Molecular and computational studies of ADGRG6 reveal its extracellular CUB domain as a key regulator of cAMP signaling

Ethan Dintzner*, Sumit J. Bandekar*, Katherine Leon, and Demet Araç†

Department of Biochemistry and Molecular Biology, University of Chicago Neuroscience Institute, Institute for Biophysical Dynamics, and Center for Mechanical Excitability, University of Chicago, Chicago, IL, 60637, USA

*These authors contributed equally to this work.

†To whom correspondence should be addressed, arac@uchicago.edu

Short Title: Regulation of ADGRG6 by its CUB domain

Keywords: adhesion GPCR, molecular dynamics, ADGRG6, cell signaling, extracellular region

Abstract

ADGRG6 is a member of the adhesion G Protein-coupled receptor (aGPCR) family that regulates Schwann cell myelination among other developmental processes. ADGRG6 is associated with developmental diseases and cancers and ADGRG6 signaling activity is regulated by its multidomain extracellular region (ECR). Previous work shows that the ADGRG6 ECR adopts a compact conformation mediated by its Complement C1r/C1s, Uegf, Bmp1 (CUB) domain. An alternatively spliced form of ADGRG6 (+ss) has higher signaling levels and does not adopt the compact conformation, suggesting that the CUB domain regulates ADGRG6 signaling activity. However, mechanistic detail of how the CUB domain regulates ADGRG6 signaling is lacking. Here, we show that deletion of the ADGRG6 CUB domain decreases signaling levels of the receptor. Using microsecond-length molecular dynamics simulations, we demonstrate that a cancer-associated variant C94Y disrupts the compact conformation of ADGRG6. C94Y exposes the CUB ligand binding site to solvent without introducing any change to the binding site itself. In cell-based assays, C94Y signals significantly higher than the wild-type receptor. Another cancer associated variant, Y96A, removes a critical residue in the canonical CUB domain ligand binding site and this variant inhibits ADGRG6 signaling. Several CUB domains bind to collagen-like structures using the same surface which prominently features Y96. Taken together with the literature, our results strongly suggest that the CUB domain regulates ADGRG6 activity through its ligand binding surface, either through intramolecular binding to the ADGRG6 ECR, or to a putative ADGRG6 ligand.

Introduction

Multicellular organisms rely on cell-surface receptors to mediate the intercellular communication required for critical biological processes (*1*, *2*). Cell-surface receptors use their extracellular regions (ECRs) to sense signals and transmit this information to inside the cell. For instance, the ECRs of several G Protein-Coupled Receptors (GPCRs) propagate signals across the membrane through binding of small molecule or protein ligands (*3*). Furthermore, the ECRs of many cell surface receptors are attractive drug targets (*4*). Despite their importance, the molecular mechanisms of how ECRs regulate receptor function remain incompletely understood.

Adhesion G Protein-Coupled Receptors (aGPCRs) are a diverse and understudied family of GPCRs with critical roles in synapse formation, myelination of the nervous system, angiogenesis, neutrophil activation, embryogenesis, and more (*5*–*8*). While other families of GPCRs represent over a third of FDA-approved drugs (*9*), there are no approved drugs targeting aGPCRs despite their roles in developmental diseases, conditions of the nervous system, and cancer (*8*, *10*, *11*). aGPCRs contain characteristically large and complex ECRs, several of which have been shown to bind ligands or modulate the GPCR transmembrane domain (*12*–*19*). Alternative splicing often regulates the binding of aGPCRs to their ligands, such as with ADGRL binding to TENs, which regulates synapse specificity (*17*, *20*) and ADGRG1 binding to phosphatidylserine during synaptic pruning (*16*). Alternative splice forms of the ECR can also modulate signaling of the 7TM region in several aGPCRs, including ADGRG6 which is the focus of this work (*15*, *18*, *21*).

ADGRG6 is critical for peripheral nervous system (PNS) myelination, skeletal development, angiogenesis, and heart development, among other functions (*22*–*25*). ADGRG6 function in PNS myelination is dependent on its ability to elevate intracellular cyclic adenosine monophosphate (cAMP) levels (*26*). *adgrg6*-null zebrafish and mice display severe myelination defects, a phenotype which is rescued by treatment by forskolin, a chemical activator of adenylyl cyclase, or cAMP analogs (*26*, *27*). ADGRG6 is suggested to be causative in the developmental conditions arthrogryposis multiplex congenita (AMC), where patients with point mutations in the ADGRG6 ECR display loss of PNS myelination (*28*). Additionally, mutations in ADGRG6 are associated with adolescent idiopathic scoliosis (AIS) (*29*–*32*). Notably, AIS mutations can alter the alternative

splicing of ADGRG6 in its ECR and this could alter signaling and contribute to disease progression (*18*, *31*). ADGRG6 is also associated with cancer (*33–35*). Although ADGRG6 is strongly connected to human development and disease through its ECR, the molecular mechanisms underlying the function of the ADGRG6 ECR are incompletely understood.

ADGRG6 has a multidomain ECR consisting of the Complement C1r/C1s, Uegf, and Bmp1 (CUB), Pentraxin (PTX), sperm protein, enterokinase and agrin (SEA), Hormone Receptor (HormR), and GPCR autoproteolysis inducing (GAIN) domains (Fig. 1A). CUB domains bind to a diverse set of ligands, typically using several anionic residues coordinating a Ca^{2+} ion and a key tyrosine residue as part of the interaction surface (*18*, *36–38*). The ligand typically extends a lysine from its surface which engages in cation-pi interactions with the conserved tyrosine and interacts with the anionic residues on the CUB domain. For ADGRG6, a tandem of the CUB-PTX domains is sufficient to bind its putative ligand Collagen IV (*39*). Pentraxin domains typically multimerize and they can bind small molecule ligands (*40–42*). SEA domains commonly have autoproteolysis sites within them and this has been shown for human ADGRG6 (*18*). HormR domains are homologous to hormone binding receptors, and the GAIN domain is common to almost all aGPCRs and contains the tethered agonist for the 7TM region (*18*, *43*). Following the ECR is a GPCR seven-transmembrane region (7TM) which couples to $G\alpha_s$.

ADGRG6 is alternatively spliced at exon 6 leading to versions of the protein product either lacking (-ss) or including (+ss) a 23-residue splice insert between the PTX and SEA domains (Fig. 1A) (*18*). The ECR of -ss ADGRG6 is found in a compact conformation where the CUB domain engages its canonical ligand binding surface to interact with HormR (Fig. 1B). This involves a key lysine from the 'ligand', which is the HormR domain, interacting with anionic residues from CUB, including E89 and D134 which coordinate a Ca^{2+} ion. The lysine also has cation-pi interactions with the conserved tyrosine Y96 (Fig. 1C). Other residues, such as F135, are involved for a total of 700 $Å^2$ of surface area buried from solvent.

The +ss ADGRG6 displays higher cAMP signaling than the -ss version (*18*). Furthermore, the splice insert profoundly affects the conformation of the ECR. The +ss ECR contains 23 extra residues in a loop which rests between the CUB and HormR domains in the structure but is distal to the

canonical CUB interaction surface. Also, the +ss ECR is found in a series of extended conformations where the CUB domain is removed from its HormR site and exposed to solvent. Since the absence or presence of the splice insert in ADGRG6 affects both ECR conformation and signaling activity, it is suggested that the conformation of the ECR regulates 7TM activity (*18*). However, the mechanism of this remains unclear. Further work is needed to fully decode the molecular basis by which the ECR regulates the functions of ADGRG6. This will shed light on how ECR mutations in ADGRG6 contribute to disease and pave the way towards therapeutic development.

In this work, we employ a combination of cell signaling assays and microsecond-length molecular dynamics (MD) simulations to understand the mechanism by which CUB regulates the function of ADGRG6. We find that deletion of the CUB domain inhibits the capacity of ADGRG6 to elevate intracellular cAMP levels in both the –ss and +ss splice forms. Using MD simulations, we show that the WT ADGRG6 ECR adopts multiple compact conformations where the CUB ligand binding site remains bound to HormR. We demonstrate that the cancer associated variant C94Y completely removes the CUB domain from the HormR binding site but does not affect the canonical CUB ligand binding surface. This variant displays significantly higher signaling than WT ADGRG6. Another cancer-associated variant, Y96A, removes a critical residue in the CUB domain ligand binding site and does not access the higher signaling state, similar to the previously known variant D134A/F135A. Taken together, these results suggest that exposure of the CUB domain to solvent activates ADGRG6 signaling as long as the CUB ligand binding site is not affected. Our results are consistent with a central role for the CUB domain ligand binding site in regulating ADGRG6 function.

Results

Deletion of the CUB domain inhibits ADGRG6 cAMP signaling

We used a previously optimized assay to measure basal signaling of full-length ADGRG6, which raises intracellular cAMP levels (*18*). As previously shown (*18*), WT +ss ADGRG6 signals at a higher level than the -ss receptor (Fig. 1 D-F). To test whether the CUB domain has a regulatory effect on ADGRG6 signaling, receptor activity was measured for both full length (WT) and CUB-deleted (ΔCUB) receptors, using both the -ss and +ss isoforms. We observed that deletion of the CUB

domain in both isoforms resulted in significantly less cAMP elevation than their WT counterparts. This demonstrates that the CUB domain is required for the basal signaling of ADGRG6.

MD simulations of the ADGRG6 -ss ECR show sustained interdomain contacts involving the CUB, PTX, and SEA domains

To investigate the conformational dynamics and interdomain interactions of the ADGRG6 ECR, we performed several one-microsecond all-atom MD simulations initialized from the crystal structure of the -ss ECR of ADGRG6 (Fig. 2A). Throughout these simulations, the ADGRG6 CUB domain remained associated to the HormR binding site. However, the observed conformations were noticeably different from the crystal structure (Fig. 2B). Interface (i), involving the CUB domain contacting HormR, was mostly unchanged from the crystal structure, however, a new interface (ii) formed within 100 ns of the simulations as the PTX domain came in close proximity to SEA as well as the back of the CUB domain, and this interface remained intact for the remainder of the simulations. To quantify the extent of the interaction between CUB and the rest of the ECR (CUB:ECR), as well as PTX and ECR (PTX:ECR), we performed buried surface area (BSA) and hydrogen bond calculations (Fig. 2C, D). Throughout all simulations, the CUB domain buries approximately 1000 $Å^2$ from solvent and forms on average 10 hydrogen bonds with the rest of the ECR. The PTX and SEA/CUB domains take about 100 ns to associate, but once they do, they bury approximately 350 $Å^2$ from solvent and form on average 3 interdomain hydrogen bonds. Our CUB:ECR interface is greater than that seen in the crystal structure (700 $Å^2$) and this is likely due to the PTX domain docking onto the CUB domain on the opposite face of the canonical ligand binding site. These results suggest that the CUB interaction with HormR is robust and mediated by a network of hydrophobic and hydrophilic contacts, and that the PTX domain could associate with the SEA and CUB domains in solution.

ADGRG6 -ss adopts multiple conformations which are distinct from the crystal structure and involve large conformational changes of the CUB-PTX domains

We performed principal component analysis (PCA) on the concatenated Cα coordinates of our simulations to understand the conformational dynamics of the -ss ADGRG6 ECR. Our analysis revealed two principal components (PCs) which capture 85% of the Cα conformational variance.

The first of these two PCs corresponds to a ~40 º bend of the CUB-PTX domains about the GAIN domain (Fig. 3A). The CUB domain remains associated to the HormR binding site in all cases, despite the CUB-PTX tandem exhibiting conformational diversity. The second PC is represented by an orthogonal ~60 º bending motion of CUB, PTX, SEA, and HormR domains about the GAIN domain involving relatively equal contributions from Cαs in all domains (Fig. 3B). The projection of each simulation onto PCs 1 and 2 (Fig. 3C) shows that each simulation settles in its own region of the PC space, highlighting that ADGRG6 can adopt a multitude of conformations. We calculated the Gibbs free energy of each simulation (Fig. 3D) and observed three conformational basins. Using hierarchical clustering, we separated the PC space into three clusters. The centroids of each cluster are distinct conformations from the crystal structure (Fig. 3E), with Cα RMSDs of 13.3 Å, 15.9 Å, and 11.5 Å for centroids 1, 2, and 3 respectively. These data show that the ADGRG6 -ss ECR can adopt several conformational states, but throughout all these states the CUB domain remains docked to its binding site on HormR.

The ADGRG6 -ss ECR variant C94Y adopts a conformation where the CUB ligand binding site is exposed to solvent

We simulated the C94Y variant of -ss ADGRG6, which is present in cancer patients (*44*), using the same system as above. C94 is part of the extended interface CUB makes with HormR however it is distal from the canonical CUB ligand binding site. We thus hypothesized the C94Y variant, which inserts a bulky tyrosine residue into the position of a small cysteine residue, may disrupt this interface. In our simulation, the C94Y variant resulted in complete loss of the CUB/HormR interface observed across multiple simulations of WT ADGRG6 -ss (Fig. 4). Similar to the WT simulations, the C94Y PTX domain first made contact with the SEA domain (Fig. 4A). Following this, the CUB domain shifted away from the HormR domain and continued to complete dissociation, during which it was highly flexible. Towards the last portion of the simulation, CUB associated with the lateral side of HormR in a manner that leaves the canonical CUB ligand binding site solvent exposed. Using PCA, free energy, and BSA analysis, we found that after the CUB domain dissociates from the HormR ligand site, it travels to a low energy basin which coincides with a profound loss of BSA between CUB and the rest of the ECR (Fig. 4B, C). To obtain a direct comparison between the

dynamics sampled by C94Y ADGRG6 vs WT ADGRG6, we concatenated the simulations and performed PCA to visualize them on a combined essential space (Fig. 4D). Using hierarchical clustering, we grouped the essential space into four regions, one of which is exclusively occupied by C94Y. Taken together, these results show that the C94Y variant adopts a favorable conformational state unique from the WT receptor where the CUB ligand binding surface is exposed to solvent.

ADGRG6 point variants sourced from cancer patients affect receptor activity

To test the mechanistic hypotheses based on the MD simulations, we employed the ADGRG6 cAMP assay to test whether variants sourced from cancer patients in the CUB/HormR interface could affect receptor activity. C94Y leads to exposure of the CUB ligand binding site in our simulations without affecting any residues key to the canonical CUB ligand binding surface. We also tested Y96A, which removes a crucial and highly conserved residue in the CUB ligand binding site (*37*, *38*). Additionally, we tested P520Q, which involves a proline on HormR which contacts CUB yet is distal from the ligand binding surface. As seen previously, ADGRG6 WT +ss signals at a 10-fold level higher than empty vector (EV), a significantly higher extent than the -ss isoform which signaling 5-fold higher compared to EV. ADGRG6 C94Y was found super active in both -ss and +ss contexts (~ 30-fold over EV for both); not only is the signaling higher, but the difference between -ss and +ss is no longer observable. The Y96A –ss variant signaled at a level comparable to WT –ss. However, Y96A +ss signaling was reduced to roughly 5-fold over EV, suggesting that the event that occurs in WT +ss which leads to increased signaling is not possible in the context of Y96A. The P520Q variant behaves similarly to WT ADGRG6 in both -ss and +ss contexts.

Discussion

Despite their essential functions in development and their role in disease, aGPCRs remain an understudied subfamily of cell-surface receptors with enormous therapeutic potential (*10*, *11*). Understanding the molecular basis of aGPCR function is an important step towards therapeutic development. Within the aGPCR family, ADGRG6 has gained significant attention due to its roles in PNS myelination, angiogenesis, heart and skeletal development, as well as its association with

several developmental diseases and cancers. Here, we describe an ECR-dependent mechanism for ADGRG6 regulation that centers around the ligand binding surface of its CUB domain.

The ADGRG6 -ss ECR samples a conformational landscape defined by the intact CUB/HormR interface

Our MD simulations of ADGRG6 reveal conformational dynamics that clarifies published biophysical data. In addition to the compact conformation observed in the crystal structure of ADGRG6 -ss, negative stain electron microscopy analysis on the -ss ECR revealed five classes with distinct, yet compact, conformations (*18*). We also observed multiple compact conformations of WT -ss ADGRG6 in MD simulations, in agreement with the idea that the -ss ECR explores conformational space while keeping the CUB/HormR interface intact. The radius of gyration in solution (R_g) for ADGRG6 -ss as determined by small angle X-ray scattering is 41.1 Å and agrees well with our R_g calculated from MD, 40.5 Å. The crystal structure 6V55 has a calculated R_g of 42.6 Å (*18*). This difference could be to the PTX:ECR interface observed in our simulations which make the molecule more compact. The PTX:ECR interaction is not present in PDB 6V55 likely because in the crystal packing of 6V55, the PTX and SEA domains are wedged apart by a PTX domain from a symmetry mate. Taken together, these data suggest that the PTX:ECR interaction seen in our MD simulations may occur in solution.

An updated mechanism for ADGRG6 regulation by its ECR and the importance of the CUB ligand binding site

The CUB domain of ADGRG6 uses its canonical ligand binding surface to engage the HormR domain in the compact ECR conformation observed in the -ss crystal structure. The +ss ECR is observed in extended conformations where the CUB domain is displaced from this surface and exposed to solvent. The extended conformation of +ss correlates with higher cAMP signaling through ADGRG6, suggesting that these two events may be linked. Previous work studied the ADGRG6 variant D134A/F135A, which removes key residues in the CUB binding interface. In the context of the -ss splice form, D134A/F135A results in an in-solution radius similar to the WT +ss ECR, suggesting that this variant disrupts the CUB/HormR interaction to break open the compact conformation of ADGRG6. D134A/F135A also leads to developmental defects in zebrafish, further

highlighting the role of the CUB ligand-binding interface in ADGRG6 function (*18*). However, D134/F135A in the context of +ss ADGRG6 does not have the increased signaling of the WT +ss protein. This suggests that the conformation of the ADGRG6 ECR is not sufficient to explain the signaling differences of -ss and +ss ADGRG6 (*18*).

Our work suggests a mechanistic model where the CUB domain modulates ADGRG6 signaling through its ligand binding site (Fig. 6). For WT ADGRG6 -ss, a moderate level of signaling is available, where the CUB domain remains bound to the HormR domain even though multiple conformations are explored (Fig. 6A). The CUB domain is required for basal ADGRG6 signaling as its deletion on both the -ss and +ss receptor produces receptors which are defective in cAMP signaling and cannot bind either its intramolecular or other putative ligand (Fig 6B). In the context of WT +ss, the CUB domain is displaced from the HormR domain and its ligand binding surface is exposed to solvent, and we propose that the higher signaling activity is due to its interaction with a putative ligand through the CUB ligand binding surface (Fig 6C). C94Y also acts in a manner similar to the WT +ss receptor, but more potently. Thus C94Y and WT +ss form one class of variant which act distally to the CUB canonical ligand binding surface, but effectively remove the CUB domain from the binding site on the ADGRG6 HormR domain. The solvent exposed CUB domain then interacts with a putative ligand to access a higher signaling state.

Another class of variant is Y96A along with the previously characterized D134A/F135A (Fig. 6D). The D134/F135A was shown to displace CUB from its HormR binding site, and we speculate since Y96A is also critical in the CUB domain binding surface that it also would do so. Key residues in the CUB ligand binding site are lacking in this class of variant, so the higher level of signaling accessible to WT +ss is not accessible as they cannot bind as well to the putative ligand in the extracellular space. These variants are a less potent version of ΔCUB.

A population of AIS patients exhibit an ADGRG6 mutation which leads to decreased inclusion of exon 6 when studied in an in vitro assay to assess prevalence of splice forms (*31*). If this holds true in vivo, this means that the ADGRG6 -ss is more prevalent. Our model thus adds context to this observation; we speculate that these patients are more susceptible to AIS because their ADGRG6

is mostly -ss and thus is defective in binding the putative extracellular ligand and does not access the higher cAMP signaling state.

The identity of ADGRG6 CUB domain ligands remain mysterious

The missing piece in our proposed mechanism is the identity of putative ligands in the extracellular space which grant access to a higher level of signaling through interaction with the ADGRG6 canonical CUB domain ligand binding surface. A prime candidate is Collagen IV, which is a key component of the extracellular matrix in the Schwann cells which initiate PNS myelination. Collagen IV has been proposed as a bona fide ligand of ADGRG6 (*39*). In pulldown experiments, an ECR fragment including the CUB-PTX domain tandem of ADGRG6 binds to Collagen IV. Furthermore, addition of soluble Collagen IV peptides stimulates cAMP signaling through ADGRG6 in an ECR-dependent manner. Other CUB domain containing proteins interact with collagen-like triple helical molecules using the canonical ligand binding surface. C1s is a member of the complement cascade which uses the canonical ligand binding surface of its CUB domain to bind to the collagen-like domain of C1q (*38*). The CUB domain of mannin-binding lectin associated serine protease (MASP) interacts with the collagen-like domain of mannin-binding lectin (MBL), also through the canonical CUB ligand binding surface (*37*). Both binding interfaces employ highly similar residues to those that the ADGRG6 CUB domain uses to interact with the HormR domain (Fig. 6E), including the conserved anionic residues which coordinate calcium and the key tyrosine residue as well as the lysine extending from the ligand. The GKLGP motif which serves as the CUB ligand in the collagen repeats of C1q and MBL is present in Collagen IV as GKVGP, GKAGP, and GKPGP (*38*). When D134/F135A ADGRG6 was studied in zebrafish, we speculate that impaired ability to bind Collagen IV through the CUB domain ligand binding site contributed to the observed myelination defects (*18*) as Collagen IV is key to the myelin-producing Schwann cells of the PNS (*39*). Future studies are necessary to comprehensively understand the regulation of ADGRG6 function by its ECR, including comprehensive identification of ADGRG6 CUB domain ligands and direct studies of the putative interaction between the ADGRG6 CUB domain and Collagen IV or other discovered ligands.

Materials and methods

Molecular dynamics simulations

Software used for this work including GROMACS (*45*) was compiled and maintained by SBGrid (*46*). Simulations were initialized from PDB 6V55 (*18*). Atoms missing from the crystal structure were added with Swiss PDB Viewer. N-linked glycosylations and crystal structure waters were removed from the protein. Using GROMACS, the structure was placed in a 4216.9 nm^3 box with periodic boundary conditions such that all protein atoms are at least 13 Å from the boundaries, resulting in a 160 × 160 × 160 Å cube (Fig. 2A). The box was solvated with the SPC water model using gmx solvate and Na^+ and Cl^- counterions were added using gmx genion. The resulting model of ADGRG6 was minimized with the leapfrog integrator in step sizes of 0.1 Å until the max force was smaller than 1000 kJ mol^{-1} nm^{-1}.

The system was then subjected to a two-phase equilibration in the NVT and NPT ensembles with protein atoms under harmonic position restraints. A 100 ps NVT equilibration was carried out with step size 2 fs at 300 K with the Berendsen thermostat using a 0.1 ps coupling constant, and solvent velocities were assigned from the Boltzmann distribution. NPT equilibration was also carried out with step size 2 fs for 100 ps, and at 1 atm with Parrinello-Rahman pressure coupling (2.0 ps coupling constant) using velocities from the NVT ensemble. Simulations on the unrestrainted protein were run in the NPT ensemble for 1 μs and a 2 fs step size. Long-range electrostatics were computed with PME with a 1.6 Å Fourier spacing and 10 Å cutoff. All simulations were run with the AMBER99SB-ILDN force field. Replicate simulations were each separately solvated, minimized, and equilibrated.

Simulated systems

Genotype	Splice isoform	Simulation Length	Replicates
WT	–ss	1 μs	3
D134A/F135A	–ss	1 μs	3
C94Y	–ss	1 μs	1
G372W	–ss	1 μs	1
P520Q	–ss	1 μs	1

PC analysis and free energy landscape analysis

Trajectories from the simulations were extracted for protein atoms and output saved for every ten frames. PCs of the Cαs were calculated by concatenating replicate simulation trajectories for each condition using gmx trjcat and diagonalizing the covariate matrix with gmx covar. Trajectories were projected onto PCs using gmx anaeig and were plotted and visualized with a homemade python script. The fractions of variance f captured by PCs were checked to verify that the first two principal components captured most of the variance and verify the quality of the dimensionality reduction by computing

$$f = \frac{\{\lambda_i\}_{1 \le i \le 3N}}{\mathrm{tr}(C)}$$

where $\{\lambda_i\}_{1 \le i \le 3N}$ is the set of $3N$ eigenvalues for N α-carbons, and $\mathrm{tr}(C)$ is the trace of the covariate matrix. For visualization of trajectories over time on PC space, a Gaussian filter with $\sigma = 50$ was applied to PC1 and PC2 components of each simulation's trajectory. Free energy landscapes were obtained by binning the unfiltered projected simulations into 50×50 bins and computing the Gibbs free energy

$$G = RT\ ln \frac{N_{i,j}}{N_{max}}$$

where $N_{i,j}$ is the number of frames in the ith bin of PC1 and j^{th} bin of PC2 and N_{max} is the number of frames in the most occupied bin. Hierarchical clustering of free energy landscape wells was performed by scipy.cluster in a homemade python script using the Ward clustering method to minimize variance in the l^2 norm. Cluster centroids c_i were computed by averaging the PC1 and PC2 coordinates of frames within each cluster

$$c_i = \frac{1}{N_i} \sum_{j \in \text{cluster } i} (\text{PC1,PC2})_j$$

where N_i is the number of frames in the ith cluster. We chose the frame from the simulations whose PC coordinates are closest by l^2 norm and constructed the 3D structure as a linear combination of PC1 and PC2 using gmx anaeig. For entropy calculations, we used Schlitter's formula due to the

size of the system and emphasis on higher inertia eigenvectors to approximate an upper bound for the protein's entropy as follows

$$S_{\text{system}} \leq S_{\text{ho}} = \frac{k_B \alpha}{exp(\alpha) - 1} - ln(1 - exp(-\alpha))$$

where

$$\alpha = \frac{\hbar\omega}{k_B T}$$

S_{ho} is the entropy of the harmonic oscillator, and ω is the frequency of the harmonic oscillator related to the variance of the system. Entropy calculations were performed with gmx anaeig -entropy. Cumulative entropy plots were calculated at each time t from the covariate matrix for the portion of the simulation $[0, t]$. Subspace overlap between two different simulations with associated sets of eigenvectors $v_1, v_2, \ldots, v_{3N}$ and $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{3N}$ is found by

$$\text{overlap}(v, w) = \frac{1}{3N} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left(v_i \cdot w_j\right)^2$$

producing overlap$(\boldsymbol{v}, \boldsymbol{w}) \in [0,1]$. Computation of overlap was carried out with gmx anaeig -over.

Cloning

Full-length wild-type and mutant *D. rerio* (zebrafish) ADGRG6 (Uniprot ID: C6KFA3) constructs were cloned into pCMV5 by annealing three fragments with the Gibson assembly method. All constructs include N-terminal FLAG-tags for measuring cell-surface expression levels. pCMV5 backbone was amplified using the following primers: F: 5' – CGTCCACCAGAAAGGAGCAGTGG 3' and R: 5'– GGGAACCGGAGCTGAATGAAGCC – 3'. The first zfADGRG6 –ss Y96A fragment was mutated and amplified from full-length zfADGRG6 DNA using F: 5' – AGCACAAGGCTGCATGCGTGACCGCGTTGTCGTC - 3' and R: 5' – CCACTGCTCCTTTCTGGTGGACG 3'; the second Y96A fragment was mutated and amplified from full-length zfADGRG6 DNA using F: 5' – GGCTTCATTCAGCTCCGGTTCCC – 3' and R: 5' – GACGACAACGCGGTCAGCGATGCAGCCTTGTGCT – 3'. In a similar scheme to Y96A, the following fragment: primer combinations were utilized to amplify the –ss zfADGRG6 C94Y, P520Q, and G372W variants:

Fragment	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
C94Y, 1	CGAGGAAGCACAAGGCTACATCTATGACCGCG	CCACTGCTCCTTTCTGGTGGACG
C94Y, 2	GGCTTCATTCAGCTCCGGTTCCC	CGCGGTCATAGATGTAGCCTTGTGCTTCCTCG
P520Q, 1	CCACTATAGATGGCAAGAGAGCAGACCCACAG	CCACTGCTCCTTTCTGGTGGACG
P520Q, 2	GGCTTCATTCAGCTCCGGTTCCC	GACTGTGGGTCTGCTCTCTTGCCATCTATAGTGG
G372W, 1	GCAGGTTGTGCTTCTTGGCTTGGCTGCCCAG	CCACTGCTCCTTTCTGGTGGACG
G372W, 2	GGCTTCATTCAGCTCCGGTTCCC	CTGGGCAGCCAAGCCAAGAAGCACAACCTGC

The identity of each construct was confirmed with full-plasmid sequencing.

cAMP signaling assay

HEK293 cells (ATCC CRL-1573) 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-h incubation, the transfected cells were detached and seeded (50,000 cells per well) in a white 96-well assay plate. Following another 24-h incubation, the DMEM was replaced with 100 µL Opti-MEM (Gibco, 31985079) and incubated for 30 min. To each well was then added 1 µL GloSensor substrate and 11 µL FBS. Basal-level luminescence measurements were taken after 30 min to allow for equilibration.

Flow cytometry

HEK293T cells were transfected in a 6-well plate with 1 µg ADGRG6 DNA and 3 µL transfection reagent Fugene 6 (Promega, PRE2693) at 60-70% confluency and incubated for 48 h. The cells were detached with citric saline and washed with phosphate-buffered saline (PBS). The cells were

then washed twice with PBS supplemented with 0.1% BSA (Sigma–Aldrich, A3803). The cells were incubated with mouse anti-FLAG primary antibody (1:1000 dilution in PBS + 0.1% BSA; Sigma–Aldrich, F3165) at room temperature for 30 min and washed three times with PBS + 0.1% BSA. The cells were then incubated with donkey anti-mouse Alexa Fluor 488 secondary antibody (1:500 dilution in PBS + 0.1% BSA; Invitrogen) at room temperature for 30 min and washed three times with PBS + 0.1% BSA. Stained cells were resuspended in PBS + 0.1% BSA and at least 5000 data points were collected with a BD Accuri C6 flow cytometer. The resulting data were analyzed using FloJo v10, single HEK293T cells were gated, and mean fluorescence intensity in the 488 channel was used for comparison.

Statistical Analysis

All statistical analysis for the cAMP signaling assays was performed using GraphPad Prism Version 10. All experiments were performed in at least N=3 independent experiments each with at least five replicates. For statistical comparisons, the one-way ANOVA with Tukey's correction for multiple comparisons test was used to assess statistical significance. Each conditions are compared with every other condition to generate the statistical significance, but only some comparisons are shown for clarity.

Acknowledgements

This work was completed in part with resources provided by the University of Chicago Research Computing Center.

Author contributions

Conceptualization: ED, SJB, DA

Methodology: ED, SJB, DA

Investigation: ED, SJB, KL

Visualization: ED, SJB

Supervision: DA

Writing—original draft: ED

Writing—review & editing: ED, SJB, DA

References

1. N. King, C. T. Hittinger, S. B. Carroll, Evolution of key cell signaling and adhesion protein families predates animal origins. *Science (80-.).* **301**, 361–363 (2003).
2. N. King, M. J. Westbrook, S. L. Young, A. Kuo, M. Abedin, J. Chapman, S. Fairclough, U. Hellsten, Y. Isogai, I. Letunic, M. Marr, D. Pincus, N. Putnam, A. Rokas, K. J. Wright, R. Zuzow, W. Dirks, M. Good, D. Goodstein, D. Lemons, W. Li, J. B. Lyons, A. Morris, S. Nichols, D. J. Richter, A. Salamov, P. Bork, W. A. Lim, G. Manning, W. T. Miller, W. McGinnis, H. Shapiro, R. Tjian, I. V. Grigoriev, D. Rokhsar, The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451**, 783–788 (2008).
3. L. M. Wingler, R. J. Lefkowitz, Conformational Basis of G Protein-Coupled Receptor Signaling Versatility. *Trends Cell Biol.* **30**, 736–747 (2020).
4. D. E. Grigoriadis, S. R. J. Hoare, S. M. Lechner, D. H. Slee, J. A. Williams, Drugability of extracellular targets: Discovery of small molecule drugs targeting allosteric, functional, and subunit-selective sites on GPCRs and ion channels. *Neuropsychopharmacology* **34**, 106–125 (2009).
5. J. Hamann, G. Aust, D. Araç, F. B. Engel, C. Formstone, R. Fredriksson, R. A. Hall, B. L. Harty, C. Kirchhoff, B. Knapp, A. Krishnan, I. Liebscher, H. H. Lin, D. C. Martinelli, K. R. Monk, M. C. Peeters, X. Piao, S. Prömel, T. Schöneberg, T. W. Schwartz, K. Singer, M. Stacey, Y. A. Ushkaryov, M. Vallon, U. Wolfrum, M. W. Wright, L. Xu, T. Langenhan, H. B. Schiöth, International union of basic and clinical pharmacology. XCIV. adhesion G protein-coupled receptors. *Pharmacol. Rev.* **67**, 338–367 (2015).
6. S. Yona, H. H. Lin, W. O. Siu, S. Gordon, M. Stacey, Adhesion-GPCRs: emerging roles for novel receptors. *Trends Biochem. Sci.* **33**, 491–500 (2008).
7. T. Langenhan, X. Piao, K. R. Monk, Adhesion G protein-coupled receptors in nervous system development and disease. *Nat. Rev. Neurosci.* **17**, 550–561 (2016).
8. T. Lala, R. A. Hall, ADHESION G PROTEIN-COUPLED RECEPTORS: STRUCTURE, SIGNALING, PHYSIOLOGY, AND PATHOPHYSIOLOGY. American Physiological Society (2022). https://doi.org/10.1152/physrev.00027.2021.
9. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, D. E. Gloriam, Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
10. F. Bassilana, M. Nash, M. G. Ludwig, Adhesion G protein-coupled receptors: opportunities for drug discovery. *Nat. Rev. Drug Discov.* **18**, 869–884 (2019).
11. A. D. Bondarev, M. M. Attwood, J. Jonsson, V. N. Chubarev, V. V. Tarasov, H. B. Schiöth, Opportunities and challenges for drug discovery in modulating Adhesion G protein-coupled receptor (GPCR) functions. *Expert Opin. Drug Discov.*, 1–17 (2020).
12. K. J. Paavola, J. R. Stephenson, S. L. Ritter, S. P. Alter, R. A. Hall, The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *J. Biol. Chem.* **286**, 28914–28921 (2011).
13. A. Kishore, R. H. Purcell, Z. Nassiri-Toosi, R. A. Hall, Stalk-dependent and stalk-independent signaling by the adhesion G protein-coupled receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). *J. Biol. Chem.* **291**, 3385–3394 (2016).
14. G. S. Salzman, S. Zhang, A. Gupta, A. Koide, S. Koide, D. Araç, Stachel-independent modulation of GPR56/ADGRG1 signaling by synthetic ligands directed to its extracellular region. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10095–10100 (2017).
15. G. S. Salzman, S. D. Ackerman, C. Ding, A. Koide, K. Leon, R. Luo, H. M. Stoveken, C. G. Fernandez, G. G. Tall, X. Piao, K. R. Monk, S. Koide, D. Araç, Structural Basis for Regulation of GPR56/ADGRG1 by Its Alternatively Spliced Extracellular Domains. *Neuron* **91**, 1292–1304 (2016).
16. T. Li, B. Chiou, C. K. Gilman, R. Luo, T. Koshi, D. Yu, H. C. Oak, S. Giera, E. Johnson-Venkatesh, A. K. Muthukumar, B. Stevens, H. Umemori, X. Piao, A splicing isoform of GPR56 mediates microglial synaptic refinement via phosphatidylserine binding. *EMBO J.* **39**, 1–20 (2020).
17. J. Li, Y. Xie, S. Cornelius, X. Jiang, R. Sando, S. P. Kordon, M. Pan, K. Leon, T. C. Südhof,

M. Zhao, D. Araç, Alternative splicing controls teneurin-latrophilin interaction and synapse specificity by a shape-shifting mechanism. *Nat. Commun.* **11**, 2140 (2020).
18. K. Leon, R. L. Cunningham, J. A. Riback, E. Feldman, J. Li, T. R. Sosnick, M. Zhao, K. R. Monk, D. Araç, Structural basis for adhesion G protein-coupled receptor Gpr126 function. *Nat. Commun.* **11** (2020).
19. S. P. Kordon, P. Dutka, J. M. Adamska, S. J. Bandekar, K. Leon, S. K. Erramilli, B. Adams, J. Li, A. A. Kossiakoff, D. Araç, Isoform- and ligand-specific modulation of the adhesion GPCR ADGRL3/Latrophilin3 by a synthetic binder. *Nat. Commun.* **14** (2023).
20. A. A. Boucard, S. Maxeiner, T. C. Südhof, Latrophilins function as heterophilic cell-adhesion molecules by binding to teneurins: Regulation by alternative splicing. *J. Biol. Chem.* **289**, 387–402 (2014).
21. S. Wang, C. DeLeon, W. Sun, S. R. Quake, B. L. Roth, T. C. Südhof, Alternative splicing of latrophilin-3 controls synapse formation. *Nature* **626** (2024).
22. H. Cui, Y. Wang, H. Huang, W. Yu, M. Bai, L. Zhang, B. A. Bryan, Y. Wang, J. Luo, D. Li, Y. Ma, M. Liu, GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. *J. Biol. Chem.* **289**, 34871–34885 (2014).
23. K. R. Monk, K. Oshima, S. Jörs, S. Heller, W. S. Talbot, Gpr126 is essential for peripheral nerve development and myelination in mammals. *Development* **138**, 2673–2680 (2011).
24. C. Patra, M. J. Van Amerongen, S. Ghosh, F. Ricciardi, A. Sajjad, T. Novoyatleva, A. Mogha, K. R. Monk, C. Mühlfeld, F. B. Engel, Organ-specific function of adhesion G protein-coupled receptor GPR126 is domain-dependent. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16898–16903 (2013).
25. C. Patra, K. R. Monk, F. B. Engel, The multiple signaling modalities of adhesion G protein-coupled receptor GPR126 in development. *Recept. Clin. Investig.* **1**, 79 (2014).
26. K. R. Monk, S. G. Naylor, T. D. Glenn, S. Mercurio, J. R. Perlin, C. Dominguez, C. B. Moens, W. S. Talbot, A G protein-coupled receptor is essential for schwann cells to initiate myelination. *Science (80-.).* **325**, 1402–1405 (2009).
27. A. Mogha, A. E. Benesh, C. Patra, F. B. Engel, T. Schöneberg, I. Liebscher, K. R. Monk, Gpr126 functions in schwann cells to control differentiation and myelination via G-protein activation. *J. Neurosci.* **33**, 17976–17985 (2013).
28. G. Ravenscroft, F. Nolent, S. Rajagopalan, A. M. Meireles, K. J. Paavola, D. Gaillard, E. Alanio, M. Buckland, S. Arbuckle, M. Krivanek, J. Maluenda, S. Pannell, R. Gooding, R. W. Ong, R. J. Allcock, E. D. F. Carvalho, M. D. F. Carvalho, F. Kok, W. S. Talbot, J. Melki, N. G. Laing, Mutations of GPR126 Are Responsible for Severe Arthrogryposis Multiplex Congenita. *Am. J. Hum. Genet.* **96**, 955–961 (2015).
29. I. Kou, Y. Takahashi, T. A. Johnson, A. Takahashi, L. Guo, J. Dai, X. Qiu, S. Sharma, A. Takimoto, Y. Ogura, H. Jiang, H. Yan, K. Kono, N. Kawakami, K. Uno, M. Ito, S. Minami, H. Yanagida, H. Taneichi, N. Hosono, T. Tsuji, T. Suzuki, H. Sudo, T. Kotani, I. Yonezawa, D. Londono, D. Gordon, J. A. Herring, K. Watanabe, K. Chiba, N. Kamatani, Q. Jiang, Y. Hiraki, M. Kubo, Y. Toyama, T. Tsunoda, C. A. Wise, Y. Qiu, C. Shukunami, M. Matsumoto, S. Ikegawa, Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat. Genet.* **45**, 676–679 (2013).
30. J. F. Xu, G. hai Yang, X. H. Pan, S. J. Zhang, C. Zhao, B. S. Qiu, H. F. Gu, J. F. Hong, L. Cao, Y. Chen, B. Xia, Q. Bi, Y. P. Wang, Association of GPR126 gene polymorphism with adolescent idiopathic scoliosis in Chinese populations. *Genomics* **105**, 101–107 (2015).
31. E. Xu, W. Shao, H. Jiang, T. Lin, R. Gao, X. Zhou, A Genetic Variant in GPR126 Causing a Decreased Inclusion of Exon 6 Is Associated with Cartilage Development in Adolescent Idiopathic Scoliosis Population. *Biomed Res. Int.* **2019** (2019).
32. G. Liu, S. Liu, M. Lin, X. Li, W. Chen, Y. Zuo, J. Liu, Y. Niu, S. Zhao, B. Long, Z. Wu, N. Wu, G. Qiu, Genetic polymorphisms of GPR126 are functionally associated with PUMC classifications of adolescent idiopathic scoliosis in a Northern Han population. *J. Cell. Mol. Med.* **22**, 1964–1971 (2018).
33. S. Nik-Zainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S.

Morganella, M. R. Aure, O. C. Lingjærde, A. Langerød, M. Ringnér, S. M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. J. Hooijer, S. J. Jang, D. R. Jones, H. Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J. Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodríguez-González, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van Den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. V. t. Veer, A. Tutt, S. Knappskog, B. K. T. Tan, J. Jonkers, Å. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. Van De Vijver, J. W. M. Martens, A. L. Børresen-Dale, A. L. Richardson, G. Kong, G. Thomas, M. R. Stratton, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

34. S. Garinet, G. Pignot, S. Vacher, C. Le Goux, A. Schnitzler, W. Chemlali, N. Sirab, N. B. Delongchamps, M. Zerbib, M. Sibony, Y. Allory, D. Damotte, I. Bieche, High prevalence of a hotspot of noncoding somatic mutations in intron 6 of GPR126 in bladder cancer. *Mol. Cancer Res.* **17**, 469–475 (2019).
35. S. W. Piraino, S. J. Furney, Identification of coding and non-coding mutational hotspots in cancer genomes. *BMC Genomics* **18**, 1–17 (2017).
36. C. Gaboriaud, L. Gregory-Pauron, F. Teillet, N. M. Thielens, I. Bally, Ǵerard J. Arlaud, Structure and properties of the Ca 2+ -binding CUB domain, a widespread ligand-recognition unit involved in major biological functions. *Biochem. J.* **439**, 185–193 (2011).
37. A. R. Gingras, U. V. Girija, A. H. Keeble, R. Panchal, D. A. Mitchell, P. C. E. Moody, R. Wallis, Structural basis of mannan-binding lectin recognition by its associated serine protease MASP-1: Implications for complement activation. *Structure* **19**, 1635–1643 (2011).
38. U. V. Girija, A. R. Gingras, J. E. Marshall, R. Panchal, M. A. Sheikh, P. Gál, W. J. Schwaeble, D. A. Mitchell, P. C. E. Moody, R. Wallis, Structural basis of the C1q/C1s interaction and its central role in assembly of the C1 complex of complement activation. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13916–13920 (2013).
39. K. J. Paavola, H. Sidik, J. B. Zuchero, M. Eckart, W. S. Talbot, Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci. Signal.* **7**, 1–10 (2014).
40. B. Kieslich, R. H. Weiße, J. Brendler, A. Ricken, T. Schöneberg, N. Sträter, The dimerized pentraxin-like domain of the adhesion G protein–coupled receptor 112 (ADGRG4) suggests function in sensing mechanical forces. *J. Biol. Chem.* **299**, 105356 (2023).
41. D. Thompson, M. B. Pepys, S. P. Wood, The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure* **7**, 169–177 (1999).
42. S. E. Koistoe, B. H. Ridha, V. Bellotti, N. Wang, C. V. Robinson, S. J. Crutch, G. Keir, R. Kukkastenvehmas, J. R. Gallimore, W. L. Hutchinson, P. N. Hawkins, S. P. Wood, M. N. Rossor, M. B. Pepys, Molecular dissection of Alzheimer's disease neuropathology by depletion of serum amyloid P component. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7619–7623 (2009).
43. D. Araç, A. A. Boucard, M. F. Bolliger, J. Nguyen, S. M. Soltis, T. C. Südhof, A. T. Brunger, A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO J.* **31**, 1364–1378 (2012).
44. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander, N. Schultz, The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
45. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).
46. A. Morin, B. Eisenbraun, J. Key, P. C. Sanschagrin, M. A. Timony, M. Ottaviano, P. Sliz, Collaboration gets the most out of software. *Elife* **2013**, 1–6 (2013).

Figures

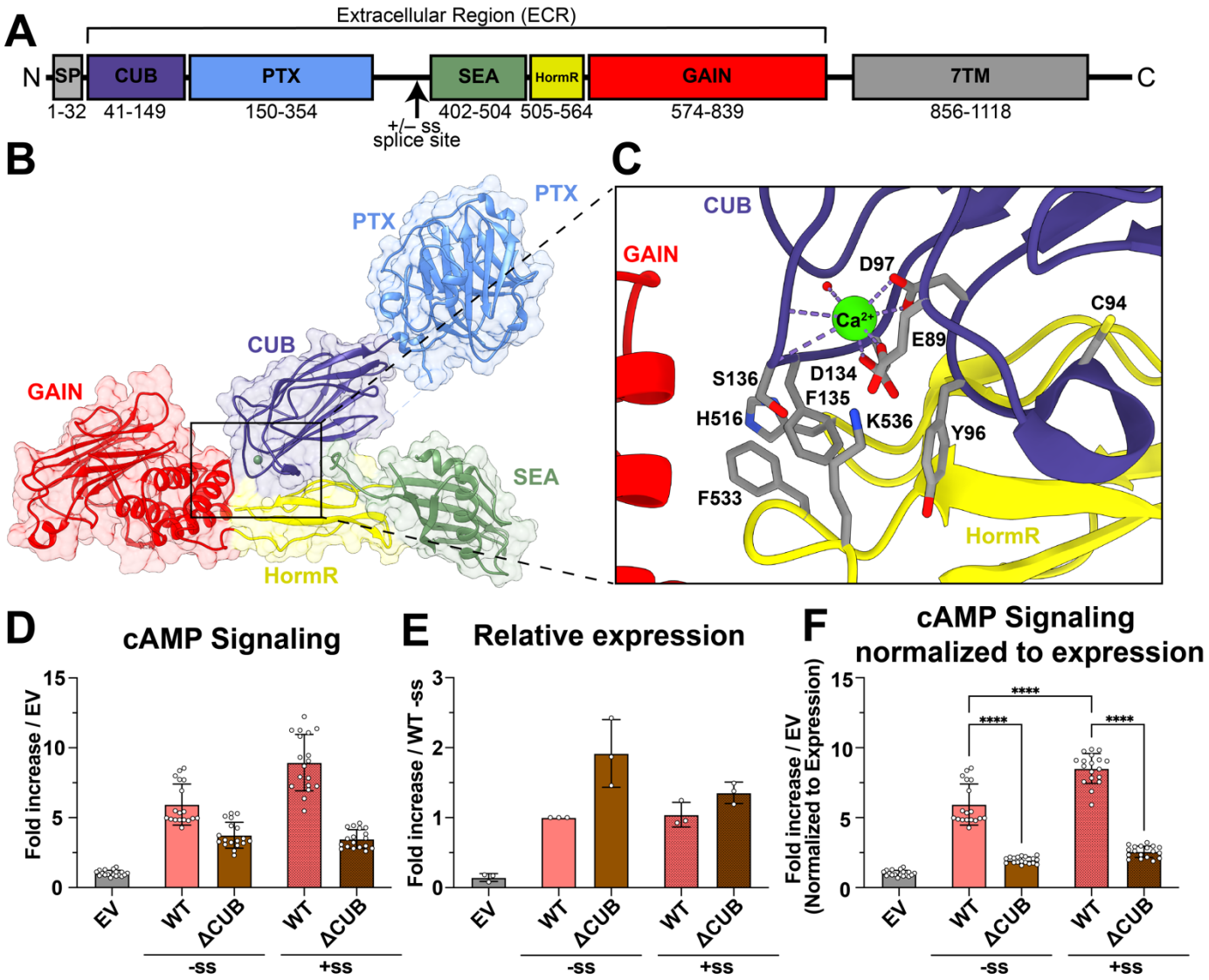

Fig. 1: Deletion of the CUB domain inhibits ADGRG6-dependent cAMP signaling. (A) Linear domain diagram for ADGRG6, SP denotes the signal peptide. The position of the splice site is designated by an arrow. (B) ADGRG6 –ss ECR crystal structure (PDB 6V55) depicted using cartoon along with surface representation. (C) Close-up of the CUB/HormR interface from the crystal structure. (D) Basal cAMP signaling of ADGRG6 constructs shown as fold change over empty vector (EV). Variants are grouped by absence (-ss) or presence (+ss) of the ECR splice insert. (E) Relative cell surface expression levels for ADGRG6 constructs shown as fold increase over WT -ss. These are measured using flow cytometry against the N-terminal FLAG tag. (F) Basal cAMP signaling fold change over EV normalized to cell-surface expression. **** - p<0.0001; one-way ANOVA test with Tukey's correction for multiple comparisons. Each condition was compared to every other condition, but not all comparisons are shown for clarity. N=3 independent experiments were performed with six replicates each.

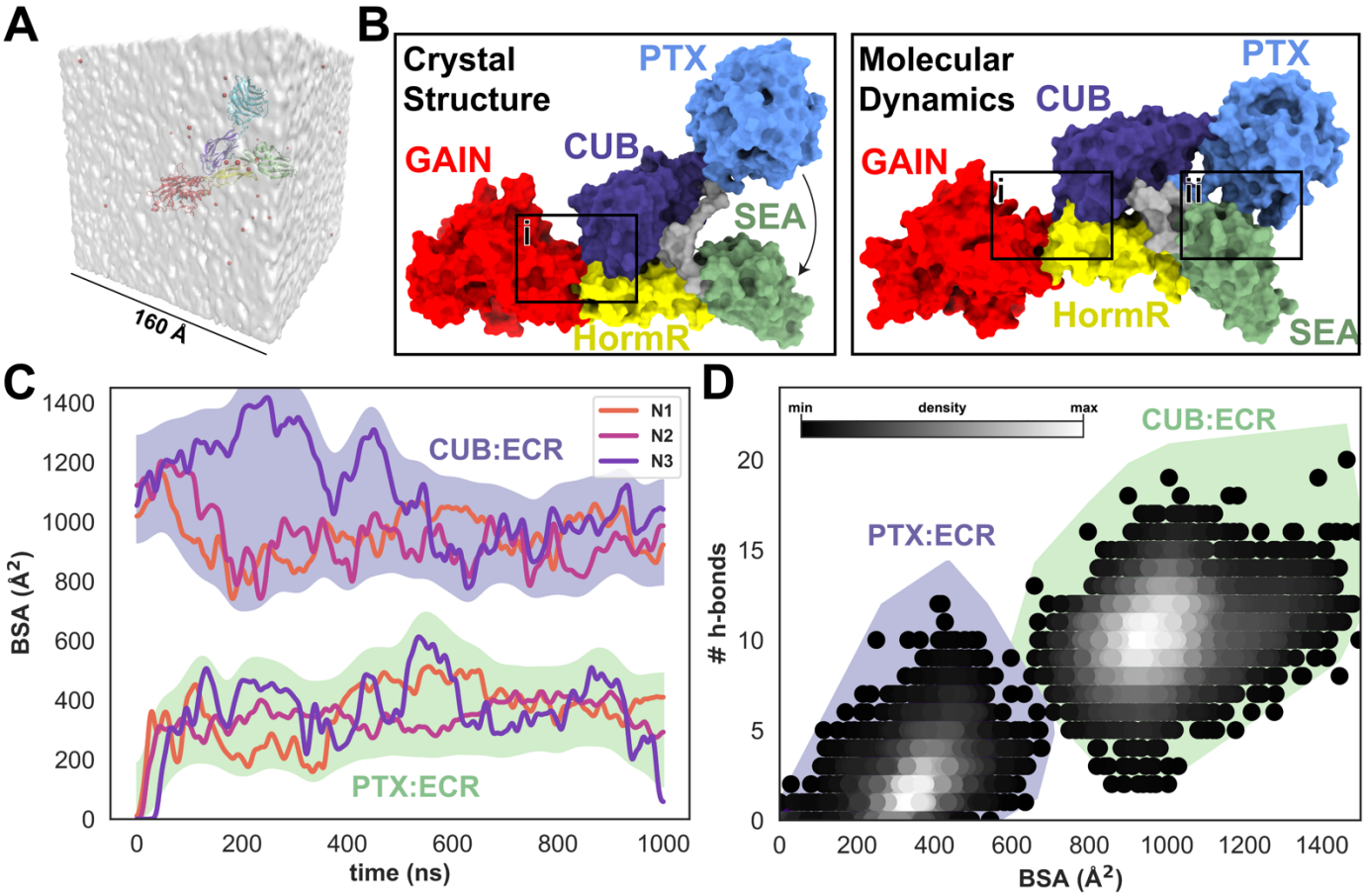

Fig. 2: Molecular dynamics simulations of ADGRG6 –ss show sustained interdomain contacts involving the CUB, PTX, and SEA domains. (A) System simulated for each MD experiment, with cubic water box shown and sodium counterions shown as red spheres. (B) Surface representation of ADGRG6 from the crystal structure with the CUB:ECR interface (i) shown as a black rectangle. A representative frame from the MD simulations with the CUB:ECR interface (i) and PTX:ECR interface (ii) shown as black rectangles. (C) Buried surface area (BSA) of CUB:ECR and PTX:ECR interfaces over time. Colored lines show each individual simulation, and the shaded areas represent the upper and bounds of all simulations. (D) The number of hydrogen bonds versus the BSA of the described interfaces. Shading represents the density of simulation frames that occupy a given location in the plot.

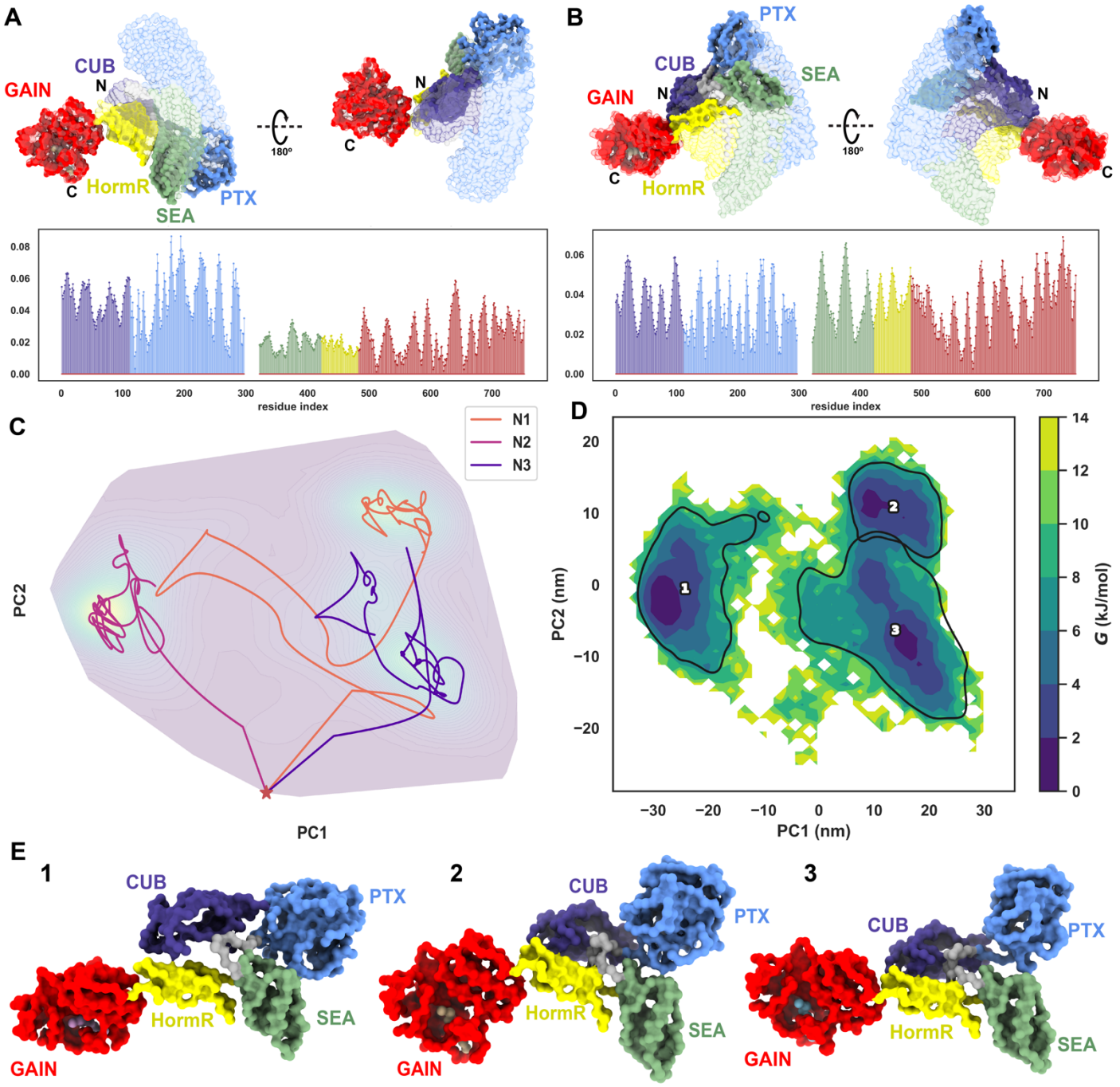

Fig. 3: Principal component analysis reveals multiple conformations of the ADGRG6 –ss ECR with the CUB domain ligand binding site docked onto HormR. Principal component analysis was performed on the coordinates of concatenated simulations. (A) Extreme projections of PC1, which is characterized by a 40 ° bending movement of CUB and PTX are shown as darker surface representations and other frames are shown as transparent representations. Below, the length of eigenvectors plotted for each residue shows the contribution of each residue to the motion captured in the PC. (B) Extreme projections of PC2, which is characterized by a 60 ° rotation movement of CUB-PTX-SEA-HormR about the GAIN domain, are shown as darker surface representations and other frames are shown as transparent representations. Below, the length of eigenvectors plotted for each residue shows the contribution of each residue to the motion captured in the PC. (C) Colored line indicates the simulation trajectories in essential space for each simulation. The red star is a projection of the starting point of the simulation, 6V55, onto this PC space. (D) Essential space colored using free energy estimates of each coordinate. Black boundaries show clusters obtained using hierarchical clustering and numbers are paced at the centroid of each cluster. (E) The three centroids from panel D are shown as surface representations to illustrate conformational changes across ADGRG6 -ss ECR simulations.

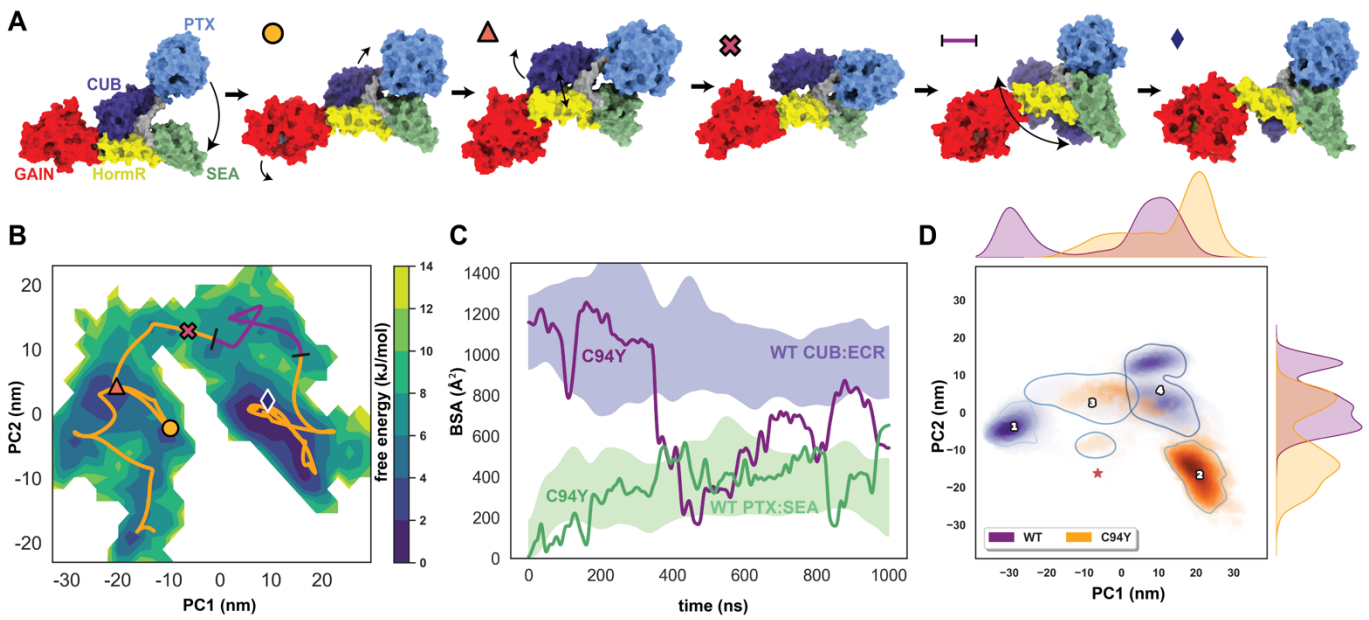

Fig. 4: The C94Y variant leads the ADGRG6 –ss ECR to adopt a favorable conformation with the CUB domain ligand binding site exposed. (A) Representative frames of C94Y -ss in MD show conformational transitions throughout the simulation. First PTX docks to SEA, and then CUB dissociates from HormR. After CUB samples several conformations, it docks against the side of HormR with its ligand binding surface exposed. (B) PCA reveals two PCs. A colored line indicates the simulation trajectory in essential space with different shapes corresponding to the frames shown in panel A. The PC space is colored using free energy estimates of each coordinate in essential space. (C) Buried surface area (BSA) of CUB:ECR and PTX:SEA interfaces over time. Colored lines show the C94Y simulation, and the shaded areas represent the upper and bounds of all WT simulations. (D) Essential space comparison of the WT (purple) and C94Y simulations (orange), obtained from performing PCA on concatenated WT and C94Y simulations. Blue boundaries show clusters obtained using hierarchical clustering. 1D histograms represent density of projections of each simulation along each PC. The red star is a projection of the starting point of the simulation, 6V55, onto this PC space.

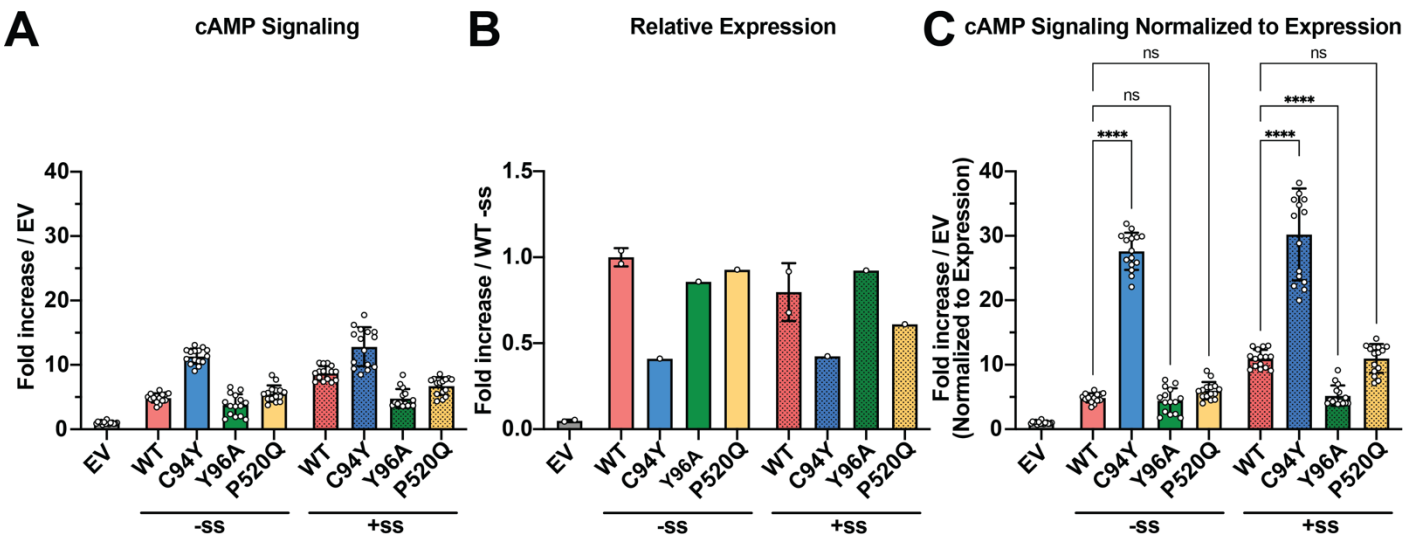

Fig. 5: Point variants sourced from cancer patients modulate ADGRG6 cAMP signaling. (A) Basal cAMP signaling of ADGRG6 constructs shown as fold change over empty vector (EV). Variants are grouped by absence (-ss) or presence (+ss) of the ECR splice insert. (B) Relative cell surface expression levels for ADGRG6 constructs shown as fold change over WT -ss. These are measured using flow cytometry against the N-terminal FLAG tag. (C) Basal cAMP signaling fold change over EV normalized to cell-surface expression. **** - $p<0.0001$, ns - not significant; one-way ANOVA test with Tukey's correction for multiple comparisons. Each condition was compared to every other condition, but not all comparisons are shown for clarity. WT -ss vs. WT +ss comparison results in ****, $p<0.0001$ but is not shown for clarity. N=3 independent experiments were performed with six replicates each.

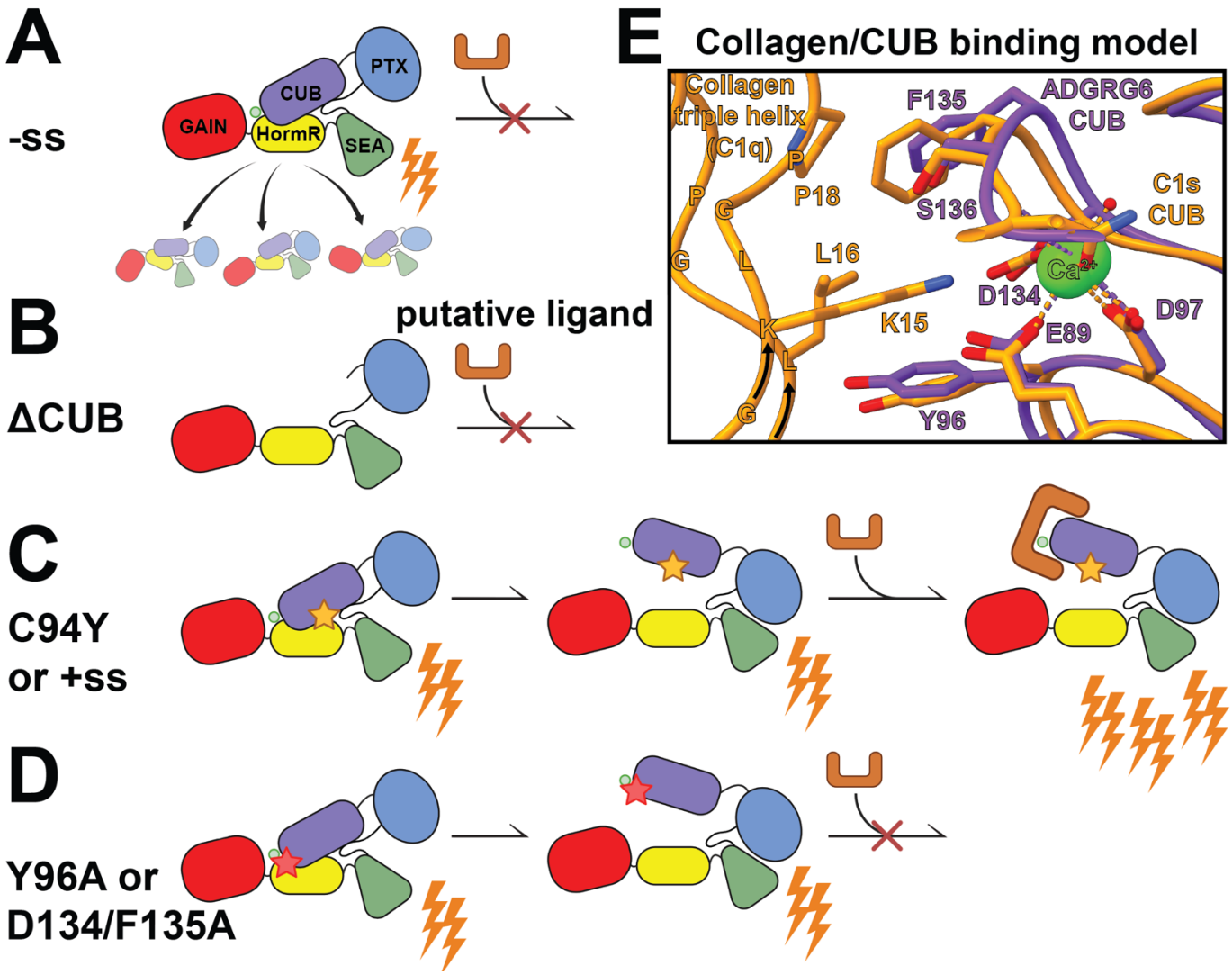

Fig. 6: Mechanistic model for regulation of ADGRG6 signaling by its CUB domain. (A) For the ADGRG6 -ss receptor, a variety of conformations are sampled, all which keep the CUB/HormR interface intact. This leads to a moderate level of cAMP signaling (two lightning bolts) and this receptor is defective in binding a putative extracellular ligand (brown U-like shape). (B) The ΔCUB receptor is unable to either bind the HormR site or a putative extracellular ligand and is defective in cAMP signaling. (C) Variants which affect the CUB/HormR interface which are distal to the canonical CUB ligand binding surface displace the CUB domain from HormR and then they can access a higher signaling state (six lightning bolts) upon binding of a putative ligand. (D) Variants which directly affect the canonical ligand binding interface displaces the CUB domain from HormR, but they cannot bind a putative extracellular ligand and thus they cannot access the higher signaling state. (E) Speculative model for how the CUB domain of ADGRG6 could interact with a collagen like ligand, perhaps Collagen IV. In orange is the C1s CUB domain bound to the collagen triple helix of C1q. In purple is the ADGRG6 CUB domain. Important residues are shown as sticks and the calcium ion is shown as a green sphere. On the collagen triple helix backbone, the GKLGP motif is shown with arrows indicating N ---> C.