The EMBO Journal (2012) 31, 1364–1378 | © 2012 European Molecular Biology Organization | Some Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

EMBO open

A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis

Demet Araç[1,2,3,*], Antony A Boucard[2], Marc F Bolliger[2], Jenna Nguyen[1,2,3], S Michael Soltis[4], Thomas C Südhof[1,2] and Axel T Brunger[1,2,3,*]

[1]Howard Hughes Medical Institute, Stanford, CA, USA, [2]Department of Molecular and Cellular Physiology, Stanford, CA, USA, [3]Departments of Neurology and Neurological Sciences, Photon Science, and Structural Biology, Stanford University, Stanford, CA, USA and [4]Macromolecular Crystallographic Group, The Stanford Synchrotron Radiation Lightsource, SLAC, Stanford University, Stanford, CA, USA

The G protein-coupled receptor (GPCR) Proteolysis Site (GPS) of cell-adhesion GPCRs and polycystic kidney disease (PKD) proteins constitutes a highly conserved autoproteolysis sequence, but its catalytic mechanism remains unknown. Here, we show that unexpectedly the ~40-residue GPS motif represents an integral part of a much larger ~320-residue domain that we termed GPCR-Autoproteolysis INducing (GAIN) domain. Crystal structures of GAIN domains from two distantly related cell-adhesion GPCRs revealed a conserved novel fold in which the GPS motif forms five β-strands that are tightly integrated into the overall GAIN domain. The GAIN domain is evolutionarily conserved from tetrahymena to mammals, is the only extracellular domain shared by all human cell-adhesion GPCRs and PKD proteins, and is the locus of multiple human disease mutations. Functionally, the GAIN domain is both necessary and sufficient for autoproteolysis, suggesting an autoproteolytic mechanism whereby the overall GAIN domain fine-tunes the chemical environment in the GPS to catalyse peptide bond hydrolysis. Thus, the GAIN domain embodies a unique, evolutionarily ancient and widespread autoproteolytic fold whose function is likely relevant for GPCR signalling and for multiple human diseases.

The EMBO Journal (2012) **31,** 1364–1378. doi:10.1038/emboj.2012.26; Published online 14 February 2012
Subject Categories: signal transduction; structural biology
Keywords: adhesion GPCRs; autoproteolysis; latrotoxin; polycystic kidney disease-1; synapse

Introduction

Autoproteolysis is a widely used intramolecular post-translational process that often converts inactive protein precursors into biologically active mature proteins. For instance, the hedgehog pathway, one of the fundamental signal transduction pathways in animal development, is activated only if the signalling domain is cleaved off from the autoproteolytic domain of hedgehog (Lee *et al*, 1994). Among autoproteolytic domains/motifs described in various proteins (including hedgehog, N-terminal nucleophile (Ntn) hydrolases, nucleoporin98, protein inteins, and pyruvoyl enzymes; Perler *et al*, 1997), the G protein-coupled receptor (GPCR) Proteolysis Site (GPS) is one of the few autoproteolysis-mediating domains/motifs found in membrane proteins. The GPS motif is conserved in all 33 human cell-adhesion GPCRs and all 5 human polycystic kidney disease (PKD) proteins (Krasnoperov *et al*, 1997; Hughes *et al*, 1999; Ichtchenko *et al*, 1999; Chang *et al*, 2003; Lin *et al*, 2004); representing one of the most widespread and conserved autoproteolytic domains/motifs. Recent studies indicate emerging physiological and pathological roles for these GPS-containing membrane proteins in cancer (Kan *et al*, 2010), PKD (Qian *et al*, 2002), brain development (Piao *et al*, 2004), myelination of neurons (Monk *et al*, 2009), central nervous system angiogenesis (Kuhnert *et al*, 2010), attention-deficit/hyperactivity disorder (Arcos-Burgos *et al*, 2010), cell polarity in neural development (Chae *et al*, 1999; Usui *et al*, 1999; Shima *et al*, 2004; Langenhan *et al*, 2009), Usher syndrome 2 (Ebermann *et al*, 2009), mouse male infertility (Davies *et al*, 2004), and immunity (see Yona *et al*, 2008 for a review), but the mechanism of GPS-mediated autoproteolysis remains unclear.

Cell-adhesion GPCRs are characterized by long N-terminal extracellular sequences comprising multiple domains, and constitute a poorly studied class of GPCRs that is the second largest family of GPCRs in humans (Fredriksson *et al*, 2003). Most cell-adhesion GPCRs are orphan receptors with no known ligands. They are found in animals, protozoa, and alveolates, predating many other important signalling molecules. The CIRL/Latrophilins (here abbreviated as CLs, CL1–3) and brain angiogenesis inhibitors (BAIs, BAI1–3) are two distant subfamilies of cell-adhesion GPCRs that both are primarily expressed in neurons (Krasnoperov *et al*, 1997; Lelianova *et al*, 1997; Nishimori *et al*, 1997; Sugita *et al*, 1999). CL1 is a calcium-independent receptor for α-latrotoxin, a black widow spider toxin that triggers massive neurotransmitter release from neurons and neuroendocrine cells (Krasnoperov *et al*, 1997; Lelianova *et al*, 1997; Sugita *et al*, 1998; Sudhof, 2001; Deak *et al*, 2009). BAIs initially received their name because an extracellular fragment of BAI1 inhibited neovascularization in experimental angiogenesis (Nishimori *et al*, 1997; Kaur *et al*, 2005). Recently, C1q-like proteins, a family of secreted proteins, were found to bind to BAI3 and to regulate synapse formation and/or maintenance (Bolliger *et al*, 2011). CLs and BAIs are now known to be two of the most mutated genes in human cancers (especially in squamous lung carcinoma and lung adenocarcinoma) and may act as tumour suppressor genes (Wood *et al*, 2007; Jones *et al*, 2008; Parsons *et al*, 2008; Kan *et al*, 2010). On the other

*Corresponding authors. D Araç or AT Brunger, Howard Hughes Medical Institute, Stanford University, J.H. Clark Center, E300-C, 318 Campus Drive, Stanford, CA 94305-5432, USA. Tel.: +650 736 1714; Fax: +650 736 1961; E-mail: demeta@stanford.edu or Tel.: +1 650 736 1031; Fax: +1 650 745 1463; E-mail: brunger@stanford.edu

Received: 26 September 2011; accepted: 16 January 2012; published online: 14 February 2012

hand, PKD family proteins are not GPCRs, but they contain large extracellular sequences, multiple transmembrane regions and a GPS motif similar to cell-adhesion GPCRs (Ponting *et al*, 1999). PKD1 is a 500-kDa 11-pass membrane protein with a large N-terminal extracellular region that is cleaved at the GPS motif (Qian *et al*, 2002). It is mutated in 90% of the autosomal dominant PKD (ADPKD) patients (Rossetti *et al*, 2001). A mouse mutant with uncleavable PKD1 exhibits abnormal kidney development (Yu *et al*, 2007).

In cell-adhesion GPCRs and PKD proteins, the GPS motif is always located at the end of their long N-terminal extracellular regions, immediately before the first transmembrane helix of the respective protein (Krasnoperov *et al*, 1997; Sugita *et al*, 1998; Ichtchenko *et al*, 1999; Chang *et al*, 2003; Lin *et al*, 2004; Figure 1A). The GPS motif is a conserved sequence of ~40 amino acids containing canonical cysteine and tryptophan residues. The GPS is referred to as a 'domain' in the Pfam database although no three-dimensional structure is available, and it is unclear whether it folds autonomously. Most, if not all, cell-adhesion GPCRs undergo autoproteolysis in the GPS between a conserved aliphatic residue (usually a leucine) and a threonine, serine, or cysteine residue (L↓T/S/C) (Krasnoperov *et al*, 1997, 2002b; Nechiporuk *et al*, 2001; Lin *et al*, 2004). Although previous studies suggested that the GPS motif on its own promotes autoproteolysis (Krasnoperov *et al*, 2002a), it is hard to imagine how such a short sequence could fold into a catalytically active autoproteolytic domain given the stability of peptide bonds. Interestingly, in cell-adhesion GPCRs the GPS motif appears to be preceded by a conserved so-called 'stalk' region, which in turn is often preceded by a Hormone Receptor (HormR) domain, but the role of these sequences remains unclear.

To shed light on the GPS-mediated autoproteolysis of cell-adhesion GPCRs and PKD proteins, and to explore the significance of the stalk and GPS sequences and their relation to the HormR domain, we performed structural and functional studies of fragments containing these regions. Here, we show that the GPS motif is not an autonomously folded unit but rather part of a much larger evolutionarily conserved domain that we refer to as the GPCR-Autoproteolysis INducing (GAIN) domain. We obtained crystal structures of fragments from two distantly related cell-adhesion GPCRs (CL1 and BAI3) that include the GAIN domain and, additionally, the HormR domain. These structures revealed a novel fold with a high degree of structural conservation of even distant GAIN domains. Functionally, the entire GAIN domain is required and sufficient for autoproteolysis. Sequence analyses showed that the GAIN domain is an ancient domain that exists in primitive ancestor organisms, and is conserved in all cell-adhesion GPCRs and all PKD1-related proteins. Our structures of the cleaved CL1 and uncleaved BAI3 GAIN domains, in combination with mutagenesis experiments, reveal the structural determinants for autoproteolysis and suggest a plausible mechanism of cleavage. Taken together, our data describe a newly identified, evolutionarily conserved domain that is present in 38 different human proteins and functions as an autoproteolytic fold.

Results

Structures of CL1 and BAI3

To test the significance of the sequence arrangements of the HormR domain, the stalk region and the GPS motif, and to explore the mechanism of the GPS-mediated autoproteolysis, we determined the structures of fragments containing these regions from two distantly related cell-adhesion GPCRs, CL1 and BAI3 (Figure 1A). Experimental phases for the CL1 and BAI3 structures were independently obtained from sulphur and iodine single anomalous diffraction (SAD) data at 1.9 and 2.3 Å resolution, respectively (Supplementary Table S1, Materials and methods). Both structures reveal two domains (Figure 1B and C; Supplementary Figure S1A and B): at the N-terminus is the smaller ~70-residue HormR domain, and at the C-terminus is a larger ~320-residue domain that we refer to as the GAIN domain; both structures are similar, but with some significant differences (see below and Supplementary Figure S1C). Surprisingly, the GPS motif is an integral part of the GAIN domain (comprising the last five β-strands). Thus, the GPS motif itself does not constitute an autonomously folded unit, but rather forms a single folded domain together with the so-called 'stalk'.

DALI (Holm and Sander, 1993) searches (as of January 2012) indicated that the GAIN domain is a newly identified fold, and that our CL1 and BAI3 crystal structures represent the first high-resolution structures of this domain. The N-terminal subdomain A of the GAIN domain is composed of six α-helices, while the C-terminal subdomain B consists of a twisted β-sandwich including 13 β-strands and 2 small α-helices. The high similarity of the GAIN domain structures (root-mean square deviation (r.m.s.d.) = 1.38 Å; Supplementary Figure S1C) from two distantly related cell-adhesion GPCRs (24% sequence identity) suggests structural conservation of this novel domain across the entire family of cell-adhesion GPCRs.

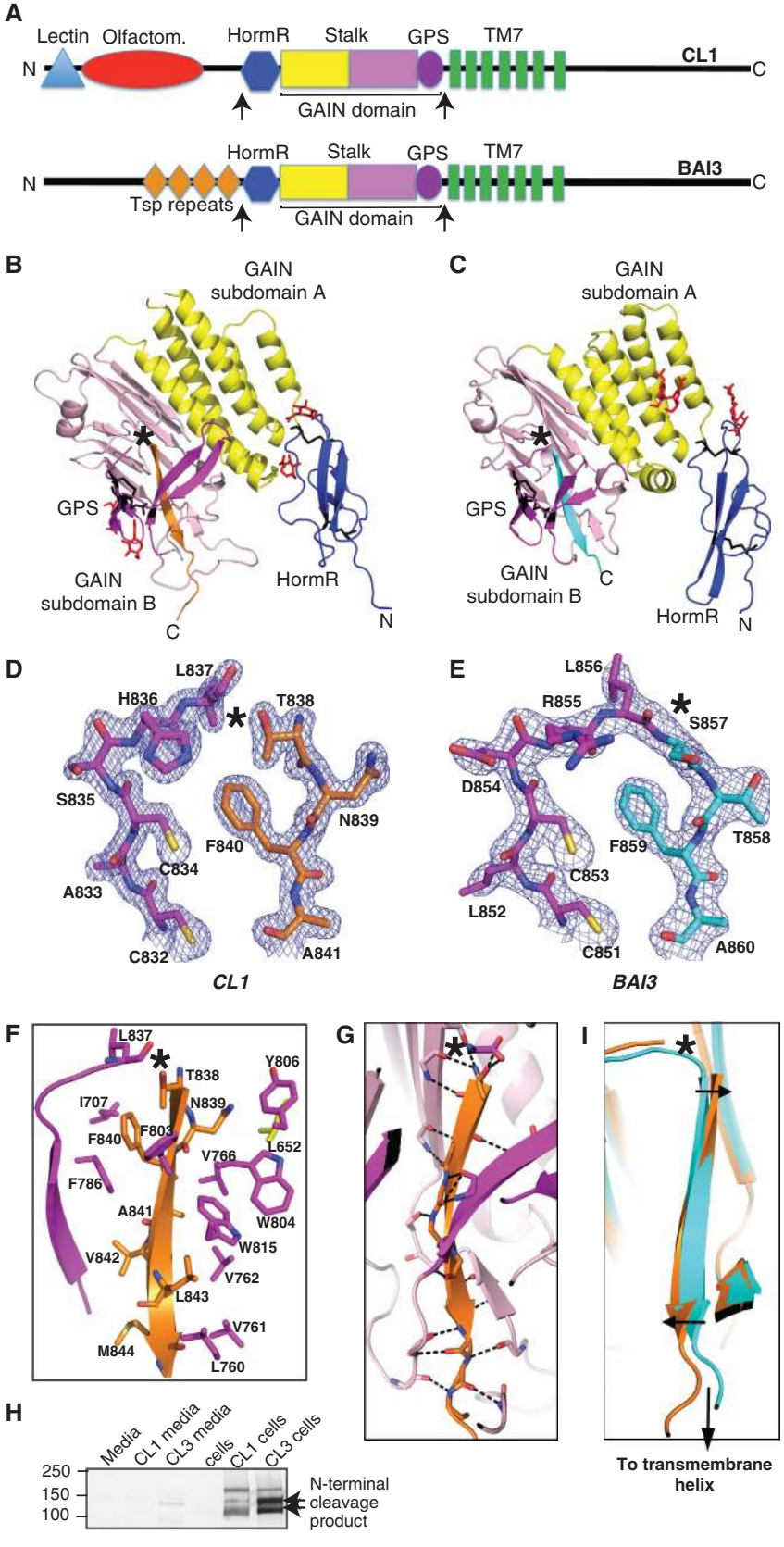
The GAIN domain functions as an autoproteolytic domain

To determine whether the GAIN domain is the minimally required region for autoproteolysis, we designed deletion mutants of full-length CL1: CL1-GPS which lacks the HormR and GAIN domains except for the GPS motif; CL1-HormR_GPS in which all residues of the GAIN domain except for the GPS motif are deleted; CL1-ΔSubA in which only subdomain A of the GAIN domain is deleted; and CL1-ΔHormR in which only the HormR domain is deleted (Figure 2A). We expressed the deletion mutants in HEK293 cells, and analysed their autoproteolysis by immunoblotting of cell lysates, and their plasma membrane transport by indirect immunofluorescence of non-permeabilized cells. Our results show that the entire GAIN domain with both subdomains A and B is required for autoproteolysis (Figure 2B and C; Supplementary Figure S2A). Consistent with the identification of the GPS motif as an integral part of the GAIN domain, the GPS motif is not functional by itself (Figure 2B and C), presumably because it is unfolded when expressed without the other GAIN domain sequences. A protein fragment composed of only the GAIN domain of CL1 underwent autoproteolysis, demonstrating that the GAIN domain is not only necessary but also sufficient for autoproteolysis in CL1 (Figures 2D and 5D; Supplementary Figure S2B). Indeed, the electron density map of the crystal structure shows that the crystallized HormR-GAIN fragment of CL1 is cleaved at the expected position (Figure 1D; Krasnoperov *et al*, 1997). In addition, the GAIN domain of GPR56 (a different cell-adhesion GPCR, mutations of which cause bilateral frontoparietal polymicrogyria (BFPP); Piao

et al, 2004) is sufficient for autoproteolysis (Figure 2E and F; Supplementary Figure S2C), suggesting that the GAIN domain mediates autoproteolysis in other cell-adhesion GPCRs as well.

The GAIN domain is an ancient conserved fold

To study the evolution of GAIN domains, we searched for the presence of GAIN domains in the genomes of primitive evolutionarily distant organisms, and identified genes encod-



ing such domains. Strikingly, structure-based BLAST searches revealed that the genome of *Dictyostelium discoideum*, an evolutionarily ancient organism that arose before animals emerged, encodes a GAIN domain (Figure 3A), although it lacks hedgehog autoproteolytic domains as well as signalling and adhesion domains found in higher eukaryotes, such as HormR, Wnt-signalling components, cadherins, immunoglobulins, and integrins (Eichinger *et al*, 2005). Moreover, a unicellular ciliate, *Tetrahymena thermophila*, encodes more than eight GAIN domains. Thus, the GAIN domain is among the most ancient GPCR-related extracellular domains. Computational analysis (McGuffin *et al*, 2000; Pei *et al*, 2007) of ancient GAIN domain sequences showed that their predicted secondary structure pattern is identical to that of the CL1 and BAI3 GAIN domain (Figure 3B; Supplementary Table S2). Moreover, their GPS motifs including the putative cleavage sites are well conserved (Figure 3A), whereas those of the autoproteolytic domain of nucleoporin98 (which also exists in ciliates) are not. Furthermore, sequence alignments of cell-adhesion GPCRs and PKD proteins revealed that all cell-adhesion GPCRs and PKD proteins are predicted to have a GAIN domain with identical predicted secondary structure patterns to that of the CL1 and BAI3 crystal structures (Figure 3B; Supplementary Table S3). Apart from the shared GAIN domain, PKD proteins are unrelated to cell-adhesion GPCRs but also undergo autoproteolysis (Ponting *et al*, 1999), indicating that the GAIN domain has been utilized by at least two different membrane protein families during evolution. These results suggest that the GAIN domain evolved prior to the evolutionary emergence of animals, that it is one of the first functional autoproteolytic domains, and thus represents an ancient feature that is not directly connected to multicellularity.

Autoproteolysis does not lead to the dissociation of the cleavage products

In general, autoproteolysis domains produce two protein fragments that then dissociate (e.g., hedgehog and protein inteins). Strikingly, the cleaved β-strand in the CL1 structure is well ordered with full occupancy and thermal factors comparable to the surrounding protein, indicating no sign of dissociation (Figure 1D). The cleaved β-strand is tightly bound to the rest of the GAIN domain involving an extensive network of conserved interstrand hydrogen bonds and primarily hydrophobic side-chain interactions (Figure 1F and G). Thus, we predict that dissociation of the cleaved β-strand would be energetically unfavourable, and would conceivably lead to unfolding of the remaining part of the GAIN domain. Consistent with this structural prediction, the N-terminal cleavage product of full-length CL1 and CL3 in transfected HEK293 cells remained bound to the rest of the protein on the cell surface, and was not secreted into the media (Figure 1H). Thus, the GAIN domain forms a tightly associated heterodimer upon proteolysis.

Different from CL1, the electron density of the crystal structure of the BAI3 GAIN domain showed no evidence of cleavage (Figure 1E, Supplementary Figure 7B), and BAI3 expressed in transfected HEK293 cells was uncleaved (Figure 4A). To address the question whether the BAI3 GAIN domain is functional, we raised antibodies to BAI proteins and showed that endogenous BAI proteins in brain are cleaved (Figure 4C). The conditions of BAI expression (i.e., expression levels, glycosylation machinery, and reductive environment) are very different between insect cells (used for expression of the crystallized proteins) and HEK293 cells on the one hand, and neurons on the other hand. Thus, these differences in post-translational processing and environment could be the reason why BAI3 is uncleaved when expressed in insect or HEK293 cells, but cleaved in the brain. Nevertheless, uncleaved BAI3 was efficiently transported to the plasma membrane in HEK293 cells, indicating that cleavage at the GPS is not required for surface transport of a cell-adhesion GPCR (Figure 4B; see also below). In contrast, autoproteolysis of the autoproteolytic protein nucleoporin98 is required for its transport to the nuclear envelope (Hodel *et al*, 2002).

Unique structural features of the GAIN domain enable autoproteolysis

The autocatalytic scissile bond in the GPS of the GAIN domain (star, Figure 5A and B) is positioned at a sharply kinked loop between the last two β-strands of the GAIN domain. In the structure of the uncleaved BAI3 GAIN domain, the autocatalytic scissile bond (Figure 1E) has a distorted geometry (phi = 174.0, psi = −158.2 degrees). Furthermore, numerous residues contribute to the unique structural features of the GAIN domain that localize the scissile bond at the tip of the sharply kinked loop (Figure 5A and B; Supplementary Figure S3A and B). The kink is created due to three structural features of the protein: (1) Two conserved disulphide bonds (C834-C820 and C832-C801 in CL1) between neighbouring β-strands (Supplementary Figure S3A) keep this β-turn in place. (2) The cleaved β-strand has conserved residues involved in numerous hydrophobic interactions with other conserved residues (Figure 1F). (3) The side chain of

Figure 1 Structures of CL1 and BAI3. (**A**). Diagram of CL1 and BAI3 showing the domains suggested by the SMART protein domain prediction server. No domain is predicted for the stalk region (coloured yellow and light pink). Note that GPS is defined as a separate domain in the Pfam database (magenta). In contrast, our crystal structures (see **B** and **C**) now show that the GPS motif is not an autonomously folded domain, but rather an integral part of the GAIN domain. The stalk region is also part of the GAIN domain. The HormR, TM7, lectin, olfactomedin, and thrombospondin motifs are coloured blue, green, cyan, red, and orange, respectively. The boundaries of the crystal structures are indicated with arrows. (**B**, **C**) Cartoon representation of the crystal structure of the HormR and GAIN domains of CL1 (PDB accession code 4DLQ) (**B**) and BAI3 (PDB accession code 4DLO) (**C**). The HormR domain is coloured blue, the GAIN subdomain A yellow, the GAIN subdomain B light pink, and the GPS motif, which is part of subdomain B, magenta. The cleaved β-strand of CL1 and the corresponding uncleaved β-strand of BAI3 are coloured orange and cyan, respectively. Disulphide bonds and carbohydrates are shown as black and red sticks, respectively. (**D**) $2mF_o-DF_c$ electron density map of the indicated residues around the cleavage site in CL1, contoured at 2σ. (**E**) $2mF_o-DF_c$ electron density map of the corresponding uncleaved site in BAI3, also contoured at 2σ. (**F**) Showing the hydrophobic packing interactions between the cleaved β-strand and the surrounding residues in CL1. (**G**) Showing the hydrogen bonds between the cleaved β-strand and the surrounding residues in CL1. In all panels, the cleavage site in CL1 and the corresponding peptide bond in BAI3 are marked with a black star. (**H**) The majority of the extracellular regions of CL1 and CL3 remain attached to the cells upon cleavage. HEK293 cells were transfected with full-length CL1 and CL3 constructs. Equal ratio of cell lysates and media were loaded on the gel and immunoblotted for the N-terminal Flag epitope. (**I**) The cleaved β-strand of the GAIN domain undergoes a conformational change upon autoproteolysis (the direction of the movement is indicated by arrows).

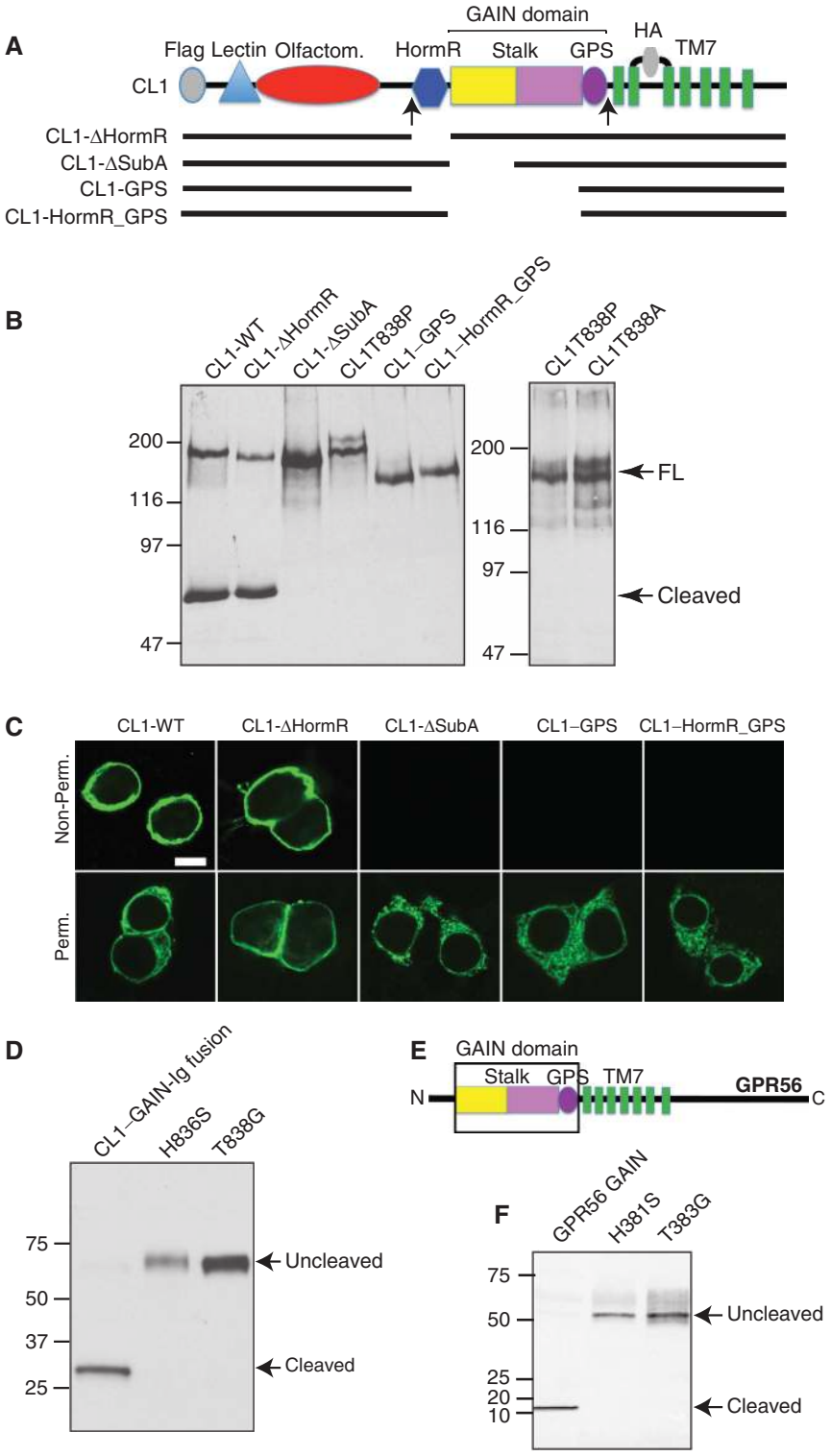


Figure 2 The GAIN domain is an autoproteolytic domain. (**A**) Diagram showing the CL1 deletion mutants that we tested in full-length CL1. The constructs were tagged with Flag and Haemagglutinin A (HA) epitopes (grey), both facing the extracellular side. (**B**) Effect of domain deletions on cleavage of CL1 constructs in HEK293 cells. For the cleavage assay, cell extracts were immunoblotted for the HA epitope. If cleavage at the GPS site occurs, a C-terminal 70 kDa fragment is produced. FL represents full length. (**C**) Effect of the domain deletions on cell-surface transport of CL1 in HEK293 cells. Transfected HEK293 cells expressing the indicated CL1 proteins were analysed by immunocytochemistry using antibodies to the HA tag; cells were either analysed with or without permeabilization. Cells were visualized by confocal microscopy using an Alexa-488 secondary antibody. (**D**) A construct encoding only the GAIN domain of CL1 as an Ig-fusion protein is autoproteolysed in transfected HEK293 cells whereas the two autoproteolysis site mutants are not cleaved (blotted for Ig). (**E**) Schematic domain diagram of GPR56. (**F**) A construct encoding only the GAIN domain of GPR56 as a myc tagged PDGF receptor transmembrane domain-fusion protein is autoproteolysed whereas the two autoproteolysis site mutants are not cleaved (blotted for myc-epitope).

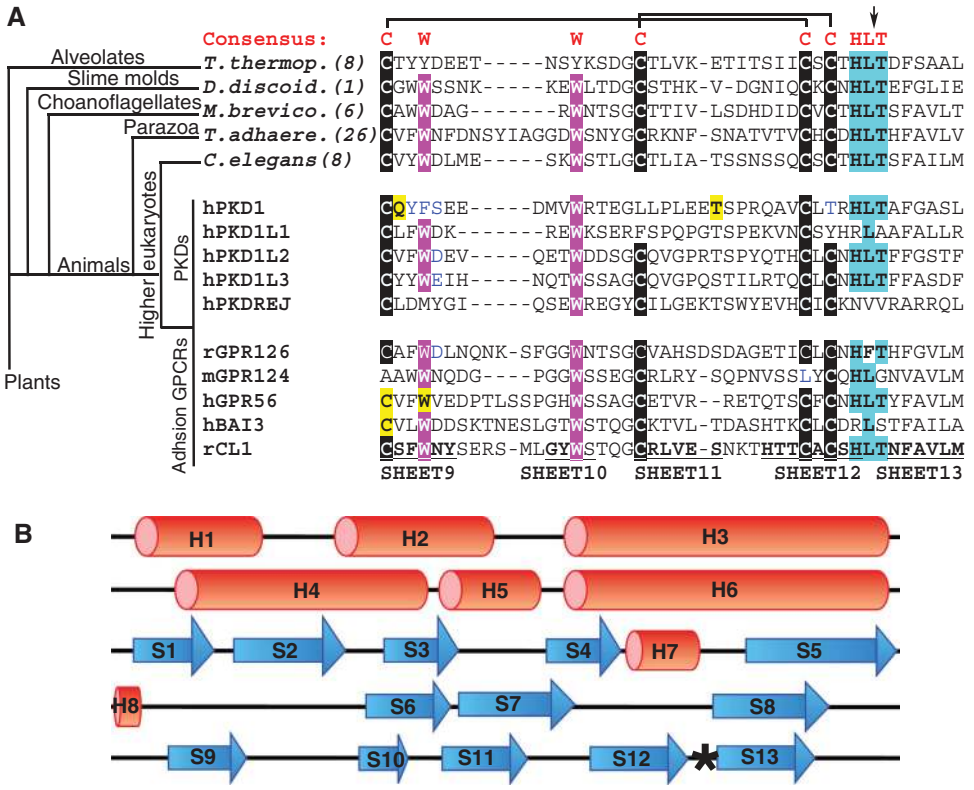


Figure 3 The GAIN domain is an ancient domain that exists in primitive organisms. The structure of the GAIN domain was used to discover this evolutionary relationship. First, BLAST searches were performed to find other GPS motifs. Then, among this group of proteins, the search pattern was extended by including approximately 400 residues upstream of the GPS motif. These sequences were aligned by PROMALS and the secondary structure patterns corresponding to each sequence were predicted by PSIPRED. Finally, the predicted secondary structure patterns were checked against the crystal structure of the GAIN domain and only those that had the same pattern were identified as GAIN domains. (**A**) Showing the sequence alignment of GPS motifs from evolutionary ancient organisms (top, *Tetrahymena thermohila* (gi89298346), *Dictyostelium discoideum* (gi66815909), *Monosiga brevicollis* (gi16752408), *Trichoplax adhaerens* (gi19600360), *Caenorhabditis elegans* (gi11553241)), PKD homologues (middle), and cell-adhesion GPCRs (bottom). The evolutionary relationship between the ancient organisms is described as a tree. The minimum number of GAIN domains in each organism's genome is indicated in parenthesis. The conserved cysteines, tryptophans, and cleavage site residues are highlighted black, magenta, and cyan, respectively. Disease mutations (ADPKD, cancer, and BFPP) are highlighted yellow. The cleavage site is indicated with an arrow. Disulphide bonds are shown as black lines. (**B**) Predicted consensus secondary structure pattern for representative ancient GAIN domains, cell-adhesion GPCRs, and PKD-related proteins shown in Supplementary Tables S2 and S3.

L837 in CL1 (L856 in BAI3) at the cleavage site is trapped in a conserved hydrophobic pocket (Supplementary Figure S3B). Comparison of the cleaved and uncleaved GAIN domain structures revealed a small tilt of the cleaved β-strand upon autoproteolysis that relaxes the sharp kink (Figure 1I).

To understand the mechanism of GAIN-domain catalysed autoproteolysis, we performed extensive mutagenesis of the CL1 GAIN domain (Figure 5C). Among others, three mutations (H836S, L837A, and T838G) impaired autoproteolysis but did not interfere with surface transport (and consequently, protein folding) (Figure 5C and D; Supplementary Figure S4), indicating that these residues are directly involved in CL1 autoproteolysis (also providing further evidence that autoproteolysis is not required for trafficking of GAIN domain-containing proteins). Consistent with a previously proposed mechanism for GPS autoproteolysis (Lin *et al*, 2004), H836 may act as a general base (although the general base is likely different in BAI3 since that residue is an Arginine; Figure 5A and B). The general base then removes the hydroxyl hydrogen from the side chain of T838 (Serine in BAI3) that in turn makes a nucleophilic attack on the carbonyl carbon of L837 (conserved in both CL1 and BAI3). Note that the T838P (Krasnoperov *et al*, 2002b) and T838A mutants are unfolded (Figure 5D; Supplementary Figure S4), but replacing threonine with a glycine enables proper folding, presumably by decreasing rigidity at the cleavage site.

Disease-related mutations in PKD1 and CL GAIN domains

The GAIN domain of PKD1 is a hot spot for mutations associated with ADPKD, one of the most common monogenic diseases (1 in 1000) (Qian *et al*, 2002; Figure 6A). In addition, *CL* and *BAI* genes are among the most frequently mutated genes in human cancers, with mutation frequencies of 3 and 12%, respectively (Kan *et al*, 2010). Their GAIN domains are, again, a hot spot for cancer mutations (Figure 6C; Supplementary Figure S6A). We tested the ADPKD and cancer mutations in GAIN domains for their effect on autoproteolysis and plasma membrane localization. Almost all ADPKD mutations in the GAIN domain of PKD1 abolished autoproteolysis, either by directly interfering with catalysis or by altering their membrane localization (presumably, by impairing folding) (Figure 6A and B; Supplementary Figure S5), indicating an essential role for the GAIN domain. However, cancer mutations in CL1 and CL3 did not affect autoproteolysis or membrane localization (Figure 6C and D; Supplementary Figure S6), suggesting that the mechanism by which these mutations lead to disease is different from that of the ADPKD mutations. Since most cancer mutations are

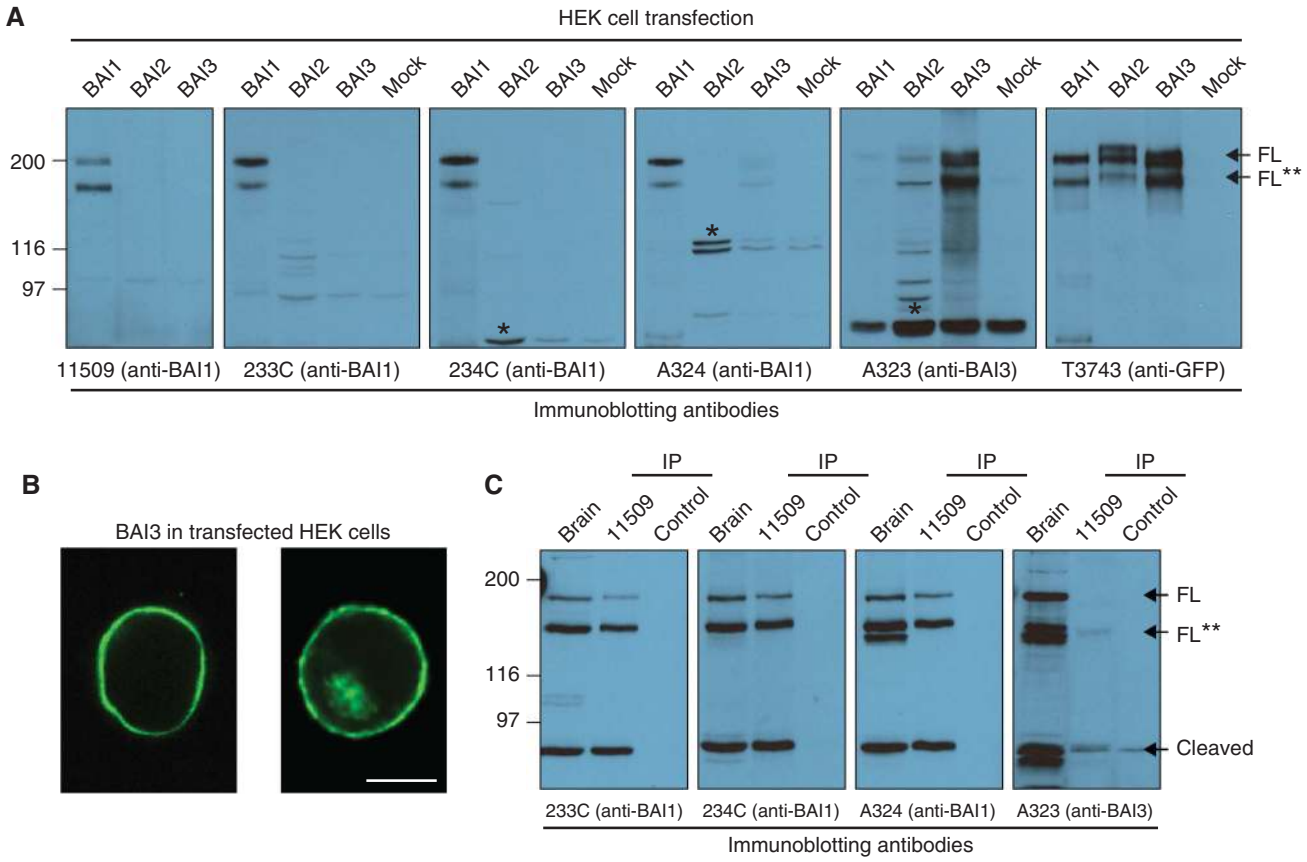


Figure 4 BAI3 GAIN domain is functional but its autoproteolysis is not required for surface transport. (**A**) Immunoblots of the lysates from HEK293 cells transfected with BAI1, BAI2, or BAI3 constructs tagged with mVenus in their C-terminal tails. Sample amounts were adjusted for immunoblotting to obtain comparable signals for BAIs; therefore, unspecific bands exhibit different intensities (black star). Blots were probed with antibodies raised against C-terminal epitopes of BAI1 and BAI3. Only full-length (FL) BAIs (200 kDa) and N-terminally truncated BAIs of unknown physiological relevance (FL**) (~170 kDa) were detected, but no products of autoproteolytic cleavage (expected mass: ~100 kDa) were observed. (**B**) Images of non-permeabilized HEK293 cells transfected with mVenus-tagged BAI3 cDNA demonstrate cell-surface membrane localization of uncleaved BAI3 (scale bar = 10 μm). (**C**) Immunoblots of mouse brain lysates and of brain immunoprecipitates (IPs) obtained with the BAI1-specific antibody 11509 (control: IP without 11509) with the antibodies used in (**A**). A C-terminal cleavage product of 72 kDa was detected.

located on the surface of the CL GAIN domains, they may be involved in the interaction of the GAIN domain with a ligand or with the nearby transmembrane helices. Indeed, it is the GAIN domain of CL1 that interacts with its long-known binding partner α-latrotoxin, a black widow spider toxin (Sudhof, 2001; Figure 7), demonstrating the ability of the CL1 GAIN domain to function as a protein-binding domain in addition to an autoproteolytic domain.

Cell-adhesion GPCRs exhibit structural similarity to hormone receptors

Although the HormR domain is the second most frequently observed domain in cell-adhesion GPCRs (found in 12 of 33 human cell-adhesion GPCRs), no hormones have yet been identified to bind cell-adhesion GPCRs. HormR domain-containing cell-adhesion GPCRs are homologous to Class B GPCRs that always include a HormR domain (Lagerstrom and Schioth, 2008). Remarkably, our HormR domain crystal structures from CL1 and BAI3 revealed an unusually high structural similarity (0.7 Å r.m.s.d.) to the genuine hormone-binding domain of the corticotrophin releasing factor receptor (CRFR), a member of the class B GPCRs, supporting the hypothesis that cell-adhesion GPCRs may be hormone receptors (Figure 8A).

Superposition of the CL1 structure with the structure of the CRFR HormR domain bound to corticotrophin releasing factor (CRF), a 41 amino-acid peptide hormone, reveals that a similar hormone could not bind to CL1 HormR in the conformation observed in the crystal structure since the GAIN domain is blocking the homologous hormone binding site on the HormR domain of CL1 (Figure 8B). Clearly, hormone binding would require a conformational change in CL1 to expose the putative hormone binding site. In class B GPCRs, the HormR domain precedes the transmembrane helices and is juxtaposed to the membrane. According to the two-domain model of class B GPCR activation, the interaction of the extracellular hormone-binding domain with the hormone promotes the interaction of the hormone with the transmembrane helices, leading to the activation of the receptor (Hoare, 2005). However, the mechanism of cell-adhesion GPCR activation upon hormone binding may be different, because the GAIN domain lies between the HormR domain and the transmembrane helices.

Discussion

In this study, we aimed to obtain a structural understanding of the autoproteolysis function of the GPS motif that is

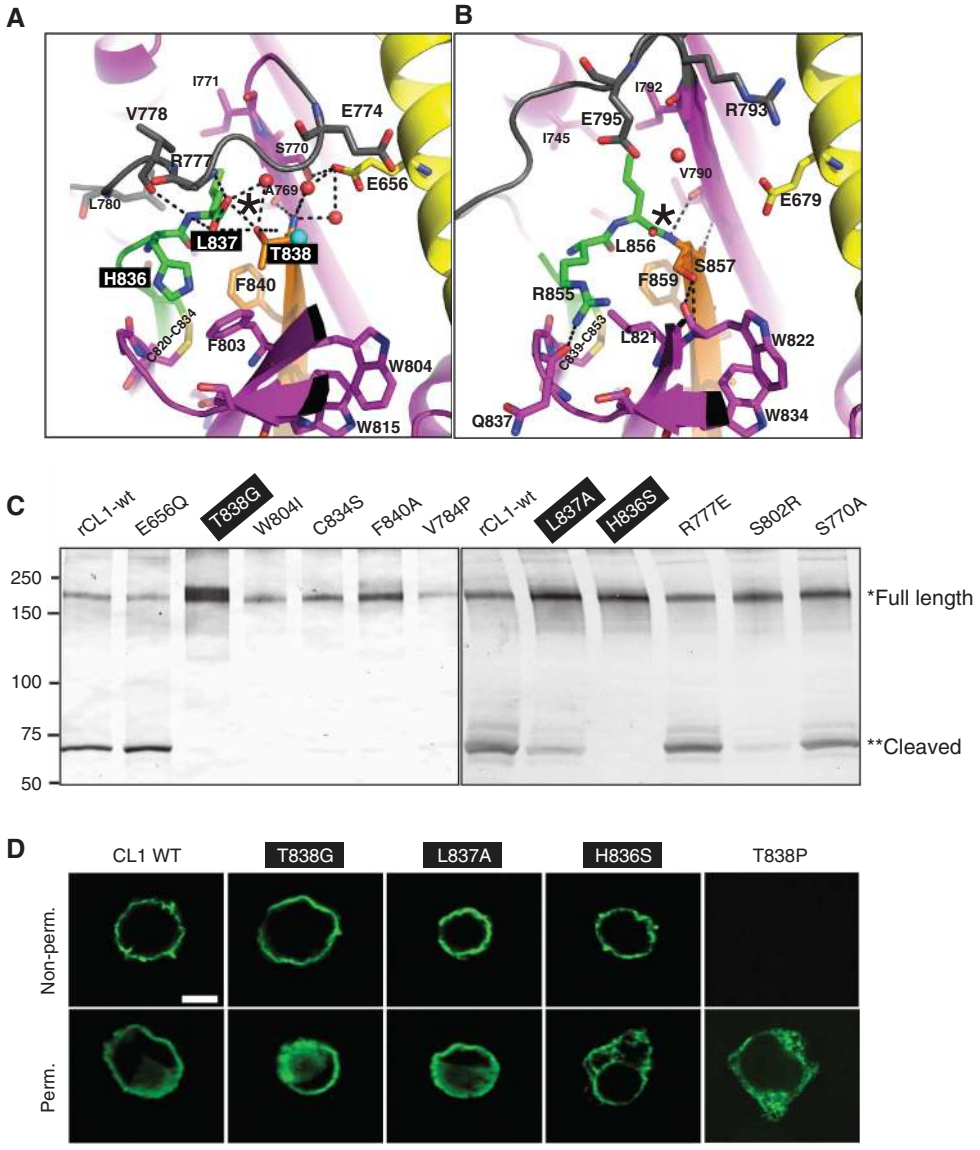


Figure 5 Mutagenesis of the CL1 autoproteolysis site. (**A**) Close-up view of the cleavage site in CL1 (black star). (**B**) The corresponding site in uncleaved BAI3 (black star). The colour coding is similar to that in Figure 1, except that the β-strands 12, 13, and the top loop are coloured green, orange, and grey, respectively. Residues involved in autoproteolysis (see mutagenesis experiments) are highlighted black. Water molecules and a sulphate ion are shown as red and cyan spheres, respectively. Hydrogen bonds are shown as black dashes. (**C**) Results of the cleavage assay of CL1 mutants (same assay as described in Figure 2B). (**D**) Effect of selected cleavage mutations on cell-surface transport of CL1 in HEK293 cells. Uncleaved, but properly cell-surface localized (i.e., folded), mutants are highlighted black.

characteristically found in cell-adhesion type GPCRs and PKD proteins. Towards this end, we here determined the crystal structures of fragments containing the GPS motif, the stalk region and the HormR domain from two distantly related cell-adhesion GPCRs, CL1 and BAI3. Unexpectedly, we found that a single large domain termed the GAIN domain spans the GPS motif and the stalk region (Figure 1), and that the GPS motif by itself is not a functional domain, but only a small part of the much larger GAIN domain. The GAIN domain is a novel domain, and our CL1 and BAI3 crystal structures now define this domain family.

The CL1 deletion experiments showed that the GPS motif is not functional by itself (Figure 2). Instead the entire GAIN domain is both required and sufficient for autoproteolysis. Thus, the GAIN domain functions as an autoproteolytic fold. Strikingly, structure-based database searches revealed that primitive organisms, such as *Dictyostelium discoideum* that arose early in evolution before animals emerged, encode GAIN domains although they lack most other autoproteolytic domains, important adhesion and signalling domains and critical signalling pathways (Figure 3; Supplementary Table S2). These ancient GAIN domains have conserved GPS motifs and are examples of one of the most ancient functional autoproteolytic domains. Computational analysis of the GAIN domain sequences showed that all cell-adhesion GPCRs and all PKD proteins have a GAIN domain with identical predicted secondary structure patterns to that of the CL1 and BAI3 crystal structures showing that the GAIN domain is widespread and conserved in higher eukaryotes as well as in ancient organisms (Supplementary Table S3). Our cleaved and uncleaved GAIN domain structures, in combination with extensive mutagenesis of the CL1 cleavage site,

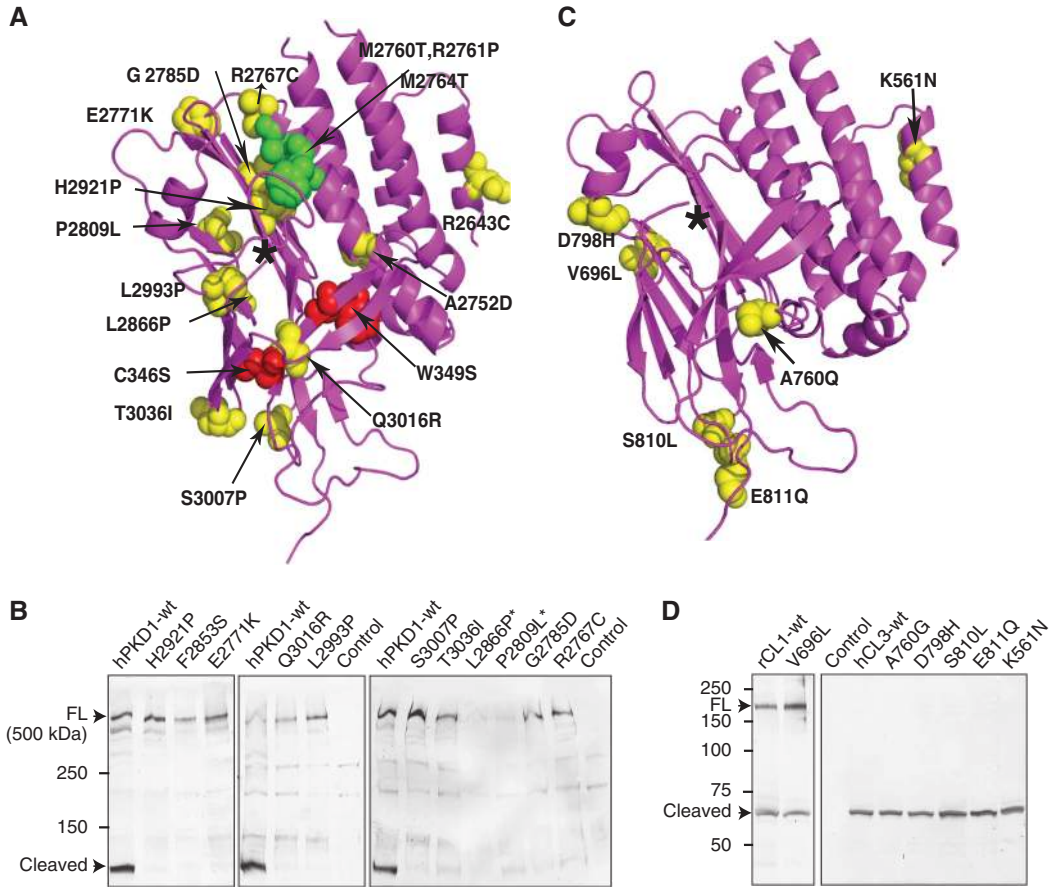


Figure 6 The effect of ADPKD and cancer mutations on autoproteolysis. (**A**) The ADPKD substitution mutations (yellow spheres) and gene conversion mutations (green spheres) in the human *PKD1* gene; and BFPP substitution mutations (red spheres) in the human *GPR56* gene are superposed on the structure of the CL1 GAIN domain. The cleavage site is indicated with a black star. (**B**) Effect of ADPKD mutations on the cleavage of the full-length hPKD1 construct that has an N-terminal GFP and a C-terminal Flag tags. HEK293 cells were transfected with the wild-type and mutant hPKD1 constructs and cell extracts were immunoblotted for the Flag epitope. A C-terminal 130 kDa fragment is produced following autoproteolysis in the GAIN domain. The approximate size of hPKD1 is indicated. The mutants that poorly expressed are labelled with black stars. (**C**) The cancer substitution mutations (yellow spheres) in the human *CL1* and *CL3* genes marked in the structure of the CL1 GAIN domain. (**D**) Effect of cancer mutations on the cleavage of the full-length rat CL1 and human CL3 that has tags similar to those shown in Figure 2A. Cell extracts were immunoblotted for the HA epitope. A C-terminal 70 kDa fragment is produced following cleavage. FL represents full length.

reveal the unique structural features of the GAIN domain that enable self-cleavage and shed light on the mechanism of autoproteolysis. The discovery of the GAIN domain also revealed a possible role for the GAIN domain in a variety of diseases, including cancer, ADPKD, and BFPP. We found that nearly all ADPKD mutations interfere with autoproteolysis of the GAIN domain, whereas cancer mutations have no effect on autoproteolysis or cell-surface localization. Thus, the mechanism by which the cancer mutations lead to disease are unrelated to autoproteolysis and the GAIN domain likely has other functions (such as interacting with ligands or transmembrane helices) in addition to the autoproteolysis function.

In summary, our experiments demonstrate that the GAIN domain is an ancient autoproteolytic domain that exhibits unique properties and is present in a large number of GPCRs and PKD proteins. Although the proteolysis of membrane proteins by sheddases and regulated intramembrane proteolysis (RIP) proteases is well studied (Blobel, 2005; Wang *et al*, 2006) the autoproteolysis of a membrane protein is poorly understood. GAIN domain is one of the few domains to mediate self-cleavage of a membrane protein. Two properties of the GAIN domain-mediated autoproteolysis make it unique and intriguing: (1) All GAIN domains that we analysed (from ancient organisms to higher eukaryotes) always immediately precede the N-terminal transmembrane helix (Figure 9). Additionally, HormR domains, when present, always precede the N-terminus of the GAIN domain. (2) In contrast to most other autoproteolytic domains that cause the release of the cleaved mature protein, autoproteolysis of the GAIN domain does not lead to the dissociation of the cleaved extracellular sequences containing the GAIN domain from the membrane-embedded regions of the protein. These observations naturally lead to the hypothesis that the GAIN domain is functionally and/or structurally associated with the membrane domains and that the GAIN domain-mediated autoproteolysis has a complex mechanism of action.

The HormR/GAIN/transmembrane helix sequence arrangement is unrelated to the autoproteolysis function of the GAIN domain although it may be significant for the regulation of receptor signalling via intramolecular interactions. Indeed, deletion of the extracellular sequences of GPR56, a cell-adhesion GPCR with an extracellular GAIN domain that is preceded only by a short non-GAIN domain

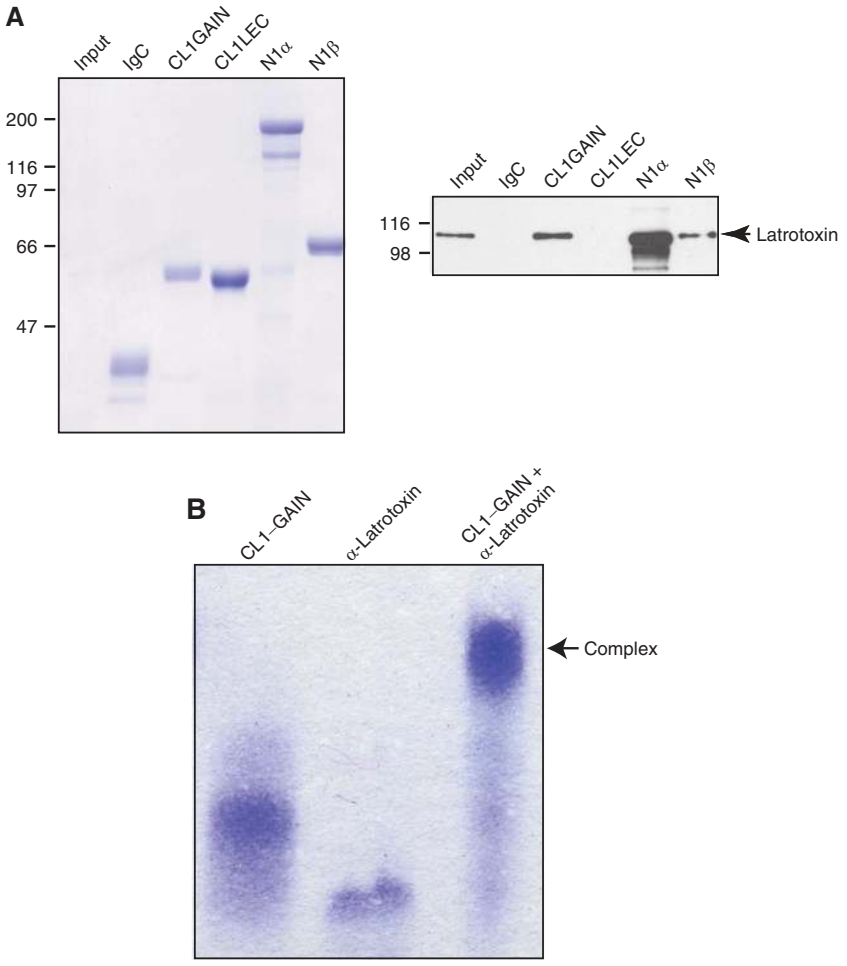


Figure 7 The GAIN domain of CL1 binds to α-latrotoxin. (**A**) Ig-fusion proteins of CL1 comprising the lectin domain (CL1LEC), the GAIN domain (CL1GAIN), or control (IgC) were expressed in HEK293 cells and purified on protein A sepharose beads along with Ig-fusion proteins of Neurexin-1α and Neurexin-1β (N1α and N1β). Recombinant myc–α-latrotoxin produced in bacteria was added. Proteins were visualized by Coomassie-blue staining (left) or immunoblotting using anti-myc antibody (right). (**B**) Native gel showing the formation of a complex of CL1-GAIN domain and the 20 ankyrin repeats of α-latrotoxin in a calcium-independent manner. Both proteins were expressed in insect cells as His-tagged proteins, secreted to the media, and purified to >99% purity by affinity tag purification followed by size-exclusion chromatography.

sequence (Figure 2E), causes constitutive activation of the GPCR (Paavola *et al*, 2011), suggesting a possible inhibitory role for the GAIN domain. In the GAIN domain, the cleaved β-strand is separated from the first transmembrane helix by 5–6 linker residues (Figure 9), positioning the GAIN and HormR domains into close proximity to the transmembrane helices. The small conformational change that we observe upon autoproteolysis (Figure 1I) may be critical for the relative orientation of the GAIN domain with respect to the transmembrane helices. Binding of a hormone to the HormR domain or binding of an endogenous ligand to the GAIN domain may be potential regulators of signalling via the receptor. Our work here redefines the poorly studied cell-adhesion GPCR class to share a large unique and widespread autoproteolytic domain that may be involved in downstream signalling.

Materials and methods

Protein expression and purification

The HormR and GAIN domains of rat CL1 (residues 460–849) and human BAI3 (residues 498–868) were cloned into the *Xba*1–*Not*1 and the *Bam*H1–*Not*1 sites of the pAcGP67A vector, respectively (CL1 = 43 kDa plus 5 carbohydrates; BAI3 = 41.7 kDa plus 3 carbohydrates). A 8 × His tag was added at the C-terminus for affinity purification. Spodoptera frugiperda (Sf9) cells (Invitrogen) were transfected with pAcGP67A carrying the gene and linearized AcNPV DNA (Sapphire Baculovirus DNA; Orbigen) using Cellfectin (Invitrogen). Baculovirus was amplified in Sf9 cells in 10% (v/v) fetal bovine serum containing SF900-II medium (Gibco).

Large-scale protein expression of CL1 and BAI3 was performed by infection of Trichoplusia ni (High Five) cells in Insect-Xpress medium (Biowhitaker) at a cell density of 2×10^6 cells/ml with an infection course of 72 h. The media containing the secreted, glycosylated, recombinant protein was concentrated and dialysed using a tangential flow filtration system (PALL) and subsequently purified with Ni-NTA resin (Qiagen) at 4°C. The proteins were eluted with 200 mM imidazole (pH 7.2). CL1 and BAI3 were further purified by size-exclusion chromatography using a Superdex 200 column (Amersham Pharmacia Biotech) in buffer containing 10 mM Hepes (pH 7.2) and 150 mM NaCl. The peak corresponding to the proteins was then concentrated to 10.7 mg/ml (for CL1) and 17 mg/ml (for BAI3) in 30 kDa Centricon (Millipore) concentrators and used for crystallization trials.

α-Latrotoxin (residues 413–1080) with N-terminal GST and Myc tags in a pGEX-KG vector was expressed in Origami cells. The expression and purification were described in Li *et al* (2005). The latrotoxin construct used for native gel experiments corresponds to the 20 ankyrin repeats of latrotoxin with a C-terminal His tag. The protein was expressed in High Five cells and secreted protein was purified by affinity chromatography followed by size-exclusion chromatography.

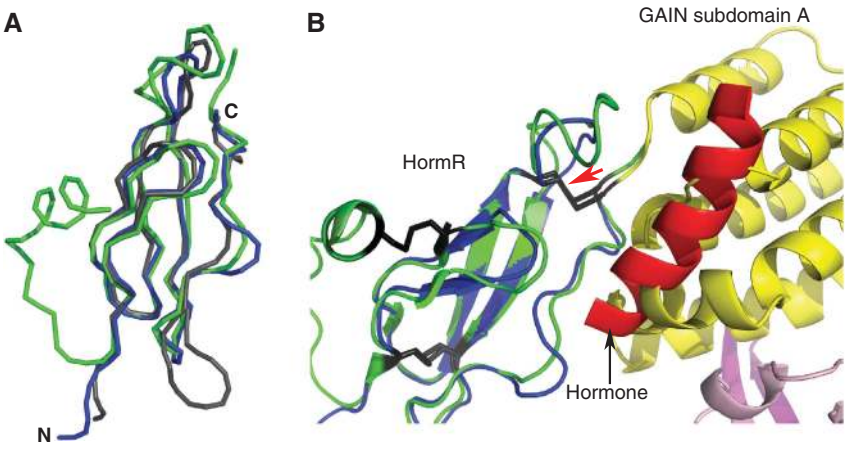


Figure 8 The HormR domains of cell-adhesion GPCRs exhibit structural similarity to genuine hormone-binding domains. (**A**) Superposition of the HormR domains of CL1 (blue), BAI3 (grey), and CRFR (green) (PDB ID 3ehu). The structure of the CL1 HormR domain is superposed on the structure of the CRFR HormR domain (green). (**B**) Binding of a putative hormone (red α-helix) to the CL1 HormR domain based on hormone (corticotrophin releasing factor) binding to CRFR would produce a clash with the GAIN domain. Thus, a conformational change would be required in order to make the putative hormone binding site accessible to the hormone. The colour coding of CL1 is the same as in Figure 1A. Disulphide bonds are shown as black sticks. The disulphide bond close to the hormone binding site is indicated with a red arrow.

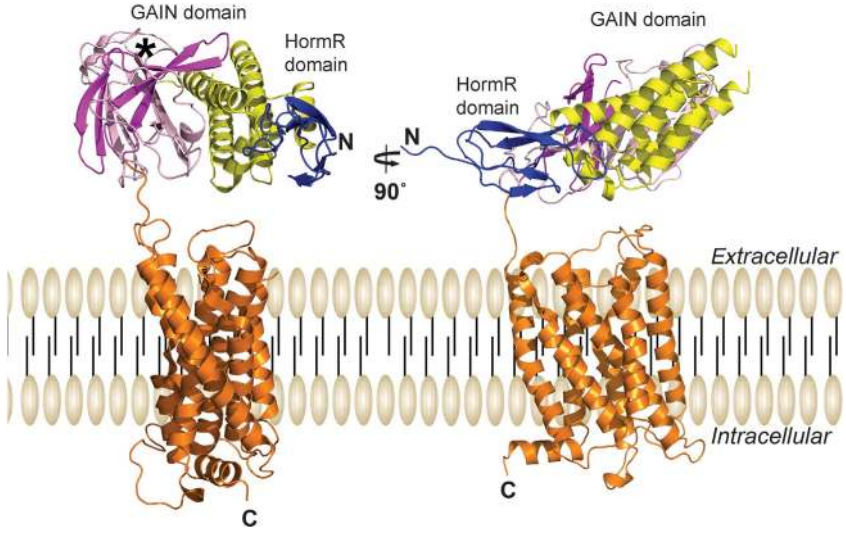


Figure 9 The GAIN and HormR domains are in close proximity to GPCR transmembrane helices. The short linker between the GAIN domain and the first GPCR transmembrane helix brings the GAIN and HormR domains close to the transmembrane domain, restricting possible arrangements between the domains. A model of CL1 is shown with a plausible arrangement of the GAIN and HormR domains. The model represents only one of the possible orientations of the GAIN domain with respect to the transmembrane helices. The seven-pass transmembrane domain (orange) was modelled by homology using the crystal structure of the β2 adrenergic GPCR structure (PDB ID 2RH1) (Cherezov *et al*, 2007; Rosenbaum *et al*, 2007), and the GAIN and HormR domains were taken from the crystal structure of CL1. The cleavage site is indicated with a black star.

The GAIN domain of rat CL1 (residues Thr533–Ile849) was cloned into a pCMV5 vector and fused to Ig (26 kDa) at the C-terminus and carried an HA tag at the N-terminus. The GAIN domain of mouse GPR56 (residues Q115–E395) was cloned into the pDisplay vector with N-terminal HA and C-terminal myc tags and has the PDGFR single-spanning transmembrane domain at the C-terminus. The full-length human PKD construct carried N-terminal GFP and C-terminal Flag tags. Full-length human CL3 was synthesized by codon optimization and carried Flag and HA tags at identical positions as in CL1.

Crystallization of CL1 and BAI3

Initial screens for crystallization of CL1 were carried out using 96-well format kits (Hampton Index, Salt and Crystal Screen HT, Hampton Research) on a Phoenix crystallization robot (Art Robbins Instruments). Automatic plate scanning by a CrystalPro 2 imaging system (TriTek Corporation) identified initial hits in the Wizard 1–2 screen (Emerald Biosystems). After optimization, crystals of CL1 were grown using hanging drop vapour diffusion at 20°C in 24-well format. In all, 1 μl of 10.7 mg/ml protein in 10 mM HEPES pH 7.2, 150 mM NaCl was mixed with equal volume of mother liquor consisting of 1.26 M $(NH_4)_2SO_4$, 0.1 M sodium acetate pH 4.5, 0.2 M NaCl, and equilibrated against 1 ml mother liquor. Crystals grew to full size within 3 days. Crystals were cryoprotected by transferring the crystals into mother liquor with 10, 20, and 28% glycerol gradually. The best crystals diffracted to $d_{min} = 1.85$ Å. Crystals of CL1 formed in space group $P3_121$ with one molecule per asymmetric unit and 67% solvent content.

Crystallization hits for BAI3 were obtained with a 96-well format kit (Protein Complex, Nextal) at 20°C. In all, 17 mg/ml protein in 10 mM HEPES, 150 mM NaCl pH 7.2 was mixed with equal volume of mother liquor consisting of 15% PEG6000, 0.1 M sodium citrate

pH 5.5. Crystals grew to full size within 2 days and were cryoprotected with 25% glycerol. Crystals diffracted to $d_{min} = 2.3$ Å. Crystals of BAI3 formed in space group $C222_1$ with two molecules per asymmetric unit and 63% solvent content.

Structure determination and refinement of CL1

For the CL1 structure, experimental phase information was obtained from sulphur SAD data. This structure was the first structure that was determined at the Stanford Synchrotron Radiation Lightsource by sulphur SAD phasing using an optimized data collection strategy for a long wavelength at 2 Å, and one of a few *de-novo* structures that has been determined by sulphur SAD at this resolution. More than 40 crystals were screened to find the best diffracting large crystal. A test radiation damage experiment was performed on a different crystal to determine the effect of radiation damage and to optimize the parameters for the experiment. Three data sets, each comprising 360° and offset by 20°, were collected on a single long crystal by translating the beam along the long edge of the crystal. The crystal was placed in a helium stream operating at 65 K to minimize background scatter at long wavelength and to further limit radiation damage. Data collection was carried out in inverse beam mode with a wedge angle of 0.5° and an exposure time of 8 s. Total data collection lasted 9 h. The data were collected at a wavelength of 2 Å, which limited the resolution to 2.3 Å (diffraction was observed to 1.8 Å resolution at a wavelength of 1 Å). The resolution limit decreased from 2.3 to 2.7 Å at the end of one 360° data collection as a result of radiation damage. All data sets were indexed, integrated, scaled, and combined in HKL2000 using the scale anomalous and high absorption options (Otwinowski, 1997). The error model was not optimized during scaling. In all, 25 sulphur sites were found (17 sites within the protein and 8 in sulphate ions from the crystallization condition) by SHELXD (Schneider and Sheldrick, 2002). SHELXD used a high-resolution cutoff of 2.7 Å, at which the I/σ was ~50. The sulphur sites were refined using SHARP (Vonrhein *et al*, 2007). Density modification was performed by DM (Cowtan and Zhang, 1999). Among all space groups tested, only space group $P3_121$ yielded easily interpretable electron density maps. The density modified electron density maps showed good connectivity of secondary structure elements and cleavage at the GPS site (Supplementary Figure S7A). The sulphur sites corresponded to the expected side chains when compared within the density modified electron density map. Automatic model building was done with Phenix Resolve (Terwilliger, 2003) for two thirds of the structure. The rest of the model building was carried out manually using Coot (Emsley and Cowtan, 2004), and refinement performed with Phenix.refine (Adams *et al*, 2010) using the MLHL target function with experimental phases. The resulting model was then refined against the higher resolution native data set at 1.85 Å resolution. The subsequent refinement consisted of alternating rounds of positional minimization, individual restrained thermal factor refinement, both using the MLF maximum likelihood target function, and model building. Water molecules and N-linked carbohydrates were added. A summary of the data collection and refinement statistics is given in Supplementary Table S1 (PDB accession code 4DLQ).

Structure determination and refinement of BAI3

BAI3 crystals were soaked into three solutions containing the crystallization condition, 0.25 M sodium iodide, and 10, 20, and 30% glycerol immediately before they were frozen in liquid nitrogen. The total soak time was about 1–2 min. Only one in ten crystals survived this treatment. The presence of increased glycerol, thus viscosity, in the soak solutions prevented the crystals from cracking. Data collection was done in inverse beam mode with a wedge angle of 25° and an exposure time of 5.5 s. The derivative crystals were not isomorphous to the native crystals, so the native data set could not be used for isomorphous phasing. The diffraction data were collected at a wavelength of 1.6 Å, resulting in diffraction data to 2.7 Å resolution. Data processing was similar to that of CL1. The BAI3 structure was independently determined with experimental phases obtained by iodine-SAD data (a density modified map of the iodine-SAD data around the GPS site is shown in Supplementary Figure S7B, clearly indicating lack of cleavage in contrast to the CL1 structure). SHELXD used a high-resolution cutoff of 4 Å, at which the I/σ was ~50, and it found 6 iodine sites. There were two molecules in the asymmetric unit and non-crystallographic symmetry (NCS) restraints were used during initial refinement cycles using Phenix.refine but were removed at later stages. A summary of the data collection and refinement statistics is given in Supplementary Table S1 (PDB accession code 4DLO).

Cleavage and plasma membrane localization experiments

Antibodies. Mouse monoclonal anti-HA antibody was from Covance and rabbit polyclonal anti-Flag antibody was from Sigma.

Mutagenesis and expression constructs. CL1 deletion constructs were engineered by PCR oligonucleotide amplification. pCMV-CL1 HA Flag was generated as follows: The Flag epitope containing constructs were obtained by subcloning *Hin*dIII–*Xba*I PCR fragment from pCMV-CL1 into the same site of pCMV-NL1-Flag. The Flag epitope was fused at Leu25 of CL1. A single HA epitope was inserted by PCR between Tyr918 and Glu919 in the second extracellular loop of CL1. Deletion constructs were generated using pCMV-CL1 HA Flag construct as a backbone. Deletion constructs are as follow: pCMV-CL1-ΔHormR HA Flag was generated by deleting Phe479–Cys532. pCMV-CL1-GPS HA Flag was generated by deleting Phe479–Glu792. pCMV-CL1-GPS_HormR HA Flag was generated by deleting Pro535–Glu792. pCMV-CL1-ΔSubA HA Flag was generated by deleting Pro535–Val666. pCMV-IgCL1 LEC was generated by amplifying Met1–Tyr131 PCR fragment and subcloning into *Eco*RI–*Sal*I of pCMV-Ig. pCMV-IgCL1 GAIN was generated by subcloning the PCR-amplified fragment of Thr533–Ile849 into *Sac*II–*Sal*I of pDisplay then by excising *Eco*RI–*Sal*I fragment and subcloning into the same sites of pCMV-Ig. Therefore, the resulting pCMV-IgCL1 GAINR construct also bears an N-terminal HA epitope. pCMV IgN1a-1 and pCMVN1b were described elsewhere (Ushkaryov *et al*, 1994; Sugita *et al*, 2001). pCMV-CL1 T838A and pCMV-CL1 T838P were generated by PCR mutagenesis. PKD mutagenesis was performed by cutting out the wild-type human PKD with *Bsi*WI restriction enzyme, pasting the small fragment into a carrier plasmid. After site-directed mutagenesis was performed, the mutated fragments were subcloned back into the original plasmid. CL3 and CL1 mutagenesis was performed by the Qiagen site-directed mutagenesis kit.

Cell culture and transfection. For immunoblotting and immunocytochemistry experiments, HEK293T cells were cultured in 6-well plates until they reached 70–80% confluency. The cells were then transfected using 4 μl of Fugene 6® (Roche Diagnostics) and 2 μg of respective CL1 DNA constructs. For immunoblotting experiments, cells were harvested 48 h post transfection. For immunocytochemistry experiments, the cells were trypsinized 24 h post transfection and splitted into 12-well plates containing coverslips that have been previously coated with 0.5 mg/ml poly-L-lysine in borate buffer. The cells were grown on the coverslips for an additional 24 h before further experiments were done. For recombinant Ig-protein expression, HEK cells were transfected using calcium phosphate with chloroquine and 20 μg of cDNA corresponding to the various Ig proteins. Media was then harvested 4 days post transfection.

Immunocytochemistry and image acquisition. Cells transfected with CL1 constructs were washed once with PBS and fixed with 4% paraformaldehyde for 10 min on ice. Cells were washed again three times with cold PBS and incubated at room temperature for 30 min in a blocking solution containing 3% BSA in PBS with or without 0.1% Triton X-100 meaning that cells were permeabilized or not, respectively. Mouse anti-HA antibody was then added (1:500 ratio) and the incubation was prolonged for another 2 h. Cells were washed three times with blocking solution and incubated for 1 h at room temperature with anti-mouse Alexa-488 fluorescent antibody to label CL1 receptors. Cells were finally washed again three times with blocking solution and once with water before mounting on slides using media containing DaPI for nuclear staining. Slides were then analysed by confocal microscopy. Images were acquired using the confocal microscope Leica TCS2. The same confocal acquisition settings were applied to all samples of the experiment. Collected z-section images were analysed blindly using Leica confocal software.

Immunoblotting procedures. Transfected HEK cells were harvested 2 days after transfection and solubilized for 1 h at 4°C in buffer containing 20 mM Hepes-NaOH pH 7.4, 0.1 mM EDTA, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$ and 1% Triton. Insoluble material was removed by centrifugation. SDS–PAGE sample buffer was added to the supernatant and the resulting sample was loaded on

6% SDS–PAGE. Gels were then transferred onto nitrocellulose membranes and processed using standard procedures. Membranes were probed with a mouse monoclonal anti-HA antibody or a rabbit polyclonal anti-Flag antibody followed by Horseradish peroxidase-coupled secondary antibody, incubated with ECL reagents and revealed on X-ray films.

Ig-fusion protein purification and pull-down experiments

The soluble Neurexin and CL1 Ig-fusion proteins were affinity purified from transfected HEK293 cell supernatant using protein-A Sepharose 4 Fast Flow (Amersham Pharmacia Biotech AB, Sweden) to bind the human-IgG portion and washed to remove unbound proteins. The immobilized Ig-fusion proteins were then incubated 16 h at 4°C with the eluted 3 μg GST–myc–Latrotoxin in buffer containing 20 mM Hepes buffer, pH 7.4; 0.1 mM EDTA, 150 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$ and 1% (v/v) Triton X-100. The beads were then washed, incubated in boiling SDS–PAGE sample buffer and loaded on 6% SDS–PAGE. SDS–PAGE gels were either transferred onto nitrocellulose membrane and immunoblotted with anti-myc antibody or stained using Coomassie Brilliant Blue R-250 (Bio-Rad).

BAI cleavage experiments

Antibodies. Goat antiserum 11509 and rabbit antisera 233C and 234C were generated against a synthetic peptide (N-CEKAGA TIPLVGQDIIDLQTEV-C; mouse BAI1). Rabbit antisera A324 and A323 were raised against recombinant GST-fusion proteins (P1447→V1582 of mouse BAI1 and R1151→E1348 of human BAI3, respectively). T3743 was raised against GFP in rabbit.

Cell culture. HEK293T cells were plated in 6-well plates. At 70–80% confluency, they were transfected with 1 μg expression plasmid (FuGENE 6 from Roche), and then cultivated for 2 days. Cells were lysed in extraction buffer (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, protease inhibitors), homogenized, and the soluble fractions were collected after centrifugation.

Immunoprecipitation. Two mouse brains were homogenized in 8 ml buffer (25 mM HEPES pH 7.2, 320 mM sucrose, protease inhibitors). After centrifugation at 1500 *g* for 10 min, supernatant S1 was centrifuged at 150 000 *g* for 60 min. Pellet P2 was then homogenized in 6 ml lysis buffer (50 mM Tris–HCl pH 7.5, 100 mM NaCl, 1% Triton X-100, protease inhibitors), and centrifuged again at 150 000 *g* for 60 min. In all, 3 ml of supernatant S3 was incubated with 18 μl antiserum 11509 for 60 min. Then, 180 μl protein G sepharose (GE Healthcare) was added, and reactions were incubated for 3 h. The beads were washed four times with lysis buffer, and the proteins were eluted in SDS buffer.

Immunoblotting. Proteins were resolved on 6% acrylamide gels and transferred onto nitrocellulose membranes (Whatman). After a blocking step, membranes were incubated overnight with the primary antibody diluted in blocking solution (5% milk powder, 2.5% goat or rabbit serum, 0.1% Tween-20, in TBS). After washing with TBS containing 0.1% Tween-20, blots were incubated with a peroxidase-conjugated secondary antibody (Cappel) for 45 min. After washing, immunoreactive bands were detected by ECL (GE Healthcare).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal. org).

Acknowledgements

We thank E Özkan for crystallographic advice and discussion; M Etherton, X Piao, F Qian for α-latrotoxin, GPR56, and PKD1 cDNAs. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. This study was also supported by a postdoctoral fellowship from Canadian Institutes of Health Research (CIHR, to AAB) and a grant from the NIMH (R37 MH52804-08 to TCS). D Araç is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation.

Author contributions: DA designed and performed experiments and wrote the paper. AAB, DA, MFB, JN, and MS performed experiments. TCS and ATB provided guidance and wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221

Arcos-Burgos M, Jain M, Acosta MT, Shively S, Stanescu H, Wallis D, Domene S, Velez JI, Karkera JD, Balog J, Berg K, Kleta R, Gahl WA, Roessler E, Long R, Lie J, Pineda D, Londono AC, Palacio JD, Arbelaez A *et al* (2010) A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Mol Psychiatry* **15:** 1053–1066

Blobel CP (2005) ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol* **6:** 32–43

Bolliger MF, Martinelli DC, Sudhof TC (2011) The cell-adhesion G protein-coupled receptor BAI3 is a high-affinity receptor for C1q-like proteins. *Proc Natl Acad Sci USA* **108:** 2534–2539

Chae J, Kim MJ, Goo JH, Collier S, Gubb D, Charlton J, Adler PN, Park WJ (1999) The Drosophila tissue polarity gene starry night encodes a member of the protocadherin family. *Development* **126:** 5421–5429

Chang GW, Stacey M, Kwakkenbos MJ, Hamann J, Gordon S, Lin HH (2003) Proteolytic cleavage of the EMR2 receptor requires both the extracellular stalk and the GPS motif. *FEBS Lett* **547:** 145–150

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318:** 1258–1265

Cowtan KD, Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* **72:** 245–270

Davies B, Baumann C, Kirchhoff C, Ivell R, Nubbemeyer R, Habenicht UF, Theuring F, Gottwald U (2004) Targeted deletion of the epididymal receptor HE6 results in fluid dysregulation and male infertility. *Mol Cell Biol* **24:** 8642–8648

Deak F, Liu X, Khvotchev M, Li G, Kavalali ET, Sugita S, Sudhof TC (2009) Alpha-latrotoxin stimulates a novel pathway of Ca2+-dependent synaptic exocytosis independent of the classical synaptic fusion machinery. *J Neurosci* **29:** 8639–8648

Ebermann I, Wiesen MH, Zrenner E, Lopez I, Pigeon R, Kohl S, Lowenheim H, Koenekoop RK, Bolz HJ (2009) GPR98 mutations cause Usher syndrome type 2 in males. *J Med Genet* **46:** 277–280

Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P *et al* (2005) The genome of the social amoeba Dictyostelium discoideum. *Nature* **435:** 43–57

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132

Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main

families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63:** 1256–1272

Hoare SR (2005) Mechanisms of peptide and nonpeptide ligand binding to class B G-protein-coupled receptors. *Drug Discov Today* **10:** 417–427

Hodel AE, Hodel MR, Griffis ER, Hennig KA, Ratner GA, Xu S, Powers MA (2002) The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Mol Cell* **10:** 347–358

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233:** 123–138

Hughes J, Ward CJ, Aspinwall R, Butler R, Harris PC (1999) Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Hum Mol Genet* **8:** 543–549

Ichtchenko K, Bittner MA, Krasnoperov V, Little AR, Chepurny O, Holz RW, Petrenko AG (1999) A novel ubiquitously expressed alpha-latrotoxin receptor is a member of the CIRL family of G-protein-coupled receptors. *J Biol Chem* **274:** 5491–5498

Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR *et al* (2008) Core signalling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321:** 1801–1806

Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L *et al* (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466:** 869–873

Kaur B, Brat DJ, Devi NS, Van Meir EG (2005) Vasculostatin, a proteolytic fragment of brain angiogenesis inhibitor 1, is an anti-angiogenic and antitumorigenic factor. *Oncogene* **24:** 3632–3642

Krasnoperov V, Bittner MA, Mo W, Buryanovsky L, Neubert TA, Holz RW, Ichtchenko K, Petrenko AG (2002a) Protein-tyrosine phosphatase-sigma is a novel member of the functional family of alpha-latrotoxin receptors. *J Biol Chem* **277:** 35887–35895

Krasnoperov V, Lu Y, Buryanovsky L, Neubert TA, Ichtchenko K, Petrenko AG (2002b) Post-translational proteolytic processing of the calcium-independent receptor of alpha-latrotoxin (CIRL), a natural chimera of the cell adhesion protein and the G protein-coupled receptor. Role of the G protein-coupled receptor proteolysis site (GPS) motif. *J Biol Chem* **277:** 46518–46526

Krasnoperov VG, Bittner MA, Beavis R, Kuang Y, Salnikow KV, Chepurny OG, Little AR, Plotnikov AN, Wu D, Holz RW, Petrenko AG (1997) alpha-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron* **18:** 925–937

Kuhnert F, Mancuso MR, Shamloo A, Wang HT, Choksi V, Florek M, Su H, Fruttiger M, Young WL, Heilshorn SC, Kuo CJ (2010) Essential regulation of CNS angiogenesis by the orphan G protein-coupled receptor GPR124. *Science* **330:** 985–989

Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **7:** 339–357

Langenhan T, Promel S, Mestek L, Esmaeili B, Waller-Evans H, Hennig C, Kohara Y, Avery L, Vakonakis I, Schnabel R, Russ AP (2009) Latrophilin signaling links anterior-posterior tissue polarity and oriented cell divisions in the C. elegans embryo. *Dev Cell* **17:** 494–504

Lee JJ, Ekker SC, von Kessler DP, Porter JA, Sun BI, Beachy PA (1994) Autoproteolysis in hedgehog protein biogenesis. *Science* **266:** 1528–1537

Lelianova VG, Davletov BA, Sterling A, Rahman MA, Grishin EV, Totty NF, Ushkaryov YA (1997) Alpha-latrotoxin receptor, latrophilin, is a novel member of the secretin family of G protein-coupled receptors. *J Biol Chem* **272:** 21504–21508

Li G, Lee D, Wang L, Khvotchev M, Chiew SK, Arunachalam L, Collins T, Feng ZP, Sugita S (2005) N-terminal insertion and C-terminal ankyrin-like repeats of alpha-latrotoxin are critical for Ca2 + -dependent exocytosis. *J Neurosci* **25:** 10188–10197

Lin HH, Chang GW, Davies JQ, Stacey M, Harris J, Gordon S (2004) Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J Biol Chem* **279:** 31823–31832

McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16:** 404–405

Monk KR, Naylor SG, Glenn TD, Mercurio S, Perlin JR, Dominguez C, Moens CB, Talbot WS (2009) A G protein-coupled receptor is essential for Schwann cells to initiate myelination. *Science* **325:** 1402–1405

Nechiporuk T, Urness LD, Keating MT (2001) ETL, a novel seven-transmembrane receptor that is developmentally regulated in the heart. ETL is a member of the secretin family and belongs to the epidermal growth factor-seven-transmembrane subfamily. *J Biol Chem* **276:** 4150–4157

Nishimori H, Shiratsuchi T, Urano T, Kimura Y, Kiyono K, Tatsumi K, Yoshida S, Ono M, Kuwano M, Nakamura Y, Tokino T (1997) A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. *Oncogene* **15:** 2145–2150

Otwinowski ZMW (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276:** 307–326

Paavola KJ, Stephenson JR, Ritter SL, Alter SP, Hall RA (2011) The GPR56 N-terminus controls receptor signaling activity. *J Biol Chem* **286:** 28914–28921

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz Jr LA, Hartigan J *et al* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* **321:** 1807–1812

Pei J, Kim BH, Tang M, Grishin NV (2007) PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res* **35:** W649–W652

Perler FB, Xu MQ, Paulus H (1997) Protein splicing and autoproteolysis mechanisms. *Curr Opin Chem Biol* **1:** 292–299

Piao X, Hill RS, Bodell A, Chang BS, Basel-Vanagaite L, Straussberg R, Dobyns WB, Qasrawi B, Winter RM, Innes AM, Voit T, Ross ME, Michaud JL, Descarie JC, Barkovich AJ, Walsh CA (2004) G protein-coupled receptor-dependent development of human frontal cortex. *Science* **303:** 2033–2036

Ponting CP, Hofmann K, Bork P (1999) A latrophilin/CL-1-like GPS domain in polycystin-1. *Curr Biol* **9:** R585–R588

Qian F, Boletta A, Bhunia AK, Xu H, Liu L, Ahrabi AK, Watnick TJ, Zhou F, Germino GG (2002) Cleavage of polycystin-1 requires the receptor for egg jelly domain and is disrupted by human autosomal-dominant polycystic kidney disease 1-associated mutations. *Proc Natl Acad Sci USA* **99:** 16981–16986

Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK (2007) GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318:** 1266–1273

Rossetti S, Strmecki L, Gamble V, Burton S, Sneddon V, Peral B, Roy S, Bakkaloglu A, Komel R, Winearls CG, Harris PC (2001) Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. *Am J Hum Genet* **68:** 46–63

Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58:** 1772–1779

Shima Y, Kengaku M, Hirano T, Takeichi M, Uemura T (2004) Regulation of dendritic maintenance and growth by a mammalian 7-pass transmembrane cadherin. *Dev Cell* **7:** 205–216

Sudhof TC (2001) alpha-Latrotoxin and its receptors: neurexins and CIRL/latrophilins. *Annu Rev Neurosci* **24:** 933–962

Sugita S, Ichtchenko K, Khvotchev M, Sudhof TC (1998) alpha-Latrotoxin receptor CIRL/latrophilin 1 (CL1) defines an unusual family of ubiquitous G-protein-linked receptors. G-protein coupling not required for triggering exocytosis. *J Biol Chem* **273:** 32715–32724

Sugita S, Khvochtev M, Sudhof TC (1999) Neurexins are functional alpha-latrotoxin receptors. *Neuron* **22:** 489–496

Sugita S, Saito F, Tang J, Satz J, Campbell K, Sudhof TC (2001) A stoichiometric complex of neurexins and dystroglycan in brain. *J Cell Biol* **154:** 435–445

Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* **59:** 38–44

Ushkaryov YA, Hata Y, Ichtchenko K, Moomaw C, Afendis S, Slaughter CA, Sudhof TC (1994) Conserved domain structure of beta-neurexins. Unusual cleaved signal sequences in receptor-like neuronal cell-surface proteins. *J Biol Chem* **269:** 11987–11992

Usui T, Shima Y, Shimada Y, Hirano S, Burgess RW, Schwarz TL, Takeichi M, Uemura T (1999) Flamingo, a seven-pass transmem-

brane cadherin, regulates planar cell polarity under the control of Frizzled. *Cell* **98:** 585–595

Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364:** 215–230

Wang Y, Zhang Y, Ha Y (2006) Crystal structure of a rhomboid family intramembrane protease. *Nature* **444:** 179–180

Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z *et al* (2007) The genomic landscapes of human breast and colorectal cancers. *Science* **318:** 1108–1113

Yona S, Lin HH, Siu WO, Gordon S, Stacey M (2008) Adhesion-GPCRs: emerging roles for novel receptors. *Trends Biochem Sci* **33:** 491–500

Yu S, Hackmann K, Gao J, He X, Piontek K, Garcia-Gonzalez MA, Menezes LF, Xu H, Germino GG, Zuo J, Qian F (2007) Essential role of cleavage of Polycystin-1 at G protein-coupled receptor proteolytic site for kidney tubular structure. *Proc Natl Acad Sci USA* **104:** 18688–18693

