Review

Matching structure with function: the GAIN domain of Adhesion-GPCR and PKD1-like proteins

Simone Prömel[1*], Tobias Langenhan[2*], and Demet Araç[3*]

[1] Institute of Biochemistry, Molecular Biochemistry, Medical Faculty, University of Leipzig, Johannisallee 30, 04103 Leipzig, Germany
[2] Department of Neurophysiology, Institute of Physiology, University of Würzburg, Röntgenring 9, 97070 Würzburg, Germany
[3] Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA

Elucidation of structural information can greatly facilitate the understanding of molecular function. A recent example is the description of the G-protein-coupled receptor (GPCR) autoproteolysis-inducing (GAIN) domain, an evolutionarily ancient fold present in Adhesion-GPCRs (aGPCRs) and polycystic kidney disease 1 (PKD1)-like proteins. In the past, the peculiar autoproteolytic capacity of both membrane protein families at the conserved GPCR proteolysis site (GPS) had not been described in detail. The physiological performance of aGPCRs and PKD1-like proteins is thought to be regulated through the GPS, but it is debated how. A recent report provides pivotal details by discovery and analysis of the GAIN domain structure that incorporates the GPS motif. Complementary studies have commenced to analyze physiological requirements of the GAIN domain for aGPCR function, indicating that it serves as the linchpin for multiple receptor signals. Structural analysis and functional assays now allow for the dissection of the biological duties conferred through the GAIN domain.

Why GAIN domains are important for study

The GAIN domain is a recently discovered fold which hallmarks the two protein families aGPCRs and PKD1-like proteins. The members of both families are transmembrane receptors whose signaling modes are still only poorly understood. In this review, we discuss the structure of the GAIN domain and its integral part, the GPS, and its impact on receptor function. This review also highlights potential pharmacological implications resulting from the emerging understanding of the role of the GAIN domain in its different protein contexts. This venture is timely because both protein families appear responsible for the infliction of diverse diseases. Studying GAIN domain function will thus hopefully shed light onto the etiology and therapy of various human ailments.

Corresponding authors: Langenhan, T. (tobias.langenhan@uni-wuerzburg.de); Araç, D. (arac@uchicago.edu).
Keywords: autocatalytic cleavage; GPCR proteolysis site (GPS); receptor activity; disease-associated mutations.
[*] These authors contributed equally to this work.

0165-6147/$ – see front matter
 http://dx.doi.org/10.1016/j.tips.2013.06.002

Adhesion-class GPCRs and PKD1-like proteins are hallmarked by the GAIN domain

aGPCRs belong to the GPCR superfamily that is the largest group of transmembrane receptors and the most commonly targeted group of molecules for treatment of human diseases. Although aGPCRs constitute, with 33 paralogs in humans [1–3], the second largest of the five subgroups of GPCRs, they are the least understood. Recent studies demonstrate their essential functions in immunology [4–6], development [7,8], and neurobiology [9–11]. Evident and putative roles in brain function and disorders such as bilateral frontoparietal polymicrogyria (BFPP) [11], Usher syndrome [10], and attention-deficit and hyperactivity disorder (ADHD) [9] have been uncovered for some members of this class, making them interesting potential drug targets.

aGPCRs display structural features unique to their class, indicating that they signal via different mechanisms [3]. Unlike other GPCRs, they have large extracellular regions that are autoproteolytically cleaved from their seven-transmembrane (7TM) regions at a conserved GPS [12,13]. The GPS motif is well described and recognized as a vital element for receptor function [14–16]. Previously, the region preceding the GPS motif was designated as a 'stalk' or 'mucin-like' region [16,17]. Owing to the absence of crystallographic data, the structural properties of the GPS motif and the stalk initially remained unknown until recently, when the crystal structures of the region that spans the GPS motif and the stalk have been obtained from two distantly related aGPCRs. Structures of latrophilin 1 and brain angiogenesis inhibitor 3 (BAI3) have revealed a previously unidentified novel conserved domain. Surprisingly, the ~40-residue spanning GPS motif is not a separately folded domain but represents an integral part of this much larger ~320-residue domain that was termed the GAIN domain [18] (Figure 1A). Intriguingly, both the GPS motif and the stalk region, which is not unstructured but indeed composed of numerous structural elements, are constituent components of the GAIN domain. The crystal structures of the GAIN domain further showed that the GPS motif cannot function by itself, but needs to be within the context of the GAIN domain to mediate autoproteolysis.

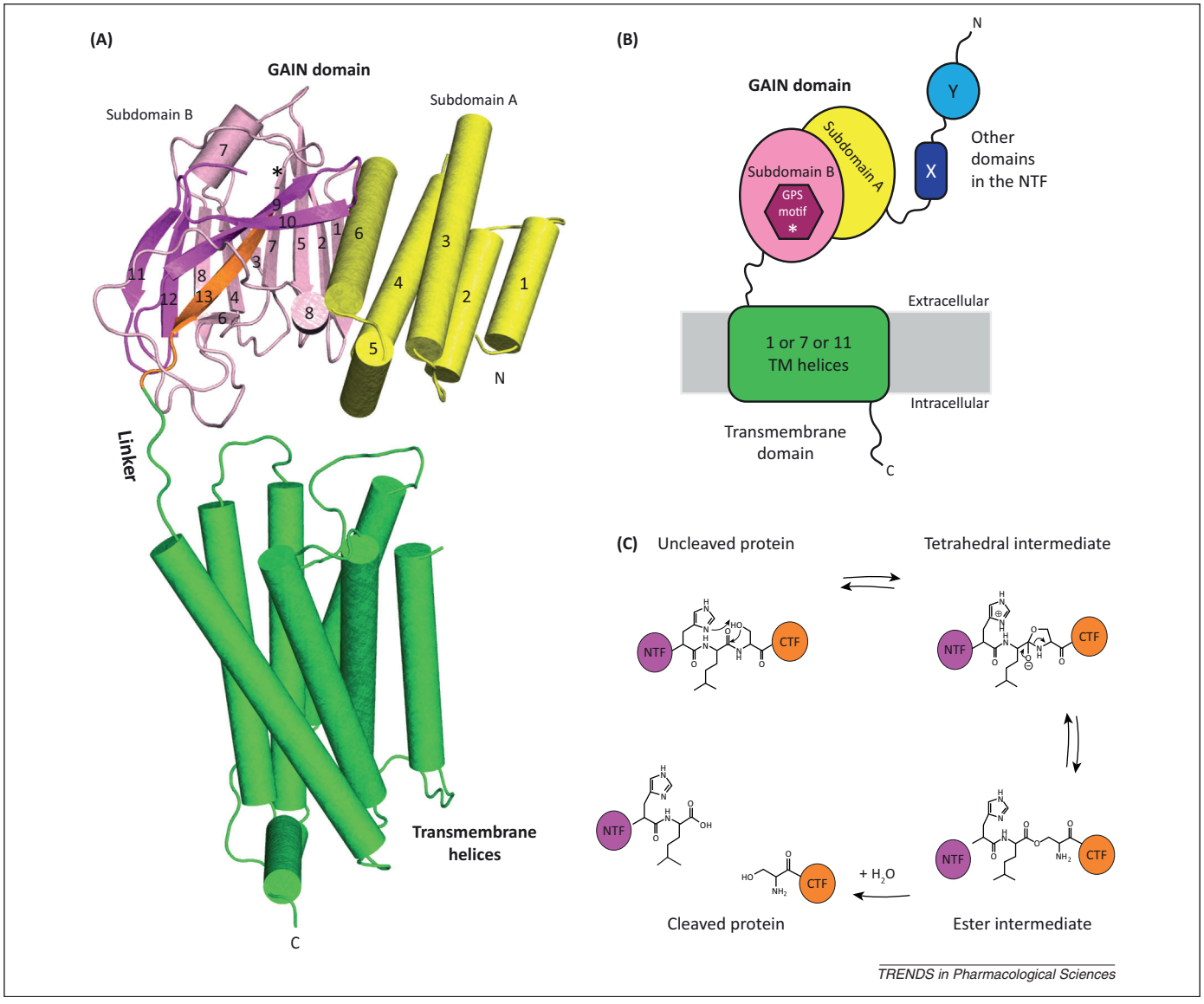

Figure 1. The GAIN domain structure. **(A)** Ribbon diagram of the rat latrophilin 1 GAIN domain and its subdomains A (yellow) and B (pink and magenta) in one possible orientation to the modeled transmembrane helices (green). The last β-strand of the GAIN domain (orange) is autoproteolyzed but stays tightly associated with the rest of the GAIN domain. The GPS motif, which is a part of subdomain B, is colored magenta. The short linker between the GAIN and the 7TM domains brings the GAIN domain in close proximity to the transmembrane helices. The seven-pass transmembrane domain (green) was modeled by homology using the crystal structure of the β2 adrenergic GPCR structure (PDB ID 2RH1). **(B)** Cartoon representation of a GAIN domain-containing protein. Note that the green box substitutes for any transmembrane domain design, to which GAIN domains are found to be appended to with 1TM, 7TM, or 11TM helices. **(C)** The proposed mechanism for the autocatalytic cleavage at the GPS. A histidine or another general base withdraws a proton from the hydroxyl group of a serine or threonine at position +1. The resulting negatively charged oxygen makes a nucleophilic attack on the carbonyl group of the residue at position –1 (e.g., a leucine), yielding a tetrahedral intermediate and subsequently an ester intermediate. This ester is then hydrolyzed to yield the cleavage products. Abbreviations: GAIN, G-protein-coupled receptor autoproteolysis-inducing domain; GPCR, G-protein-coupled receptor; GPS, GPCR proteolysis site; 1TM, single transmembrane; 7TM, seven transmembrane; 11TM, eleven transmembrane; CTF, C-terminal fragment; NTF, N-terminal fragment.

The GAIN domain is unique in that it is the only domain that exists in all members of the class of aGPCRs in humans (except GPR123 which bears no extracellular domains at all), indicating an essential role in aGPCR function. The conservation of the primary sequence of the GAIN domain decreases from the C terminus to the N terminus, and the GPS motif is the most conserved region of the domain. Despite the low sequence identity between the GAIN domains of latrophilin 1 and BAI3 (24%), the very high similarity of the domain structures indicate that the 3D structure is more strictly conserved than the primary amino acid sequence.

Intriguingly, GAIN domains are also found in all five members of human PKD1-like proteins and in sea urchin sperm receptors [19–21]. The GAIN domain spans a part of the previously described receptor for the egg jelly (REJ) homology region (of the sea urchin sperm receptor protein) and the GPS motif in these proteins [18]. PKD1 is thought to function as a mechanosensor that activates tubular differentiation in kidneys by activating intracellular signaling pathways via its interaction with PKD2 [22,23]. Approximately 90% of autosomal dominant polycystic kidney disease patients have a mutation in their PKD1 protein. Although PKD1 family proteins are not GPCRs, they also contain multiple transmembrane helices and large extracellular regions (up to 300 kDa) that are autoproteolyzed similar to aGPCRs [19]. In contrast to aGPCRs, which have 7TM helices, PKD1 has 11, whereas sea urchin

Box 1. Hallmarks of GAIN domains

- Newly discovered domain evolutionarily conserved from slime molds and tetrahymena to mammals.
- The only extracellular domain that exists in all aGPCRs throughout vertebrates and invertebrates (except GPR123), and in all members of the PKD1 protein family.
- Always immediately N-terminal to the TM domain (1, 7, or 11 TM helices) within any protein layout.
- The ~40-residue GPS motif is an integral part of the ~320-residue GAIN domain.
- The entire domain is required and sufficient for autoproteolysis.
- Does not dissociate upon autoproteolysis.
- Mutated in human diseases such as polycystic kidney disease, BFPP, and cancer.

sperm receptor has only 1. The fact that GAIN domains independently occur in three otherwise unrelated protein families suggests a common evolutionary origin of this domain and indicates a general utility of the GAIN domain in the function of these proteins, although this biological function is elusive at the moment.

Moreover, primitive organisms that arose early in evolution before animals emerged (such as slime molds and tetrahymena) encode GAIN domains in spite of their lack of most other autoproteolytic domains and signaling proteins [16]. However, whereas in higher organisms GAIN domains appear in context with 7TM or 11TM domains, in aGPCRs and PKD1-like molecules, respectively, lower metazoa contain aGPCR precursors with very short N-terminal regions and no GAIN domains at all [24]. Instead, in these simple organisms, GAIN domains feature in other protein architectures. For example *Dictyostelium discoideum*, an early ancestor to animals, encodes a protein (accession number Q54XQ9) with a GAIN domain N-terminal to a single transmembrane helix. These observations show that the GAIN domain is widespread and conserved in higher eukaryotes as well as in ancient organisms, and resides N-terminal to a variety of transmembrane domain layouts.

Intriguingly, however, a GAIN domain always immediately precedes the first transmembrane helix by a short linker irrespective of the overall protein and transmembrane architecture, raising the possibility that the GAIN domain regulates receptor signaling via intramolecular interactions with the nearby transmembrane helices (Figure 1A, see below). The signaling pathways, the mechanism of activation, and specific interactions probably vary in different GAIN domain-containing proteins. A summary of the hallmark features of GAIN domains are presented in Box 1.

The GAIN domain: a complex fold sufficient and necessary for autoproteolysis

The GAIN domain of the aGPCR latrophilin 1 contains an N-terminal subdomain A that is composed of six α-helices, and a C-terminal subdomain B that consists of a twisted β-sandwich including 13 β-strands and 2 small α-helices, of which the last 5 β-strands constitute the GPS motif (Figure 1A,B).

Studies of the GAIN domain show that it is both required and sufficient for autoproteolysis [13,18]. The self-cleavage occurs at the GPS and takes place in the endoplasmic reticulum [14,25]. An autoproteolysis mechanism has previously been suggested at first for the aGPCR EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2) [13] and for PKD1 [26] based on the sequence similarity of the cleavage site to the N-terminal nucleophile hydrolases. The structural details of the GAIN domain are consistent with this mechanism; however, more evidence, based on biochemical, structural, and spectrophotometric studies, is needed to confirm the validity of the suggested autoproteolytic reaction. The proposed mechanism suggests that a general base at position –2 respective to the cleavage site (such as a histidine in rat latrophilin 1 or a water molecule) retracts a proton from the hydroxyl group of a threonine/serine at position +1 to yield a negatively charged oxygen (Figure 1C) [13,26]. This oxygen makes a nucleophilic attack on the carbonyl carbon of the residue at position –1 (a leucine in rat latrophilin 1) forming a tetrahedral intermediate followed by an ester intermediate that eventually produces the cleaved protein. In most GAIN domains, the cleavage site is HL↓T/S. Autoproteolysis can occur only if the residue at position +1 is a threonine, serine, or cysteine. However, the residues at positions –2 and –1 may vary in different GAIN domains.

The two GAIN domain structures, one in the cleaved form (latrophilin 1) and the other in the uncleaved form (BAI3), allow insights in the unique structural features that enable autoproteolysis [18]. The scissile bond is located in a kinked, short loop between the last two β-strands of the GAIN domain and lies close to the core of the protein away from the surface (Figure 1A). For autoproteolysis to occur, first, the GAIN domain has to fold properly and form the right chemical environment around the scissile bond. Numerous residues and structural features of the GAIN domain need to be properly positioned to enable self-cleavage, and small variations of sequence and structure between different GAIN domains probably influence the efficiency of autoproteolysis. Intriguingly, the cleaved latrophilin 1 structure shows that autoproteolysis does not cause the dissociation of the cleaved β-strand from the rest of the GAIN domain [18]. Dissociation is prevented by the strong hydrogen bonds and hydrophobic interactions the cleaved β-strand is involved in. This is confirmed by experimental evidence, which found that a high fraction of cleaved aGPCRs and PKD1-like molecules exist as homogeneric dimers at the plasma membrane of the transfected cells [12,18,20]. Structural considerations support the lack of aGPCR fragment dissociation after autoproteolysis, as removal of the cleaved β-strand from the rest of the GAIN domain is energetically unfavorable because it will leave a large hole in the core of the protein that will render the domain unstable and probably cause it to collapse instantly and irreversibly onto itself [18]. Finally, in contrast to GAIN domain-carrying proteins, precedent examples of autoproteolytic proteins are cleaved in a linker region between two domains, rather than in the core of a domain, with the purpose of separating them (i.e., hedgehog, N-terminal nucleophilic hydrolases, inteins) [27,28].

By contrast, other studies suggest that the N-terminal fragment (NTF) and C-terminal fragment (CTF) of aGPCRs, which originate from self-cleavage, are able to

dissociate under specific conditions as reassociation of NTF with CTF seems reversible and might be triggered by ligand binding [29,30]. Hence, the functional consequences of autoproteolysis in the core of the GAIN domain without it resulting in fragment dissociation require further investigations in the future.

Autoproteolysis at the GPS motif: one way of controlling receptor activity?

The biological function of self-cleavage within the GAIN domain is controversial. Cleavage in close proximity to the membrane, which may release the NTF, yields a wide spectrum of possibilities to control receptor activity. Cleavage could be necessary for receptor processing and stability, or could serve to regulate receptor transport and expression [31]. Moreover, it is conceivable that liberation of the N terminus could expose an interaction site for binding partners or lead to receptor activation/inhibition and thus signaling, similar to protease-activated receptors (PARs) of the GPCR superfamily [32]. However, at present, there is no evidence supporting this signaling model for aGPCRs.

Several research groups have argued that cleavage at the GPS is essential for receptor trafficking. Mutations at the cleavage site were reported to lead to impaired protein processing and membrane trafficking [20,25,31]. However, it is important to determine whether faulty processing is due to abolished cleavage or whether an individual mutation impairs the domain structure (e.g., by unfolding the protein; see example substitution mutations that impair autoproteolysis but also abolish proper domain folding in [18]). Indeed, it was shown that uncleaved but properly folded forms of BAI3 and latrophilin 1 can properly localize to the plasma membrane demonstrating that autoproteolysis is not needed for trafficking [18]. It is possible that the GPS as a structural motif of the GAIN domain might be essential for receptor signaling rather than mediating autoproteolysis only.

It has been suggested that cleavage is necessary for some receptors to mediate receptor function. For example, cleavage of the aGPCR EMR2 is essential for correct migration of HT1080 cells *in vitro* [29]. Moreover, *Pkd1* knock-in mice with disabled cleavage in polycystin-1 die from uremia shortly after birth caused by cystic dilation of the kidneys. However, only a subset of defects of *Pkd1* knockout mice [33] is observed in cleavage-deficient mice. Although these studies demonstrate the importance of autoproteolysis at the GPS for distinct aspects of receptor activity, the latter one suggests that it is not essential for the whole spectrum of its biological functions. This has also been shown for the latrophilin homolog LAT-1 in *Caenorhabditis elegans*, in which abolished autoproteolysis does not affect receptor function [34]. Moreover, signaling of the aGPCR GPR133 via a G protein is independent of self-cleavage [35]. Further, aGPCRs that are not cleaved at the GPS motif at all, such as GPR111 and GPR115 [36], support the notion that self-cleavage is not the sole role of the GAIN domain. Sequence analyses revealed that two human aGPCRs (GPR124 and GPR125) lack the threonine/serine residue at +1 position required for cleavage, strongly suggesting that these aGPCRs are not autoproteolytically processed by default [34].

Because autoproteolysis is rarely fully efficient, both versions – cleaved and noncleaved – can be detected in sampled tissues [18,37,38]. The ability to self-cleave and thus the cleavage efficiency does not only depend on the GAIN domain but also on the receptor and the cell type [18,37,39]. A naturally occurring variation in cleavage efficiency, which could be subject to receptor glycosylation [39], might indicate a possible means and need for cleavage regulation, which in turn may tune aGPCR or PKD1 signaling. It is probable that each GAIN domain conformation could have a specific cellular function and the ratio of a cleaved to an uncleaved receptor may be essential for the fine balance of aGPCR-dependent cellular signaling or adhesion.

Another possible means to modulate aGPCR activity is the liberation of the NTF after autoproteolysis. Removal of NTF by external stimuli (e.g., shear stress as in the case of the aGPCR CD97) can result in internalization of the remaining CTF and thus abrogate receptor signaling similar to PARs [40–42]. In another scenario, the NTF might mediate cell non-autonomous functions that are independent of the CTF and are located at a distance from the aGPCR-expressing cell [14,34,43–45].

In summary, cleavage at the GAIN domain is not a feature that all GAIN domain-containing receptors possess. Thus, autoproteolysis might not be the only function of the GAIN domain. Cleavage may be essential for receptor activity and its regulation in some receptors only, whereas in others it may not occur or be dispensable. As the GPS motif is highly conserved and a persistent strong selective pressure can be detected [34], the motif on its own or as part of the GAIN domain appears essential for receptor function.

One possible reason to account for the evolutionary conservation of the GAIN domain is its potential role as an interaction interface with partner molecules. Previous research demonstrated that the GAIN domains of several aGPCRs can directly engage with partner molecules. Latrophilin 1 is the receptor for α-latrotoxin, a component of the black widow spider toxin that triggers massive neurotransmitter release from nerve terminals and neuroendocrine cells [12,46–48]. A recent study showed that the GAIN domain of latrophilin 1 is the region that mediates α-latrotoxin binding [18], although evidence for an endogenous binding partner of the GAIN domain is still lacking. GPR56 has one of the shortest extracellular domains among all aGPCRs, containing only a GAIN domain preceded by a short, 100-residue long region that appears to be unstructured. The binding sites for the two reported interaction partners of GPR56 (collagen III [49] and transglutaminase 2 [50]) map onto its GAIN domain. It can be speculated that ligand binding to the GAIN domain of GPR56 may lead to downstream signaling. Approximately a dozen interaction partners for aGPCRs have been identified to date, but not all interaction sites on the receptors have been determined [51–59]. It will be interesting to see where these and newly discovered partners will bind and how binding affects receptor signaling.

GAIN domain in human disease

Several studies showed that the human GAIN domain is a hotspot for disease-associated mutations (Figure 2)

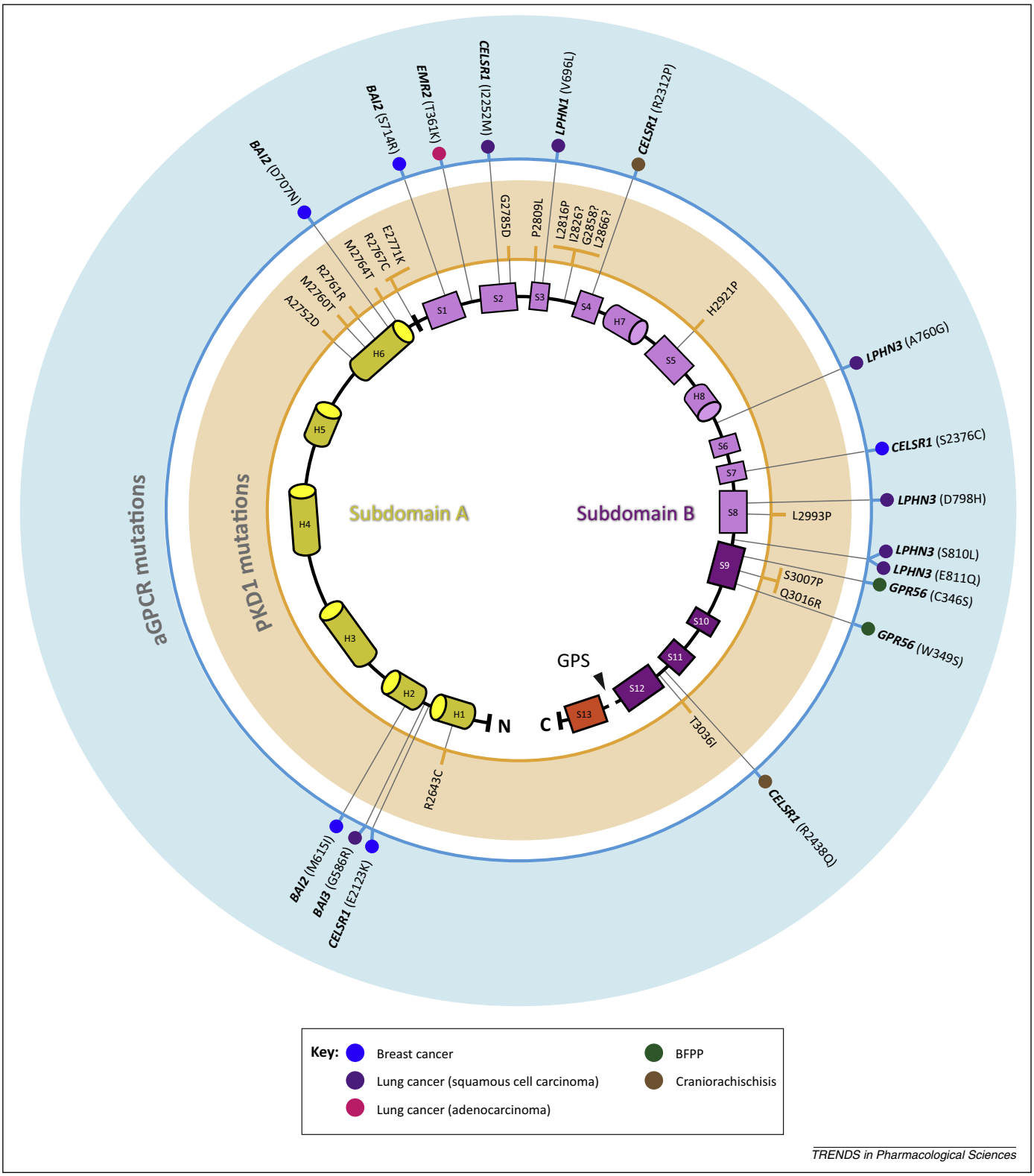

Figure 2. Disease-associated mutations in the GAIN domain. Positions of point mutations in aGPCRs (outer shell) and PKD1 (middle shell) associated with human diseases (see legend) mapped onto the GAIN domain secondary structural elements, that is, α-helices (H) and β-strands (S, inner shell). Color scheme corresponds to Figure 1A. Two hotspots of mutations can be identified: one at the border of subdomain A (helix 6) and subdomain B (strands 1–4), the other in subdomain B around the border of the GPS motif (strands 8 and 9). Abbreviations: BFPP, bilateral frontoparietal polymicrogyria; GAIN, G-protein-coupled receptor autoproteolysis-inducing domain; aGPCR, Adhesion-G-protein-coupled receptor; GPS, GPCR proteolysis site; PKD1, polycystic kidney disease 1.

[3,10,11,60–64]. Missense mutations in GPR56 cause BFPP, a hereditary disease causing severe malformation of the cerebral cortex [11,63]. Mutations in the GAIN domain of CELSR1 lead to craniorachischisis, the result of failure in neural tube closure [64]. Furthermore, point mutations in the PKD1 GAIN domain cause polycystic kidney disease [65–67]. Also various types of cancer have been associated with point mutations in the GAIN domain in members of the BAI, EMR, CELSR, and latrophilin aGPCR families [61]. Interestingly, the majority of the mutations are clustered in

two discrete areas of the GAIN domain (Figure 2, our own analysis). The first cluster is located at the border of subdomain A (helix 6) and subdomain B (strands 1–4), the second one is found in subdomain B at the border of the GPS motif (strands 8 and 9). Although disease causality has not been tested for every mutation, the significant clustering of these mutations within the GAIN domain highlights the importance of these regions. It can be speculated that the mutations impair binding of interaction partners or abolish structural features essential for signaling because only a few of the described missense mutations have been shown to affect protein processing or membrane trafficking, respectively [11,18,31,63,64]. Genetically, however, it is unclear thus far whether this peculiar mutation profile reflects particularly vulnerable regions of the GAIN domain (with high selective pressure on them) leading to receptor failure and consequently cell dysfunction. Alternatively, the sampled mutations might enrich because of the low selective pressure on these domain regions, and carriers survive their mutational load as the receptor function remains fully or partially intact. Mutations in other areas of the GAIN domain might thus not be detectable due to strong negative selection of affected receptor mutants leading to zygotic lethality. This will be an important aspect to follow up by bioinformatical and phenotype analyses.

Although the disease mutations in the PKD1 GAIN domain abolish autoproteolysis either directly by interfering with the proteolytic reaction or indirectly by misfolding the protein, the cancer mutations on latrophilins and BAIs have no effect on autoproteolysis or protein trafficking [18], suggesting that the mechanism through which GAIN domain mutations cause disease is different. It is likely that the GAIN domain might be involved in more than one function (such as autoproteolysis, ligand binding, interaction with transmembrane helices, etc.; see below for further possible roles of the GAIN domain). Thus, pharmacological approaches to design drugs targeting the GAIN domain may need to be customized for each GAIN domain-containing protein (Box 2).

The GAIN domain as the central element in models of receptor activity

Three properties of the GAIN domain make it unique and intriguing. (i) All human aGPCRs (except GPR123) have a GAIN domain and no other extracellular domain is shared by all receptor class members. (ii) All GAIN domains analyzed thus far always immediately precede the N-terminal transmembrane helix by a short linker and are in close association with the signaling transmembrane domains. (iii) In contrast to most other autoproteolytic domains, upon autoproteolysis, the GAIN domain remains attached to the membrane-embedded regions of the protein [14,16,18,30,68]. These observations naturally lead to the hypothesis that the GAIN domain may regulate receptor signaling via intramolecular interactions with the transmembrane helices of the receptor [18,34,69,70]. Intriguingly, it has also been suggested by several groups that the 7TM domain of aGPCRs is a GAIN domain interactor, posing the question whether the GAIN domain can inhibit or activate receptor signaling [18,34,70].

Several biological models have emerged recently to describe aGPCR activity and the potential impact of the GAIN domain, which differ from canonical GPCR activation scenarios. The best-studied and structurally most-understood model for the activation of GPCRs, mainly the rhodopsin class receptors, suggests that a ligand binds to a groove formed by the extracellular loops between the 7TM helices and induces a conformational change that leads to intracellular signaling [68,69]. Such ligands can be a small compound, odor, hormone, neurotransmitter, etc. A second well-understood GPCR activation mechanism is the two-domain model of secretin/class B GPCR activation, in which the C terminus of the peptide binds to the extracellular hormone binding domain of the receptor and promotes the interaction of the N terminus of the peptide hormone with the transmembrane helices, leading to activation of the receptor [71–73] An exception to these commonly accepted mechanisms is the PAR and the thyrotropin receptor of the rhodopsin family that requires the cleavage and dissociation of the extracellular domain by an exogenous protease such as thrombin [72].

Recently, additional activation mechanisms via the extracellular domains of GPCRs have been reported and the idea of extracellular domains regulating receptor activation either in an inhibitory or excitatory manner was reiterated. For example, in the melanocortin-4 receptor [73] and the GPR61 receptor [74] of the rhodopsin class, the extracellular domain acts as an intramolecular agonist of the receptor keeping it constitutively active. By contrast, in the glucagon receptor of the secretin class, the extracellular domain acts as an inhibitor of activation due to its interaction with the third extracellular loop of the 7TM domain [75].

Activity of a few GPCRs in the adhesion class was suggested to be subject to regulation by their extracellular domains. For example, removal of the entire extracellular region of GPR56 triggers enhanced signaling ability compared with the wild type GPR56, and thus the extracellular region may have an inhibitory role in receptor activation consistent with a tethered negative regulator model (Figure 3A) [70,76]. However, whether such direct inhibitory function of the GAIN domain is a general feature of aGPCR modulation or whether conversely the GAIN domain can activate 7TM-dependent signals as in other GPCRs (a tethered positive regulator model, Figure 3B) is currently debated. Alternatively, the GAIN domain may indirectly regulate binding of possible small ligands to

Box 2. Pharmacological perspectives on GAIN domain modulation

GAIN domains are potential mediators of a multitude of actions in their native protein architectures. Consequently, drugs ought to interfere with either of these functions by:

- binding to GAIN domain surface areas that
  - (i) interact with the transmembrane helices
  - (ii) interact with GAIN domain ligands
  - (iii) interact with other extracellular domains within the NTF of aGPCRs or PKD1-like molecules (e.g., the HormR domain)
  - (iv) interfere with autoproteolytic cleavage
- binding to GAIN-interacting structures such as
  - (i) the extracellular loops of the transmembrane helices
  - (ii) ligands
  - (iii) extracellular domains within the NTF of aGPCRs or PKD1-like molecules.

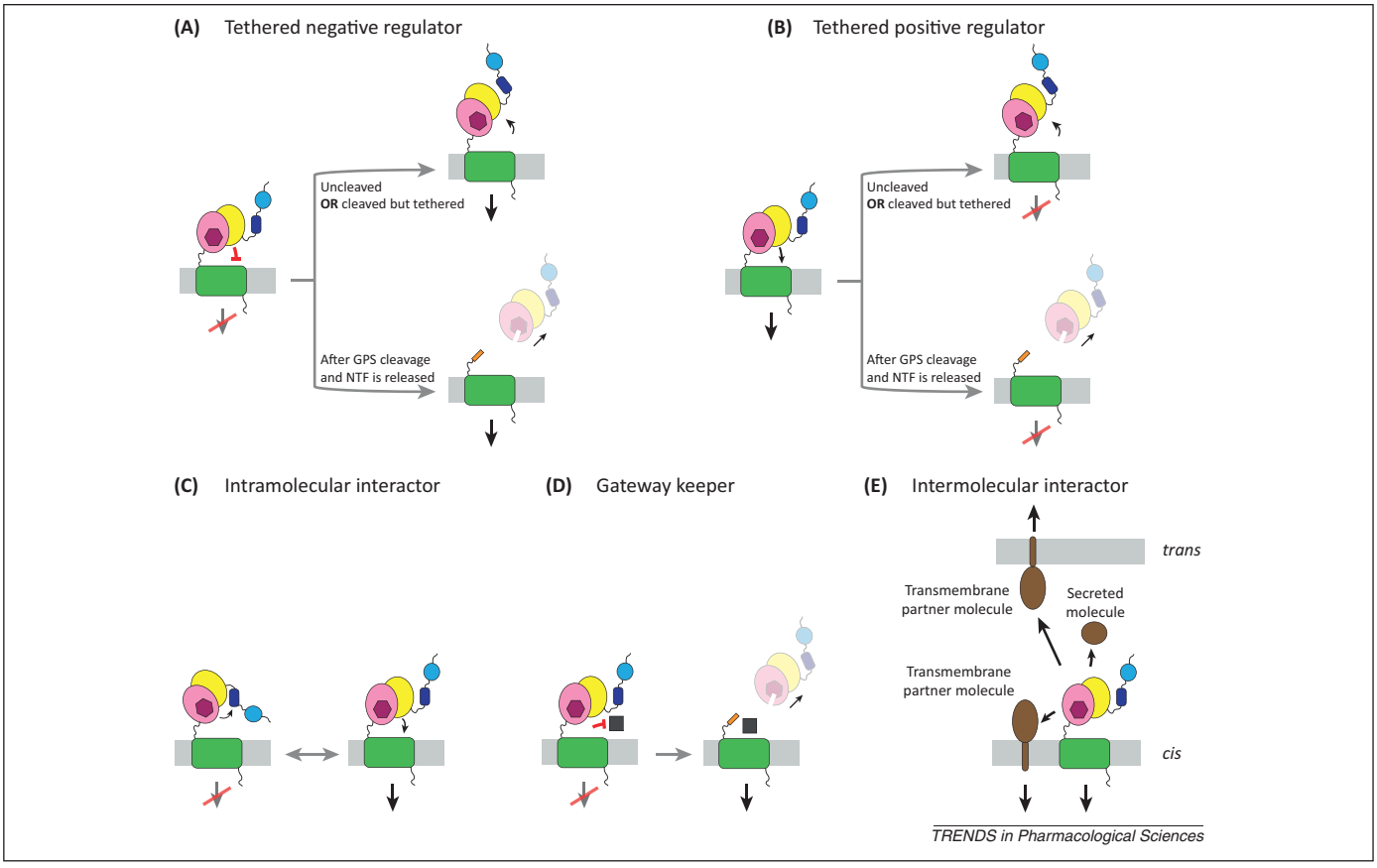

Figure 3. Possible models for the role of the GAIN domain in receptor signaling. Various signaling mechanisms have been proposed. Other signaling mechanisms are also possible. **(A)** The GAIN domain interacts with the 7TM domain serving as a tethered negative regulator. Cleavage or conformational changes, possibly by binding of another molecule, releases the interaction and the receptor is able to signal [70,76]. **(B)** Conversely, the GAIN domain might act as a tethered positive regulator of the 7TM domain leading to receptor signaling. By cleavage or through conformational changes, this interaction is disrupted and signaling is terminated [34]. **(C)** The GAIN domain may have intramolecular interactions with the other components in the extracellular region of the receptor. Such intramolecular interactions may regulate GAIN domain function or vice versa. Support for this model is provided by the crystal structures of HormR/GAIN domains from two divergent receptors [18]. **(D)** The presence of the GAIN domain nearby the extracellular loops of the transmembrane helices may block the binding of ligands to the transmembrane helices and suppress receptor activation. This model is currently unsupported by experimental evidence. **(E)** The GAIN domain is an interaction partner for another molecule transducing the signal upon GAIN domain contact. This potential interaction partner may be another transmembrane protein in the same or neighboring cell, or it may be a secreted protein. Binding of this partner to the GAIN domain may also regulate aGPCR signaling [34,49]. Abbreviations: GAIN, G-protein-coupled receptor autoproteolysis-inducing domain; 7TM, seven transmembrane; aGPCR, Adhesion-G-protein-coupled receptor; HormR, hormone binding domain; NTF, N-terminal fragment.

transmembrane helices via steric hindrance (gateway keeper model, Figure 3D). Another possibility is that the GAIN domain can bind to protein ligands in the extracellular matrix (*cis* or *trans*) and couple adhesion to receptor signaling (intermolecular interactor model, Figure 3E). These ligands may directly affect receptor activation. It is also possible that intramolecular interaction of the GAIN domain with other extracellular domains, such as the nearby hormone binding (HormR) domain as observed in the crystal structures of both latrophilin 1 and BAI3 [18], might be another way for GAIN domains to regulate aGPCR function (intramolecular interactor model, Figure 3C).

Research on the latrophilin homolog LAT-1 in *C. elegans* has extended the above models, which suppose that the 7TM domain-bearing CTF is the sole signal generator of aGPCRs similar to rhodopsin class GPCRs, and that NTF–CTF interplay governs control of 7TM activity. In addition to this CTF-dependent function, the membrane-anchored extracellular domain of LAT-1 containing the GAIN domain is also required for an activity that is independent from the CTF (Figure 3E) [34] (e.g., for the engagement with a coreceptor on the same cell membrane or to activate receptors on the surface of neighboring cells). Hence, the GAIN domain with its GPS motif acts as a bimodal switch for at least two receptor activities [34].

Finally, the 'split personality' model adds another intriguing twist to the conceptual ideas on aGPCR signaling by suggesting that reassociation of NTF with CTF is triggered by ligand binding [20,29]. The 'split personality' hypothesis suggests that a cleaved NTF of a given aGPCR is not exclusively bound to re-engage with its native CTF, but can also interact with the CTF of other aGPCRs [77]. For example, latrophilins can form functionally active chimeric complexes with GPR56. Thus, cleavage at the GPS might lead to a large recombinatorial repertoire of hybrid aGPCR variants, thereby connecting various receptor inputs to flexible outputs via intracellular signaling pathways. However, this cross-interaction does not seem to be applicable for all aGPCRs [34]. Further, it remains to be determined how the homogeneric or heterogeneric NTF–CTF reassociation occurs on a molecular level, as the crystal structure of the GAIN domain strongly suggests that reassociation of the cleaved products is physically and energetically not possible because the GAIN domain is destined to become destabilized irreversibly, if the cleaved strand is removed from its core [18].

Certainly, even more complex models on GAIN domain function in the context of aGPCR signaling are feasible, which require further investigations. These may involve but are not limited to allosteric pathways and removal of the extracellular sequences upon ligand binding.

The GAIN domain as a pharmacological target

Importantly, the emerging understanding on aGPCR roles in diseases and their signaling mechanisms has begun to open perspectives of clinical relevance on this under-researched GPCR class. These opportunities promise great pharmacological potential given that GPCRs are the target for approximately 40% of all drugs in the pharmaceutical industry [2]. As the GAIN domain very likely plays a critical role for the signaling of its native proteins, aGPCRs or PKD1-like molecules, it is destined to become a central target for drugs designed to interfere with their activity. Our current state of knowledge on GAIN domain functions allows for a scope of concepts to modulate GAIN-bearing protein signaling (Box 2). One major consideration in these efforts will be the low conservation of surface residues between GAIN domains, whereas their secondary and tertiary structural architectures appear highly conserved. This constitutes challenges and opportunities. Challenges, because GAIN domains probably bind different ligands via their different surfaces that trigger the execution of similar actions through their archetypical domain layout, rendering the search for general GAIN domain modulators tedious or obsolete. Opportunities, as the variety of GAIN domain interfaces will help to achieve pharmacological selectivity in the modulation of individual GAIN domains and ensure a low degree of interference with nontargeted GAIN-mediated signals. Random library searches for interactors might thus be a good starting point for drug design.

Concluding remarks

The GAIN domain is a newly discovered fold hallmarking two transmembrane protein families: aGPCRs and PKD1-like proteins. Until recently, only its integral part, the GPS motif, and cleavage at this motif have been the focus of research. The GAIN domain comprises a unique and highly conserved domain located in close proximity to the membrane, and therefore it is conceivable that it is involved in novel aspects of transmembrane signaling. The GAIN domain structure and initial functional characterization provides the chance for a better understanding of the signaling mechanisms of both protein families. Their members play essential roles in biological processes and mutations in their genes can lead to severe pathologies. Thus, understanding the structure–function relationships encoded in the GAIN domain bears great potential to uncover the role of aGPCRs and PKD1-like proteins in a wide scope of physiological duties and the etiology of associated diseases.

Acknowledgments

This work was funded by a junior research grant from the Medical Faculty, University of Leipzig (977000-145) to S.P. T.L acknowledges support from the Deutsche Forschungsgemeinschaft (DFG LA 2861/1-1; SFB 1047/1 A05) and the IZKF Würzburg (Z-3/12).

References

1 Fredriksson, R. *et al.* (2003) There exist at least 30 human G-protein-coupled receptors with long Ser/Thr-rich N-termini. *Biochem. Biophys. Res. Commun.* 301, 725–734
2 Lagerstrom, M.C. and Schioth, H.B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* 7, 339–357
3 Langenhan, T. *et al.* (2013) Sticky signaling – adhesion class G protein-coupled receptors take the stage. *Sci. Signal.* 6, re3
4 Galle, J. *et al.* (2006) Individual cell-based models of tumor–environment interactions: multiple effects of CD97 on tumor invasion. *Am. J. Pathol.* 169, 1802–1811
5 Hamann, J. *et al.* (1996) Structure of the human CD97 gene: exon shuffling has generated a new type of seven-span transmembrane molecule related to the secretin receptor superfamily. *Genomics* 32, 144–147
6 Lin, H.H. *et al.* (2005) The macrophage F4/80 receptor is required for the induction of antigen-specific efferent regulatory T cells in peripheral tolerance. *J. Exp. Med.* 201, 1615–1625
7 Langenhan, T. *et al.* (2009) Latrophilin signaling links anterior–posterior tissue polarity and oriented cell divisions in the *C. elegans* embryo. *Dev. Cell* 17, 494–504
8 Tissir, F. *et al.* (2005) Protocadherin Celsr3 is crucial in axonal tract development. *Nat. Neurosci.* 8, 451–457
9 Arcos-Burgos, M. *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
10 Bonnet, C. *et al.* (2011) Complete exon sequencing of all known Usher syndrome genes greatly improves molecular diagnosis. *Orphanet J. Rare Dis.* 6, 21
11 Piao, X. *et al.* (2004) G protein-coupled receptor-dependent development of human frontal cortex. *Science* 303, 2033–2036
12 Krasnoperov, V.G. *et al.* (1997) α-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron* 18, 925–937
13 Lin, H.H. *et al.* (2004) Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *J. Biol. Chem.* 279, 31823–31832
14 Gray, J.X. *et al.* (1996) CD97 is a processed, seven-transmembrane, heterodimeric receptor associated with inflammation. *J. Immunol.* 157, 5438–5447
15 Shashidhar, S. *et al.* (2005) GPR56 is a GPCR that is overexpressed in gliomas and functions in tumor cell adhesion. *Oncogene* 24, 1673–1682
16 Stacey, M. *et al.* (2002) EMR4, a novel epidermal growth factor (EGF)-TM7 molecule up-regulated in activated mouse macrophages, binds to a putative cellular ligand on B lymphoma cell line A20. *J. Biol. Chem.* 277, 29283–29293
17 Fredriksson, R. *et al.* (2002) Novel human G protein-coupled receptors with long N-terminals containing GPS domains and Ser/Thr-rich regions. *FEBS Lett.* 531, 407–414
18 Araç, D. *et al.* (2012) A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO J.* 31, 1364–1378
19 Ponting, C.P. *et al.* (1999) A latrophilin/CL-1-like GPS domain in polycystin-1. *Curr. Biol.* 9, R585–R588
20 Qian, F. *et al.* (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. U.S.A.* 99, 16981–16986
21 Hughes, J. *et al.* (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
22 Gabow, P.A. (1993) Autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* 329, 332–342
23 Hughes, J. *et al.* (1995) The polycystic kidney disease 1 (*PKD1*) gene encodes a novel protein with multiple cell recognition domains. *Nat. Genet.* 10, 151–160
24 Krishnan, A. *et al.* (2012) The origin of GPCRs: identification of mammalian like rhodopsin, adhesion, glutamate and frizzled GPCRs in fungi. *PLoS ONE* 7, e29817
25 Krasnoperov, V. *et al.* (2002) Post-translational proteolytic processing of the calcium-independent receptor of α-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
26 Wei, W. *et al.* (2007) Characterization of cis-autoproteolysis of polycystin-1, the product of human polycystic kidney disease 1 gene. *J. Biol. Chem.* 282, 21729–21737
27 Lee, J.J. *et al.* (1994) Autoproteolysis in hedgehog protein biogenesis. *Science* 266, 1528–1537
28 Perler, F.B. *et al.* (1997) Protein splicing and autoproteolysis mechanisms. *Curr. Opin. Chem. Biol.* 1, 292–299
29 Huang, Y.S. *et al.* (2012) Activation of myeloid-specific adhesion-GPCR EMR2 via ligation-induced translocation and interaction of receptor subunits in lipid raft microdomains. *Mol. Cell. Biol.* 32, 1408–1420
30 Volynski, K.E. *et al.* (2004) Latrophilin fragments behave as independent proteins that associate and signal on binding of LTX(N4C). *EMBO J.* 23, 4423–4433
31 Jin, Z. *et al.* (2007) Disease-associated mutations affect GPR56 protein trafficking and cell surface expression. *Hum. Mol. Genet.* 16, 1972–1985
32 Coughlin, S.R. (2000) Thrombin signalling and protease-activated receptors. *Nature* 407, 258–264
33 Yu, S. *et al.* (2007) Essential role of cleavage of polycystin-1 at G protein-coupled receptor proteolytic site for kidney tubular structure. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18688–18693
34 Prömel, S. *et al.* (2012) The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. *Cell Rep.* 2, 321–331
35 Bohnekamp, J. and Schoneberg, T. (2011) Cell adhesion receptor GPR133 couples to Gs protein. *J. Biol. Chem.* 286, 41912–41916
36 Prömel, S. *et al.* (2012) Characterization and functional study of a cluster of four highly conserved orphan adhesion-GPCR in mouse. *Dev. Dyn.* 241, 1591–1602
37 Iguchi, T. *et al.* (2008) Orphan G protein-coupled receptor GPR56 regulates neural progenitor cell migration via a G α 12/13 and Rho pathway. *J. Biol. Chem.* 283, 14469–14478
38 Usui, T. *et al.* (1999) Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled. *Cell* 98, 585–595
39 Hsiao, C.C. *et al.* (2009) Site-specific N-glycosylation regulates the GPS auto-proteolysis of CD97. *FEBS Lett.* 583, 3285–3290
40 Karpus, O.N. *et al.* (2013) Shear stress-dependent downregulation of the adhesion-G protein-coupled receptor CD97 on circulating leukocytes upon contact with its ligand CD55. *J. Immunol.* 190, 3740–3748
41 Trejo, J. *et al.* (1998) Termination of signaling by protease-activated receptor-1 is linked to lysosomal sorting. *Proc. Natl. Acad. Sci. U.S.A.* 95, 13698–13702
42 Hammes, S.R. *et al.* (1999) Shutoff and agonist-triggered internalization of protease-activated receptor 1 can be separated by mutation of putative phosphorylation sites in the cytoplasmic tail. *Biochemistry* 38, 9308–9316
43 Kaur, B. *et al.* (2005) Vasculostatin, a proteolytic fragment of brain angiogenesis inhibitor 1, is an antiangiogenic and antitumorigenic factor. *Oncogene* 24, 3632–3642
44 Kaur, B. *et al.* (2009) Vasculostatin inhibits intracranial glioma growth and negatively regulates in vivo angiogenesis through a CD36-dependent mechanism. *Cancer Res.* 69, 1212–1220
45 Vallon, M. and Essler, M. (2006) Proteolytically processed soluble tumor endothelial marker (TEM) 5 mediates endothelial cell survival during angiogenesis by linking integrin α(v)β3 to glycosaminoglycans. *J. Biol. Chem.* 281, 34179–34188
46 Lelianova, V.G. *et al.* (1997) α-Latrotoxin receptor, latrophilin, is a novel member of the secretin family of G protein-coupled receptors. *J. Biol. Chem.* 272, 21504–21508
47 Sugita, S. *et al.* (1998) α-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
48 Krasnoperov, V. *et al.* (1999) Structural requirements for α-latrotoxin binding and α-latrotoxin-stimulated secretion. A study with calcium-independent receptor of α-latrotoxin (CIRL) deletion mutants. *J. Biol. Chem.* 274, 3590–3596
49 Luo, R. *et al.* (2011) G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12925–12930
50 Xu, L. *et al.* (2006) GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9023–9028
51 Bolliger, M.F. *et al.* (2011) The cell-adhesion G protein-coupled receptor BAI3 is a high-affinity receptor for C1q-like proteins. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2534–2539
52 Boucard, A.A. *et al.* (2012) High-affinity neurexin binding to the cell-adhesion G-protein coupled receptor CIRL1/Latrophilin-1 produces an intercellular adhesion complex. *J. Biol. Chem.* 287, 9399–9413
53 Hamann, J. *et al.* (1998) Characterization of the CD55 (DAF)-binding site on the seven-span transmembrane receptor CD97. *Eur. J. Immunol.* 28, 1701–1707
54 Hamann, J. *et al.* (1996) The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF). *J. Exp. Med.* 184, 1185–1189
55 O'Sullivan, M.L. *et al.* (2012) FLRT proteins are endogenous latrophilin ligands and regulate excitatory synapse development. *Neuron* 73, 903–910
56 Park, D. *et al.* (2007) BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature* 450, 430–434
57 Silva, J.P. *et al.* (2011) Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12113–12118
58 Wandel, E. *et al.* (2012) Thy-1 (CD90) is an interacting partner for CD97 on activated endothelial cells. *J. Immunol.* 188, 1442–1450
59 Wang, T. *et al.* (2005) CD97, an adhesion receptor on inflammatory cells, stimulates angiogenesis through binding integrin counterreceptors on endothelial cells. *Blood* 105, 2836–2844
60 Besnard, T. *et al.* (2012) Non-USH2A mutations in USH2 patients. *Hum. Mutat.* 33, 504–510
61 Kan, Z. *et al.* (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466, 869–873
62 Ke, N. *et al.* (2008) Biochemical characterization of genetic mutations of GPR56 in patients with bilateral frontoparietal polymicrogyria (BFPP). *Biochem. Biophys. Res. Commun.* 366, 314–320
63 Piao, X. *et al.* (2005) Genotype–phenotype analysis of human frontoparietal polymicrogyria syndromes. *Ann. Neurol.* 58, 680–687
64 Robinson, A. *et al.* (2012) Mutations in the planar cell polarity genes *CELSR1* and *SCRIB* are associated with the severe neural tube defect craniorachischisis. *Hum. Mutat.* 33, 440–447
65 Li, H. *et al.* (2003) PDGF-D is a potent transforming and angiogenic growth factor. *Oncogene* 22, 1501–1510
66 Qian, F. *et al.* (1997) PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat. Genet.* 16, 179–183
67 Reynolds, D.M. *et al.* (1999) Aberrant splicing in the *PKD2* gene as a cause of polycystic kidney disease. *J. Am. Soc. Nephrol.* 10, 2342–2351
68 Rosenbaum, D.M. *et al.* (2007) GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *Science* 318, 1266–1273
69 Rasmussen, S.G. *et al.* (2011) Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* 477, 549–555
70 Paavola, K.J. *et al.* (2011) The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *J. Biol. Chem.* 286, 28914–28921
71 Hoare, S.R. (2005) Mechanisms of peptide and nonpeptide ligand binding to class B G-protein-coupled receptors. *Drug Discov. Today* 10, 417–427
72 Traynelis, S.F. and Trejo, J. (2007) Protease-activated receptor signaling: new roles and regulatory mechanisms. *Curr. Opin. Hematol.* 14, 230–235
73 Ersoy, B.A. *et al.* (2012) Mechanism of N-terminal modulation of activity at the melanocortin-4 receptor GPCR. *Nat. Chem. Biol.* 8, 725–730
74 Toyooka, M. *et al.* (2009) The N-terminal domain of GPR61, an orphan G-protein-coupled receptor, is essential for its constitutive activity. *J. Neurosci. Res.* 87, 1329–1333
75 Koth, C.M. *et al.* (2012) Molecular basis for negative regulation of the glucagon receptor. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14393–14398
76 Yang, L. *et al.* (2011) GPR56 regulates VEGF production and angiogenesis during melanoma progression. *Cancer Res.* 71, 5558–5568
77 Silva, J.P. *et al.* (2009) Functional cross-interaction of the fragments produced by the cleavage of distinct adhesion G-protein-coupled receptors. *J. Biol. Chem.* 284, 6495–6506