act as a cell-surface receptor or form part of a large membrane-associated complex that is capable of signaling by several different pathways to control cell proliferation and differentiation. To further aid the understanding of this enigmatic protein, we have surveyed all the intracellular regions for potential domains that may suggest novel functions and identify further avenues for experiment.

The current model of polycystin-1 topology suggests that there are four intracellular regions that are large enough to contain a discrete protein domain. These regions are between transmembrane (TM) helices TM1 and TM2 (residues 3,096–3,280), TM3 and TM4 (residues 3,344–3,558), TM5 and TM6 (residues 3,603–3,668) and the carboxy-terminal region after TM11 (residues 4,105–4,302). The high sequence conservation seen in these regions between the human and *Fugu* polycystin-1 suggest that they are functionally important. We have used each of these regions as a query for the sequence comparison program PSI-BLAST, using an expectation-value (*E*-value) threshold of 0.001 [7]. For each of the four intracellular regions, PSI-BLAST returned the known PKD1 orthologues from human, mouse and Fugu. Only one region returned significant matches with PSI-BLAST. The first intracellular region between TM1 and TM2, which represents the most strongly conserved sequence region of PKD1 between human and *Fugu* [2], was found to match to 67 other sequences in SWISS-PROT version 37 and TrEMBL version 9 [8]. These sequences include mammalian lipoxygenases, triacylglycerol lipase and lipoprotein lipase. The common feature found in

Figure 2



A multiple sequence alignment of PLAT domains. The colouring scheme is that used in ClustalX. The secondary structures for the known structures are shown below the sequences; E denotes β-strand.

all these alignments was topological and sequence similarity to a β-sandwich domain. We call this new protein domain the PLAT domain (after polycystin-1, lipoxygenase and alpha toxin). The copy of the PLAT domain found in polycystin-1 therefore identifies an important new region of the protein.

The three-dimensional structure of the PLAT domain is known for human pancreatic lipase [9], rabbit 15-lipoxygenase [10] and alpha toxin from *Clostridium perfringens* [11]. The domain is a β-sandwich composed of two sheets of four strands each. The sequence relationship of the alpha toxin to polycystin-1 can be demonstrated by using the sequence of the PLAT domain from the known structure as a query for PSI-BLAST. The search essentially converges to the same family, including the polycystin-1 PLAT domain. Soybean lipoxygenase L-1 [12] contains a domain structurally related to the PLAT domains. It is more distant in sequence to the rest of the family; PSI-BLAST is able to find relationships to the rest of the family for only a few sequences. Although structural similarities were noticed between these structures, it was not suggested that they share a common ancestor [11].

The most highly conserved regions in the alignment of known PLAT domains (Figure 2) coincide with the β-strands. Most of the highly conserved residues are buried residues. An exception to this is a surface lysine or arginine that occurs on the surface of the fifth β-strand of all the eukaryotic PLAT domains. In pancreatic lipase, the lysine in this position forms a salt bridge with the procolipase protein. The conservation of a charged surface residue may indicate the location of a conserved ligand-binding site within the PLAT domain.

The importance of PLAT domains is underlined by mutations that lead to human disease. Mutations in lipoprotein lipase lead to chylomicronaemia and mutations in triacylglycerol lipase lead to hepatic lipase deficiency [13]. In pancreatic lipase the PLAT domain is involved in binding to the procolipase protein. This interaction is required to bring the enzymatic active site of the lipase into close contact with its lipid substrate. In 15-lipoxygenase, a protein composed of an amino-terminal β-sandwich (PLAT) and a carboxy-terminal catalytic domain, the PLAT domain may function to localise the enzyme near its membrane or lipoprotein sequestered substrates, by analogy to the lipase–procolipase protein–protein interaction. It is also possible that the PLAT domain of 5-lipoxygenase, another member of the mammalian lipoxygenase family, mediates an interaction with the 5-lipoxygenase activating protein (FLAP), an integral membrane protein. For alpha toxin and plant lipoxygenases, it has been suggested that the PLAT domain interacts directly with the membrane in a Ca^{2+} dependent manner. Although a Ca^{2+}-binding region has been predicted for alpha toxin from crystallographic data and similarity to eukaryotic calcium-binding C2 domains [11], the conserved residues that form this region are not present in the PLAT domains identified in polycystin-1 or pancreatic lipase. PLAT domains may therefore be involved in protein–protein and protein–lipid interactions. The presence of the PLAT domain in the first cytoplasmic loop of polycystin-1 suggests that this region is important in mediating interactions with other membrane protein(s) involved in polycystin-1 function.

References

1. Qian F, Watnick T, Onuchic L, Germino G: The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell* 1996, 87:979-987.
2. Sandford R, Sgotto B, Aparicio S, Brenner S, Vaudin M, Wilson RK, Chissoe S, Bateman A, Chothia C, Hughes J, *et al.*: Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. *Proc Natl Acad Sci USA* 1997, 6:1483-1489.
3. Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan JL, Gamble V, Harris PC: The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* 1995, 10:151-160.
4. Qian F, Germino F, Cai Y, Zhang X, Somlo S, Germino G: PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat Genet* 1997, 16:179-183.
5. Arnould T, Kim E, Tsiokas L, Jochimsen F, Gruning W, Chang J, Walz G: The polycystic kidney disease 1 gene product mediates protein kinase C alpha-dependent and c-Jun N-terminal kinase-dependent activation of the transcription factor AP-1. *J Biol Chem* 1998, 273:6013-6018.
6. Parnell S, Magenheimer B, Maser R, Rankin C, Smine A, Okamoto T, Calvet J: The polycystic kidney disease-1 protein, polycystin-1, binds and activates heterotrimeric G-proteins in vitro. *Biochem Biophys Res Commun* 1998, 251:625-631.
7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389-3402.
8. Bairoch A, Apweiler R: The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res* 1999, 27:49-54.
9. van Tilbeurgh H, Egloff MP, Martinez C, Rugani N, Verger R, Cambillau C: Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature* 1993, 362:814-820.
10. Gillmor SA, Villasenor A, Fletterick R, Sigal E, Browner MF: The structure of mammalian 15-lipoxygenase reveals similarity to the lipases and the determinants of substrate specificity. *Nat Struct Biol* 1998, 4:1003-1009.
11. Naylor CE, Eaton JT, Howells A, Justin N, Moss DS, Titball RW, Basak AK: Structure of the key toxin in gas gangrene. *Nat Struct Biol* 1998, 5:738-746.
12. Minor W, Steczko J, Stec B, Otwinowski Z, Bolin JT, Walter R, Axelrod B: Crystal structure of soybean lipoxygenase L-1 at 1.4 Å resolution. *Biochem* 1996, 35:10687-10701.
13. Hegele R, Tu L, Connelly P: Human hepatic lipase mutations and polymorphisms. *Hum Mutat* 1992, 1:320-324.
14. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer ELL: Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res* 1999, 27:260-262.
15. Pfam: Protein families database of alignments and HMMs. http://www.sanger.ac.uk/Software/Pfam/

Addresses: *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. †Cambridge Institute for Medical Research, Addenbrooke's Hospital, Cambridge CB2 2XY, UK.
E-mail: agb@sanger.ac.uk

Supplementary material

A latrophilin/CL-1-like GPS domain in polycystin-1

C.P. Ponting, K. Hofmann and P. Bork

Current Biology 26 August 1999, 9:R585–R588

The WSC domain

A PSI-BLAST [S1,S2] search of the region of PKD1 lying between the leucine-rich repeat flanking regions and the first PKD domain of PKD1 (amino acid residues 181–272) using an expect value (*E*-value) inclusion threshold of $E < 0.01$ revealed significant similarity to two regions of *Trichoderma harzianum* β-1,3-exoglucanase [S3] ($E = 1 \times 10^{-5}$ and 6×10^{-3}) by iteration 1, and nine additional proteins (Figure S1) by convergence after five iterations. Independent evidence for these domain homologues and those discussed below was provided by generalised profile analysis [S4], in which the significance was better than $p < 0.01$ in all cases.

We term these homologues WSC domains, after *S. cerevisiae* WSC1 (cell-wall integrity and stress-response component 1; also called Slg1 and Hcs77), WSC2, WSC3 and WSC4 proteins, which each contain a single such domain [S5]. *S. cerevisiae* WSC1–3 proteins are localised to the plasma membrane and function upstream of the PKC1–MPK1 pathway [S5–S7]. The WSC domains in these proteins are predicted to be extracellular and are amino-terminal to serine- and threonine-rich regions, a single predicted transmembrane sequence and divergent carboxy-terminal cytoplasmic sequences. The functions of these WSC domains are unknown but the WSC1–3 proteins are suggested to act as sensors of environmental stress [S5–S7], and WSC4 (also called Yhc8p) is implicated in protein translocation to the endoplasmic reticulum [S8].

Tandem WSC domains are also found in a hypothetical human protein (KIAA0523) that appears to be a member of a novel sulphotransferase subfamily, and in the *T. harzianum* β-1,3-exoglucanase [S3]. The latter occurrences indicate that the PKD1 WSC domain, and its homologues in general, might possess a carbohydrate-binding role. This would complement the predicted protein-, lipoprotein- and carbohydrate-binding functions conferred by the PKD, LDLα and C-type lectin domains of PKD1.

The LH2 domain

The cytoplasmic region between the first and second transmembrane regions of PKD1 was found to contain a β-barrel domain, homologous to a non-catalytic domain of

Figure S1

Multiple sequence alignments of WSC domains. Amino acid residues are coloured according to a 80% consensus (calculated using the program Consensus [S13]: + indicates positively charged residues (H, K and R, green); – indicates negatively charged residues (D and E, green); a indicates aromatic residues (F, H, W and Y, highlighted in yellow); b indicates big residues (E, F, I, K, L, M, Q, R, W and Y, grey or yellow); c indicates charged residues (D, E, H, K and R, green); h indicates hydrophobic residues (A, C, F, H, I, L, M, V, W and Y, highlighted in yellow); l indicates aliphatic residues (I, L and V, highlighted in yellow); o indicates alcohol residues (S and T, pink), p indicates polar residues (D, E, H, K, N, Q, R, S and T, dark blue); s indicates small residues (A, C, S, T, D, N, V, G and P, light blue); u indicates tiny residues (A, G and S, light blue). Residues that are predicted to form disulphide bridges are shown as white-on-black. Predicted [S14] secondary structures are indicated below the alignment (e/E, extended or β-strand structure; h/H, helix); lowercase letters represent predictions that have expected accuracies of > 72% and uppercase letters represent predictions that have expected accuracies of > 82%. Residue numbers and GenBank identifiers are shown following the alignment. Abbreviations: Fr, *Fugu rubripes*; Hp, *Hansenula polymorpha*; Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Th, *Trichoderma harzianum*; β-1,3exoGluc, β-1,3 exoglucanase.

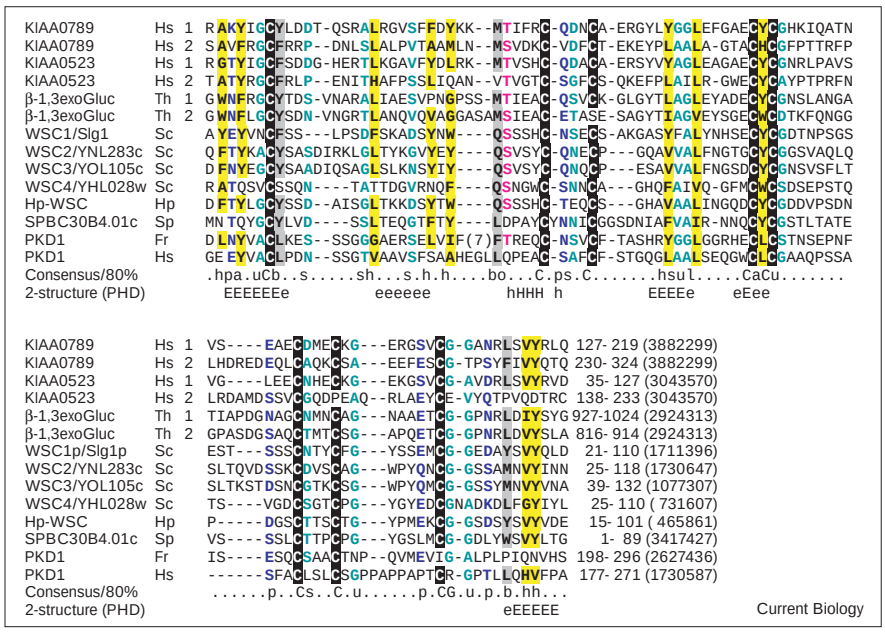


Figure S2

Multiple alignment of LH2 domains, coloured according to the consensus scheme of Figure S1. Predicted [S14] secondary structures are indicated below the alignment (e/E, extended or β-strand structure; h/H, helix); lowercase letters represent predictions that have expected accuracies of > 72% and uppercase letters represent predictions that have expected accuracies of > 82%. Residue numbers and GenBank identifiers are shown following the alignment. Abbreviations: Ce, *Caenorhabditis elegans*; Fr, *Fugu rubripes*; Gm, *Glycine max* (soybean); Hs, *Homo sapiens*; Mm, *Mus musculus*; Oc, *Oryctolagus cuniculus* (rabbit); Rn, *Rattus norvegicus*; LOX1, arachidonate 15-lipoxygenase; LIP2, pancreatic lipase-related protein 2.

lipoxygenases [S9]. A PSI-BLAST search [S1,S2] using PKD1 residues 3,096–3,280 as query revealed significant similarity ($E < 10^{-6}$) by iteration 1 to the amino-terminal domain of mammalian arachidonate 5-lipoxygenases (Figure S2). This domain has an eight-stranded β-barrel fold [S9] and is highly similar in structure to the carboxy-terminal domains of mammalian lipases [S10,S11] and the amino-terminal domains of plant 15-lipoxygenases [S12] (Figure S2). We term these domains lipoxygenase homology 2 (LH2) domains. The current model for the function of LH2 domains is that they facilitate binding of lipase and lipoxygenase substrates to the enzymes' active sites [S9]. Thus the presence of this domain in the first cytoplasmic loop of PKD1 suggests a role in lipid-mediated modulation of PKD1 function.

Supplementary references

S1. Altschul SF, Madden TL, Shaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389-3402.

S2. Altschul SF, Koonin EV: Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci* 1998, 23:444-447.

S3. Cohen-Kupiec R, Broglie KE, Friesem D, Broglie RM, Chet I: Molecular characterization of a novel β-1,3-exoglucanase related to mycoparasitism of ***Trichoderma harzianum***. *Gene* 1999, 226:147-154.

S4. Bucher P, Karplus K, Moeri N, Hofmann, K: A flexible search technique based on generalized profiles. *Comput Chem* 1996, 20:3-24.

S5. Verna J, Lodder A, Lee K, Vagts A, Ballester R: A family of genes required for maintenance of cell wall integrity and for the stress response in ***Saccharomyces cerevisiae***. *Proc Natl Acad Sci USA* 1997, 94:13804-13809.

S6. Gray JV, Ogas JP, Kamada Y, Stone M, Levin DE, Herskowitz I: A role for the Pkc1 MAP kinase pathway of ***Saccharomyces cerevisiae*** in bud emergence and identification of a putative upstream regulator. *EMBO J* 1997, 16:4924-4937.

S7. Jacoby JJ, Nilius SM, Heinisch JJ: A screen for upstream components of the yeast protein kinase C signal transduction pathway identifies the product of the ***SLG1*** gene. *Mol Gen Genet* 1998, 258:148-155.

S8. Mamoun CB, Beckerich J-M, Gaillardin C, Kepes F: Disruption of ***YHC8***, a member of the ***TSR1*** gene family, reveals its direct involvement in yeast protein translocation. *J Biol Chem* 1999, 274:11296-11302.

S9. Gillmor SA, Villasenor A, Fletterick R, Sigal E, Browner MF: The structure of mammalian 15-lipoxygenase reveals similarity to the lipases and the determinants of substrate specificity. *Nat Struct Biol* 1997, 4:1003-1009.

S10. Roussel A, Yang Y, Ferrato F, Verger R, Cambillau C, Lowe M: Structure and activity of rat pancreatic lipase-related protein 2. *J Biol Chem* 1998, 273:32121-32128.

S11. Winkler FW, D'Arcy A, Hunziker W: Structure of human pancreatic lipase. *Nature* 1990, 343:771-774.

S12. Boyington JC, Gaffney BJ, Amzel LM: The three-dimensional structure of an arachidonic acid 15-lipoxygenase. *Science* 1993, 260:1482-1486.

S13. Brown N, Lai J: Consensus. http://www.bork.embl-heidelberg.de/Alignment/consensus.html

S14. Rost B, Sander C: Prediction of protein structure at better than 70% accuracy. *J Mol Biol* 1993, 232:584-599.