Structural basis for adhesion G protein-coupled receptor Gpr126 function

Katherine Leon[1,2], Rebecca L. Cunningham[3], Joshua A. Riback[1,4], Ezra Feldman[1], Jingxian Li[1,2], Tobin R. Sosnick[1,5], Minglei Zhao[1], Kelly R. Monk[3,6], Demet Araç[1,2*]

[1]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, 60637, USA.

[2]Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior, The University of Chicago, Chicago, IL, 60637, USA.

[3]Department of Developmental Biology, Washington University School of Medicine, St. Louis, MO, 63110, USA.

[4]Graduate Program in Biophysical Sciences Program, The University of Chicago, Chicago, IL, 60637, USA.

[5]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637, USA.

[6]Vollum Institute, Oregon Health & Science University, Portland, OR, 97239, USA.

*Corresponding author:

Demet Araç: arac@uchicago.edu

Abstract

Many drugs target the extracellular regions (ECRs) of cell-surface receptors. The large and alternatively-spliced ECRs of adhesion G protein-coupled receptors (aGPCRs) have key functions in diverse biological processes including neurodevelopment, embryogenesis, and tumorigenesis. However, their structures and mechanisms of action remain unclear, hampering drug development. The aGPCR Gpr126/Adgrg6 regulates Schwann cell myelination, ear canal formation, and heart development; and *GPR126* mutations cause myelination defects in human. Here, we determine the structure of the complete zebrafish Gpr126 ECR and reveal five domains including a previously-unknown domain. Strikingly, the Gpr126 ECR adopts a closed conformation that is stabilized by an alternatively spliced linker and a conserved calcium-binding site. Alternative splicing regulates ECR conformation and receptor signaling, while mutagenesis of the newly-characterized calcium-binding site abolishes Gpr126 function *in vivo*. These results demonstrate that Gpr126 ECR utilizes a multi-faceted dynamic approach to regulate receptor function and provide novel insights into ECR-targeted drug design.

Keywords: adhesion GPCR, GPR126, ADGRG6, myelination, Schwann cell development, G-protein signaling, alternative splicing

Introduction

Multicellular organisms rely on cellular communication to carry out critical biological processes, and numerous cell-surface receptors utilize their extracellular regions (ECRs) to modulate these cellular-adhesion and signaling events. For example, the ECRs of integrins, epidermal growth factor receptor (EGFR), and several G protein-coupled receptors (GPCRs) change conformation upon ligand binding, which propagates signals across the membrane[1,2,3,4,5,6,7,8,9]. Targeting the essential ECRs of receptors with antibody-like drugs to trap the ECRs in distinct conformations, or to modulate ECR-ligand interactions has been an effective way to treat diseases caused by defective proteins. Currently, the anti-cancer drug cetuximab targets EGFR to prevent an activating extended ECR conformation[10], and the drug etrolizumab blocks ligand binding to the ECRs of integrins in order to treat inflammatory bowel diseases[11]. Remarkably, earlier this year, the migraine preventive drug erenumab, which blocks ligand binding to the ECR of calcitonin receptor-like receptor, became the first antibody drug against a GPCR to be approved by the Food and Drug Administration[12,13]. Despite these and other breakthroughs, there are many essential receptors in the human genome that are not currently drugged, including the 32 adhesion GPCRs (aGPCRs), a diverse and understudied family of GPCRs with critical roles in synapse formation, angiogenesis, neutrophil activation, embryogenesis, and more[14,15,16].

Like all GPCRs, aGPCRs have canonical signaling seven-transmembrane (7TM) domains[17,18]. However, unlike most other GPCRs, aGPCRs have large ECRs which can extend up to almost 6000 amino acids (aa) and consist of various adhesion domains that mediate cell-cell and cell-matrix interactions[19]. In addition, during biosynthesis, aGPCRs are uniquely autoproteolysed within a conserved GPCR Autoproteolysis INducing (GAIN) domain of the ECR that is juxtaposed to the 7TM[20], resulting in a fractured receptor that nevertheless remains tightly associated at the cell surface[21,22].

Although their protein architectures remain largely unknown, functional studies have shown that aGPCR ECRs can regulate receptor function and that antibody-like synthetic proteins that target the ECRs can modulate downstream signaling[9,22,23,24,25]. A current model for aGPCR regulation suggests that transient interactions between the ECR and 7TM directly regulate receptor signaling[9,22,23,24,25]. There are also numerous reports that aGPCRs use their ECRs to mediate functions in a 7TM-independent manner[26,27,28,29,30]. Another non-mutually exclusive model for aGPCR activation posits that ligand binding to the ECR can exert force and cause dissociation at the autoproteolysis site, revealing a tethered peptide agonist, which then activates the receptor[31,32,33,34]. Clearly, the ECRs of aGPCRs have significant and diverse roles but remain poorly understood at a molecular level due to the scarcity of structural information, such as interdomain interactions, protein architecture, and identities of extracellular domains, which would provide insight into their mechanisms of action.

Gpr126/Adgrg6 is one of the better studied aGPCRs and is essential for Schwann cell (SC) myelination and other functions[35,36,37]. In vertebrate peripheral nervous system (PNS) development, the myelin sheath surrounding axons is formed by SCs and functions to facilitate rapid propagation of action potentials[38]. Disruption of myelination is associated with disorders such as Charcot-Marie-Tooth disease, which is characterized by muscle weakness[39,40]. In *gpr126* mutant zebrafish, SCs fail to express genes critical for myelination during development and are not able to myelinate axons due to deficient G-protein signaling. Additional studies have shown that this regulatory function of Gpr126 is conserved in mammals[41,42] and that Gpr126 also plays a role in myelin maintenance through communication with the cellular prion protein[43]. In humans, GPR126 mutations are linked to several cancers and other diseases[44,45,46,47,48], including adolescent idiopathic scoliosis[49] and arthrogryposis multiplex congenita, a disorder characterized by multiple joint contractures[50]. Furthermore, Gpr126 is required for inner ear development in zebrafish[35] and GPR126 is required for heart development in mouse[37], and it has been shown that the latter function is ECR-dependent and does not require the 7TM[30]. While the biological significance of Gpr126 has become indisputable over recent years, the molecular mechanisms underlying Gpr126 functions remain unclear.

Gpr126 has a large ECR consisting of 839 aa. Prior to the current study, four domains in the ECR of Gpr126 had been identified through sequence-based bioinformatics: Complement C1r/C1s, Uegf, Bmp1 (CUB), Pentraxin (PTX), Hormone Receptor (HormR), and GAIN[20,51,52]. However, a 150 aa region between PTX and HormR, could not be assigned to a known structural fold. Furin, a Golgi-localized protease, is reported to cleave human and mouse GPR126 in this region[51], although any effect on protein architecture is unclear because of the unspecified structure. In addition, alternative splicing occurs in *Gpr126*/*GPR126*, resulting in Gpr126/GPR126 isoforms that vary in their ECRs[51,53]. Alternative splicing of exon 6 was observed in human and zebrafish[30,51], producing isoforms that either include (S1 isoform, henceforth referred to as +ss) or exclude (S2 isoform, henceforth referred to as -ss) a 23 aa segment found within the unknown region between PTX and HormR. A genetic variant in *GPR126* leading to decreased inclusion of exon 6 was recently found to be associated with adolescent idiopathic scoliosis[54]. Thus, determining the ECR structure, conformation, and other possible unexplored features will be instrumental in understanding Gpr126 function.

In this study, we determined the high-resolution crystal structure of the full-length ECR of zebrafish Gpr126, which reveals five domains, including a newly identified Sperm protein, Enterokinase and Agrin (SEA) domain, in which furin-mediated cleavage would occur in the human and mouse homologs. Intriguingly, the ECR is in an unexpected closed conformation that is reminiscent of the inactive closed conformation of the ECRs from EGFR and integrin families. This closed conformation is sustained by an alternatively spliced linker, while insertion of the alternatively spliced site gives rise to dynamic open-like ECR conformations and increases downstream signaling. A second feature that also mediates the closed conformation is a newly-identified calcium-binding site at the tip of the ECR. Strikingly, zebrafish carrying point mutations at this site have both myelination defects and malformed ears, demonstrating the critical role of the ECR in Gpr126 function *in vivo*. These results altogether show that the ECR of Gpr126 has multifaceted roles in regulating receptor function, a feature that is likely true for other aGPCRs, and that will form the basis for further investigations in the efforts to drug aGPCRs.

Results

Structure of the full-length ECR of Gpr126

To determine the structure of the ECR of Gpr126, the full-length ECR (-ss) from zebrafish Gpr126 (T39-S837) was expressed and purified from insect cells using the baculovirus expression system. Zebrafish Gpr126 (Figure 1A) has high sequence identity (47%) to its human homolog but its ECR has a fewer number of N-linked glycosylation sites (15 predicted in zebrafish, 26 in human) and no furin cleavage site (Supplementary Figure 1A), and thus yields a more homogeneous sample (Supplementary Figures 1B, C). Crystals of both native and selenomethionine (SeMet)-labeled zebrafish Gpr126 ECR (-ss) were obtained and diffracted to 2.4 Å (Supplementary Figure 1D), and the structure was determined by SeMet single-wavelength anomalous diffraction (SAD) phasing (Supplementary Table 1).

The structure, with overall dimensions of 110 Å x 80 Å x 35 Å, revealed the presence of five domains (Figures 1B, C), of which only four were identified previously. The N-terminal region of the protein is composed of the CUB domain followed very closely by the PTX domain. The 150 aa unknown region after the PTX domain was revealed to be a 22 aa linker that is partially disordered, the 23 aa alternatively spliced region (not present in crystal structure construct), and a structured domain which spans 105 aa and was identified as a SEA domain through the Dali server[55]. The Gpr126 SEA domain adopts a ferredoxin-like alpha/beta sandwich fold, common to SEA domains from other proteins. Interestingly, analysis of the structure as well as sequence alignments between zebrafish and human showed that furin cleavage in humans would occur in the SEA domain (Supplementary Figure 1A). Finally, the SEA domain is followed by the HormR and GAIN domains, the latter of which is autoproteolyzed as expected (Supplementary Figure 1E). The HormR and GAIN domain structures are similar to previously-solved HormR+GAIN domain structures from other aGPCRs[9,20], with the exception of the relative orientation between HormR and GAIN. There is a 90º rotation of the HormR domain with respect to the GAIN domain (Supplementary Figure 1F) in Gpr126 compared to previously-solved HormR+GAIN structures from rLphn1 and hBAI3[20]. In addition, Gpr126 was observed to have at least ten sites of glycosylation throughout all domains of the ECR except the PTX domain (Figure 1B).

Gpr126 (-ss) ECR adopts a closed conformation

Unexpectedly, the structure revealed a compact, closed conformation where the most N-terminal CUB domain interacts with the more C-terminal HormR and GAIN domains (Figure 1B). To ensure that this conformation is not a crystallization artifact, we utilized both negative-stain electron microscopy (EM) and small-angle X-ray scattering (SAXS) to confirm that the closed confirmation is observed for Gpr126 in solution. Negative-stain 2D class averages of Gpr126 ECR showed a V-shaped protein architecture (Figure 1D). The individual domains in the 2D class averages were assigned according to size and are consistent with the closed architecture of the crystal structure. In addition, we measured the radius of gyration (R_g) of the ECR using SAXS to confirm that the closed conformation exists in solution. The observed R_g (41.1 $\pm$ 0.1 Å) is consistent with the calculated R_g of the closed conformation crystal structure model (42.6 Å) and inconsistent with that of an extended model of Gpr126 ECR in which the CUB domain points away, rather than toward, the center of the molecule (R_g = 52.2 Å) (Supplementary Figure 1G). Taken together, these results: show that Gpr126 ECR is in a closed conformation in solution, demonstrate that this conformation is not an artifact of crystal-packing contacts, and suggest that this closed conformation may play an important role in Gpr126 function.

As the closed conformation of Gpr126 (-ss) ECR was shown to exist both in solution and in the crystal lattice, we next wanted to explore the interactions that contribute to this protein architecture. Close examination of the crystal structure revealed two interaction sites that mediate the closed conformation, the first of which is a direct interaction between domains that are at opposite ends of the ECR and the second is an indirect interaction formed between two domains through a loop that holds them together (Figure 2A).

First, a direct interaction exists at the tip of the CUB domain (close to the N-terminus), which points inward towards the center of the molecule and lies in the interface between GAIN and HormR. Residues in the HormR domain (H516, F533, P534, Y535) interact with each other through pi-pi stacking (sandwich), promoting interaction with F135 on the CUB domain through additional (T-shaped) pi-pi stacking to stabilize the CUB-HormR interaction (Figure 2B).

Surprisingly, examination of the 2Fo-Fc electron density map showed that there is density within the CUB domain at this interface that does not belong to any amino acid residue (Supplementary Figure 2A). This density is coordinated by the side-chain groups of E89, D97 (bidentate) and D134, main-chain carbonyl groups of S136 and V137, as well as a water molecule for a complex with coordination number 7 in a pentagonal bipyramid geometry (Figure 2C). The geometry and distances between the density and the coordinating residues in Gpr126 are consistent with calcium coordination[56]. Several CUB domains from extracellular proteins are reported to coordinate calcium, including Gpr126[57], and some have been discovered to use this coordination to mediate ligand binding[57,58,59,60,61] (Supplementary Figure 2B). For example, the C1s protein uses its CUB calcium-binding site to bind to ligand C1q and initiate the classical pathway of complement activation[61], and the Lujo virus recognizes a calcium-binding site on the CUB domain of the neurophilin-2 receptor in order to gain cell entry[59]. The calcium-coordinating residues are all conserved in the Gpr126 CUB domain (among GPR126 proteins from various species (Figure 2D) as well as among calcium-binding CUB domains from other proteins (Supplementary Figure 2C)), suggesting that the density is indeed calcium. Importantly, the calcium coordination aligns the coordinating residues E89 and D134 on the surface of the CUB domain such that they can interact with K536 on the HormR domain (Figure 2C), contributing to the closed conformation.

In addition to the direct CUB-HormR interaction, a second interaction site is formed by a disulfide-stabilized loop which provides a bridge between the CUB and HormR domains. Although 13 (C355-A367) of the 22 aa (C355-P376) in the linker region are disordered in the structure, the rest were able to be resolved and they form a small loop stabilized by a disulfide bond between C369 and C375 (Figure 2E). This loop is located directly N-terminal to the SEA domain and is inserted between the CUB and HormR domains, effectively bridging the two domains and likely contributing to the stabilization of the closed conformation. The cysteines that form the disulfide bond are conserved among all except four of the 94 species analyzed in this study (Supplementary Figure 2D and Supplementary Table 2), suggesting that this disulfide bond plays an important role in Gpr126 function. The five residues (ASGLG) flanked by the cysteines are small and flexible, accommodating the formation of the disulfide loop as well as insertion into the small pocket between CUB and HormR.

Alternative splicing modulates Gpr126 ECR conformation

Gpr126 is alternatively spliced, producing several isoforms that may modulate protein function. Skipping of exon 6 results in deletion of 23 aa in zebrafish (28 aa in human) and is of particular interest because these amino acids reside in the previously unknown region of Gpr126 ECR. The 23 aa region is rich in serine/threonine residues (10 out of 23) and contains a predicted N-linked glycosylation site, which suggests that this region may be a highly O- and N-link glycosylated stalk. From analysis of the crystal structure (-ss isoform, in which the 23 aa are deleted), we determined that the splice site is directly between the regions encoding the disulfide-stabilized loop and the SEA domain (Figure 3A). Because the disulfide-stabilized loop makes contacts that are important for the closed conformation of Gpr126 ECR (-ss) (Figure 2E), we hypothesized that the (+ss) isoform would disrupt the closed conformation and have a different, more open conformation.

To test whether Gpr126 ECR (+ss) and (-ss) have different conformations, the two proteins were purified and analyzed using negative stain EM. Single particles were classified into 2D class averages and the class averages were further categorized into groups to facilitate interpretation of different conformations. The class averages for the (-ss) isoform, categorized into five main orientations (Figure 3B), were consistent with the closed conformation of the crystal structure (Figure 1B). However, the class averages for the (+ss) isoform showed a diverse population of ECR molecules, as they contain additional more open-like conformations (group vi, 21% of particles, Figure 3D) as well as closed conformations that were observed in the (-ss) isoform (Figures 3C, D). Furthermore, individual (+ss) particles showed the presence of open conformations (Figure 3E), including a fully extended conformation which could not be classified into a distinct class average during image processing. These results are consistent with our hypothesis that the (+ss) ECR conformation is different from that of (-ss) and suggest that the addition of 23 aa extends the linker in (+ss), likely disrupting the indirect and direct CUB-HormR interactions and preventing the stable closed conformation that is observed in (-ss) (Figure 3F).

The negative stain EM data is consistent with SAXS experiments showing that the R_g of zebrafish Gpr126 ECR (+ss) is larger than that of (-ss) with a more dramatic change in R_g observed between the human GPR126 isoforms (Supplementary Figures 3A-F and Supplementary Table 3). Size exclusion chromatography elution profiles for both zebrafish and human constructs also showed that (+ss) elutes earlier compared to (-ss), indicative of a larger size and different shape (Supplementary Figures 3G, H).

Alternative splicing modulates Gpr126 receptor signaling

To determine whether the two isoforms also exhibit different levels of signaling, receptor activity was measured for both isoforms using a G protein signaling assay. Human GPR126 has been shown previously to couple to and activate $G\alpha_s$, leading to production of cAMP[42]. Therefore, we used a cAMP signaling assay in which HEK293 cells were co-transfected with a full-length zebrafish Gpr126 construct and a reporter luciferase that emits light upon binding to cAMP. Cell-surface expression levels of the constructs were quantified by flow cytometry analysis of cells stained by antibodies against N-terminal FLAG-tags (Figure 4A and Supplementary Figures 4A, B), and basal signaling results (Figure 4B) were normalized to expression level (Figure 4C).

Cells transfected with either (-ss) or (+ss) Gpr126 had higher cAMP levels compared to cells transfected with an empty vector (EV) (Figure 4C), demonstrating that basal activity of Gpr126 can be detected in this assay. As a positive control, a synthetic peptide agonist that targets the 7TM activated zebrafish Gpr126 and human GPR126 signaling to a level consistent with similar, previously-published experiments on human GPR126[32] (Supplementary Figure 4C) and did not activate signaling in EV-transfected cells.

However, the closed-conformation (-ss) Gpr126 signaled significantly less compared to the more dynamic (+ss) Gpr126 (Figure 4C), and this result was consistent between both zebrafish and human constructs (Figures 4D-F). This suggests that the additional amino acids in the linker region of the ECR as a result of alternative splicing plays a role in modulating the activity of Gpr126 and that the ECR of Gpr126 is coupled to receptor signaling. Taken together with the negative stain EM results, the (-ss) and (+ss) Gpr126 isoforms are distinct in terms of ECR conformation dynamics as well as G protein signaling activity.

In addition, we mutated calcium-binding site residues D134A/F135A in the (-ss) isoform, which we predicted would disrupt the closed conformation. Using negative stain EM, we observed open ECR conformations for this construct (Supplementary Figure 4D-F), similar to the wild-type (+ss) isoform (Supplementary Figure 4G). The calcium-binding site mutation did not increase or decrease the cAMP signaling for the (-ss) Gpr126 isoform, which suggests that the ECR conformation is not solely responsible for regulation of receptor signaling. However, the same mutation in the (+ss) isoform resulted in lower cAMP levels compared to wild-type (+ss) (Supplementary Figure 4H). Cell-surface expression levels of these mutant Gpr126 constructs in HEK293 cells were similar or higher than wild-type constructs, excluding the possibility that lower signaling was due to improper protein folding or trafficking (Supplementary Figure 4H). Altogether, these results might be explained by a complex, rather than a simple and straightforward, model of regulation for receptor signaling and suggest a possible functional role for the calcium-binding site.

Calcium-binding site is critical for PNS myelination *in vivo*

Functional sites on proteins are usually highly evolutionarily conserved. We used the ConSurf server[62] to perform surface conservation analyses on a diverse set of 94 Gpr126 protein sequences. The conservation score for each residue was mapped onto the Gpr126

ECR structure (Supplementary Figure 5A), which revealed that the most conserved domain in the ECR is the CUB domain. Importantly, the calcium-binding site is absolutely the most highly conserved patch within the CUB domain and within the entire Gpr126 ECR (Figure 5A). The calcium-binding site is universally conserved among all species analyzed, which suggests that the calcium-binding site has an essential role in Gpr126 function.

We next wanted to test whether the residues in the calcium-binding site are important for Gpr126 function *in vivo*. Gpr126 has previously been shown to regulate both PNS myelination and ear development in zebrafish through elevation of cAMP[35,36]. Zebrafish *gpr126* mutations that impair G protein signaling result in abolished myelination of the peripheral axons by SC and cause “puffy” ears[28,32,35,36,63]. GPR126 has been shown to have a role in heart development in mouse[37], supported by additional studies in zebrafish[28,30,63]. Gpr126 activity in zebrafish can be readily measured by analyzing the expression of *myelin basic protein (mbp)*, which encodes a major structural component of the myelin sheath and is essential for PNS myelination, and by assessing ear and heart morphologies of the fish. To determine whether the calcium-binding site is important for these functions, two amino acids in the site, D134 and F135, were targeted and mutated to alanines using CRISPR/Cas9-mediated homologous recombination. D134 directly coordinates the calcium ion and F135 is an adjacent hydrophobic residue which forms one arm of the calcium-binding pocket (Figures 2C, 5A). As a result, the mutant zebrafish, *gpr126^{stl464}*, harbor D134A and F135A mutations (Figure 5B, Supplementary Figure 5B). These mutations created a BstUI restriction enzyme site, which was used to genotype individual zebrafish (Figure 5C). Expression of *gpr126* is unaffected in *gpr126^{stl464}* mutants (Supplementary Figure 5C, D). Strikingly, compared to wild-type siblings, the *gpr126^{stl464}* mutant zebrafish developed the puffy ears (Figures 5D, E) that are indicative of a defect in Gpr126-mediated G protein signaling, though they do not appear to have heart defects (Supplementary Figure 5E-T). In addition to the ear phenotype, mutant zebrafish did not express *mbp*, indicative of failed PNS myelination (Figures 5F, G, Supplementary Figures 5U, V). These results show that D134 and F135 in the calcium-binding pocket of Gpr126 are essential for ear and SC development *in vivo*.

Identification of a proteolytic SEA domain in human GPR126

As mentioned earlier, the previously unknown region in the Gpr126 ECR contains a structured domain, which we revealed to be a SEA domain (Figure 6A). Gpr126 SEA superimposes well over known SEA domains from Mucin-1 and Notch-2[64,65], which are cleaved (via autoproteolysis and furin, respectively), both in the same loop between beta-strand 2 and beta-strand 3 (Figures 6B, C). Although the GPR126 furin cleavage site is conserved in many mammals and birds (Supplementary Table 2, Supplementary Figure 1A), with a consensus sequence of (R/K)-X-K-R↓, it is not conserved in zebrafish Gpr126. Using sequence alignments (Figure 6D) and homology modeling, we mapped the furin-cleavage site in human GPR126 (Figure 6E, Supplementary Figure 6A) to the same loop that is cleaved in Mucin-1 and Notch-2, suggesting that SEA domain cleavage plays similar roles in each of these proteins. Consistent with a previous study[51], R468A mutations abolish furin cleavage in both human GPR126 (-ss) and (+ss) isoforms (Supplementary Figures 6B, C). In addition, these mutant GPR126 constructs were transfected into HEK293 cells and were detected on the cell surface (Supplementary Figure 6D), and therefore, the importance of furin cleavage is likely not primarily important for proper expression and trafficking.

To our knowledge, SEA and GAIN are the only known protein domains that are proteolyzed and remain associated even after proteolysis. In proteins like Mucins and Notch, the cleaved SEA domain remains intact[65,66] and shear forces likely unfold the domain and separate the protein into two fragments[67,68]. The Gpr126 SEA domain shows several noncovalent interdomain interactions, particularly between all four of the beta-strands that form a beta-sheet (Figure 6F). The separation of the human GPR126 furin-cleaved SEA domain into two fragments does not readily occur immediately following cleavage as the cleaved protein resists separation when purified by size exclusion chromatography (Supplementary Figure 6B), similar to the aforementioned SEA domains as well as to GAIN domain autoproteolysis. Instead, the two fragments likely stay associated non-covalently until a disruptive event, such as ligand binding and mechanical force, unfolds the SEA domain and leads to separation or shedding of the region N-terminal to the furin-cleavage site (CUB, PTX, linker, half of SEA) and the C-terminal region (half of SEA, HormR, GAIN, 7TM).

Discussion

aGPCRs make up the second largest family of GPCRs with 32 members in humans and are essential for numerous biological processes such as synapse formation, cortex development, neutrophil activation, angiogenesis, embryogenesis, and many more. Recent studies have shown that the ECRs of aGPCRs play important roles in these functions, however, the relative lack of information about the structures of ECRs and their mechanisms of activation hampers further studies toward drugging these receptors. Here we show that the large ECR of Gpr126, an aGPCR with critical functions in PNS myelination, ear development, and heart development, adopts an unexpected closed conformation where the most N-terminal CUB domain interacts with the more C-terminal HormR domain. The structure of the Gpr126 ECR revealed that the closed conformation is mediated through a calcium-binding site as well as a disulfide-stabilized loop. Interestingly, the residues involved in these intramolecular interactions are highly conserved among Gpr126 sequences, including that of zebrafish, raising questions about their role in Gpr126 function.

Since *gpr126* is alternatively spliced in the region encoding the ECR, we examined the functional differences between isoforms. Alternative splicing in proteins is an important mechanism to greatly expand the functional capacity of metazoan genomes, and its regulatory role in brain function has been repeatedly demonstrated. For instance, DSCAMs, protocadherins, calcium channels, neurexins, and neuroligins have been shown to use alternative splicing for diversifying their functions[69,70,71,72]. It is also proposed that alternative splicing may cause a large conformational change in the ECR of the synaptic protein teneurin, since alternative splicing allows the protein to act as a switch in regulating ligand binding despite the ligand-binding site being distant from the seven aa alternatively spliced site[73]. Our negative stain EM and SAXS results suggest that alternative splicing between the regions encoding the PTX and SEA domains in *gpr126* perturbs the closed conformation and generates a population of ECR conformations that range from closed to extended (Figure 7, left). Several of the inserted residues resulting from alternative splicing are predicted to be sites of glycosylation. These glycosylation sites as well as the state of the other glycosylation sites may contribute to the change in ECR conformation. Our signaling assay results also show that alternative splicing leads to changes in basal receptor activity, which suggests that the architecture and conformation of aGPCR ECRs play more important roles in their functions than previously thought. However, the signaling assay results showing that the change of Gpr126 ECR conformation is not solely responsible for changes in signaling may be confusing and contradictory. Rather, a more complex model that combines changes in ECR conformation with exposure of potential functional sites due to these changes may be key for alternative splicing-mediated regulation.

Importantly, we identified the calcium-binding site in Gpr126 as a potential functional site. Our *in vivo* results showed that zebrafish carrying two point-mutations in the calcium-binding site have defective SC and ear development, suggesting that the calcium-binding site is essential for the *in vivo* functions of Gpr126 (Figure 7, right). Since a subset of CUB domains from other proteins coordinate calcium in order to mediate ligand-binding[57], one possibility for the critical function of the calcium-binding site in Gpr126 may be to act as a ligand-binding site as well, although future experiments will need to be performed to validate this hypothesis.

The structure also revealed the presence of a SEA domain. In human and other species, a furin cleavage site is mapped to this domain but this cleavage site is not conserved in zebrafish. Therefore, the function of the furin cleavage may play a role in GPR126 that is not conserved in zebrafish. Cleaved SEA domains from other proteins have been shown to stay intact until a force is applied and pulls apart the fragments[67,68]. Similarly, GPR126 may regulate its activity by furin-dependent shedding in addition to the established GAIN-autoproteolysis-dependent shedding. Moreover, the released extracellular fragments may act as diffusible ligands and bind to other cell-surface receptors, but further studies need to be done to test this model. Other aGPCRs that have SEA domains in their ECRs include ADGRF1/GPR110 and ADGRF5/GPR116[74,75]. Although these SEA domains are not cleaved by furin, they do contain the GSVVV (or GSIVA) motif that leads to autoproteolytic cleavage in the same loop (between beta-strand 2 and beta-strand 3) that is cleaved by furin in GPR126. Therefore, SEA domain cleavage, whether by autoproteolysis or by furin, is a common feature in several aGPCRs and may have similar roles in regulating receptor function.

Taken together, our results suggest that Gpr126 is a complex protein that makes use of its many domains to regulate its function. In addition to the autoproteolysis-dependent activation mechanism (Supplementary Figure 7A), Gpr126 uses other mechanisms to regulate its function including modulation of the ECR conformation. In the closed conformation, Gpr126 signals less compared to when the ECR is in a more dynamic, open conformation, which may be regulated by alternative splicing (Figure 7, left). Alternative splicing which deletes the CUB domain[30] may also regulate receptor function (Supplementary Figure 7B). Mutation of the calcium-binding site leads to signaling defects *in vitro* and to ear and PNS defects *in vivo*

(Figure 7, right)). In addition, furin cleavage may allow GPR126 another mode of activation that is common to other receptors and adhesion GPCRs.

The Gpr126 closed conformation and hidden calcium-binding site is conceptually similar to EGFR. EGFR is in a closed, compact inactive conformation until ligand binding leads to a conformational change that extends the protein and reveals a hidden functional site that is important for its activation[10,76]. Because this mechanism is key for drugging EGFR, the conceptual similarity provides an opportunity to also drug Gpr126. Drugs that alter the ECR conformation of Gpr126 or block functional sites, such as the calcium-binding site, may be useful for treating Gpr126-associated diseases. The ECRs of other aGPCRs are major players in mediating receptor functions as well. For example, using its ECR, ADGRA2/GPR124 regulates isoform-specific Wnt signaling[77,78,79,80], the *C. elegans* ADGRL1/LAT-1 controls cell division planes during embryogenesis, and ADGRB1/BAI1 and ADGRL3/Lphn3 mediate synapse formation through interaction with other cell-surface proteins[81,82,83,84]. Thus, the ECRs of other aGPCR family members are also promising drug targets to treat numerous diseases once mechanistic details about their regulatory functions are understood.

Methods

Cloning and purification of Gpr126/GPR126 from insect cells

The ECRs (residues T39-S837) of zebrafish Gpr126 and ECRs (residues C38-A853) of human GPR126, along with C-terminal 8XHis-tags, were cloned into the pAcGP67a vector. A baculovirus expression system was used to express Gpr126/GPR126 ECRs in High Five cells as previously described[20]. SeMet-labeled Gpr126 (-ss) ECR was expressed as previously described[85]. The proteins were purified using nickel-nitrilotriacetic agarose resin (Qiagen) and size-exclusion chromatography (Superdex 200 10/300 GL; GE Healthcare).

X-ray crystallography

Purified Gpr126 (-ss) ECR (both native and SeMet-labeled) was crystallized at 3 mg mL^{-1} in 50 mM potassium dihydrogen phosphate, 20% (w/v) PEG 8000. Both native and SeMet-labeled datasets were collected to 2.4 Å at the Advanced Photon Source at Argonne National Laboratory (beamline 23-ID-D). The data sets were processed with HKL2000 and an initial model was determined by SAD phasing using Crank2 in CCP4. Refinement was performed with both REFMAC5 (CCP4) and phenix.refine (PHENIX).

Negative stain electron microscopy

Purified Gpr126 (-ss), (+ss), and (-ss) D134A/F135A ECR constructs were diluted to ~5 ug mL^{-1} and applied to an EM grid (Electron Microscopy Sciences, CF400-Cu,) using a conventional negative-stain protocol[86]. The sample was imaged on a Tecnai G2 F30 operated at 300 kV. Gpr126 -ss (6565 particles), +ss (2529 particles), and -ss D134A/F135A (3916 particles) were processed using EMAN2[87].

Small-angle X-ray scattering

SAXS measurements were performed at the Advanced Photon Source at Argonne National Laboratory (beamline 18-ID). with an in-line SEC columns (Superdex 200 or Biorad EnRich 5-650 10-300) equilibrated with 20 mM HEPES, pH 7.4, and 150 mM NaCl. Data was analyzed using autorg and datgnom using the commands ‘‘autorg –sminrg 0.55 –smaxrg 1.1’’ and ‘‘datgnom ‘1’.dat -r ‘2’ – skip ‘3’ -o ‘1’.out,’’ respectively, where ‘1’ is the file name, ‘2’ is the R_g determined by autorg, and ‘3’ is the number of points removed at low q as determined from autorg. SAXS curves of molecular models were generated with Crysol version 2.83[88].

cAMP signaling assay

Full-length, truncated, and mutant Gpr126 constructs were cloned into pCMV5. All constructs include N-terminal FLAG-tags for measuring cell-surface expression levels. HEK293 cells were seeded in 6-well plates with Dulbecco's Modified Eagle Medium (DMEM; Gibco, 11965092) supplemented with 10% FBS (Sigma-Aldrich, F0926). At 60-70% confluency, the cells were co-transfected with 0.35 µg Gpr126 DNA, 0.35 µg GloSensor reporter plasmid (Promega, E2301), and 2.8 µL transfection reagent Fugene 6 (Promega, PRE2693). After a 24-hour incubation, the transfected cells were detached and seeded (50,000 cells per well) in a white 96-well assay plate. Following another 24-hour incubation, the DMEM was replaced with 100 µL Opti-MEM (Gibco, 31985079) and incubated for 30 minutes. To each well was then added 1 µL GloSensor substrate and 11 µL FBS. Basal-level luminescence measurements were taken after 30 minutes to allow for equilibration. For activation assays, the cells were then treated with either 1mM p14 synthetic peptide (GenScript, THFGVLMDLPRSASEKEK-Biotin)_or vehicle DMSO for 15 minutes. Measurements were taken with a Synergy HTX BioTeck plate reader at 25°C.

Flow cytometry to measure cell-surface expression of Gpr126

HEK293 cells were transfected as previously described and incubated for 24 hours. The cells were then detached and seeded in a 24-well plate. Following another 24-hour incubation, the cells were detached and stained with mouse anti-FLAG primary antibody (1:1000 dilution; Sigma-Aldrich, F3165) and donkey anti-mouse Alexa Fluor 488 secondary antibody (1:500 dilution; Invitrogen, A21202).

Zebrafish rearing

Zebrafish were maintained in the Washington University Zebrafish Consortium Facility (http://zebrafish.wustl.edu), and the following experiments were performed according to

Washington University animal protocols. The *gpr126^{stl464}* zebrafish were generated within the wild-type AB* background. All crosses were either set up as pairs or harems and embryos were raised at 28.5°C in egg water (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$). Larvae were staged at days post fertilization (dpf). *Gpr126^{stl464}* larvae can be identified at 4 dpf by a puffy ear phenotype.

Genotyping

To identify carriers of the *gpr126^{stl464}* allele, the following primers were used to amplify the 381 base pair (bp) locus of interest: F: 5'-GTTGTCGTCAAGACCGGCAC-3' and R: 5'-TCCACCTCCCAGCTACAATTCC-3'. After amplification by PCR, the product was digested with either DrdI (NEB) at 37°C or BstUI (NEB) at 60°C, and then run on a 3% agarose gel. The mutation both disrupts a DrdI binding site and introduces a BstUI binding site. DrdI cleaves wild-type PCR product into 275 and 105 bp products, and the mutant product is 380 bp. BstUI cleaves mutant PCR product into 274 and 106 bp products, and the wild-type product is 380 bp. We recommend using BstUI for genotyping. Any larvae identified with the puffy ear phenotype were always genotyped as *gpr126^{stl464}* homozygous mutant (n=20/20).

Guide RNA synthesis

Potential gRNA templates were generated by CHOPCHOP (http://chopchop.cbu.uib.no/). The chosen forward and reverse oligonucleotides, 20 bps upstream of the PAM sequence, were ordered with additional nucleotides added to the 5' end to permit cloning into the pDR274 vector[89]. The oligonucleotide forward sequence used was: 5' - tag gAC TTT AGT GTC CAA AAG AA - 3' and oligonucleotide reverse sequence used was: 5'- aaa cTT CTT TTG GAC ACT AAA GT – 3'. 2 uM of each oligonucleotide was mixed in annealing buffer (10 mM Tris, pH 8, 50 mM NaCl, 1 mM EDTA) and incubated at 90° C for 5 minutes, then cooled to 25° C over a

45 minute time interval. The pDR274 vector was linearized with BsaI and oligonucleotides were ligated into the vector with T4 ligase (NEB) for 10 minutes at room temperature. The ligation reaction was transformed into competent cells and then plated on kanamycin LB plates. Selected colonies were grown, mini-prepped (Zyppy Plasmid Kits, Zymo Research), and Sanger sequenced. The gRNA DNA sequence was then PCR amplified from 50 ng μl^{-1} of the plasmid with Phusion (NEB) and the following primers: F: 5'-GTTGGAACCTCTTACGTGCC-3' and R: 5'-AAAAGCACCGACTCGGTG-3'. The PCR product was digested with DpnI at 37°C for 1 hour, heat inactivated at 80°C for 20 minutes, and then purified with a Qiagen PCR Purification column. RNA was synthesized with a MEGAscript T7 Transcription Kit (Ambion).

Design of ssODN and microinjections

One-cell stage wild-type embryos were injected with either 2 or 3 nl of a solution containing ~132 ng μl^{-1} gRNA, ~148 ng μl^{-1} of *Cas9* mRNA (obtained from the Hope Center Transgenic Core at Washington University in St. Louis), and 60 ng μl^{-1} of the ssODN. The 150 bp ssODN was ordered from IDT and contained a 5 bp mutation (uppercase): 5'-atcataaacatacccttgcttgtaactgatatggaagcctttcttttggacactCGCCGcggagttaaagaaaacctccatcacatttccagtggagttgag-3'. Please note that an extra C (bolded), beginning after exon 3, is present in the ssODN that is not present in the *gpr126* reference sequence. The extra nucleotide was not integrated into the *stl464* mutants. At 1 dpf, embryos were genotyped for disruption of the wild-type DrdI binding site and screened for the characteristic *gpr126* puffy ear mutant phenotype. Mutations that were successfully transmitted to the F1 offspring were screened for by restriction enzyme digest analysis. Mutant bands were gel extracted (Qiagen Gel Extraction Kit) and Sanger sequenced to identify the incorporation of the ssODN containing the mutation of interest.

Whole mount in situ hybridization

1 dpf larvae were treated with 0.003% phenylthiourea to inhibit pigmentation until fixation in 4% paraformaldehyde at 4 dpf. Whole mount *in situ* hybridization was performed as previously described[90]. Previously characterized riboprobes were utilized[36,91]. For *mbp*, larvae were scored for either presence or absence of signal expression along the PLLn.

Data availability

The accession number for the coordinates and diffraction data for the Gpr126 (-ss) ECR crystal structure reported in this paper is PDB: XXXX. The SASBDB IDs for the SAXS experimental data are: SASDFT9, SDSDFU9, SDSDFV9, SASDFW9, SASDFX9. The source data underlying Figures 4B, C, E, F and Supplementary Figures 4C, H, and 6B, C are provided as a Source Data file.

References

1. Chun L, Zhang WH, Liu JF. Structure and ligand recognition of class C GPCRs. *Acta pharmacologica Sinica* **33**, 312-323 (2012).

2. de Graaf C*, et al.* Extending the Structural View of Class B GPCRs. *Trends in biochemical sciences* **42**, 946-960 (2017).

3. Hu P, Luo BH. Integrin bi-directional signaling across the plasma membrane. *Journal of cellular physiology* **228**, 306-312 (2013).

4. Kovacs E, Zorn JA, Huang Y, Barros T, Kuriyan J. A structural perspective on the regulation of the epidermal growth factor receptor. *Annual review of biochemistry* **84**, 739-764 (2015).

5. Rana R*, et al.* Structural insights into the role of the Smoothened cysteine-rich domain in Hedgehog signalling. *Nature communications* **4**, 2965 (2013).

6. Shimaoka M, Springer TA. Therapeutic antagonists and conformational regulation of integrin function. *Nature Reviews Drug Discovery* **2**, 703 (2003).

7. Sun Z, Guo SS, Fassler R. Integrin-mediated mechanotransduction. *The Journal of cell biology* **215**, 445-456 (2016).

8. Luo BH, Carman CV, Springer TA. Structural basis of integrin regulation and signaling. *Annual review of immunology* **25**, 619-647 (2007).

9. Salzman GS*, et al.* Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. *Neuron* **91**, 1292-1304 (2016).

10. Li S, Schmitz KR, Jeffrey PD, Wiltzius JJ, Kussie P, Ferguson KM. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer cell* **7**, 301-311 (2005).

11. Ley K, Rivera-Nieves J, Sandborn WJ, Shattil S. Integrin-based therapeutics: biological basis, clinical use and new drugs. *Nature reviews Drug discovery* **15**, 173-183 (2016).

12. Traynor K. FDA approves licensing of erenumab-aooe to prevent migraine. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists* **75**, 929-930 (2018).

13. Shi L*, et al.* Pharmacologic Characterization of AMG 334, a Potent and Selective Human Monoclonal Antibody against the Calcitonin Gene-Related Peptide Receptor. *The Journal of pharmacology and experimental therapeutics* **356**, 223-231 (2016).

14. Hamann J*, et al.* International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors. *Pharmacological reviews* **67**, 338-367 (2015).

15. Yona S, Lin H-H, Siu WO, Gordon S, Stacey M. Adhesion-GPCRs: emerging roles for novel receptors. *Trends in biochemical sciences* **33**, 491-500 (2008).

16. Langenhan T, Piao X, Monk KR. Adhesion G protein-coupled receptors in nervous system development and disease. *Nature reviews Neuroscience* **17**, 550-561 (2016).

17. Manglik A, Kruse AC. Structural Basis for G Protein-Coupled Receptor Activation. *Biochemistry* **56**, 5628-5634 (2017).

18. Paavola KJ, Hall RA. Adhesion G Protein-Coupled Receptors: Signaling, Pharmacology, and Mechanisms of Activation. *Molecular Pharmacology* **82**, 777-783 (2012).

19. Langenhan T, Aust G, Hamann J. Sticky signaling--adhesion class G protein-coupled receptors take the stage. *Science signaling* **6**, re3 (2013).

20. Arac D*, et al.* A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *The EMBO journal* **31**, 1364-1378 (2012).

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

22. Paavola KJ, Stephenson JR, Ritter SL, Alter SP, Hall RA. The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *The Journal of biological chemistry* **286**, 28914-28921 (2011).

23. Kishore A, Hall RA. Disease-associated extracellular loop mutations in the adhesion G protein-coupled receptor G1 (ADGRG1; GPR56) differentially regulate downstream signaling. *The Journal of biological chemistry* **292**, 9711-9720 (2017).

24. Kishore A, Purcell RH, Nassiri-Toosi Z, Hall RA. Stalk-dependent and Stalk-independent Signaling by the Adhesion G Protein-coupled Receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). *The Journal of biological chemistry* **291**, 3385-3394 (2016).

25. Salzman GS, Zhang S, Gupta A, Koide A, Koide S, Arac D. Stachel-independent modulation of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 10095-10100 (2017).

26. Koh JT*, et al.* Extracellular fragment of brain-specific angiogenesis inhibitor 1 suppresses endothelial cell proliferation by blocking alphavbeta5 integrin. *Experimental cell research* **294**, 172-184 (2004).

27. Muller A*, et al.* Oriented Cell Division in the C. elegans Embryo Is Coordinated by G-Protein Signaling Dependent on the Adhesion GPCR LAT-1. *PLoS genetics* **11**, e1005624 (2015).

28. Petersen SC*, et al.* The adhesion GPCR GPR126 has distinct, domain-dependent functions in Schwann cell development mediated by interaction with laminin-211. *Neuron* **85**, 755-769 (2015).

29. Promel S*, et al.* The GPS motif is a molecular switch for bimodal activities of adhesion class G protein-coupled receptors. *Cell reports* **2**, 321-331 (2012).

30. Patra C*, et al.* Organ-specific function of adhesion G protein-coupled receptor GPR126 is domain-dependent. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16898-16903 (2013).

31. Demberg LM, Rothemund S, Schoneberg T, Liebscher I. Identification of the tethered peptide agonist of the adhesion G protein-coupled receptor GPR64/ADGRG2. *Biochemical and biophysical research communications* **464**, 743-747 (2015).

32. Liebscher I*, et al.* A tethered agonist within the ectodomain activates the adhesion G protein-coupled receptors GPR126 and GPR133. *Cell reports* **9**, 2018-2026 (2014).

33. Stoveken HM, Hajduczok AG, Xu L, Tall GG. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 6194-6199 (2015).

34. Wilde C*, et al.* The constitutive activity of the adhesion GPCR GPR114/ADGRG5 is mediated by its tethered agonist. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **30**, 666-673 (2016).

35. Geng FS*, et al.* Semicircular canal morphogenesis in the zebrafish inner ear requires the function of gpr126 (lauscher), an adhesion class G protein-coupled receptor gene. *Development (Cambridge, England)* **140**, 4362-4374 (2013).

36. Monk KR*, et al.* A G protein-coupled receptor is essential for Schwann cells to initiate myelination. *Science (New York, NY)* **325**, 1402-1405 (2009).

37. Waller-Evans H*, et al.* The orphan adhesion-GPCR GPR126 is required for embryonic development in the mouse. *PloS one* **5**, e14047 (2010).

38. Pereira JA, Lebrun-Julien F, Suter U. Molecular mechanisms regulating myelination in the peripheral nervous system. *Trends in neurosciences* **35**, 123-134 (2012).

39. d'Ydewalle C, Benoy V, Van Den Bosch L. Charcot-Marie-Tooth disease: emerging mechanisms and therapies. *The international journal of biochemistry & cell biology* **44**, 1299-1304 (2012).

40. Nave KA, Sereda MW, Ehrenreich H. Mechanisms of disease: inherited demyelinating neuropathies--from basic to clinical research. *Nature clinical practice Neurology* **3**, 453-464 (2007).

41. Monk KR, Oshima K, Jors S, Heller S, Talbot WS. Gpr126 is essential for peripheral nerve development and myelination in mammals. *Development (Cambridge, England)* **138**, 2673-2680 (2011).

42. Mogha A*, et al.* Gpr126 functions in Schwann cells to control differentiation and myelination via G-protein activation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 17976-17985 (2013).

43. Kuffer A*, et al.* The prion protein is an agonistic ligand of the G protein-coupled receptor Adgrg6. *Nature* **536**, 464-468 (2016).

44. Garinet S*, et al.* High prevalence of a hotspot of non-coding somatic mutations in intron 6 of GPR126 in bladder cancer. *Molecular cancer research : MCR*, (2018).

45. Nik-Zainal S*, et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).

46. Piraino SW, Furney SJ. Identification of coding and non-coding mutational hotspots in cancer genomes. *BMC genomics* **18**, 17 (2017).

47. Qin X*, et al.* Genetic Variant of GPR126 Gene is Functionally Associated With Adolescent Idiopathic Scoliosis in Chinese Population. *Spine* **42**, E1098-e1103 (2017).

48. Terzikhan N*, et al.* Heritability and genome-wide association study of diffusing capacity of the lung. *The European respiratory journal* **52**, (2018).

49. Kou I*, et al.* Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nature genetics* **45**, 676-679 (2013).

50. Ravenscroft G*, et al.* Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. *American journal of human genetics* **96**, 955-961 (2015).

51. Moriguchi T, Haraguchi K, Ueda N, Okada M, Furuya T, Akiyama T. DREG, a developmentally regulated G protein-coupled receptor containing two conserved proteolytic cleavage sites. *Genes to cells : devoted to molecular & cellular mechanisms* **9**, 549-560 (2004).

52. Stehlik C, Kroismayr R, Dorfleutner A, Binder BR, Lipp J. VIGR--a novel inducible adhesion family G-protein coupled receptor in endothelial cells. *FEBS letters* **569**, 149-155 (2004).

53. Bjarnadottir TK, Geirardsdottir K, Ingemansson M, Mirza MA, Fredriksson R, Schioth HB. Identification of novel splice variants of Adhesion G protein-coupled receptors. *Gene* **387**, 38-48 (2007).

54. Xu E, Shao W, Jiang H, Lin T, Gao R, Zhou X. A Genetic Variant in GPR126 Causing a Decreased Inclusion of Exon 6 Is Associated with Cartilage Development in Adolescent Idiopathic Scoliosis Population. *BioMed research international* **2019**, 4678969 (2019).

55. Holm L, Laakso LM. Dali server update. *Nucleic acids research* **44**, W351-355 (2016).

56. Harding MM. The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. *Acta crystallographica Section D, Biological crystallography* **56**, 857-867 (2000).

57. Gaboriaud C, Gregory-Pauron L, Teillet F, Thielens NM, Bally I, Arlaud GJ. Structure and properties of the Ca(2+)-binding CUB domain, a widespread ligand-recognition unit involved in major biological functions. *The Biochemical journal* **439**, 185-193 (2011).

58. Andersen CB, Madsen M, Storm T, Moestrup SK, Andersen GR. Structural basis for receptor recognition of vitamin-B(12)-intrinsic factor complexes. *Nature* **464**, 445-448 (2010).

59. Cohen-Dvashi H, Kilimnik I, Diskin R. Structural basis for receptor recognition by Lujo virus. *Nature microbiology* **3**, 1153-1160 (2018).

60. Gingras AR*, et al.* Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: implications for complement activation. *Structure (London, England : 1993)* **19**, 1635-1643 (2011).

61. Venkatraman Girija U*, et al.* Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation. *Proceedings of the*

National Academy of Sciences of the United States of America **110**, 13916-13920 (2013).

62. Ashkenazy H*, et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **44**, W344-350 (2016).

63. Paavola KJ, Sidik H, Zuchero JB, Eckart M, Talbot WS. Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Science signaling* **7**, ra76 (2014).

64. Gordon WR, Vardar-Ulu D, Histen G, Sanchez-Irizarry C, Aster JC, Blacklow SC. Structural basis for autoinhibition of Notch. *Nature structural & molecular biology* **14**, 295-300 (2007).

65. Macao B, Johansson DG, Hansson GC, Hard T. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nature structural & molecular biology* **13**, 71-76 (2006).

66. Gordon WR*, et al.* Effects of S1 cleavage on the structure, surface export, and signaling activity of human Notch1 and Notch2. *PloS one* **4**, e6613 (2009).

67. Pelaseyed T, Zäch M, Petersson Å C, Svensson F, Johansson DGA, Hansson GC. Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell protective device. *The FEBS journal* **280**, (2013).

68. van Putten JPM, Strijbis K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. *Journal of innate immunity* **9**, 281-299 (2017).

69. Aoto J, Martinelli DC, Malenka RC, Tabuchi K, Südhof TC. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. *Cell* **154**, 75-88 (2013).

70. Irimia M*, et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511-1523 (2014).

71. Thalhammer A*, et al.* Alternative Splicing of P/Q-Type Ca(2+) Channels Shapes Presynaptic Plasticity. *Cell reports* **20**, 333-343 (2017).

72. Fuccillo MV*, et al.* Single-Cell mRNA Profiling Reveals Cell-Type-Specific Expression of Neurexin Isoforms. *Neuron* **87**, 326-340 (2015).

73. Li J*, et al.* Structural Basis for Teneurin Function in Circuit-Wiring: A Toxin Motif at the Synapse. *Cell* **173**, 735-748.e715 (2018).

74. Sandberg A, Johansson DG, Macao B, Hard T. SEA domain autoproteolysis accelerated by conformational strain: energetic aspects. *Journal of molecular biology* **377**, 1117-1129 (2008).

75. Lum AM, Wang BB, Beck-Engeser GB, Li L, Channa N, Wabl M. Orphan receptor GPR110, an oncogene overexpressed in lung and prostate cancer. *BMC cancer* **10**, 40 (2010).

76. Cho HS, Leahy DJ. Structure of the extracellular region of HER3 reveals an interdomain tether. *Science (New York, NY)* **297**, 1330-1333 (2002).

77. Eubelen M*, et al.* A molecular mechanism for Wnt ligand-specific signaling. *Science (New York, NY)* **361**, (2018).

78. Vallon M*, et al.* A RECK-WNT7 Receptor-Ligand Interaction Enables Isoform-Specific Regulation of Wnt Bioavailability. *Cell reports* **25**, 339-349.e339 (2018).

79. Zhou Y, Nathans J. Gpr124 controls CNS angiogenesis and blood-brain barrier integrity by promoting ligand-specific canonical wnt signaling. *Developmental cell* **31**, 248-256 (2014).

80. Cho C, Smallwood PM, Nathans J. Reck and Gpr124 Are Essential Receptor Cofactors for Wnt7a/Wnt7b-Specific Signaling in Mammalian CNS Angiogenesis and Blood-Brain Barrier Regulation. *Neuron* **95**, 1056-1073.e1055 (2017).

81. Jackson VA*, et al.* Structural basis of latrophilin-FLRT interaction. *Structure (London, England : 1993)* **23**, 774-781 (2015).

82. Jackson VA*, et al.* Super-complexes of adhesion GPCRs and neural guidance receptors. *Nature communications* **7**, (2016).

83. Lu YC*, et al.* Structural Basis of Latrophilin-FLRT-UNC5 Interaction in Cell Adhesion. *Structure (London, England : 1993)* **23**, 1678-1691 (2015).

84. Tu YK, Duman JG, Tolias KF. The Adhesion-GPCR BAI1 Promotes Excitatory Synaptogenesis by Coordinating Bidirectional Trans-synaptic Signaling. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **38**, 8388-8406 (2018).

85. Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* **30**, 21-32 (2009).

86. Booth DS, Avila-Sakar A, Cheng Y. Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. *Journal of visualized experiments : JoVE*, (2011).

87. Tang G*, et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157**, 38-46 (2007).

88. Petoukhov MV*, et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of applied crystallography* **45**, 342-350 (2012).

89. Hwang WY*, et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013).

90. Cunningham RL, Monk KR. Whole Mount In Situ Hybridization and Immunohistochemistry for Zebrafish Larvae. *Methods in molecular biology (Clifton, NJ)* **1739**, 371-384 (2018).

91. Lyons DA*, et al.* erbb3 and erbb2 are essential for schwann cell migration and myelination in zebrafish. *Current biology : CB* **15**, 513-524 (2005).

Acknowledgments

We thank Engin Özkan for assistance with crystal structure determination and for insightful discussions, the He lab for the use of their luminescence plate reader, and other members of the Araç lab for helpful discussions. We also thank the staff at the University of Chicago Advance Electron Microscopy Facility, the Advanced Photon Source (APS) at Argonne National Labs (ANL), the Washington University Zebrafish Consortium Facility, and Fernanda Coelho for assistance with *in situ* hybridization. Crystal diffraction and SAXS data were collected at the APS, GM/CA 23-ID-D (supported by NCI ACB-12002 and the NIGMS AGM-12006 grants) and BioCAT 18-ID (supported by NIH grant 9 P41 GM103622-18), respectively. This research used resources of the APS, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by ANL under Contract No. DE-AC02-06CH11357. The Eiger 16M detector was funded by an NIH–Office of Research Infrastructure Programs, High-End Instrumentation Grant (1S10OD012289-01A1). Use of the Pilatus 3 1M detector was provided by grant 1S10OD018090-01 from NIGMS. This work was supported by NIH grants R01-GM120322 (D.A.), R01-NS079445 (K.R.M.), and T32GM007183 (K.L.). R.L.C. was supported by the National Science Foundation Graduate Research Fellowship (DGE-1745038). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Author Contributions

K.L. cloned, expressed, purified proteins (with assistance from E.F.), carried out bioinformatic and biochemical characterizations, performed crystallography experiments (with assistance from J.L.) and structure determination, performed cAMP-based signaling assays, and collected and analyzed negative stain EM data (with assistance from M.Z.). R.L.C. and K.R.M. designed and analyzed and R.L.C. performed zebrafish experiments. J.A.R., T.R.S., and K.L. designed and performed SAXS experiments. K.L. and D.A. designed all experiments, interpreted results, and wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figure Legends

Figure 1. Crystal structure of the full extracellular region of Gpr126. (A) Domain organization of Gpr126, indicating the ECR and 7TM regions. The unknown region includes a splice site. Cleavage sites (furin cleavage, autoproteolysis) are indicated by dashed lines. Domains are colored dark blue (CUB), cyan (PTX), grey (unknown region), yellow (HormR), red (GAIN) and purple (7TM). Domain boundaries are indicated below. SP indicates signal peptide. (B) Structure of the full ECR of (-ss) Gpr126. Domains are colored as in (A) except for the newly identified SEA domain (green). Domains are numbered (1-5) from N to C-terminus. Calcium ion in CUB domain is indicated as a green sphere. Dashed lines represent disordered residues. N-linked glycans are shown as green sticks. (C) Schematic of full-length Gpr126. The previously unknown region (SEA domain and linker region) is labeled. Autoproteolysis in GAIN domain is indicated by an asterisk (*) and the last beta-strand of the GAIN domain is colored grey. (D) Representative negative stain EM 2D class average of Gpr126 (-ss) ECR. Scale bar (white) represents 50 Å. Domains are assigned and colored according to color scheme noted above. The dashed line represents the linker region.

Figure 2. The closed conformation of Gpr126 is mediated by CUB-HormR-linker interactions. (A) Structure of the full ECR of (-ss) Gpr126. (B) Close-up view of the CUB-HormR interface. Resides at the interface are shown as sticks. The calcium ion is shown as a bright green sphere. (C) Close-up view of the calcium-coordination site within CUB domain. The water molecule is shown as a blue sphere. The residues are shown as sticks. CUB residues are colored dark blue and HormR residue is colored yellow. Residue labels are colored according to their roles in CUB-HormR interaction: red (E89, D97, D134) represents calcium coordination by side chain residue, blue (S136, V137) represents calcium coordination

by main chain carbonyl group, purple (F135) represents a hydrophobic residue in CUB-HormR interface, and orange (Y61) represents a residue that stabilizes calcium-coordinating residue D97. Calcium coordination is shown as bright green dashed lines. CUB-HormR interaction is shown as yellow dashed lines. The interaction between Y61 and D97 is shown as a magenta dashed line. (D) Sequence alignment of partial Gpr126 CUB domain from various species, highlighting important conserved residues: calcium-coordinating residues by side chain group (red), calcium-coordinating residues by main-chain carbonyl (blue), a tyrosine residue that stabilizes a calcium-coordinating residue (orange), and a hydrophobic phenylalanine residue in the CUB-HormR interface (purple). (E) Close-up view of the disulfide-stabilized loop inserted between CUB and HormR domains. The disulfide bond is colored bright orange and is indicated by an arrow. The dashed line represents disordered residues in the linker region.

Figure 3. Alternative splice isoforms of Gpr126 modulate ECR conformation. (A) Schematic diagram of Gpr126 splice isoforms generated by including (+ss) or excluding (-ss) exon 6. Residues encoded by exon 6 are colored magenta. Grey asterisks indicate potential O-linked glycosylation sites and the black asterisk indicates a predicted N-linked glycosylation site. The conserved disulfide bond in the linker is colored yellow. (B, C) Negative stain EM 2D class averages for -ss (B) and +ss (C) ECR constructs. Class averages are categorized according to similar orientations: (i, ii, iii, iv, v and vi). (i, ii, iii, iv) are observed in both –ss and +ss isoforms. (vi) represents open-like conformations (>50° angle) that are observed only in the +ss isoform. (v) represents unidentifiable miscellaneous views. Scale bars (white) represents 50 Å. (D) Quantification of percentage of particles per category for both isoforms. (E) Representative individual particles for both isoforms. Yellow arrows point to particles which are not in a closed conformation. (F) ECR conformations based on negative stain EM are

depicted as cartoons. The splice site is shown in magenta. Black arrows with dashed lines indicate dynamic ECR conformation.

Figure 4. Alternative splice isoforms of Gpr126 modulate receptor signaling. (A) Cell-surface expression levels for empty vector (EV), zebrafish Gpr126 splice isoforms, measured using flow cytometry to detect binding of anti-FLAG antibody to cells expressing FLAG-tagged Gpr126. The Gpr126 cell-surface expression levels are normalized to the control EV signal. Error bars are not shown because expression levels are presented as median fluorescence intensities of 10,000 cells for each population of transfected cells, for a single flow cytometry experiment representative of at least three independent experiments. (B) Basal signaling measured by the cAMP signaling assay. Data are shown as fold increase over EV of RLU (relative luminescence units). (C) Basal cAMP signaling normalized to cell-surface expression. ns, $P > 0.05$; *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$; by one-way ANOVA and Tukey's multiple comparisons test. Data in (B) and (C) are presented as mean ± SEM, n = 3, and are representative of at least three independent experiments. (D-F) Same as (A-C) but for human GPR126 splice isoforms. Source data are provided as a Source Data file.

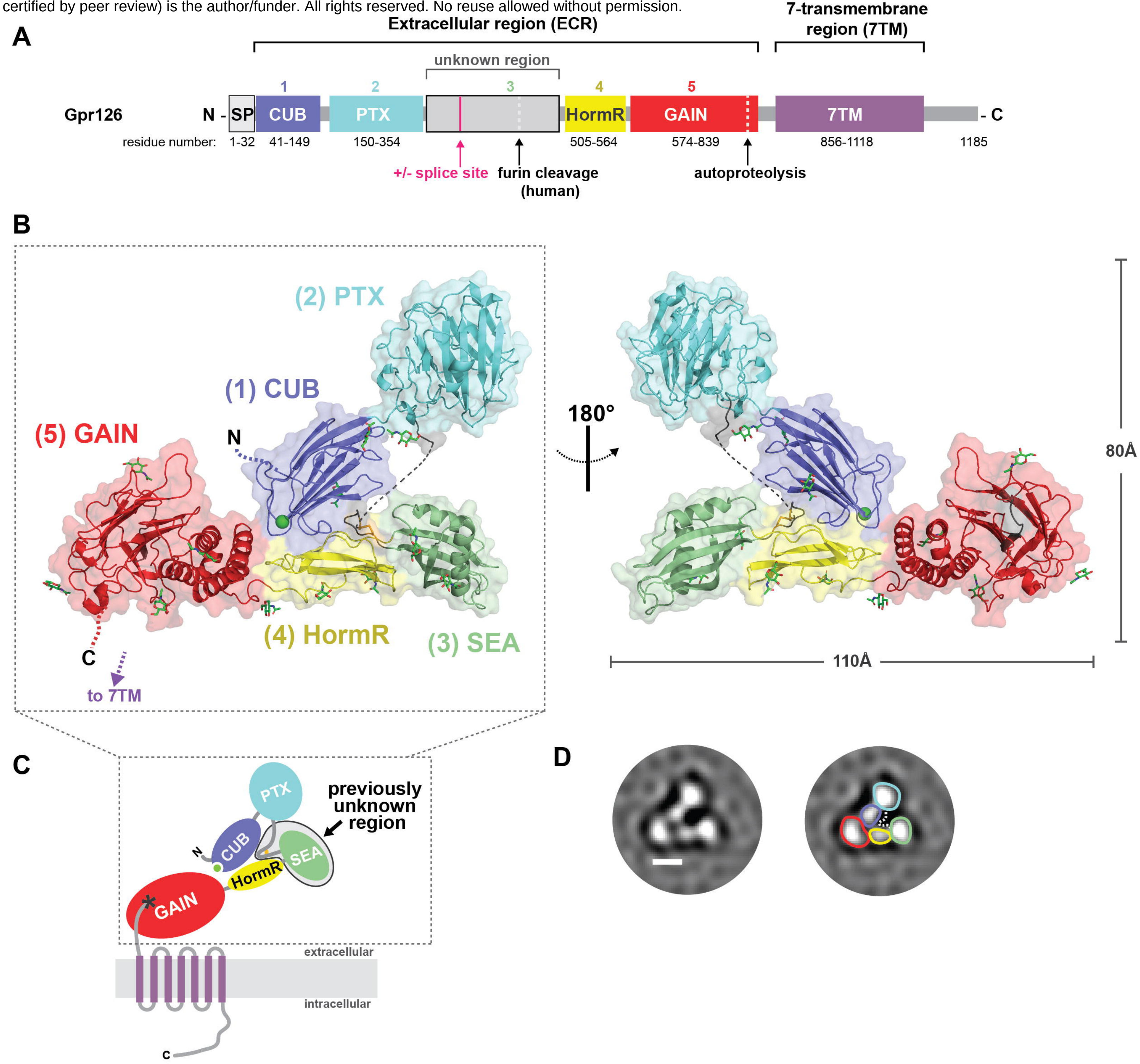
Figure 5. Zebrafish with mutations in the calcium-binding pocket in CUB domain show defective ear and Schwann cell development (A) Surface conservation analysis of CUB domain. The calcium-binding site is circled in magenta. D134 and F135 are indicated by arrows. (B) D134 and F135 were both mutated to alanines through homologous recombination of a 150 bp ssODN containing a 5 bp mutation (red nucleotides). (C) Genotyping assay for the *gpr126*^{stl464} lesion. The 5 bp mutation introduces a BstUI restriction enzyme binding site. (D) 4 dpf wild-type larva compared to (E) 4 dpf *gpr126*$^{stl464/stl464}$ larva with puffy ears (arrowheads). (F) 4 dpf wild-type larvae express *mbp* throughout the posterior lateral line nerve (PLLn,

arrowhead), whereas (G) 4 dpf *gpr126*$^{stl464/stl464}$ larva lack *mbp* expression along the PLLn (arrowhead). Asterisks indicate CNS.

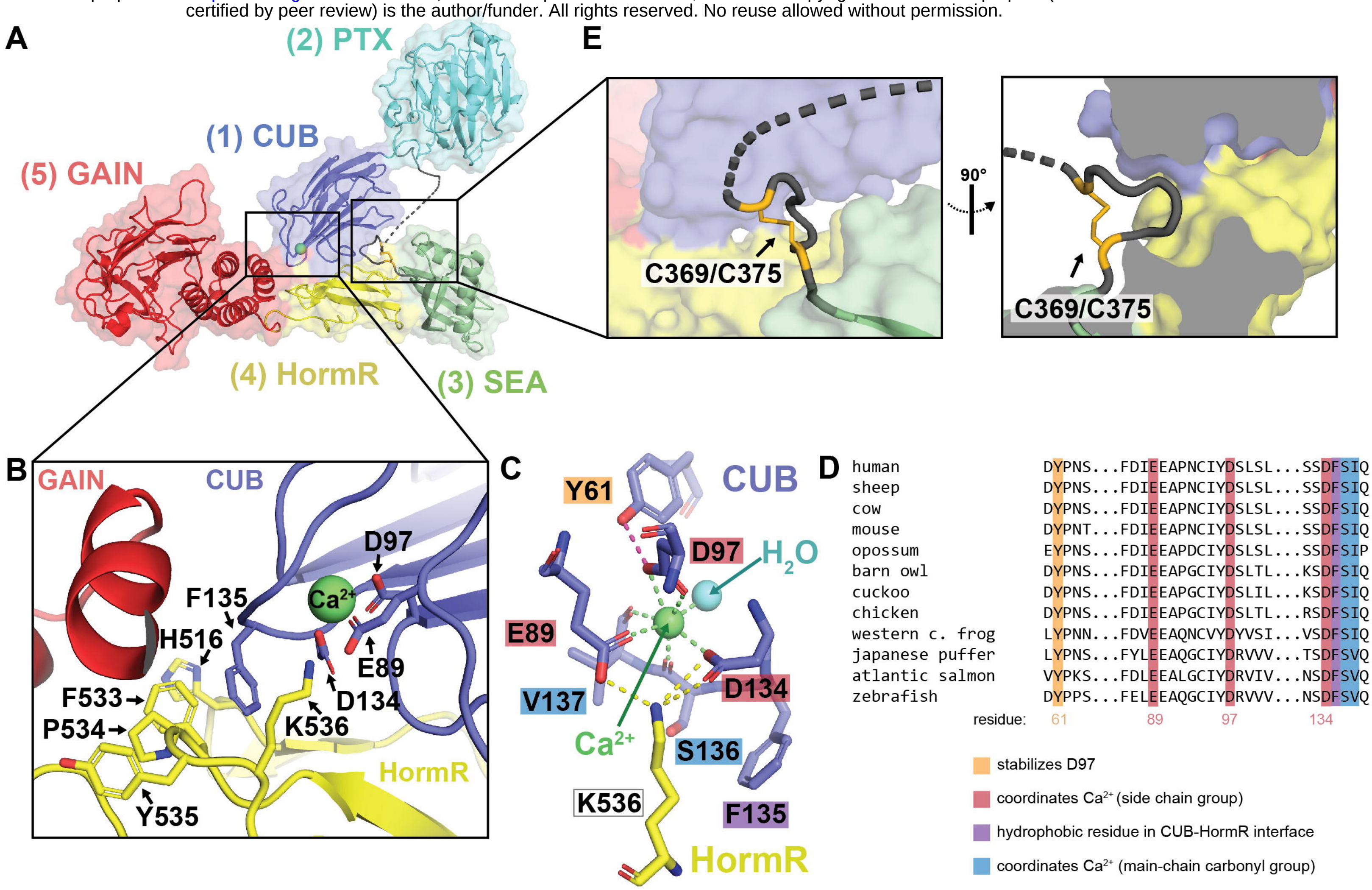
Figure 6. Identification of a proteolytic SEA domain in human GPR126. (A) Crystal structure of the SEA domain from zebrafish Gpr126. (B) (left) NMR structure of Mucin-1 SEA domain (PDB: 2ACM) and (right) Gpr126 SEA domain superimposed over Mucin-1 SEA domain. The loop containing the autoproteolysis site in Mucin-1 is indicated by a yellow arrow. The root-mean-square deviation (RMSD) of atoms between overlaid structures is 2.761 Å. (C) (left) Crystal structure of the Notch2 SEA domain (PDB: 2OO4) and (right) Gpr126 SEA domain superimposed over Notch2 SEA domain. The loop containing the furin cleavage site (deleted in crystal structure construct) is indicated by an orange arrow. The RMSD of atoms between overlaid structures is 4.767 Å. (D) Sequence alignment of partial SEA domain from human Mucin-1, human Notch2, human GPR126, and zebrafish Gpr126. (D) Homology model of human GPR126 SEA model generated using SWISSMODEL. The arrow points to modelled furin-cleavage site. (E) Protein topology map of SEA domain. Furin cleavage site is indicated by red scissors. Residues N-terminal to cleavage site are dark blue and residues C-terminal to cleavage site are light blue. Dashed lines represent backbone hydrogen bonds between beta sheets.

Figure 7. Model for ECR-dependent functions of Gpr126

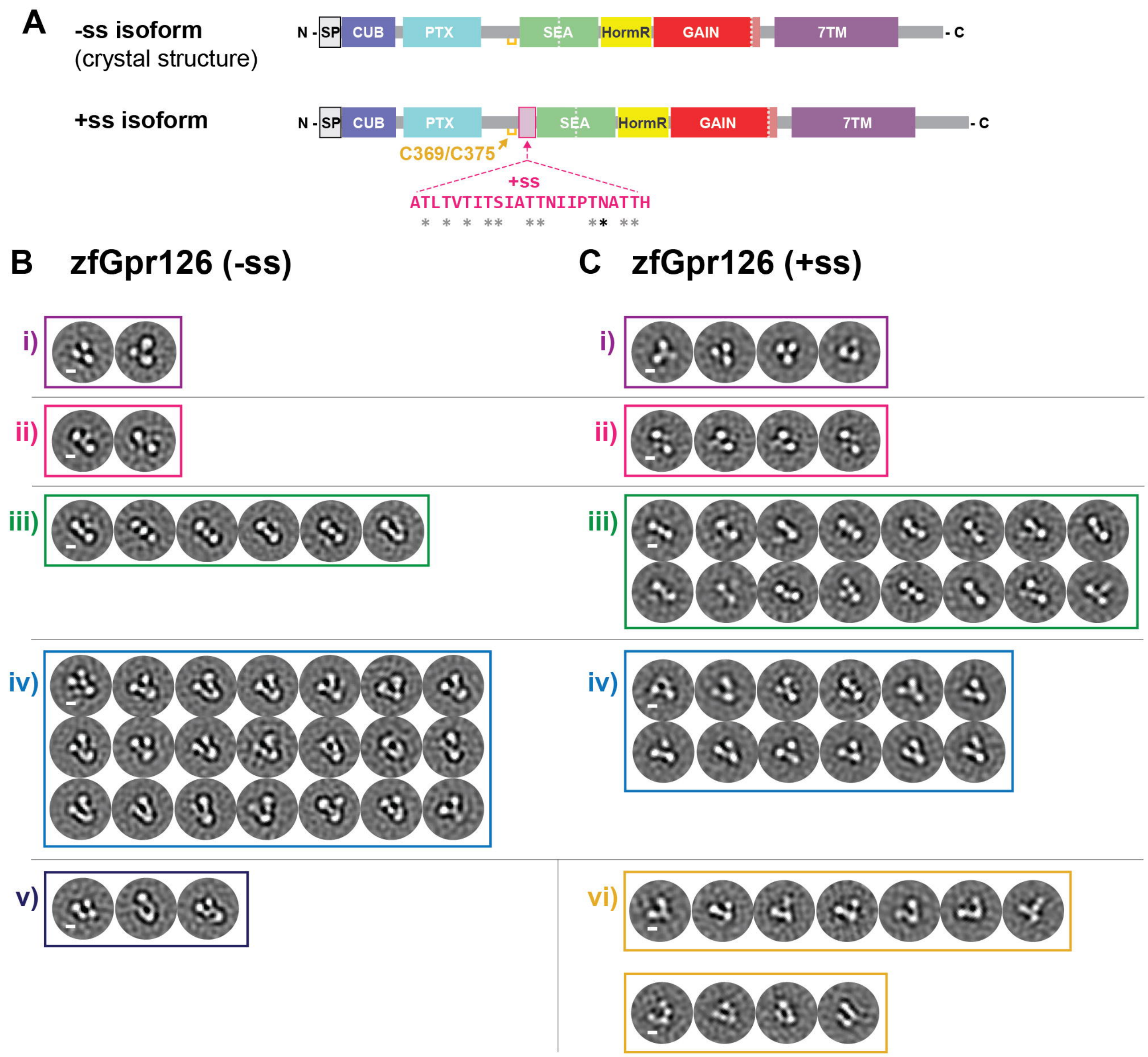
The model depicts how Gpr126/GPR126 function is regulated by its ECR. (left panel) Alternative splicing acts as a molecular switch to adopt different ECR conformations and have different basal levels of signaling. Gpr126 ECR that lacks the splice insert adopts a closed conformation and has basal activity, whereas Gpr126 ECR that includes the splice insert is more dynamic and open-like, and has enhanced basal activity. (right panel) Mutation of conserved residues within the calcium-binding site leads to defects in both myelination and ear development *in vivo*. Human GPR126 function may also be regulated by furin cleavage, indicated by a green asterisk.



Leon et al., Figure 1

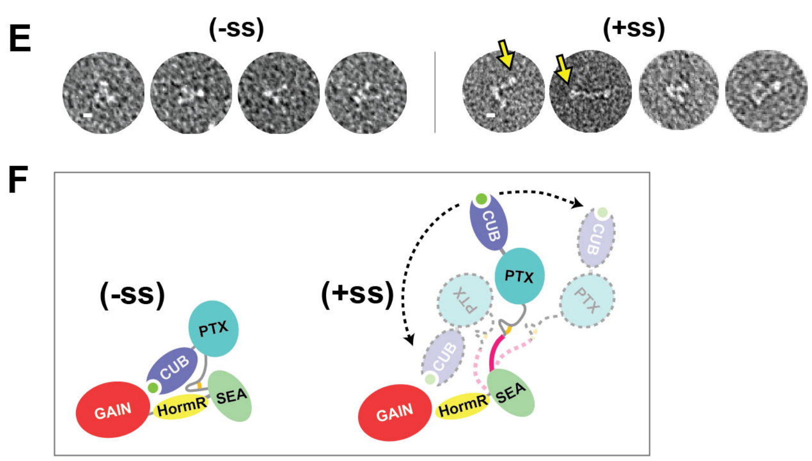


Leon et al., Figure 2

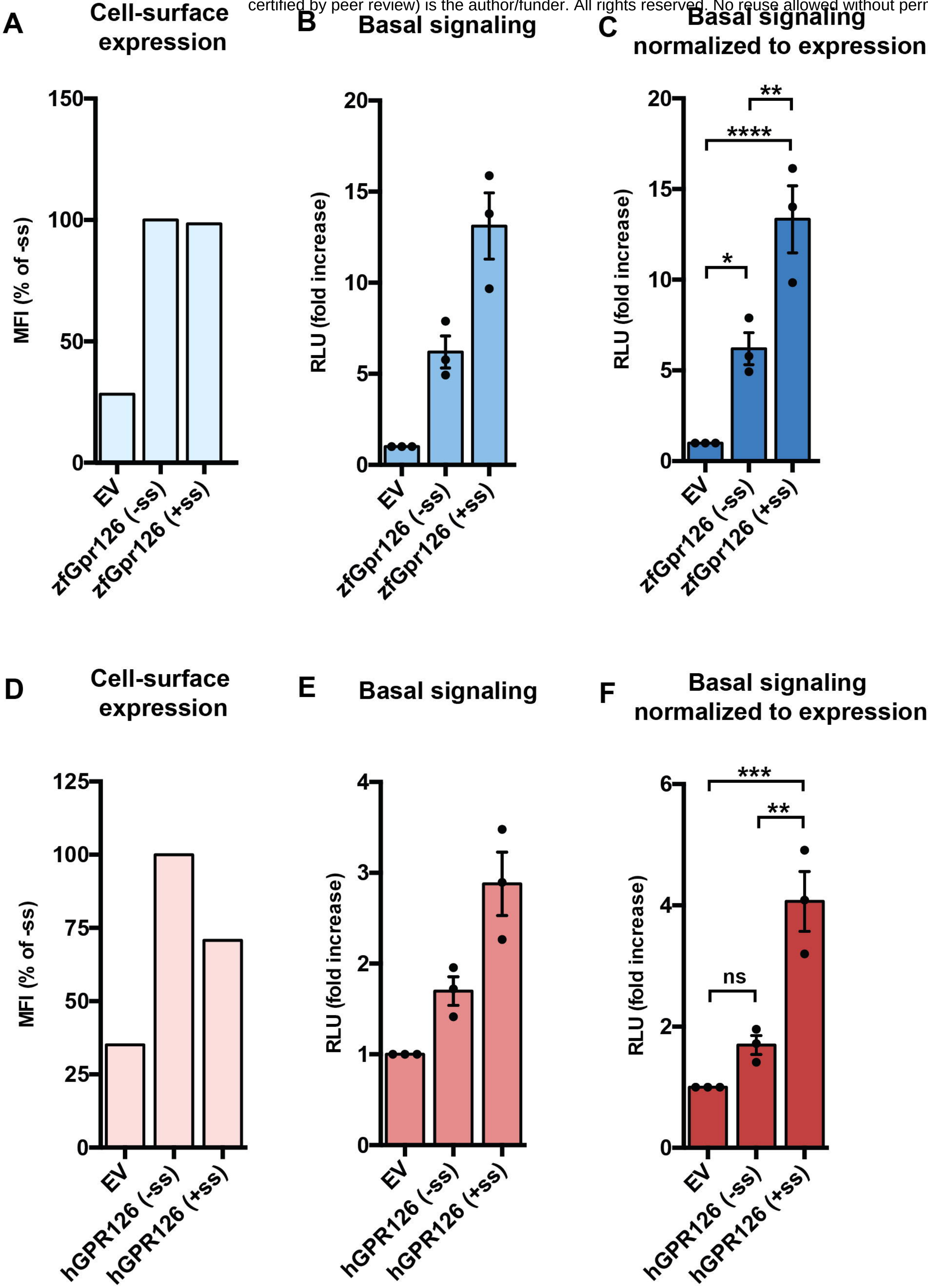


D

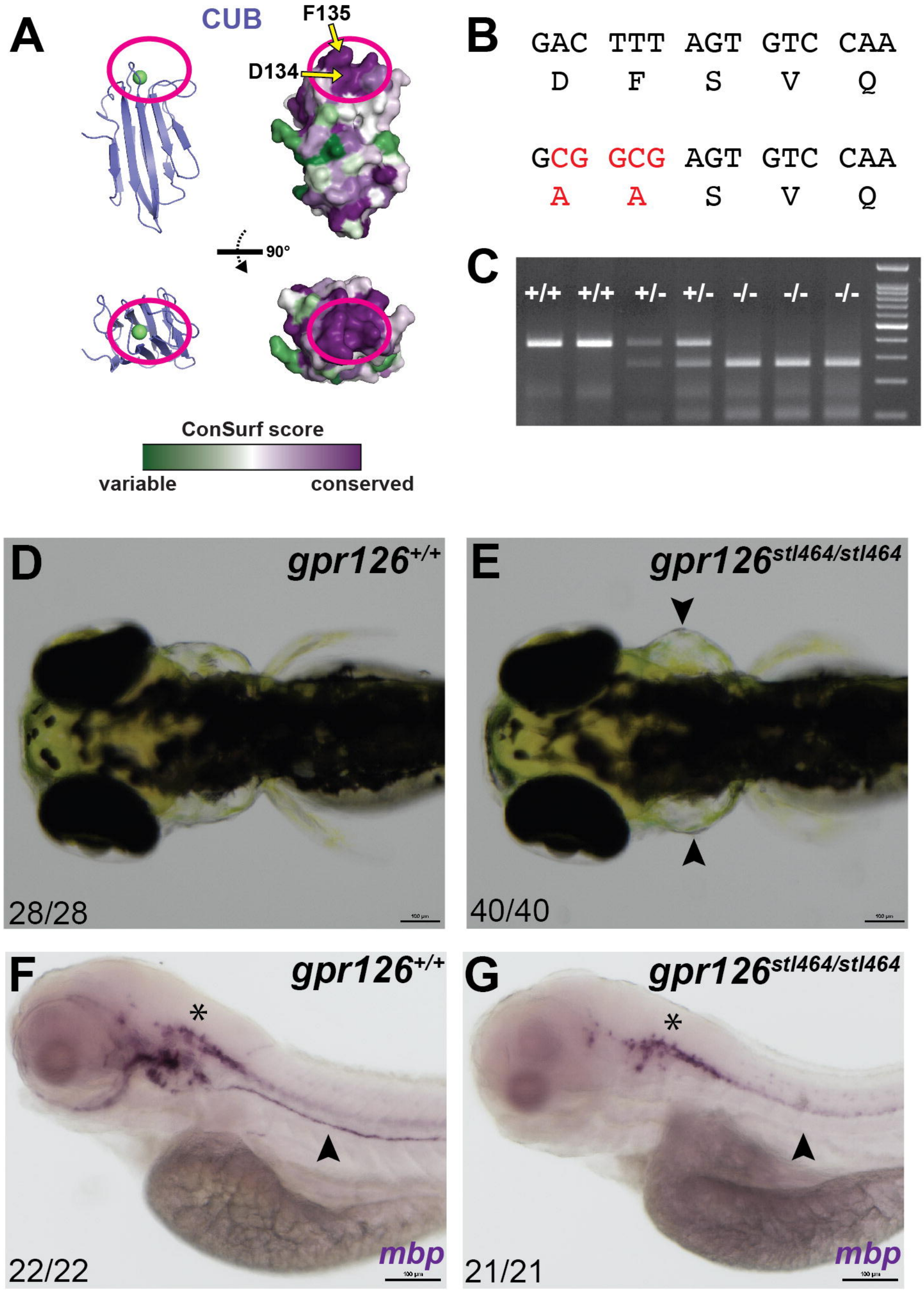
group	-ss	+ss
i	5%	11%
ii	6%	9%
iii	16%	32%
iv	64%	27%
v	9%	---
vi	---	21%



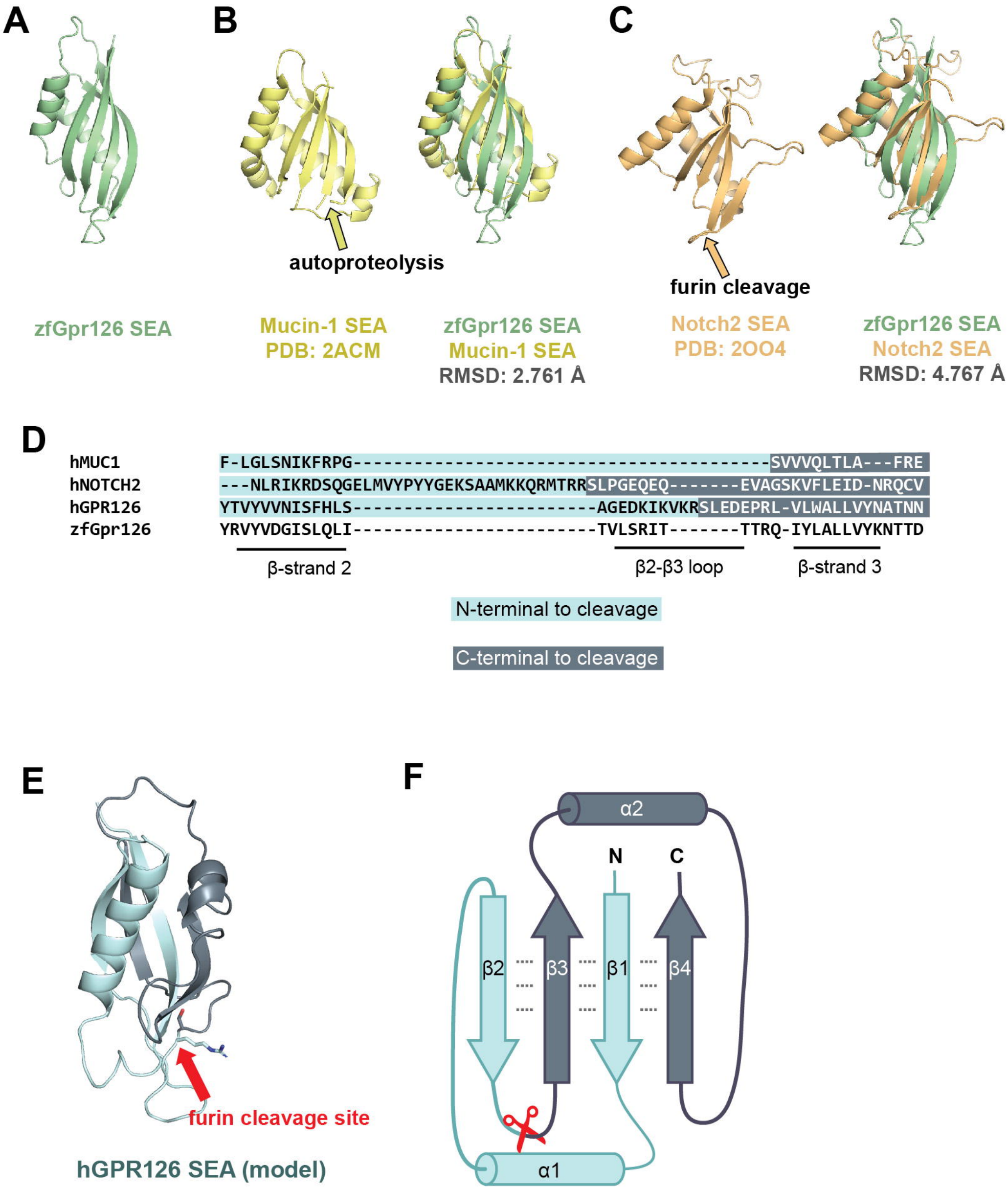
Leon et al., Figure 3



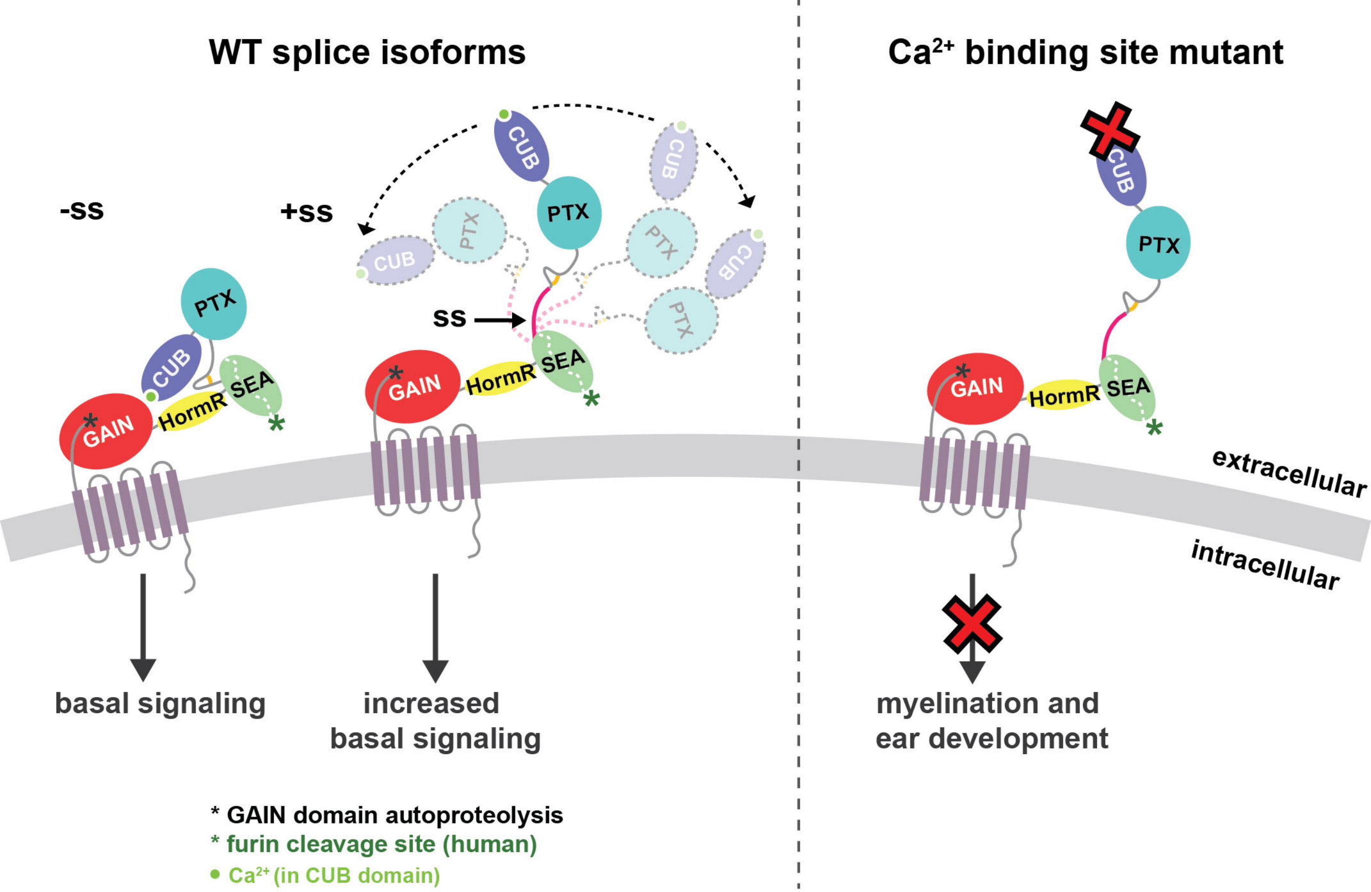
Leon et al., Figure 4



Leon et al., Figure 5



Leon et al., Figure 6



Leon et al., Figure 7